UNIVERSITÀ DEGLI STUDI DI URBINO CARLO BO

Department of Biomolecular Sciences (DiSB)

Ph.D. Course in Life Sciences, Health and Biotechnologies

XXX cycle

The biological role of the IGF-1 pool in breast cancer

SSD: BIO/13

Supervisor:

Prof. Elena Barbieri

Ph.D. student:

Dr. Serena Contarelli

Co-Advisor:

Ph.D. Giosuè Annibalini

ACADEMIC YEAR 2016-2017

ABSTRACT

The insulin-like growth factor-1 (IGF-1) is a polypeptide growth factor that is essential for normal body growth and development of several tissues. IGF-1 is implicated in the progression and risk of several malignancies, including breast cancer (BC). Alternative splicing of terminal exon 5 of the *igf-1* gene results in multiple isoforms possessing distinct carboxy-terminal extensions, called the Ea-, Eb- and Ec-domains.

The first chapter of the Thesis provides evidence for an evolutionary mechanism generating the diversity of *igf-1* splicing and expression across species. Our study highlights how the *igf-1* exon 5 originates from exonization of a Mammalian interspersed repetitive-b (MIR-b) element in mammals. The acquisition of exon 5 alters the splicing pattern of *igf-1* in mammals by generating two new isoforms: IGF-1Eb and IGF-1Ec. The evolutionary analysis of mammalian IGF-1 domains showed that E-domains are subjected to a strong evolutionary constraint on the synonymous sites, and they are enriched in disorder-promoting amino acids (i.e. intrinsically disordered), suggesting an important and novel regulatory role for these domains, not previously described.

In the second chapter, we highlighted that IGF-1 pro-hormones are not a simple inactive precursor of mature IGF-1, but are stable intermediates of their posttranslational processing. The IGF-1 pro-hormones can induce BC cell proliferation *via* the IGF-1 receptor, independently from the mature IGF-1 form. These results underline the importance of an accurate assessment of the presence of IGF-1 pro-hormones within the BC microenvironment.

The third chapter describes the mechanisms, which control the IGF-1 pro-hormones biosynthesis. We demonstrated that N-linked glycosylation regulates the stability and secretion of IGF-1Ea pro-hormone, probably ensuring proper pro-hormone folding and favoring its passage through the secretory pathway. The alternative Eb- and Ec-domains lack N-terminal glycosylation sites hence IGF-1Eb and IGF-1Ec pro-hormones were insensitive to glycosylation status of the cells. Moreover, the Eb- and Ec-domains regulate the subcellular localizations of IGF-1Eb and IGF-1Ec pro-hormones, promoting their nuclear accumulation. Thus, disordered E-domains play an important role in the structure, regulation and functioning of IGF-1.

The final chapter of the Thesis describes the data deriving from DIANA-5, and focuses on the effectiveness of modification in dietary change-associated with moderate physical activity in the prevention of BC recurrence, highlighting the importance of the lifestyle modification in the modulation of the circulating levels of IGF-1.

CONTENT

ORIGINAL PAPERS

This Thesis is based on the following original research articles, which will be referred to by their Roman numerals.

I. Annibalini G, Bielli P, De Santi M, Agostini D, Guescini M, Sisti D, Contarelli S, Brandi G, Villarini A, Stocchi V, Sette C, Barbieri E. MIR retroposon exonization promotes evolutionary variability and generates species-specific expression of IGF-1 splice variants. Biochim Biophys Acta (BBA) - Gene Regulatory Mechanisms, 2016 May; 1859(5):757-68. doi: 10.1016/j.bbagrm.2016.03.014. Epub 2016 Apr 19.

II. De Santi M, Annibalini G, Barbieri E, Villarini A, Vallorani L, Contarelli S, Berrino F, Stocchi V, Brandi G. Human IGF-1 pro-forms induce breast cancer cell proliferation via the IGF-1 receptor. Cellular Oncology, (Dordr) 2016 Apr; 39(2):149-159. doi: 10.1007/s13402-015-0263-3. Epub 2015 Dec 23.

III. Glycosylation stabilizes IGF-1Ea pro-hormone and regulates its secretion - In preparation.

IV. Italian Diana5 clinical trial: lifestyles diary food-free diet and/or Mediterranean principles and moderate physical activity influence circulating levels of unbound IGF-1 - In preparation.

V. Circulating IGF-1 and early muscle adaptive responses to an acute isoinertial exercise - In preparation.

INTRODUCTION

THE COMPLEXITY OF THE IGF-1 POOL: GENE SPLICING, REGULATION AND FUNCTION

The insulin-like growth factor 1 (IGF-1), also called somatomedin C, is a polypeptide growth factor, which is essential for normal body growth and development [1]. A variety of cellular responses are induced by IGF-1, including cell proliferation, differentiation and survival [2].

Initially, in the early 1970s, the "somatomedin hypothesis" was proposed as a model for the actions of IGF-1 on the skeleton [3]. This hypothesis postulated that the growth hormone (GH), secreted by the pituitary gland, stimulated IGF-1 synthesis in the liver and its release into the blood stream to target organs acting in an endocrine manner. This hypothesis was later challenged by subsequent findings, including a seminal study examining skeletal development in mice with liver-specific deletion of the *Igf-1* gene (LiverIGF-1−/−) [4]. In this experiment, the circulating IGF-1 level was reduced to less than 25% of normal. Despite this severe reduction, the *knockout* mice developed and grew normally, and their skeletal changes were minimal, indicating that local IGF-1 production is enough to guarantee general growth and skeletal development. The local IGF-1 production extended this hypothesis and included the autocrine-paracrine manner of IGF-1 action. Investigators have recognized that the physio-pathologic mechanisms, through which IGF-1 is regulated and secreted, are more complicated than originally believed. In most of the cellular-animal models, it was proposed that GH stabilized serum IGF-1 by promoting the formation of the ternary complex composed of IGF-1, IGF binding protein (IGFBP) 3 (IGFBP-3) and the acid-labile subunit (ALS) [3]. The bioavailability of circulating IGF-1 modulated by the IGFBPs is further described in the paragraph "*IGF-1 and binding proteins (IGFBPs)*".

In physiological conditions, circulating IGF-1 is mostly synthesized in the liver and acts as an endocrine factor. IGF-1 levels are relatively low at birth, increase during childhood, reaching peak levels in adolescence and begin to decline during the third decade of life [5]. It plays an important role in the first decades of life in normal development and

growth as a key regulator of cell proliferation and differentiation and as an apoptosis and necrosis inhibitor.

Several factors may affect the hepatic synthesis of IGF-1, including insulin, GH, age and nutrition. The growth hormone, produced by the pituitary gland, is the most important hormone involved in regulating body growth and development as well as carbohydrate and lipid metabolism, and its action can be directed to target tissues that possess specific receptors or indirectly through other factors that enhance and complement its effects. The most important mediator of GH effects is IGF-1 [6]. The production and concentration of the two hormones are positively related to each other. The connective tissue cell types that synthesize IGF-1 contain GH receptors and an increase in GH secretion stimulates IGF-1 synthesis. At the same time, an increase in IGF-1 blood concentration suppresses GH synthesis in the pituitary gland though a negative-feedback regulation that represents an important homeostatic mechanism for the maintenance of normal plasma IGF-1 concentration [3].

Other hormones participate with GH in regulating hepatic IGF-1 synthesis, including thyroxin, cortisol, estradiol and testosterone. Thyroxin enhances sensitivity to GH and can increase IGF-1 concentration in hyperthyroidism. Cortisol acts to inhibit IGF-1 synthesis, and high cortisol concentrations can lead to growth attenuation. Etradiol inhibits IGF-1 secretion in the liver by constraining GH stimulated signal transduction [7]. Testosterone enhances hepatic IGF-1 synthesis, but also alters the sensitivity of the pituitary gland to negative-feedback regulation of GH secretion, leading to an increase in GH synthesis and thus an increment in IGF-1 secretion.

Nutrient intake is another variable regulating plasma IGF-1 concentrations. The IGF-1 plasma concentrations are markedly reduced in low protein or calorie-restricted diets [8]. In adults, total caloric intake is more important than protein intake. In fact, in the presence of an adequate caloric intake, even with a low protein intake regime, there can be an increase in IGF-1 levels. Conversely, there is a threshold of caloric intake below which protein intake cannot increase the levels of IGF-1 after fasting. When caloric intake is severely reduced, the dietary content of carbohydrates and essential amino acids is critical for an optimal recovery of IGF-1 levels after fasting [9]. These conditions are associated with a marked decrease in the number of somatotropic receptors supporting the role of a receptor deficiency in the decline of the circulating IGF-1.

In the last decade, many *in vitro* and *in vivo* studies have investigated the *igf-1* gene conservative structure. Different mRNA transcripts are produced as a result of the alternative splicing of the *igf-1* gene, encoding for several IGF-1 precursor proteins also called isoforms. These IGF-1 protein isoforms can be distinguished by the structure of their extension peptides, or E-peptides, on the carboxy-terminal end and by the length of their amino-terminal signal peptides. Interestingly, it has been proposed that these pro-hormones might possess bioactivities that are distinct from those of mature IGF-1 [10].

THE GENE STRUCTURE OF IGF-1 AND ALTERNATIVE SPLICE VARIANTS

The *igf-1* gene is highly conserved among mammals and primates [11]. It is located on the long arm of chromosome 12 in humans and consists of six exons and five introns that cover about 90 kb of DNA with different promoter regions. It is widely believed that all IGF-1 biological actions are mediated by mature IGF-1, but the *igf-1* gene encodes multiple mRNA variants that differ in terms of the presence of an alternative leader sequence and polyadenylation signal [2, 12]. The gene transcription of *igf-1* is very complex due to many transcriptional and post-transcriptional modifications that give rise to several isoforms, of which six are known in the literature.

In particular, exons 1 and 2 encode for the sequence that determines the class of the protein deriving from different splicing of these two exons to the common exon 3. Transcripts starting with exon 1 are referred to as class 1, whereas class 2 transcripts use exon 2 as their leader exon. These exons form two different non-coding 5 'UTR sequences and a sequence that contains the information for a portion of the signal peptide. The expression of these two exons seems to be dependent on two different promoters that are regulated in a tissue specific manner [13]. In particular, class 2 transcripts are expressed mainly in the liver and represent the circulating IGF-1 forms. It has been shown that these forms are dispensable for fetal and postnatal growth [14], which are thought to be more GH dependent [15], whereas transcripts initiating at promoter 1 are widely expressed in many tissues representing local tissue forms [16].

The alternative splicing at the 3' end of the gene involving exons 4, 5 and 6 forms three different extension peptides in humans, called E-peptides, at the carboxy-teminal end (Fig. 1). However, all of them contain exon 3 and exon 4, which encode the mature IGF-1 peptide sequence. This sequence is composed of a total of 70 amino acids, 25 of which derive from exon 3 and 45 from exon 4, forming four domains named for their homology to insulin as B amino-terminal domain, C and A domain and D carboxy-terminal domain. This sequence represents the mature invariant peptide present in all transcripts [17].

The final part of exon 4 encodes for the first 16 amino acids of the amino-terminal portion of the IGF-1 E-peptide domain, which are common to all the three different E-peptides. The alternative splicing involving exons 4, 5 and 6, producing three different IGF-1 E-peptides in humans, produces the remaining portion of the E-peptide.

The IGF-1Ea transcripts from the splicing of exon 4 and exon 6 excluding exon 5, which represents the main isoform for the formation of mature IGF-1 produced by the liver and also in most other tissues and the most conserved isoform across species [11]. This mRNA splicing produces the Ea-peptide, which is composed of 35 amino acids. The first 16 are common peptides deriving from exon 4, whereas the remaining 19 amino acids are encoded by exon 6.

The variant, called IGF-1Eb isoform, contains exon 4 and splices with exon 5, whereas exon 6 is excluded. This transcript is rarely expressed in other species, so this isoform is often considered a human-specific splice variant [12]. It was first detected in human liver [18] and then found in other tissues such as lung carcinoma cells [10], skeletal muscle [19], prostate [20] and endometrium [21]. This variant yields the Eb-peptide, which contain 16 common peptides and 61 additional amino acids deriving from exon 5, resulting in a total of 77 amino acids.

Finally, the alternative spicing of the *igf-1* gene also generates a third transcript, human IGF-1Ec, which corresponds to IGF-1Eb in rodents. In this variant, exon 4 is joined to a partial sequence of exon 5, which in turn is joined to exon 6. It was first detected in the liver and its expression accounts for about 10% of the IGF-1 transcripts. This transcript differs from IGF-1Ea for the presence of the first 49 base pairs from exon 5 and a premature stop codon within exon 6. IGF-1Ec is generated through a cryptic IGF_{633} donor splice site located 49 bp downstream from the 5' end of exon 5, which in turn splices with the acceptor site in the intron preceding exon 6. When this cryptic IGF_{633} donor splice site is not used, the alternative splicing of exon 4 and 5 occurs and the IGF-1Eb peptide is

produced [22]. It results in a different C-terminal peptide sequence due to a read frame shift leading to an Ec-peptide composed of 16 common amino acids, 16 encoded by exon 5 and 8 by exon 6. This splice variant is also referred to as mechano-growth factor (MGF) because it has been shown to be up-regulated in response to muscle exercise and damage [23].

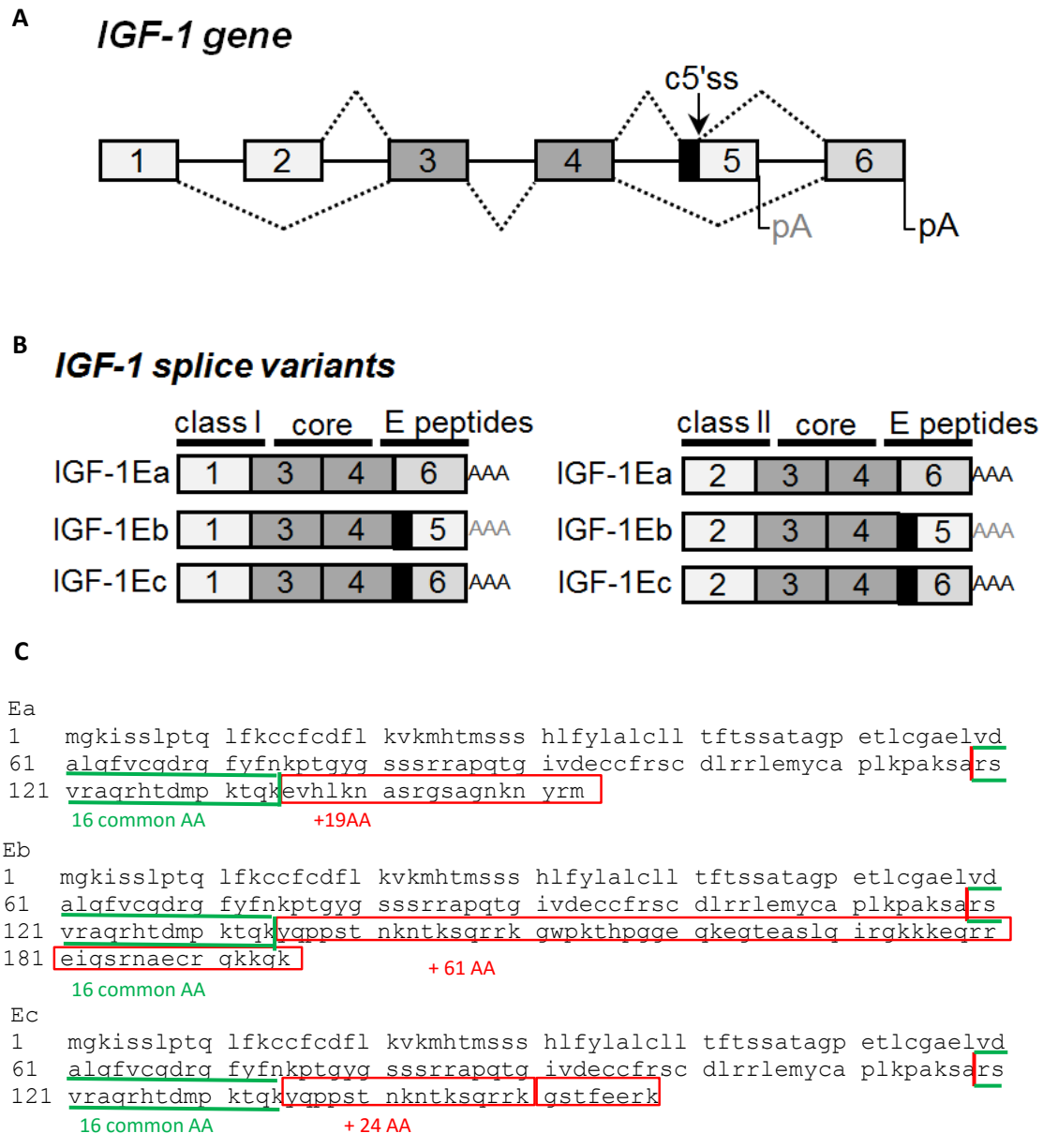


Figure 1. Schematic representation of the *igf-1* gene (A) and its splice variants (B-C). (A) Map of the *igf-1* gene showing exons (*boxes*), introns (*solid lines*), splicing options (*dashed lines*), cryptic 5' splice site (*c5'ss*) in exon 5 and poly(A) sites (*pA*). (B) Splice variants of the *igf-1* gene. Exons 1 and 2 encode for the sequence that determines the class of the protein. The mature IGF-1 is encoded by exons 3 and 4, and the three different E-peptides in humans constitute the carboxy-teminal end, and the three different E-peptides

are produced by the alternative splicing at the 3' end of the gene involving exons 4, 5 and 6. (C) Of the amino acid sequences of the three E-peptides, the first 16 common peptides deriving from exon 4 are in green and the different Ea, Eb and Ec amino acids are in red.

THE PROCESSING OF THE IGF-1 PRE-PRO-HORMONE

The translation of the *igf-1* gene gives rise to an immature IGF-1 peptide, the pre-pro-IGF-1. This precursor of the mature IGF-1, contains a signal peptide at the 5' end of the gene, the mature IGF-1 and a C-terminal E-peptide extension at the 3' end. Pre-pro-IGF-1 is subject to numerous post-translational modifications leading to mature peptide production composed of four domains and 70 amino acids. The mature sequence is highly conserved among primate species, whereas it has been shown that the sequences of both the signal peptides and the E-peptides are less strongly conserved compared to the mature IGF-1 peptide [11].

The first cleavage leads to N-terminal signal peptide removal by intracellular serine proteases facilitating the passage of the polypeptide into the endoplasmic reticulum. The resulting molecule is the pro-IGF-1 composed of the mature IGF-1 plus the E-peptide.

The pro-IGF-1 can be subject to additional processing prior to secretion, including the cleavage of the carboxy-terminal E domain resulting in the release of free mature IGF-1 and E-peptide [24] (Fig. 2). All the classes of pro-IGF-1 contain a highly conserved and unique pentabasic motif K^{65}-X-X-K^{68}-X-X-R^{71}-XX-R^{74}-X-X-R^{77}. All the E-peptides begin with amino acid 71; thus, cleavage occurs at Arg^{71} [17]. Proproteins can be processed at this specific motif by serine protease from the subtilisin-related proprotein convertase family (SPCs), a major family of endoproteolytic processing enzymes of the secretory pathway in mammals. Seven mammalian PCs have been identified, namely, PC1, PC2, furin, PC4, PC5, PACE4 and PC7, and they have been proposed as predictors of general cleavage sites. Furin appears to have a more rigorous specificity recognizing sites that contain the sequence motif R-X-[R/K]-R, whereas R-X-X-R is its minimal cleavage sequence [25].

Because SPCs are located in the secretory pathway, the process that leads to the formation of mature IGF-1 by the cleavage of the E-peptides has been shown to occur intracellularly, as expected for intracellular convertases, such as furin [24]. However, the

unprocessed pro-IGF-1 can be secreted and has been detected in conditioned media and in vivo serum [26]. This finding shows that the E domains are not cleaved intracellularly and suggests the presence of potential proprotein convertases that could process pro-IGF-1 extracellularly [27].

There is evidence of possible candidate proteases that could release the mature peptide outside of the cell when needed. The proprotein convertase subtilisin/kexin type 6, commonly known as PACE4, is expressed constitutively in muscle cells [28] and can be found in the Golgi as well as extracellularly; hence, it is a likely candidate to cleave pro-IGF-1 in both areas performing the same intracellular reaction [24].

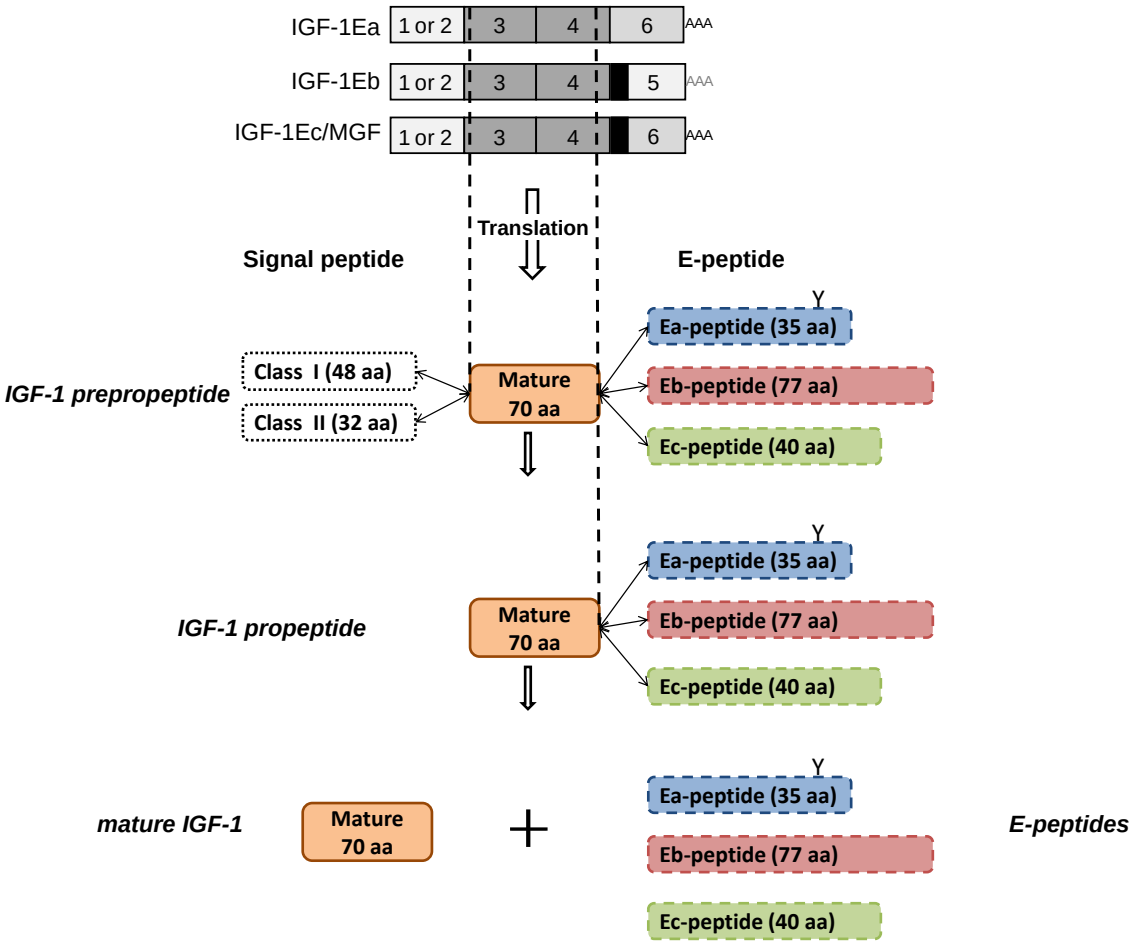


Figure 2. Processing of IGF-1 leading to the mature peptide. The *igf-1* gene is translated into the pre-pro-IGF-1, which contains a signal peptide, mature IGF-1 and a C-terminal E-peptide extension. During translation the N-terminal signal peptide is removed and the resulting molecule is the Pro-IGF-1. An additional protease cleavage separates the mature IGF-1 from the E-peptides. The Y represents the glycosylation site present within the Ea-peptide.

In addition, the secretion of unprocessed pro-IGF-1Ea isoform, both glycosylated and nonglycosylated, has been reported [27, 29].

In rodents there are two potential N-glycosylation sites at Asn^{92} and Asn^{100}, whereas the human Ea domain contains one N-linked glycosylation site at Asn^{92} [30] based on the consensus sequence Asn-X-Ser/Thr, where X represents any encoded amino acid except proline [31]. This glycosylation site is not present in the human Eb- and Ec-peptides because of the reading frame shift caused by the insertion of exon 5 in these two IGF-1 isoforms.

Considering the unique role of glycosylation in the protein biosynthesis process [31], it is possible that Ea-peptide glycosylation might play a role in interactions regarding the regulation of the bioavailability of the different species of this IGF-1 isoform (i.e., pro-IGF-1Ea, mature IGF-1 or Ea-peptide). Although glycosylation has been shown to be a critical step in the protein biosynthesis process, its significance to IGF-1 function has yet to be determined. The presence of a human N-linked glycosylation site only in Ea-peptide and not in the Ec and Eb domains might play a role in the regulation of the bioavailability of this isoform relating to its secretion and stability and possibly reflecting a different biological role of the IGF-1Ea isoform [27].

It is traditionally believed that IGF-1 post-transcriptional processing leads to the formation of the mature protein, which is the main mediator of IGF-1 actions binding IGF-1R. On the contrary, the complexity introduced by the post-transcriptional regulation and post-translational modification of the *igf-1* gene leads to the production of different IGF-1 forms that can be secreted. In particular, three forms of the IGF-1 protein could exist in the extracellular environment: mature IGF-1, non-glycosylated pro-IGF-1, and glycosylated pro-IGF-1 [27]. To date, however, it remains to be determined whether pro-IGF-1 is bioactive or simply an inactive precursor of mature IGF-1 and/or E-peptide [32].

THE BIOLOGICAL ACTIVITY OF PRO-IGF-1 AND E-PEPTIDES

The biological significance of each IGF-1 splice variant is currently unknown, and the physiological and molecular mechanisms that regulate their expression and their circulating levels are unclear [2]. It is generally assumed that the biological actions of

IGF-1 are inferred through the mature peptide, whereas different biological effects have been reported for the different IGF-1 pro-forms or for their E-peptides exogenously administrated or over-expressed in various model systems. In particular, recent studies in humans have shown a differential expression profile of the splice variants in response to various conditions and pathologies, such as skeletal muscle damage [19], endometriosis [21], prostate [20], cervix [33] and colorectal cancer [34]. Therefore, it is not just the mature IGF-1 that possesses bioactivity, and differential expression of *igf-1* gene could indicate distinct regulatory mechanisms and biological roles of the different pro-IGF-1 forms, implying that the E-peptides may promote biological effects.

A divergent action of the IGF-1 isoforms has been reported after viral-mediated expression in mouse skeletal muscle of the two different rodent pro-IGF-1 forms, IGF-1Ea and IGF-1Eb, as well as the mature IGF-1 without the E-peptides. Interestingly, it has been shown that overexpression of mature IGF-1 in skeletal muscle does not promote muscle hypertrophy in young mice, whereas both the pro-IGF-1 forms caused hypertrophy. These results suggest that the pro-IGF-1 forms are required for IGF-1 action leading to muscle hypertrophy, and the presence of E-peptides is necessary for IGF-1 action in muscle [35]. Although the two splice variants increased phosphorylation of IGF-1R, the murine IGF-1Ea overexpression resulted in increased Akt phosphorylation only, whereas the overexpression of murine IGF-1Eb in skeletal muscle activated both the PI3K/Akt and MAPK pathways [36]. These results suggest variable IGF-1 isoform–specific actions on the signaling pathways involved after IGF-1R phosphorylation.

Considering that IGF-1 isoforms differ only in terms of E-peptides, it has been suggested that E-peptides of human IGF-1 precursors may act as independent growth factors with their own biological activity [37]. Curiously, few studies have explored the activity of the Ea-peptide, whereas many investigations have focused on the other two peptides. Since it is more highly expressed than other isoforms, the Ea-peptide probably has an essential biological function also because its sequence is highly conserved in many species, while the other splice variants diverge within primates [11].

Initial studies focused on the role of human Eb-peptides and reported mitogenic activity in human bronchial epithelial cells as a result of exposure to a specific region of the human Eb domain. Furthermore, it was observed that this peptide could still induce proliferation even after IGF-1R neutralization with a specific antibody [10]. It was first suggested that the Eb-peptide mediates its effect through a specific receptor. Conversely,

it was found that the Eb-peptide was not cleaved and the pro-IGF-1Eb was not secreted but was accumulated in the nucleolus [38]. Thus, this human IGF-1Eb pro-peptide may have biological roles that are independent of the mature IGF-1 effects.

Interestingly, the role of the Ec-peptide has been the focus of several studies, mainly because of its action in skeletal muscle. Studies began at the end of the 1990s in the Goldspink Laboratory and revealed that IGF-1Ea was the only isoform expressed in resting rabbit muscles, whereas exon 5 inclusive transcripts were found to increase in rabbit muscles subjected to stretch and electrical stimulation. [39]. It was suggested that the IGF-1Eb splice variant, corresponding to the human IGF-1Ec, was responsible for the stretch induced hypertrophy. It was named Mechano-Growth Factor (MGF) because it was identified in muscle tissue in response to mechanical damage and to distinguish it from the liver forms of IGF-1 [40].

Numerous studies have investigated the role of MGF in muscle repair and survival, leading to the so-called "MGF hypothesis" (Fig. 3). The theory is based on the following findings. After skeletal muscle injury, it was found that both IGF-1Ea and MGF were produced with differential expression regulation and different time course. In particular, there was a transient increase in the splice variant containing the exon 5, which corresponds to IGF-1Eb in rodents and IGF-1Ec in humans. After some days, levels of this transcript decreased, and an up-regulation of the IGF-1Ea splice variant associated with the decline of the IGF-1Ec/MGF mRNA levels was found [23, 41]. Because of this varied expression regulation, distinct roles of these isoforms in muscle remodelling were suggested. The temporal expression of IGF-1Ec/MGF was postulated to be responsible for activating quiescent satellite cells after muscle damage and promoting myoblast proliferation. In contrast, the increase in IGF-1Ea expression appeared to correlate with myoblast differentiation and to promote the fusion of myogenic cells leading to tissue repair [23, 41]. However, it should be noted that the expression levels of MGF are normally vastly lower than those of IGF-1Ea mRNA in skeletal muscle and in other extra-hepatic tissues and the absolute mRNA levels of IGF-1Ea were always 10-fold greater than the MGF transcript levels, even when those transcript levels were elevated. In addition, there is no evidence that the mature IGF-1 and IGF-1Ea are not present in elevated amounts even during the early phase of muscle repair, when MGF is predicted to be up-regulated. Taking all these data into account, the MGF hypothesis should be viewed with extreme caution [42, 43].

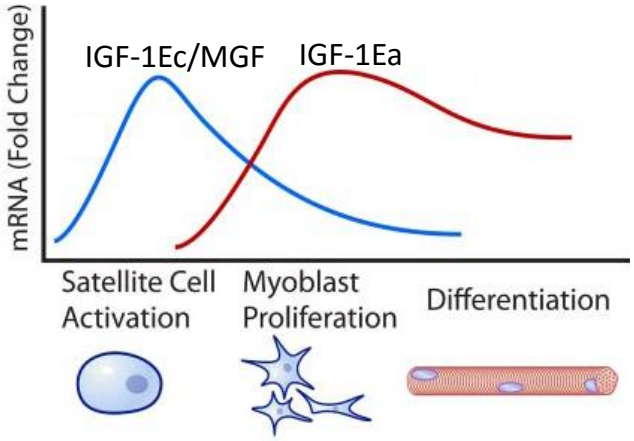


Figure 3. Schematic representation of the MGF hypothesis. The MGF hypothesis suggests that after muscle injury or exercise, the *igf-1* gene transient increases the IGF-1Eb splice variant in rabbit muscles, the so-called MGF. This specific isoform is postulated to be responsible for activating quiescent satellite cells to enter the cell cycle and become mononucleated myoblasts. Thus, MGF promotes myoblast proliferation. On the contrary, during the myoblast proliferative stage, splicing is increasingly shifted towards the IGF-1Ea splice variant, which promotes myoblast proliferation and differentiation into multinucleated fibers called myotubes. The MGF hypothesis therefore suggests that MGF is responsible for satellite cell activation and myoblast proliferation, whereas IGF-1Ea is responsible for myoblast differentiation. Although the graph shows cellular repair events that coincide with the postulated IGF-1 splice variant mRNA levels, it does not reflect the absolute mRNA levels of the two isoforms. In particular, it is essential to note that levels of IGF-1Eb transcripts are far lower than those of IGF-1Ea mRNA. All the mRNA data extrapolated should be examined with extreme caution [42].

Recently, actions regarding the bioactivity of E-peptides have been established. In particular, the biological actions of synthetic E-peptides, corresponding to the rodent Ea and Eb sequences, were compared to test the effects of these E-peptides on IGF-1R signaling. To determine whether E-peptides activate IGF-1R directly, the treatment of E-peptides alone or with mature IGF-1 in the activation of IGF-1R was tested. It was shown that E-peptides do not induce IGF-1R phosphorylation directly, but they amplify IGF-1R activation in an IGF-1-dependent manner but not when IGF-1R was inhibited. In addition, when myoblasts were treated with both mature IGF-1 and E-peptides, there was an increase in the phosphorylation of ERK1/2, but not of phospho-AKT. It was therefore proposed that E-peptides might modulate IGF-1 signaling by modulating IGF-1R downstream signaling [44].

The biological activity of pro-IGF-1 has been investigated in a recent study on murine skeletal muscle cells, particularly focused on the IGF-1Ea isoform. It was shown that the untreated cells secreted IGF-1 predominantly in the form of pro-IGF-1Ea. Moreover, both the glycosylated and non-glycosylated forms were present at high levels, but only a small portion of mature IGF-1 was present. It was shown that pro-IGF-1Ea could activate IGF-1R as well as mature IGF-1, and pro-IGF-1Ea was more potent than the mature peptide by about 20% of receptor phosphorylation when compared in an IGF-1R activation assay. On the contrary, glycosylated pro-IGF-1Ea was less efficient at receptor activation than pro-IGF-1Ea and mature IGF-1 was about 2-fold less potent than mature IGF-1. Because of its differential ability to activate IGF-1R, it was suggested that glycosylated pro-IGF-1Ea might serve as a reservoir for IGF-1, which can be stored until needed [27].

In this regard, it is known that the extracellular matrix (ECM) interacts with a range of growth factors and cytokines likely mediated by positively charged amino acid sequence motifs present in these peptides [45]. The E-peptides contain a high proportion of basic amino acids conferring a high positive charge at physiological pH and therefore pro-IGF-1 might bind to negatively charged molecules in the ECM. It has been shown that both mouse pro-IGF-1Ea and pro-IGF-1Eb bind ECM with significantly higher affinity than does mature IGF-1. Therefore, the C-terminal IGF-1 E-peptides could function by tethering pro-IGF-1 to the ECM, which presumably plays a biological role in retaining high local concentrations of IGF-1 in tissues for subsequent cleavage and receptor activation [46].

It is generally accepted that the post-translational processing of IGF-1 leads to the formation of mature protein, which is the main mediator of the IGF-1 action activating the IGF-1R. In contrast, collectively these studies suggest that E-peptides are actually translated and secreted and exist as part of pro-IGF-1. To date, however, the biological significance of the IGF-1 isoforms remains unclear.

Although much research has aimed to distinguish the E-peptides activity from that of mature peptides, it has been suggested that the E-peptides possess IGF-1 dependent activity. Recent evidence supports this hypothesis, suggesting that the E-peptides may modulate IGF-1 activity through multiple mechanisms. Thus, the retention of these sequences could control the bioavailability of IGF-1 by altering the IGF-1secretion or its association with IGFBs, modulating its power in receptor activation and improving its stabilization and localization in tissues.

IGF-1 AND BINDING PROTEINS (IGFBPS)

A family of specific proteins modulates the bioavailability and the biological activity of circulating IGF-1: the binding proteins (IGFBPs) (Fig. 4). At least six IGFBPs have been identified and well characterized [47, 48] as well as nine IGFBP-related proteins (IGFBP-rPs) [49]. These proteins belong to a family that shares the same cysteine pattern in the amino-terminal and carboxy-terminal, which is indispensable in the binding affinity of the IGF-1. Most of the circulating IGF-1 is in the form of a ternary complex of 150 KDa composed of IGFBP-3, the most abundant IGFBPs in blood, and the ALS subunit, a glycoprotein acid-labile [50]. This complex does not pass through the endothelium and acts as an inactive IGF-1 reserve, accounting for 75-80% of the total carrying capacity. It is hypothesized that binding with IGFBPs increases the half-life of IGF-1. In fact, the half-life of free IGF-1 is less than 15 minutes, whereas the ternary complex persists in circulation for a half-life of about 12-14 hours protecting the growth factor from proteolytic degradation and modulating peptide interaction with its receptor, increasing or decreasing the binding affinity [51]. The formation of this ternary complex results in most of the IGF-1 in the blood representing a stable IGF-1 reserve, while the free IGF-1 concentration in normal subjects is less than 1% compared to that of total IGF-1. The GH stimulates the secretion of IGFBP-3 and ALS and this helps to stabilize IGF-1 levels. IGFBP-3 goes through proteolysis during various catabolic processes or dysmetabolic conditions such as diabetes, and as a result, with a lower concentration of this binding protein, IGF-1 will tend to be more degraded [52]. Additionally, IGFBPs compete with receptor binding and normally have higher binding affinity to IGF-1 than the receptor does. Consequently, the binding of IGFBPs to IGF-1 prevents the ligand from interacting with the receptor and suppresses the growth factor actions [51]. The exact mechanism through which IGF-1 is released from the ternary complex into the tissue is not completely known; however, it seems that an important role is played by proteases that degrade IGFBPs by allowing the release of IGF-1 and the binding to its receptor.

It is known that IGFBP-2 is the second most abundant IGFBP. This IGFBP is not linked to ALS, and the IGF-1/IGFBP-2 complex has a rather short half-life of about 90 minutes. The serum IGFBP-2 is unsaturated and represents a reservoir capable of binding and thus carrying IGF-1. In contrast, IGFBP-1 carries only a small percentage of IGF-1. Like IGFBP-2, IGFBP-1 is generally unsaturated, it represents a potential regulator of free

IGF-1 level. The expression of this IGFBP is inhibited by insulin, whereas in fasting conditions there is a four- to five-fold increase in their expression. Following eating, the levels of these binding proteins decrease rapidly resulting in increased availability of free IGF-1 in peripheral tissues [52].

The last three binding proteins, IGFBP 4-5-6, are present in low concentrations and appear to be less important in controlling free IGF-1 blood concentration. The IGFBPs are mainly produced in the liver, and their concentrations are modulated in response to GH and by changes in nutrition, but peripheral tissues and interstitial fluids also synthesize them, which often involve additional IGF-1 transport.

Although the IGFBP family was recently expanded to include nine IGFBP-rPs that can bind IGF-1 and IGF2 [49], some investigators have challenged their inclusion due to the absence of clear phylogenetic relationships between the IGFBP-rPs and the IGFBPs [53] and the limited understanding of IGFBP-rP function [54].

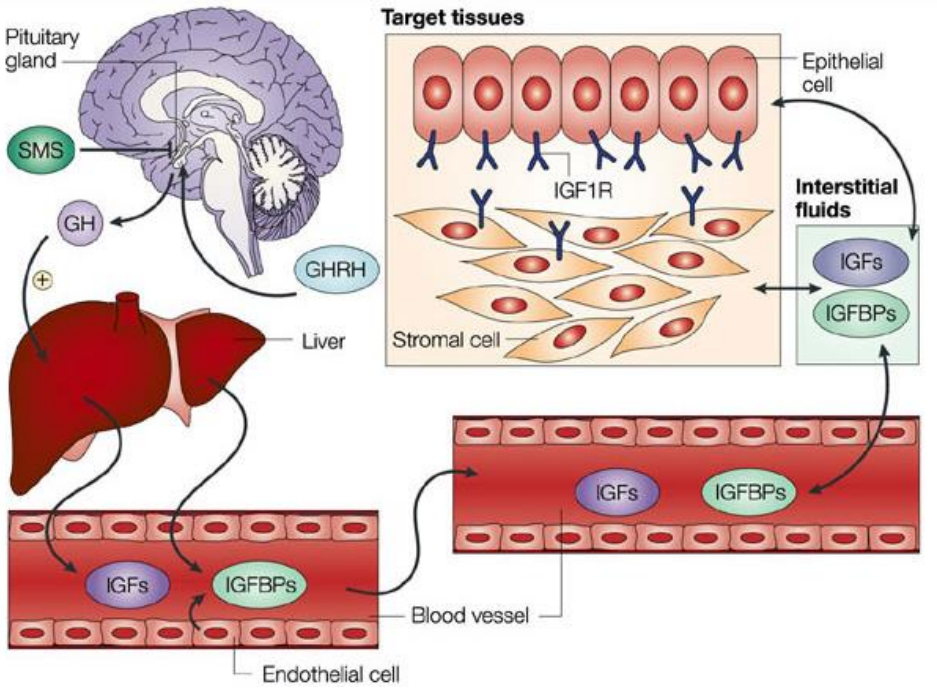


Figure 4. Regulation of circulating IGF-1 levels. Circulating IGF-1 is mostly produced by the liver acting as an endocrine factor and by extra hepatic tissues acting as autocrine and paracrine mechanisms. Growth hormone, which is produced in the pituitary gland under the control of hypothalamic factors, is the main hormone that regulates IGF-1 and IGFBPs production in the liver. The bioavailability of IGF-1 is influenced by the presence of IGFBPs, which are found in circulation and in extravascular fluid, and modulate the interactions between IGF-1 and the receptors present on the cell surfaces [55].

THE IGF-1 RECEPTOR (IGF-1R) AND SIGNALING PATHWAYS

IGF-1 mediates its biological actions on cell proliferation, differentiation and survival by binding to specific receptors present on cell surfaces. IGF-1 can interact with different receptors such as IGF-1R (or type 1 IGF receptor) and IGF-2R (or type 2 IGF receptor), insulin receptor (IR), and some atypical receptors such as the hybrid IR/IGF-1R [56]. Mature IGF-1, which is responsible for binding to the receptors, binds with high affinity to IGF-1R, which has a high degree of homology to the insulin receptor [57]. It is also able to interact with lower affinity IGF-2R, which has been shown to be identical to the cation–independent mannose 6-phosphate receptor and to IR. The IGF-1 receptor pathway shares multiple intracellular mediators with the insulin-signaling cascade stimulating glucose intake and protein synthesis in skeletal muscle. Many tissues, including skeletal muscle, express hybrid receptors, but the functional importance of these receptors remains poorly understood.

The tyrosine-kinase receptor IGF-1R is expressed in many types of cells and is a key mediator of cell growth and proliferation. The IGF-1R is a hetero-tetramic protein composed of two extracellular α-subunits specific for binding to the hormone and two transmembrane β-subunits containing the tyrosine kinase domain activity with a cluster presenting three tyrosine residues at positions 1131, 1135, and 1136 [1].

The binding of circulating IGF-1 to the cysteine-rich domain contained in the α-subunits of the receptor causes a structural rearrangement in the transmembrane β-subunits. This, in turn, induces the activation of tyrosine kinase activity and leads to the autophosphorylation of the cytoplasmic tyrosine kinase domain of the receptor, as one kinase domain phosphorylates the other.

These autophosphorylation events and conformational changes permit unrestricted access for a variety of protein substrates, including members of the insulin receptor substrate (IRS) proteins, whose function is to activate a complex signal transduction network.

Transduction involves the PI3K and AKT pathways, leading to protein synthesis, cell survival and inhibition of apoptosis, as well as the pathway of MAP kinases that stimulates cell proliferation and differentiation (Fig. 5).

Following phosphorylation, the IRS interacts with the SH2 domains of the PI3K (phosphoinositide 3-kinase) cytoplasmic protein, which actively catalyzes the phosphorylation of PIP2 (phosphatidylinositol 4,5-bisphosphate) leading to the synthesis

of PIP3 (phosphatidylinositol 3,4,5-triphosphate). The PIP3 accumulates at high concentrations and can recruit and activate PDK-1 (phosphoinositide-dependent kinases). Then PDK-1 phosphorylates another protein kinase: the AKT at Thr308 residue [58].

The activated AKT has a variety of substrates that are important to bone and muscle. It increases protein synthesis, promoting the activation of mTOR, which phosphorylates other protein substrates: p70S6K and 4E-BP (eukaryotic initiation factor 4E-binding protein) [6].

An additional pathway activated by AKT is the inhibition of GSK3 (glycogen synthase kinase-3), which results from its phosphorylation in an N-terminal serine residue. In response to IGF-1, GSK3 inhibition promotes dephosphorylation and activation of glycogen synthase, contributing to the stimulation of glycogen synthesis. Activated AKT also phosphorylates PKC (kinase-C protein). The PKC, together with AKT, increase the cell's glucose intake by facilitating the translocation of GLUT4, glucose transporters, from the intracellular vesicle to the membrane.

The AKT pathways plays a critical role in apoptosis by inhibiting BAD, which is a protein involved in the apoptosis process. When BAD proteins are not phosphorylated, they remain on the mitochondrial membrane and interact with BCL2 (B-Cell Lymphoma-2) preventing its anti-apoptotic action. When, on the other hand, BAD is phosphorylated by AKT, it is associated with a cytoplasmic protein, and is unable to interfere with the action of BCL2. In addition, AKT phosphorylates FOXO family members, promoting the export from nucleus to cytoplasm, thus reducing the induction of genes such as atrogin-1 and MuRF1, which are ubiquitin ligases involved in protein degradation [59].

Activated AKT also phosphorylates several pro-apoptotic members of the forkhead family (transcription factor), FKHRL1, FKHR, and prevents their activity. The actions of AKT diminish the expression of FasL (Fas Ligand), thus decreasing Fas-mediated apoptosis. In addition to inhibition of pro-apoptotic transcription factors, AKT activity also increases levels of anti-apoptotic proteins, including BCL2 and BCL-X and various adhesion molecules of the extracellular matrix. Activity induced by AKT also involves the expression of the anti-apoptotic transcription factor NF-kB, which enters into the nucleus and activates the transcription of anti-apoptotic genes.

The activation of the PI3K/AKT pathway leads to the transduction of multiple IGF-1 effects, such as increased glucose transport and inhibition of apoptosis through the activation of different proteins.

Another pathway activated by IGF-1R/IRS1 is the MAP kinase pathway, which is more involved in cell proliferation and migration. In this signaling, the activated IRS interacts with Shc that binds to the SH2 domain of Grb2, which in turn forms a complex with Sos, a guanine nucleotide exchange factor. This leads to activation of the small G-protein Ras and continues with a sequence of cascade phosphorylation. Subsequently, Ras activates the protein serine kinase Raf, which phosphorylates and activates MEK, leading to the phosphorylation and activation of ERK1/2 (MAPK). This complex results in the translocation of ERK1/2 to the nucleus, which phosphorylates and activates transcription factors such as elk-1 and c-jun, stimulating cell proliferation and cell cycle progression. These nuclear transcriptional factors lead to increase cyclin D1 and reduced p21 and p27 expression, stimulating cell cycle progression from G1 to S and promoting proliferation.

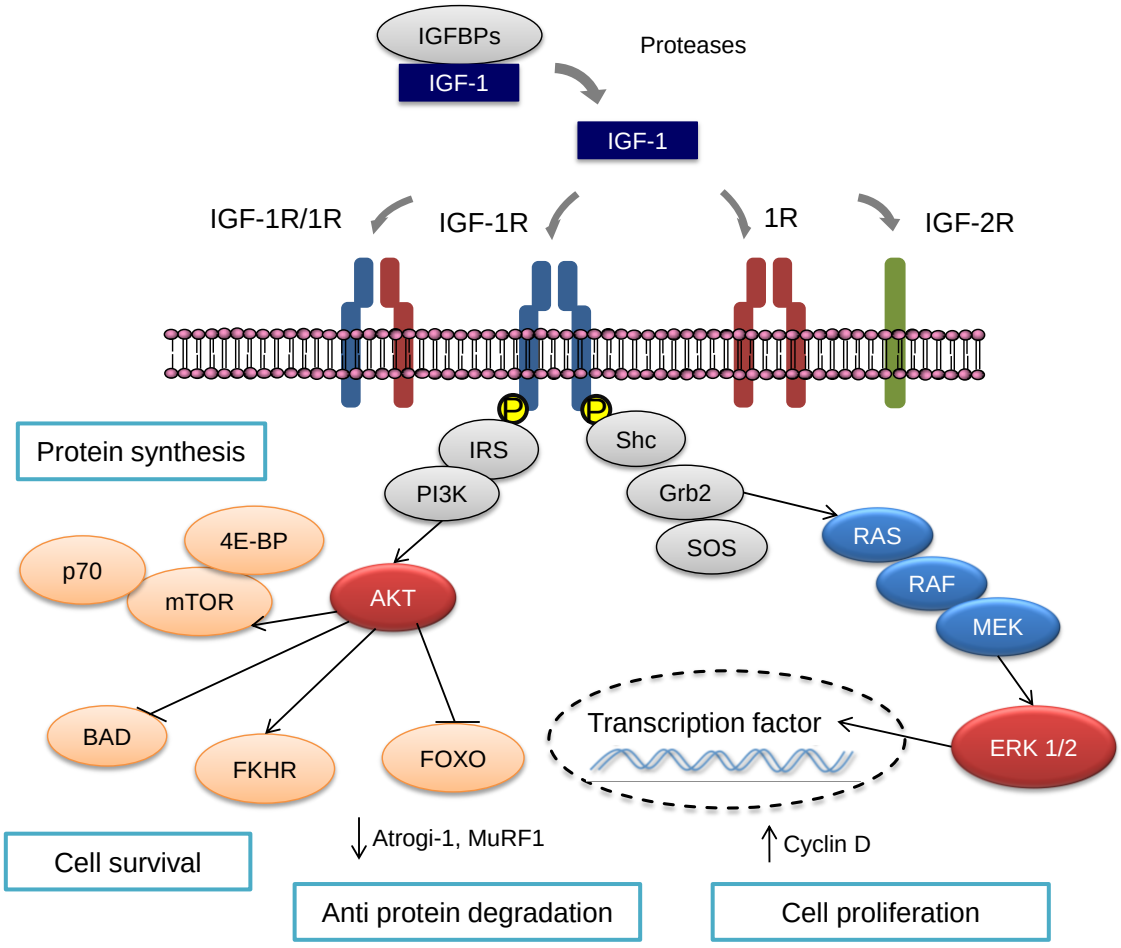


Figure 5. IGF-1 intracellular signaling. The IGF-1R comprises two extracellular α-subunits and two transmembrane β-subunits containing the tyrosine kinase domain. The binding of circulating IGF-1 induces the activation of tyrosine kinase activity of the cytoplasmic portion of the β subunits, forming binding sites for a number of signaling molecules. Transduction involves the activation of the PI3K and AKT pathway.

This leads to protein synthesis through the activation of mTOR, which stimulates p70sk6, cell survival and inhibition of apoptosis phosphorylating BAD and FOXO. When phosphorylated, FOXO is prevented from entering the nucleus and stimulating various ubiquitin ligases such as atrogin-1/MAFbx. The Shc/Grb2/SOS complex activates the MAP kinase pathway leading to cell proliferation by activating ERK1/2, which enter the nucleus to activate various transcription factors.

The signaling of IGF-1R plays an important role in both physiological and pathophysiological conditions. Under normal physiological conditions, IGF-1 increases the synthesis of DNA and proteins in cardiomyocytes, promotes myofibrenia development, and is necessary for cellular entry into the S phase. An important role is played by IGF-1 in the development of the hypertrophic response, when the expression of contractile proteins such as actin, myosin and troponin increases. Through its tyrosine kinase, with PI3K and MAPK signaling pathways, IGF-1R can also reduce the risk of heart failure by preventing apoptosis. In addition, IGF protects cells from apoptosis induced by a variety of conditions, including chemotherapy and the expression of oncogenes.

The ability to regulate apoptosis may have an impact on various types of serious illnesses in humans, including several human cancers. Hence, it is of crucial importance to gain a fuller understanding of the different activity of the pro-IGF-1s on the IGF-1R intracellular pathways and subsequently to develop strategies for cancer prevention.

Chapter 2 will focus on the results of our study indicating how IGF-1 pro-hormones can activate the IGF-1R independently of the mature IGF-1 form. These results underline the importance of an accurate assessment of the presence of IGF-1 pro-forms within the physio-pathologic context under study.

IGF-1 AND CANCER

IGF-1 is a cell growth factor that stimulates proliferation, differentiation and inhibits cellular apoptosis, all key factors associated with tumor. The circulating levels of IGF-1 were evaluated in different pathological conditions. Growth factor is the main regulator of IGF-1 plasma levels, whereas transport proteins, IGFBPs, which bind most of the circulating IGF-1, regulate bioavailability. Studies have revealed that high level of IGF-1

or altered levels of IGFBPs are associated with an increased risk of cancers such as lung [60], colon [61], prostate and breast [62]. A prospective study has shown the association with high circulation IGF-1 and lower plasma IGFBP3 concentrations and an increased risk of colorectal adenoma [63]. The cleavage from IGFBPs can results in and increased released of free IGF-1, which can then bind the receptor and promote cell proliferation. Moreover, the overexpression of the IGF-1R is evaluated in different cancer cells [64] and can be implicated in the acquisition of the transformed phenotype [65]. The concentrations of IGF-1 and its transport proteins are extremely variable between individuals and this could affect the distribution of cancer risk in the population.

The IGF-1R can regulate cell-cycle progression that may be important for acquisition of the malignant phenotype. It can promote G1-S transition by increasing cyclin D1 and cyclin-dependent kinase 4 (CDK4) leading to retinoblastoma (RB) phosphorylation and releasing the transcription factor E2F [66]. In addition, it down regulates the transcriptional inhibitor p27 and phosphatases the tumor suppressor PTEN, that has been deregulated in cancer [67]. Studies have shown an increasing expression of IGF-1 or of IGF-1R in different cancer such as breast, lung, thyroid, prostate, glioblastoma, neuroblastoma, meningioma and rhabdomyosarcoma. IGF-1 is also implicated in the development and progression of angiogenesis and can modulate the expression of the vascular endothelial growth factor (VEGF), a potent angiogenic factor [68]. Another IGF-1 action that is implicated in the development of tumor is the inhibition of apoptosis by inhibiting pro-apoptotic proteins such as BAD and inducing the expression of anti-apoptotic proteins such as BCL2. Multiple mechanisms have been shown to modulate tumor cell sensitivity to IGF-1 including increase in IGF-1 synthesis, increased IGF-1R expression, release of proteases that cleave IGFBPs.

Recent studies have investigated the biological action induced by IGF-1 isoforms in tumor tissues. In particular, a differential expression profile of IGF-1 isoforms has been documented in the development and progression of human prostate cancer (PCa). It was shown that the IGF-1Ec isoform is overexpressed in human prostate cancer tissues and in human cancer PC-3 and LNCaP cell lines not only at expression level, but also at protein level. In addition, the transcription of this specific isoform was significantly higher in PCa and in prostatic intraepithelial neoplasia (PIN) than in normal prostate tissues, whereas the normal prostate epithelial cells (HPrEC) did not express IGF-1Ec [20].

A differential expression profile of IGF-1 transcript variants between normal and tumor tissues has also been observed in other in vivo human cancers, such as endometriosis [21], cervix [33], colorectal cancer [34] and in osteosarcoma cells in vitro [69].

The expression analysis of IGF-1 isoforms in eutopic and ectopic endometrium suggest that the IGF-1Ec splice variant may be involved in the pathophysiology of endometriosis. In particular, it was revealed that all the three splice variants were expressed in both eutopic and ectopic endometrium, though it is significantly lower in endometriotic cysts both at the transcript and protein level. Different expression level was also analysed between the glandular cells of the eutopic and the ectopic endometrium, where the first not express any IGF-1 isoforms whereas the second express the IGF-1Ec isoform [21].

In addiction the analysis of the expression profile during human papillomavirus (HPV) dependent cervical carcinogenesis revealed that all the IGF-1 splice variant were up-regulated in pre-cancerous cells and in particular the IGF-1Eb expression was very high in the cancer samples [33]. Furthermore, testing the exogenous administration of the synthetic E-peptide of the pro-IGF-1Ec, consisting on the C-terminal 24 amino acids sequence of the human Ec-peptide, has been revealed in an increase in proliferation not only in PCa, PC-3 and LNCaP cells [20], in endometrial KLE cells [21] and in MG-63 osteosarcoma cells [69]. The stimulation of PCa and LNCaP cells with the synthetic MGF led to a differential activation of intracellular signaling, with greater phosphorylation of ERK 1/2 without affecting the AKT activation, suggesting, suggesting its independent mode of action of this E-peptice [20].

These data suggest the potentially different roles of different pro-IGF-1 in the pathophysiology of all those conditions. Interestingly, IGF-1 splice profile appears to be different between tumors and the state of the disease, showing a differential regulation of its splice variants. The functional meaning of IGF-1 different splice patter and how the pro-peptides are connected to different autocrine, paracrine and endocrine roles in tumor genesis remains to be clarified.

IGF-1 AND BREAST CANCER

The IGF-1 plays a significant role in human physiology, particularly in the development of many tissues, including the mammary gland. In particular IGF-1 is a key mediator of

mammary gland terminal end bud formation and ductal morphogenesis during development [70]. In contrast, many studies have been associated to the aberrant expression of the IGF-1 system in the development of several malignancies including breast cancer (BC). This is the most common malignant neoplasm in women. BC recurrences and survival are influenced by the presence of many prognostic factors, including hormone factors. IGF-1 is mainly expressed by stromal and only rarely by epithelial cells both in normal and BC tissues but its mitogenic actions are mainly expressed in the epithelium [71]. Various studies have been conducted to evaluate whether mature IGF-1 blood levels are associated with the development of BC or recurrence. Many epidemiological and prospective studies have tried to confirm the positive correlation between plasma IGF-1 levels and BC risk. A pooled data analysis of seventeen prospective studies from twelve countries by the Endogenous Hormones and BC Collaborative Group showed a positively association between circulating IGF-1 and BC risk independently by IGFBP3 levels and menopausal status in estrogen-receptor (ER+) positive tumors [72]. This result is also supported by the data analysis from the European Prospective Investigation into Cancer and Nutrition cohort [73]. Furthermore in an Italian cohort study, serum IGF-1 levels have been positively associated with increased disease risk among BRCA gene mutation (hereditary BC) [74].

Numerous studies have investigated the role of IGF-1 levels and altered circulating IGFBPs levels, in particular IGFBP3, in association with BC risk and prognosis. It has been shown that low levels of IGF-1 are associated with improved survival in BC patients [75, 76]. In addition, it has been shown that low levels of IGFBP3 increase the amount of free IGF-1 that is associated with increased mortality. This study also confirmed a better prognosis in low-level IGF-1 BC patients [76].

IGF-1R mediates the mitotic and anti-apoptotic effects of IGF-1 and has been indicated a correlation between IGF-1R expression and disease development [77]. Another study revealed that IGF-1 receptor is over-expressed in about 90% of BC cases and its level are higher in malignant cells than in normal tissue [78]. Moreover, this over-expression has been related with poor prognosis in patients with early BC [79].

In addition, IGF-1R has been involved in the metastatic progression in BC [80] and another study indicated the relation between IGF-1 and tumor progression. It favors the expression of angiogenic factors in synergy with other growth factors such as vascular endothelial growth factor (VEGF) and platelet-derived factor (PDGF). In particular, IGF-

1 promotes tumor growth and lymphatic metastases by inducing VEGF and it has been indicated a relation between high levels of IGF-1, IGFBP3 and VEGF and lymph node metastasis in patients affected by BC [81].

Although much has been investigated regarding serum IGF-1 expression, it was recently revealed by microarray analysis that the increase of IGF-1 mRNA levels within tissue samples was associated with better prognosis [82]. This finding suggests a contradictory role of circulating and tissue IGF-1 and it can be explained by the lack correlation between circulating and tissue IGF-1 levels, thus local IGF-1 expression may be a better marker of tumor compared to circulating IGF-1 levels.

Despite the large amount of studies that have investigated the role of mature IGF-1 in the BC development and recurrences, only few studies have considered the biological action of each IGF-1 isoforms involved in this tumor. Many tissues such as muscle, liver and adipose tissue synthesize IGF-1 specific isoforms, which act locally with autocrine/paracrine mechanisms and play a key role in repairing cellular damage. Not completely clear is the local contribution to the systemic IGF-1 and the potential environmental stimuli that might be involved to favour their local or systemic release. Data on survival of patients with BC indicate that subjects with a high tissue expression of one of the transcripts of IGF-1, IGF-1Ea, have a lower risk of recurrence and mortality than patients with a lower expression of the same transcript [83].

The biological role of different IGF-1 isoforms and how they are related to different autocrine, paracrine and endocrine roles in tumor genesis is not entirely explained. Indeed, the different isoforms correlate with numerous cellular responses, thus knowing which IGF-1 transcriptional variant is involved in BC could help to elucidate the effect of the IGF-1 pool in the BC development and recurrence.

AIMS OF THE THESIS

The primary aim of this Thesis was to investigate the regulation of IGF-1 alternative splicing and the biological role of IGF-1 pro-hormones in human breast cancer (BC).

The study objectives include:

a) The evaluation of the evolutionary pressure within the different coding regions of the mammalian *igf*-1 gene.

b) The comparison of the mechanisms regulating the IGF-1 exon 5 transcription and splicing between human and mouse.

c) The characterization of the IGF-1 pool and their biological activity on BC cell lines.

d) The investigation of the cellular mechanisms controlling IGF-1 pro-hormones production and secretion.

e) The determination of the effects of lifestyle modification program on circulating IGF-1 level in BC patients.

CHAPTER 1

MIR RETROPOSON EXONIZATION PROMOTES EVOLUTIONARY VARIABILITY AND GENERATES SPECIES-SPECIFIC EXPRESSION OF IGF-1 SPLICE VARIANTS

Giosuè Annibalini[1], Pamela Bielli[2], Mauro De Santi[1], Deborah Agostini[1], Michele Guescini[1], Davide Sisti[1], Serena Contarelli[1], Giorgio Brandi[1], Anna Villarini[3], Vilberto Stocchi[1], Claudio Sette[2] and Elena Barbieri[1]^

[1]Department of Biomolecular Sciences, University of Urbino Carlo Bo, 61029 Urbino, Italy.
[2]Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy; Laboratory of Neuroembryology, Fondazione Santa Lucia, 00133 Rome, Italy.
[3]Department of Preventive & Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy.
^IIM, Interuniversity Institute of Myology.

Published in Biochimica Biophysica Acta, Gene Regulatory Mechanisms 1859(5):757-68, 2016.

MIR retroposon exonization promotes evolutionary variability and generates species-specific expression of IGF-1 splice variants

Giosuè Annibalini[1]*[†], Pamela Bielli[2†], Mauro De Santi[1], Deborah Agostini[1], Michele Guescini[1], Davide Sisti[1], Serena Contarelli[1], Giorgio Brandi[1], Anna Villarini[3], Vilberto Stocchi[1], Claudio Sette[2] and Elena Barbieri[1^]

[1]Department of Biomolecular Sciences, University of Urbino Carlo Bo, 61029 Urbino, Italy.
[2]Department of Biomedicine and Prevention, University of Rome Tor Vergata, 00133 Rome, Italy; Laboratory of Neuroembryology, Fondazione Santa Lucia, 00133 Rome, Italy.
[3]Department of Preventive & Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, 20133 Milan, Italy.
^IIM, Interuniversity Institute of Myology

* To whom correspondence should be addressed. Tel:+39 0722-303418 Fax:+39 0722-303401; Email: giosue.annibalini@uniurb.it
[†]Giosuè Annibalini and Pamela Bielli equally contributed to this work.

Abstract

Insulin-like growth factor (IGF-1) -1 is a pleiotropic hormone exerting mitogenic and anti-apoptotic effects. Inclusion or exclusion of exon 5 into the IGF-1 mRNA gives rise to three transcripts, IGF-1Ea, IGF-1Eb and IGF-1Ec, which yield three different C-terminal extensions called Ea, Eb and Ec peptides. The biological significance of the IGF-1 splice variants and how the E-peptides affect the actions of mature IGF-1 are largely unknown. In this study we investigated the origin and conservation of the IGF-1 E-peptides and we compared the pattern of expression of the IGF-1 isoforms *in vivo*, in nine mammalian species, and *in vitro* using human and mouse IGF-1 minigenes.

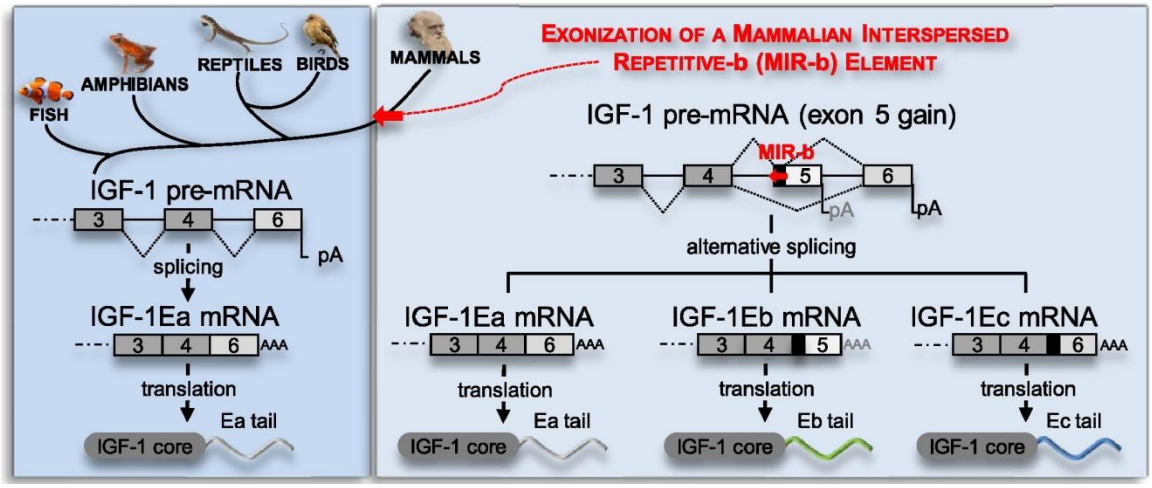
Our analysis showed that only IGF-1Ea is conserved among all vertebrates, whereas IGF-1Eb and IGF-1Ec are an evolutionary novelty originated from the exonization of a mammalian interspersed repetitive-b (MIR-b) element. Both IGF-1Eb and IGF-1Ec mRNAs were constitutively expressed in all mammalian species analyzed but their expression ratio varies greatly among species. Using IGF-1 minigenes we demonstrated that divergence in *cis*-acting regulatory elements between human and mouse conferred species-specific features to the exon 5 region. Finally, the protein-coding sequences of exon 5 showed low rate of synonymous mutations and contain disorder-promoting amino acids, suggesting a regulatory role for these domains.

In conclusion, exonization of a MIR-b element in the *IGF-1* gene determined gain of exon 5 during mammalian evolution. Alternative splicing of this novel exon added new regulatory elements at the mRNA and protein level potentially able to regulate the mature IGF-1 across tissues and species.

Keywords

IGF-1 isoforms, alternative splicing, retroposon exonization, synonymous sites, intrinsically disordered regions

Graphical abstract



1. Introduction

The mammalian insulin-like growth factor (*IGF*) *-1* gene is a single copy gene composed of six exons and five introns, which gives rise to an immature IGF-1 peptide (IGF-1 pro-peptide) containing a signal peptide at the 5' end of the gene, a core region and an E-peptide at the 3' end (Figs. 1A and 1B). Both signal peptide and E-peptide are then removed by protease cleavage to form the 70 amino acid-long mature IGF-1 peptide (IGF-1 core), which displays growth-promoting and metabolic functions.

Four of the six *IGF-1* exons are subjected to alternative splicing (Figs. 1A and 1B) [1]. Exon 1 and exon 2 are mutually exclusive first exons and generate different signal peptides. Transcripts containing exon 1 are referred to as Class I transcripts whereas those containing exon 2 are referred to as Class II transcripts. The Class II IGF-1 knockout mice indicated that class II isoforms are dispensable for fetal and postnatal growth, and the significance of the alternate signal peptides encoded by the first two IGF-1 exons was discussed in [2-3]. At the 3’ end of the gene, alternative splicing yields three different mRNA transcripts, each encoding distinct carboxyl-terminal portions of E-peptide followed by the 3'-untranslated region (3'UTR) (Figs. 1A and 1B). Thus, although alternative splicing generates different precursor peptides, it does not alter the sequence of the mature IGF-1 peptide.

Splicing of exon 4 with exon 6 yields the most common IGF-1Ea variant, which encodes the 35 amino acids-long Ea peptide. The first 16 amino acids of the Ea peptide are encoded by exon 4 and are common in all E-peptides, whereas the remaining 19 amino acids are encoded by exon 6 and are unique to this isoform. The IGF-1Eb variant is produced when exon 4 is spliced with exon 5 and encodes the Eb peptide, which contains the 16 common amino acids and 61 additional amino acids encoded by exon 5. IGF-1Eb transcript terminates in exon 5 and excludes exon 6 from the mRNA, hence it has a completely different 3’UTR compared to IGF-1Ea and IGF-1Ec. The third variant, named IGF-1Ec in humans, is generated by usage of a cryptic 5' splice site (c5’ss) (named IGF633) [4] present in exon 5, which is in turn spliced with exon 6. The c5’ss of *IGF-1* exon 5 deviates from the vertebrate consensus and is commonly used in rodents and rabbits [5-7] but rarely and in a tissue-specific manner in human [1, 4]. Notably, although this variant is named IGF-1Eb in rodents, for clarity we will use the human IGF-1Ec nomenclature throughout this manuscript, regardless of the species. The human Ec

peptide has a predicted length of 40 amino acids, with the common 16 amino acids deriving from the exon 4, 16 from exon 5 and the last 8 amino acids from exon 6. Since expression of IGF-1Ec was linked to mechanical stimuli in muscle and other mechanosensitive cells, this variant is sometime referred to as Mechano Growth Factor (MGF) [1].

Few studies have addressed the regulation of *IGF-1* alternative splicing. In human cells, splicing of exon 5 is regulated by the antagonistic activities of the serine-arginine protein splicing factor-1 (SRSF1) and heterogeneous nuclear ribonucleoproteins A1 (hnRNPA1) [8-9]. SRSF1 was proposed to increase splicing of the IGF-1Eb variant by binding to a purine-rich exonic splicing enhancer (ESE) in exon 5 and preventing the recruitment of hnRNPA1, which functions as a splicing repressor [8-9]. Notably, neither splicing factor was capable of regulating the IGF-1Ec variant in human IGF-1 minigenes [8-9]. The reason of the lack of usage of the c5'ss, and hence for the lack of IGF-1Ec isoform production, reported in these studies has yet to be clarified. Similarly, whether or not the mouse c5'ss, which is closer to the canonical splicing donor consensus, can be recognized in these experimental settings has not been tested. Furthermore, it is currently unknown whether SRSF1 and hnRNP A1 function specifically with the human IGF-1 pre-mRNA or whether they also play a role in *IGF-1* alternative splicing in other species.

To date, the fate and the biological functions of E-peptides are not entirely clear [1, 10-12]. Interestingly, gene structure comparison showed that *IGF-1* exon 6 is conserved in all vertebrates, whereas exon 5 is conserved only among mammals [13-15]. In addition, comparative studies on mammalian *IGF-1* show that both the IGF-1 core and Ea peptide are subjected to strong purifying selection, whereas sequences of the Eb and Ec peptides are more variable [16-17]. In particular it was shown that the ratio of non-synonymous (dN; amino-acid altering) to synonymous (dS; silent) substitutions is about 10-fold higher for Eb and Ec peptides compared to Ea peptide, suggesting that they have no specific roles or, at most, species-specific functions [17]. Indeed, it is usual to view poor protein sequence conservation, i.e. evolving under neutral or nearly neutral conditions, as evidence of reduced functional importance [17-18]. However, Lin and colleagues [19] recently demonstrated that the "dual-coding" DNA sequences of *IGF-1* exon 6, which is translated in two alternative reading frame to give rise to Ea and Ec peptides, showed a very low dS rate suggesting additional sequence constraints beyond those dictated by the amino acid sequence of the Ea and Ec peptides. Notably, the picture of largely neutral

evolution of synonymous substitutions in mammals has been recently challenged [20], showing that selection may constrain not only dN to preserve amino acid sequence, but also dS to preserve regulatory elements in nucleotide sequences, such as ESE, RNA secondary structures and microRNA target sites [19-21]. However, whether the dS drop of the dual-coding exon 6 has contributed to the observed local increase in dN/dS ratio of E-peptides and hence to an inaccurate estimate of evolutionary rate is still unclear.

In this study, we have analyzed the conservation and the mRNA expression pattern of IGF-1 isoforms in different mammalian species and evaluated the evolutionary pressure within the different coding regions of the mammalian *IGF-1* gene. Moreover, we have used human and mouse minigene systems to compare the mechanisms regulating the *IGF-1* exon 5 splicing between these two species.

2. Materials and Methods

2.1 Sequences and databases

Orthologous of the human *IGF-1* gene were obtained from the UCSC Genome Browser Database (http://genome.ucsc.edu; Feb. 2009 GRCh37/hg19) and Ensembl website (http://www.ensembl.org). The "CDS FASTA alignment from multiple alignments" data, derived from the "multiz100way" alignment data prepared from 100 vertebrate genomes [22], were downloaded using the Table Browser tool of the UCSC Genome Browser. Sequences were subsequently realigned using MUSCLE [23] and protein coding sequences from 27 mammalian species were extracted from these alignment datasets. Supplementary file S1 contains the IGF-1 sequence of 27 mammalian species in FASTA format and the neighbor-joining tree used in the present study. Percent nucleotide and amino acid identities between sequences were computed using the CLUSTALW procedure [24]. The splice site scores were obtained for each 5' and 3' ss sequence using the Maximum Entropy scores [25]. For Transposable elements analysis, we used RepeatMasker (http://www.repeatmasker.org) and Repbase annotations [26-28].

2.2 Human and animal tissues

Freshly frozen normal human liver (2 males, mean age 54 +/- 9 years), adipose (2 males and 2 females, mean age 52 +/- 12 years) and muscle (3 males, mean age 25 +/- 6 years) samples were provided by the complex structure of biomarkers (DOSMM) of National Cancer Institute of Milan. The macaque (*Macaca mulatta)* autoptic specimens were kindly provided by Elena Borra (Department of Neuroscience, University of Parma, Parma, Italy). Three 4-year olds macaques were tested in this study (2 female and 1 male). One-month old CD1 female mice (n=3) (*Mus Musculus*), 2-month old Young Sprague Dawley male rat (n=3) (*Rattus norvegicus*) and three 8 month-old New Zealand male rabbits (*Oryctolagus cuniculus*) (Charles River Laboratories, Milan, Italy) were housed with a 12-h light/dark cycle and free access to standard laboratory chow and water. Care and handling were in accordance with the *Guide for the Care and Use of Laboratory Animals* by Ministero della Sanità D.L. 116 (1992) and approved by the university committee for animal experiments. Animals were sacrificed with an overdose of anaesthetics (ketamine in combination with xylazine). Other tissues used in this study were collected at local slaughter during routine meat inspection: pig (*Sus scrofa*) (3 males), cow (*Bos taurus*) (3 females), sheep (*Ovis aries*) (3 female) and goat (*Capra hircus*) (2 males and 1 female) were tested in this study. The age of the animals ranged from 2 to 5 years. All tissues (about 30 mg) were immediately submerged in RNAlater stabilization solution (Qiagen, Milan, Italy) and left at least 10 min at room temperature. Then, tissues were stored at -80°C until RNA extraction.

2.3 RNA extraction and cDNA synthesis

The tissues were removed from RNAlater and transferred into a clean Eppendorf tube where the total RNA was extracted and purified using the Omega Bio-Tek E.Z.N.A.TM Total RNA kit (VWR International s.r.l., Milan, Italy) according to the manufacturer's instructions. The amount and quality of RNA were assessed with DU-640 UV Spectrophotometer (Beckman Coulter). After DNA digestion with DNase I enzyme (Qiagen, Milan, Italy) complementary DNA was synthesized from 1 μg of total RNA using Omniscript RT (Qiagen, Milan, Italy) and random hexamers or anchored oligo-dT primers where specified.

2.4 Real time PCR quantification of human, macaque and mouse IGF-1 isoforms

Serial dilution (1:4) of three recombinant plasmids containing the IGF-1Ea, IGF-1Eb and IGF-1Ec sequences of human and mouse were prepared in order to generate standard curves for plotting CT values against number of molecules. Molecules of each plasmid were calculated using the concentration of each plasmid, Avogadro's constant, the molecular weight of double-stranded DNA and the size of the target amplicon [29]. The copy number of genes was calculated in individual samples using a corresponding reference plasmid cDNA clone at known concentration. The amount of target transcripts was normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH). Percentage of IGF-1 isoforms was calculated as [(mRNA copy number of single isoform/(mRNA copy numbers of IGF1Ea+IGF1Eb+IGF1Ec)*100]. Expression of IGF-1 isoforms was compared using 2-ways analysis of variance with interactions: species and IGF-1 isoforms were used as predictive factors. In order to meet assumption of homoscedasticity percentage were arcsin radq transformed. Post-hoc analysis was performed using Bonferroni correction.

Real-time quantitative PCR was performed with two microliters of cDNA and 300 nM of of each primer in an Applied Biosystems StepOnePlus^{TM} Real Time PCR System using SYBR Select Master Mix (Applied Biosystems, Monza, Italy). The real-time PCR conditions were: 50°C for 2 min, 95°C for 2 min followed by 40 cycles of three-steps at 95°C for 15 sec, 60°C for 15 sec and 72°C for 30 sec. The specificity of the amplification products was confirmed by examining thermal denaturation plots and by sample separation in a 4% DNA agarose gel. The sequences and the annealing positions of the primers used to quantify the IGF-1 isoforms were shown in Supplementary Table S1.

2.5 3' RACE-PCR

The mRNA splicing pattern of IGF-1 was assayed by 3-RACE-PCR using an anchored oligo-dT primer (5'-AAGCAGTGGTATCAACGCAGAGTACT$_{(30)}$NV-3') as reverse transcription primer. 1 μg of total liver RNA was reverse transcribed into cDNA. To amplify the 3'-end IGF-1 variants a reverse primer corresponding to the anchor sequence of the RT primers was used in combination with the following forward primers: human, macaque, rabbit, (5'-CCTCCTCGCATCTCTTCTACCTG-3'); mouse and rat (5'-GCTATGGCTCCAGCATTCG-3'); and pig (5'-CGTGGATGAGTGCTGCTTC-3'). The

PCR reaction were as follows: 95° for 10 min; 35 cycles of three-steps at 95°C for 30 sec, 60°C for 30 sec, 72°C for 1 min and 30 sec followed by a final elongation cycle (72°C for 5 min). The PCR products were loaded on 4.0% agarose gel and DNA fragments were eluted from gel, subcloned and sequenced.

2.6 Conventional RT-PCR

RT-PCR was performed in 50 μl of reaction volume with 4 μl of cDNA, 800 nM of primers and 25 μl of 2X HotStartTaq mix (Qiagen, Milan, Italy). The sequences and the annealing positions of the primers used for RT-PCR were shown in Supplementary Table S1. RT-PCR conditions involved an initial denaturation step at 95°C for 10 min, followed by 35 cycles with denaturation step at 95°C for 30 sec, annealing at 60°C for 30 sec and extension at 72°C for 30 sec. PCR products were size-fractionated by electrophoresis on 4.0% agarose gels and visualized by ethidium bromide staining under UV light. Amplification products were purified with QIAquick gel extraction kit (Qiagen), cloned and sequenced.

2.7 Plasmid constructs

The 5' and 3'end of IGF-1 mouse minigene were amplified using primers #(1,2) and #(3,4), respectively, from C57 mice genomic DNA. After enzymatic digestion, the PCR products were cloned in KpnI/SalI and SalI/NotI restriction sites of pCI vector (Promega). The 5' and 3'end of the IGF-1 human minigene were amplified using primers #(5,6) and #(7,8), respectively, from HeLa cell genomic DNA and cloned in EcoRI/SalI and SalI/NotI restriction sites of pCI vector. The mGA*vs*TG and hTG*vs*GA IGF-1 mutant minigenes were constructed using the mega-primer strategy [30]. The mouse and human IGF-1 5' mutant ends were generated using primers #(1,9,2) or #(5,10,6) respectively. All oligonucleotide sequences are listed in Supplementary Table S1. After enzymatic digestion, the PCR products were subcloned in the corresponding wild-type minigene. PCR reactions were performed using Phusion Hot Start High-Fidelity DNA polymerase (Thermo Scientific) according to manufacturer’s instruction. All plasmids were sequenced and validated.

2.8 Cell cultures, transfections and protein extract preparation

Cell cultures, transfections and sample preparation were carried out by standard methods as previously described [31]. Briefly, human and mouse cell lines were transfected with various combination of vectors as indicated using Lipofectamine 2000 (Invitrogen). Twenty-four hours after transfection, cells were collected for RNA and protein isolation, as previously described [31]. For RNAi experiments cell were transfected twice with 60nM siRNAs using Lipofectamine RNAi Max according to manufacturer's instruction (Invitrogen). Sequences of siRNAs are listed in the Supplemental Table 1.

2.9 Splicing assay and PCR analyses

Twenty-four hours after transfection, cells were collected for RNA extraction using TRIzol (Ambion, Life Technologies) according to the manufacturer's instructions. After DNase digestion (Roche), 1 μg of total RNA was retrotranscribed using M-MLV reverse transcriptase (Promega). cDNA was used as template for conventional PCR reactions (GoTaq G2, Promega) in presence of specific primers and PCR products were analyzed on agarose gel. A minigene-specific forward primer #(11) was used to amplify minigene-derived IGF-1 isoforms. In details, primers #(11,13) were used to amplify mIGF-1Eb isoform; primers #(11,12) to amplify mouse and human IGF-1Ea and IGF-1Ec isoforms; primers #(11,14) to amplify hIGF-1Eb; primers #(11,15) and #(11,16) to amplify total mouse and human IGF-1, respectively. In splicing assay human IGF-1Ea and IGF-1Eb were amplified using primers #(11,12,14) in the same PCR reaction. Amplification of mouse and human endogenous IGF-1 isoforms were performed as above, using primer #(1) or #(5), respectively, in substitution of primer #(11). The sequences and the annealing positions of the primers are showed in Supplementary Table S1.

2.10 IGF-1 evolutionary analysis

The MG94xREV_3x4 codon model with transition/transversion bias correction and nine base frequency parameters was fitted to sequence alignments by using the HyPhy 2.2.3 package [32]. A neighbor-joining tree was built using the IGF-1 core alignment of 27 mammalian species and the tree topology along with the alignment were used as input for

IGF-1Ea, IGF-1Eb and IGF-1Ec sequence analysis (Supplementary file S1). For every branch in a phylogenetic tree, the expected number of synonymous substitutions per synonymous site (dS) and its counterpart for nonsynonymous substitutions (dN) were estimated by using maximum likelihood [33] allowing for independent dN and dS values for each branch (the local parameters option). Mean and 95% confidence interval for each dS, dN and dN-dS values are calculated using bootstrap approach with n=1000 resampling [34].

The web application of DisCons [35] was used to predict the protein disorder on the amino acid level, using IUPred long disorder prediction parameter [36].

3. Results

3.1 *IGF-1* exon 5 is conserved in Mammalia but not in other vertebrates

Previous studies demonstrated that the IGF-1Ea splice variant, which skips exon 5, is the most common E-peptide isoform expressed among vertebrates and is the predominant form in amphibians and birds [13-15]. To trace the evolutionary history of the IGF-1 isoforms that include exon 5 (i.e. IGF-1Eb and IGF-1Ec), we performed BLAST searches in mammalian and non-mammalian draft genomes using the Ensembl database and the UCSC vertebrate genome alignment provided at the UCSC Genome Browser site (Fe. 2009 GRCh37/hg19). ClustalX alignment of *IGF-1* sequences was performed along the exon 5 and exon 6 regions among 23 representative amniota vertebrates (Supplementary Fig. S1). We retrieved orthologs of the exon 5 sequence from all placental mammals (Eutheria), except Guinea pig (*Cavia aperea porcellus)*, probably because of insufficient sequencing coverage. In addition, we identified exon 5 sequences from the marsupial short-tailed opossums (*Didelphidae*) as well as from the monotremata egg-laying platypus (*Prototheria*). By contrast, we failed to identify exon 5 orthologs in all examined non-mammalian vertebrates (fish, amphibians, reptiles and birds), including turkey (*Meleagris gallopavo*), chicken (*Gallus gallus*), zebra finch (*Taeniopygia guttata*) and green anole lizard (*Anolis carolinensis*) (Supplementary Fig. S1). These data show that only exon 6 is conserved in all vertebrates and suggest that the Ea peptide represents the

ancestral IGF-1 E-peptide, whereas the exon 5-encoded Eb and Ec peptides have been acquired during evolution in the mammalian lineage.

3.2 MIR-b retroposon exonization in the *IGF-1* gene explains the origin of the alternatively spliced exon 5 in Mammalia

Transposon exonization is a major source of new exons in higher eukaryotes [37]. This process occurs when transposon-derived intronic sequences are recognized by the spliceosome and are exonized. Using the RepeatMasker web server, we searched for transposon elements in the genomic region encompassing *IGF-1* exon 5. No repetitive sequences were detected using the search tools "abblast", "rmblast" or "cross-match" whereas the recently introduced "nhmmer" search engine, which uses the Dfam database, [27-28], highlighted a transposon belonging to the Mammalian Interspersed Repetitive-b (MIR-b) element (bit score 13.6; E-value 0.0031). The *IGF-1* exon 5 displays about 60% nucleotide identity to 103 nt of MIR-b consensus sequence from Repbase (Fig. 1C). The exonized MIR-b element is inserted in the antisense orientation relative to the *IGF-1* gene. Both the polypyrimidine tract and the *IGF-1* exon 5 splice sites are embedded within the MIR-b sequence (Fig. 1C). The 3'ss of *IGF-1* exon 5 is in close proximity to the 65 nt that form the highly conserved core region of MIR-b while the c5'ss lies within this region (Fig. 1C). This configuration is similar to those described for others antisense MIR-b exons [35-36] and for disease-linked genes containing MIR pseudoexons [38].

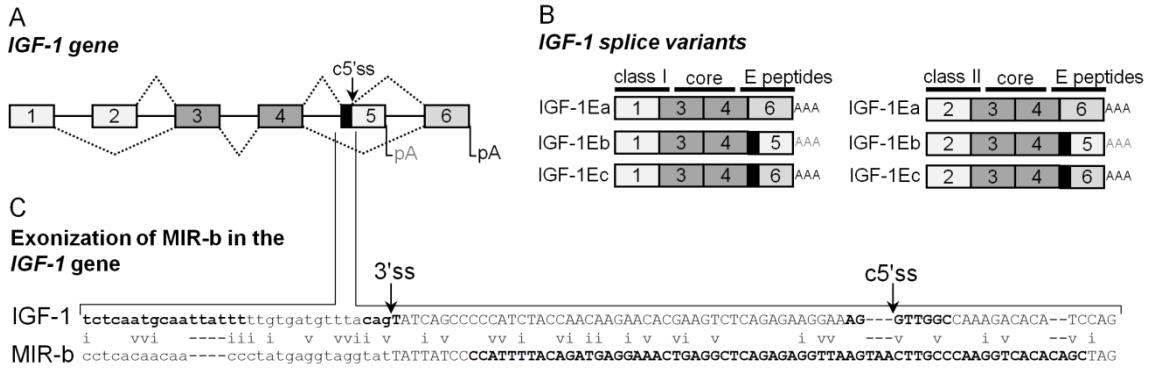


Figure 1. Schematic representation of the *IGF-1* gene (A), its splice variants (B) and the exonization of MIR-b in the *IGF-1* gene (C). (A) Map of the *IGF-1* gene showing exons (*boxes*), introns (*solid lines*), splicing options (*dashed lines*), cryptic 5' splice site (*c5'ss*) in exon 5 and poly(A) sites (*pA*). The graph is not drawn to scale. (B) Splice variants of the *IGF-1* gene; human *IGF-1* NCBI RefSeq transcripts: class I IGF-1Ea (NM_000618), class I IGF-1Eb (NM_001111285), class I IGF-1Ec (NM_001111283) and class II

IGF-1Ea (NM_001111284). (C) Pairwise alignment of the human *IGF-1* gene sequence and the MIR-b inverse consensus sequence created by RepeatMasker using "nhmmer" search engine (intronic nucleotides in lower-case letters, exonic nucleotides in upper-case letters). The polypyrimidine tract and exon 5 splice sites are shown in bold in *IGF-1* sequence; the bold region of MIR-b sequence corresponds to the 65-nt conserved central domain of MIR-b sequence. The middle line shows exact matches (spaces), gaps (dashes) and mismatches (i=transition; v=transversion).

3.3 Comparison of the MIR-derived *IGF-1* exon 5 sequence within mammalian species

Alignment of the *IGF-1* exon 5 sequence originated from exonization of MIR-b element among mammalian orthologous showed a level of nucleotide identity ranging from 56 to 62% between MIR-b consensus sequence and *IGF-1* exon 5 coding for Ec peptide (49 bp in human and 52 bp in other species) (Fig. 2A). This analysis indicates that, after exonization, the exon 5 sequence coding for the Ec peptides has been fairly well conserved among species, even though the percent of identity of nucleotide sequence was lower in rodents and logomorpha than in others species (Fig. 2A). Notably, the 3'ss of exon 5 is conserved and relatively strong, whereas the c5'ss has only a limited match to the consensus, except in mouse and rat (Figs. 2A and 2B). Interestingly, humans and murids showed a marked difference in c5'ss strength. This splice site is very weak in humans; by contrast it appears strong in mouse and rat (Fig. 2B). Further analysis of the *IGF-1* c5'ss strength among 58 mammalian genomic sequence datasets available at the UCSC genome browser showed that most mammals (79%) display medium strength, while 16% have high strength, including all members of the two rodent families Cricetidae: prairie vole (*Microtus ochrogaster*), chinese hamster (*Cricetulus barabensis griseus*), golden hamster (*Mesocricetus auratus*) and Muridae: mouse (*Mus musculus*) and rat (*Rattus norvegicus*) (Supplementary Fig. S2). These analyses suggest that processing of both IGF-1Eb and IGF-1Ec mRNA variants are expected in most mammals, whereas the weak nature of the c5'ss of human exon 5 will likely prevent its recognition by the spliceosome, yielding predominantly the IGF-1Eb isoform. By contrast, the relatively strong c5'ss of exon 5 in the Muridae Family will favour processing of the IGF-1Ec. Therefore, our analyses suggest that, under normal conditions, the IGF-1 alternative splice variants are likely expressed in a species-specific manner.

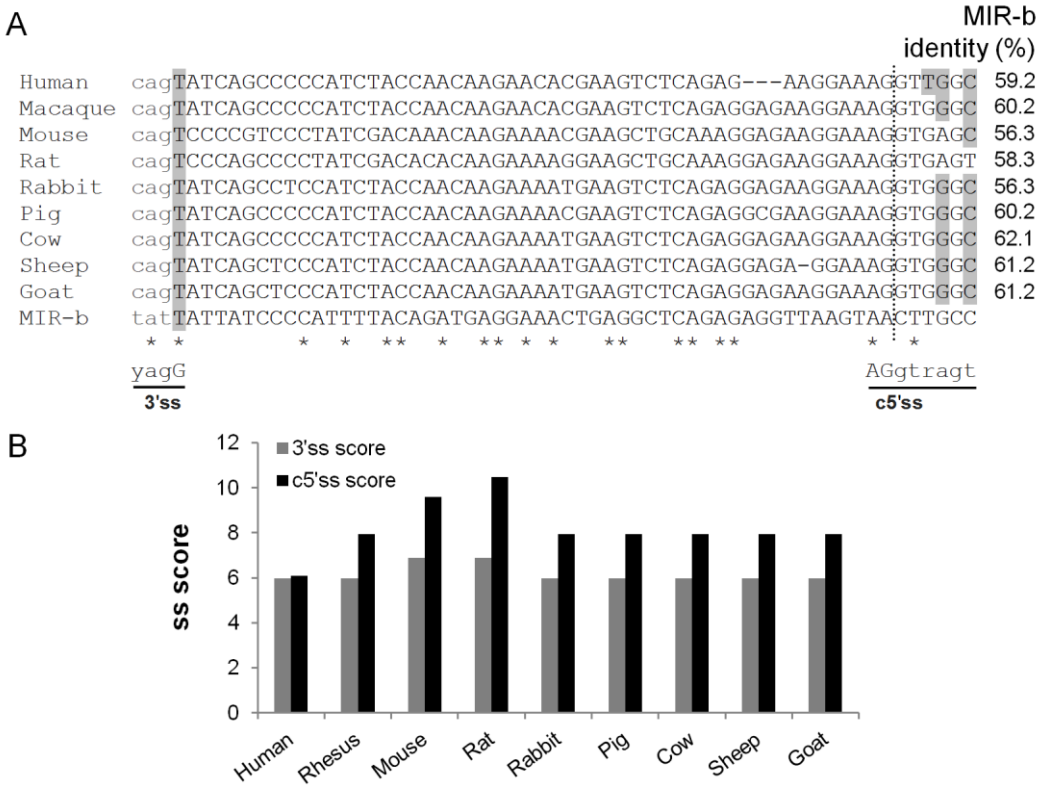


Figure 2. Comparative analyses of exon 5 splice sites among mammals. (A) Multiple alignment of MIR-derived exon 5 of *IGF-1* gene among mammals. Partial intron 4 and exon 5 sequences is shown (intronic nucleotides in lower-case letters, exonic nucleotides in upper-case letters). The nucleotides that did not match the 3'ss and 5'ss consensus sequences are highlighted in grey (y=purine; r=pyrimidine). Nucleotides conservation is marked at the lower edge with asterisks indicating full conservation relative to the MIR-b consensus sequence (lower row). The percentage of nucleotide identity between mammalian sequences and MIR-b consensus sequence is indicated under "MIR-b identity". (B) Splice site score of the 3'ss and c5'ss of *IGF-1* exon 5 was calculated using the MaxEntScan algorithm [25].

3.4 Expression pattern of IGF-1 mRNA isoforms in mammals

To test this hypothesis, we quantified by real-time PCR the IGF-1Ea, IGF-1Eb and IGF-1Ec mRNA variants in skeletal muscle, adipose tissues and liver of human (*Homo sapiens)*, macaque (*Macaca mulatta)* and mouse (*Mus musculus)*. These species represent examples of weak, intermediate and strong exon 5 c5'ss, respectively (Fig. 2B). As predicted by our analysis, the expression profile of these IGF-1 mRNA variants varies among species (Anova test; $p<0.01$) (Fig. 3A). In particular the IGF-1Ea and IGF-1Eb

isoforms are predominant in human skeletal muscle, adipose tissue and liver, whereas the IGF-1Ec isoform accounts only for 1-5% of total IGF-1 mRNAs (Fig. 3A). Conversely, IGF-1Ea and IGF-1Ec isoforms predominate in mouse tissues, whereas the IGF-1Eb isoform was barely detectable (≤1% of total IGF-1 mRNAs) (Fig. 3A), as also reported previously [39]. In macaque (*Macaca mulatta)*, the expression of IGF-1Eb was higher compared to IGF-1Ec in skeletal muscle and adipose tissue, whereas these splice variants were expressed at approximately equivalent level in the liver (Fig. 3A). Noteworthy, the same PCR primers have been used to amplify human and macaque IGF-1 mRNA variants, ruling out differences in amplification efficiencies between species.

To confirm the species-specific expression pattern of IGF-1 isoforms, we performed 3' RACE-PCRs using liver-derived cDNA from human, macaque or mouse. In addition, we analyzed liver tissues from rat (*Rattus norvegicus*; strong c5'ss), rabbit (*Oryctolagus cuniculus*; intermediate c5'ss strength) and pig (*Sus scrofa*; intermediate c5'ss strength). IGF-1Ea, IGF-1Eb and IGF-1Ec mRNA products were amplified in macaque, rabbit and pig liver, which represent examples of intermediate c5'ss (Fig. 3B and Supplementary Figs. S3A, S3B and S3C for the complete nucleotide sequences and putative termination sites). On the contrary, the IGF-1Ec and IGF-1Eb splice variants were not detected in human and Muridae liver, respectively, probably because their expression level is too low for the sensitivity of the 3' RACE-PCR analysis. Notably, human, mouse, rabbit and pig IGF-1 isoforms show multiple termination sites (Fig. 3B and Supplementary Figs. S3A, S3B and S3C). Isoform-specific PCR analysis showed that both IGF-1Eb and IGF-1Ec mRNA variants were expressed in other species characterized by an intermediate c5'ss, such as cow (*Bos taurus*), sheep (*Ovis aries*) and goat (*Capra hircus*) (Supplementary Fig. S3D).

Collectively, these results document that IGF-1Eb is more widely expressed than previously reported and point out a sharp difference in the expression of IGF-1 isoforms between human and mouse.

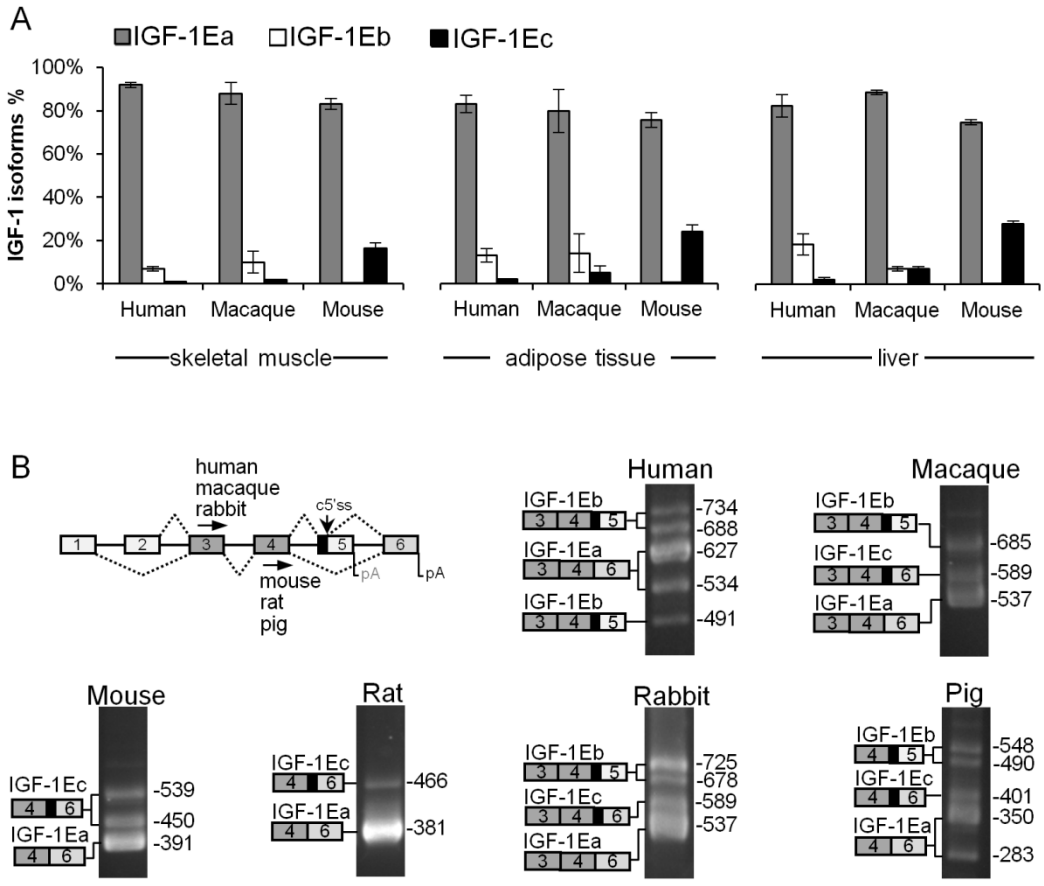


Figure 3. Expression pattern of IGF-1 mRNA isoforms within mammalian species. (A) Quantification of IGF-1Ea, IGF-1Eb and IGF-1Ec isoforms by real time RT-PCR on skeletal muscle, adipose tissue and liver of Human, Macaque and Mouse. Percentage mean (+/-SD) of IGF-1 isoforms was calculated as: [(mRNA of single isoform/(mRNAs of IGF1Ea + IGF1Eb + IGF1Ec)*100]. n=2-4. (B) Agarose gel of 3' RACE-PCR analysis performed in liver tissues from different mammalian species. PCR products were eluted from gel, subcloned and sequenced (See Supplementary Figs. S3A, S3B and S3C). A schematic representation of IGF-1 gene and the relative position of forward primers used in the 3'RACE-PCR are also shown. See Materials and Methods section for universal reverse primer information. Diagrammatic representation of splice variants is given on the left of the gel; their sizes (bp) are indicated on the right.

3.5 The mouse- and human-specific IGF-1 isoforms pattern is not dependent on the cell context

In order to analyze the species-specific splicing of exon 5 in human and mouse, we constructed minigene systems consisting of exon 4, intron 4, exon 5, part of intron 5 and exon 6 of the corresponding *IGF-1* genes (Fig. 4A). Minigene splicing assay performed in four human and three mouse cell lines showed that alternative splicing of the human

minigene (hIGF-1) did not generate the IGF-1Ec variant (Fig. 4B, *left panel* and Supplementary Fig. S4). These results are in line with the barely detectable amounts of this mRNA found by quantitative real-time PCR in human tissues (Fig. 3A). By contrast, the mouse minigene (mIGF-1) yielded mainly the IGF-1Ea and IGF-1Ec variants in all cell types analyzed (Fig. 4B, *right panel* and Supplementary Fig. S4), similarly to what observed in mouse tissues (Fig. 3A), with mostly shifted toward the IGF-1Ec isoform in mouse cell types. We suppose that these little differences might be due to the lack of part of intron 5 and/or of chromatin context, or to the presence of a different promoter in the minigene [40-41]. Collectively, these results indicate that the mouse and human IGF-1 minigenes mostly recapitulate the splicing pattern observed in tissues.

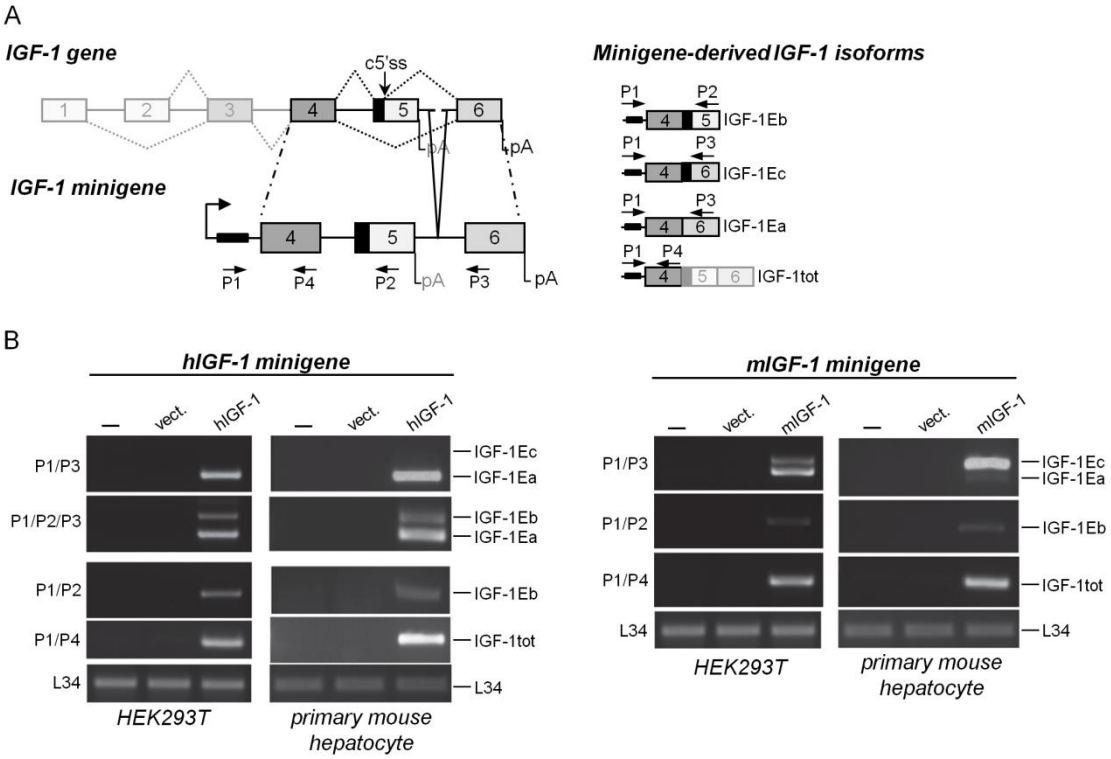


Figure 4. The expression pattern of human and mouse IGF-1 minigenes. (A) Schematic representation of IGF-1 minigenes. Exons (*boxes*), introns (*lines*), and cryptic 5' splice site (c5'ss) in exon 5 are indicated (*left panel*). A diagram of primers position used for RT-PCR analysis and the expected PCR products are also shown (right panel). A minigene-specific forward primer (P1) was used to amplify minigene-derived IGF-1 isoforms. Primers P1-P4 were used to amplify a constant region of minigene-derived IGF-1 mRNA; primers P1-P3 were used to amplify IGF-1Ea and IGF-1Ec isoforms; primers P1-P2 were used to amplify IGF-1Eb isoform. All oligonucleotide sequences are listed in Supplementary Table S1. (B) RT-PCR analysis of splicing assay performed in human HEK293T cells (*left panel*) and primary mouse hepatocytes (*right panel*) in presence of human (hIGF-1) or mouse (mIGF-1) minigenes. Untransfected cells (-) or cells transfected with the empty vector (vect.) were used as a PCR control. L34 was used as loading control.

Since the hIGF-1 and mIGF-1 splicing patterns were mostly preserved regardless of the cell context, it is likely that splicing of exon 5 is not regulated by a specific human or mouse trans-acting environment. To validate this hypothesis, we forced the expression of selected splicing factors to determine their effect on the mouse and human IGF-1 minigenes. Human *IGF-1* exon 5 splicing is regulated by competition between SRSF1 and hnRNPA1 [8-9]. Thus, we tested whether or not the differences in human and mouse exon 5 splicing could be influenced by modulation of the expression of SRSF1, hnRNP A1 and other similar splicing factors (Fig. 5 and Supplementary Fig. S5). As expected [8-9], overexpression of hnRNP A1 promoted splicing of the IGF-1Ea variant by the hIGF-1 minigene, whereas overexpression of SRSF1 favored splicing of the IGF-1Eb variant (Fig. 5B, *left panel*). However, overexpression of neither splicing factor was capable of inducing the IGF-1Ec variant from the hIGF-1 minigene (Fig. 5B, *left panel*). Moreover, splicing assays in presence of several other hnRNPs (A2, F, G, H, I and K), SR (SRSF3 and 7) and SR-like (TRA2α and 2β) proteins showed that none of them was capable of inducing the IGF-1Ec variant from the hIGF-1 minigene in HEK293T cells (Fig. 5B, *left panel*), even though most of these splicing factors influenced the IGF-1Eb/IGF-1Ea ratio (Fig. 5B, *left panel*). Similarly, depletion of hnRNPA1, SRSF1 or TRA2β did not promote splicing of IGF-1Ec isoform by the endogenous human gene (Fig. 5D, *left panel* and Supplementary Fig. S5B) and hIGF-1 minigene (Supplementary Fig. S5C). Since, splicing of IGF-1Ec is promoted by local muscle tissues damage in rodent [7], we tested whether oxidative stress condition in a human cell line (LNCaP) may promote splicing of this isoform. Notably, we observed modulation of IGF-1Eb/IGF-1Ea ratio but not splicing of IGF-1Ec (Supplementary Fig. S5D). These results strongly suggest that lack or low expression of the IGF-1Ec variant is unlikely due to deficient expression of trans-acting factors in human cells. In the case of the mIGF-1 minigene, overexpression (Fig. 5B, *right panel*) or depletion (Supplementary Fig. S5C) of most splicing factors modulated the IGF-1Ec/IGF-1Ea ratio between the two canonical variants, but none of them affected the low levels of basal expression of the IGF-1Eb variant produced by the mIGF-1 minigene. Accordingly, depletion of hnRNPA1, SRSF1 or TRA2β did not promote splicing of IGF-1Eb isoform from the endogenous mouse gene (Fig. 5D, *right panel* and Supplementary Fig. S5B). These results argue against a key role played by trans-acting factors in the alternative processing of *IGF-1* exon 5 observed in both human and mouse.

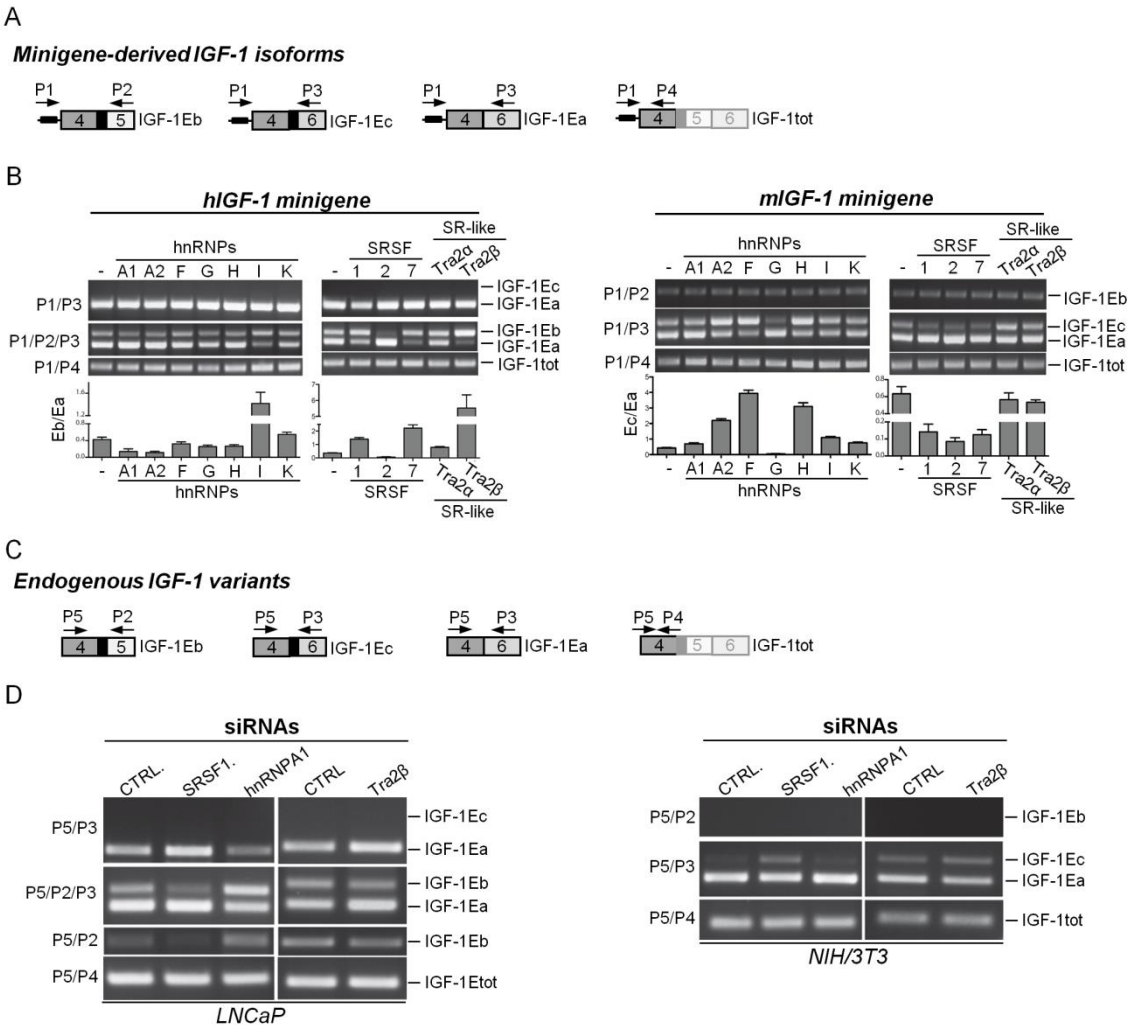


Figure 5. Modulation of splicing factors expression regulates mouse and human IGF-1 alternative splicing. (A) Diagram of primers position used to amplify minigene-derived IGF-1 isoforms. (B) RT-PCR analysis of splicing assay performed in HEK293T cell line in presence of human (*left panel*) or mouse (*right panel*) IGF-1 minigenes and the indicated splicing factors. The bar graph shows the ratio of densitometric analysis between IGF-1Eb and IGF-1Ea (*left panel*) or IGF-1Ec and IGF-1Ea (*right panel*) isoforms (mean ± SD, n=3). C) Diagram of primers position used to amplify endogenous IGF-1 isoforms. (D) RT-PCR analysis of splicing assay of endogenous IGF-1 gene performed in human LNCaP cells (left panel) and in mouse NIH/3T3 cells (*left panel*) depleted with the indicated splicing factors.

3.6 The relative strength of the c5'ss of *IGF-1* exon 5 contributes to the species-specific production of the IGF-1Ec isoform in human and mouse

Given the marked difference in strength between the mouse and human c5'ss (Fig. 2B), we hypothesized that evolutionary modifications of this splice site explains its different

usage in the two species. To verify this possibility, we replaced by site-directed mutagenesis the weak human c5'ss with the strong mouse splice site (hMUT: TGvsGA) (Fig. 6A). Notably, strengthening of human c5'ss partially recovered splicing of IGF-1Ec (Fig. 6B). By contrast, weakening of the mouse c5'ss by reverting it to the human sequence (mMUT: GAvsTG) (Fig. 6A) prevented its usage and favored selection of another cryptic 5'ss located downstream the canonical one (IGF-1Ec* in Figs. 6A and 6C). These results strongly suggest that the strength of the c5'ss in exon 5 is a crucial factor in determining the species-specific expression of the IGF-1Ec splice variant.

We noticed that humanization of the mouse c5'ss (mMUT: GAvsTG) was not sufficient to enhance splicing of the IGF-1Eb variant from the mouse minigene (Fig. 6C). Moreover, splicing was not affected by overexpression of SRSF1, which promotes the IGF-1Eb variant from the human minigene (Supplementary Fig. S6A). Similarly, unlike with the mouse minigene, overexpression of SRSF1 did not decrease the IGF-1Ec/IGF-1Ea ratio in presence of hMUT (TGvsGA) minigene (Supplementary Fig. S6A). Altogether, these results suggest that other sequence elements, in addition to the c5'ss, contribute to species-specific variations of *IGF-1* alternative splicing, possibly by conferring sensitivity to trans-acting factors.

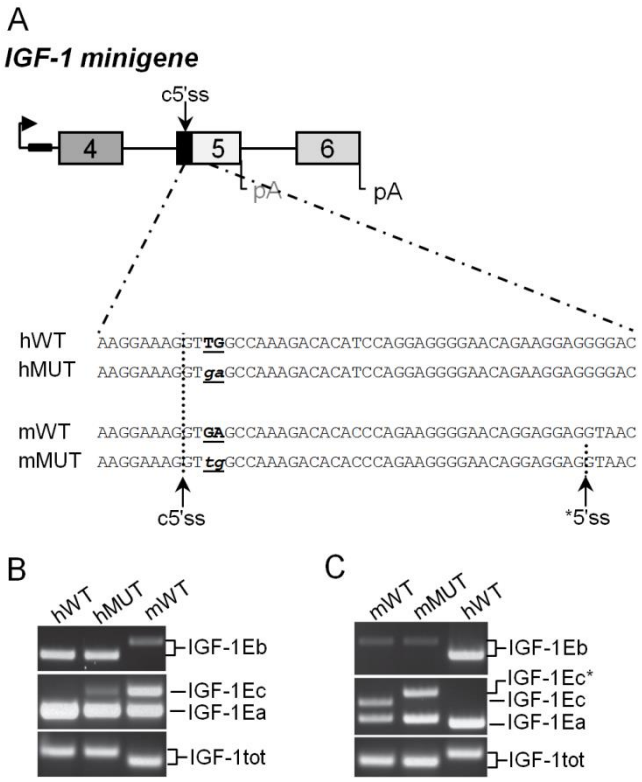


Figure 6. Comparative analysis of human and mouse IGF-1 alternative splicing on wild-type and mutated minigenes. (A) Scheme of human (hMUT: TG*vs*GA) and mouse (mMUT: GA*vs*TG) mutant minigenes. The mutated bases are underlined. Position of the canonical (c5'ss) and non-canonical (*5'ss) cryptic 5' splice site used in mMUT minigene are also shown. (B and C) RT-PCR analysis of splicing assay performed in human HEK293T cells in presence of human (B) and mouse (C) IGF-1 mutant minigenes. RT-PCR was performed using primers as indicated in Fig. 4.

3.7 Molecular evolution of mammalian IGF-1 isoforms and prediction of IGF-1 domain structural properties

The selective pressure acting on IGF-1 domains is currently largely unknown [17, 19]. The higher variability of E-peptide sequences compared to core region sequences among mammals might reflect a less stringent functional requirement for this portion of the IGF-1 protein [17]. On the other hand, the extremely low synonymous mutation rates found on the dual-coding exon 6 across 29 mammalian species [19] suggests an enrichment of conserved regulatory elements in this region of the gene [18].

To better estimate the evolutionary pressure acting on IGF-1 domains we compared the non-synonymous (*d*N) and synonymous (*d*S) substitutions among 27 mammalian protein-coding sequences of *IGF-1* [32]. As shown in Table 1, there was substantial variation in the substitution rate within the gene regions encoding the different domains. Specifically, the Ea, Eb and Ec peptides showed a marked reduction of *d*S and only a slight increase of *d*N (significant only for Eb peptide) compared to the core region. Thus, the increase of the E-peptides dN/dS ratio was only marginally caused by an increase of dN, rather, it is likely due to a significantly smaller dS.

Substitution rates of IGF-1 domains among 27 mammalian species.

IGF-1 domain	Average length (codons)	dS	dN	dN–dS
IGF-1 core	70.0	0.33 (0.23;0.44)^{a}	0.01 (0.01;0.02)^{a}	−0.31 (−0.42;−0.22)^{a*}
Ea peptide	35.0	0.13 (0.09;0.18)^{b}	0.02 (0.01;0.02)^{a}	−0.12 (−0.16;−0.08)^{b*}
Eb peptide	63.3	0.09 (0.06;0.12)^{b}	0.05 (0.03;0.06)^{b}	−0.04 (−0.06;−0.02)^{c*}
Ec peptide	41.1	0.06 (0.04;0.09)^{b}	0.02 (0.02;0.03)^{a}	−0.04 (−0.07;−0.02)^{c*}

Table 1. All rates are estimated bymaximum likelihood using theMG94xREV_3x4 (local) codon substitution model. Mean (95% confidence intervals) and P values for the neutrality tests (dN–dS expected value = 0) are estimated by using parametric bootstrap based on 1000 replicates. Multiple comparisons among IGF-1 domains were performed with Analysis of variance boostrap based. Values in the same column bearing the same letter are not significantly different. The difference dN–dS was used in place of a more common ratio dN/dS, to avoid numerical issues when dS is zero, which is possible because HyPhy permits synonymous substitution rates to vary from site to site. * $P < 0.05$.

This result confirms and extends previous findings [19], as we demonstrated strong reduction in the rate of on synonymous substitution both in the dual-coding region of exon 6 and in the single-coding region of exon 5. The strong evolutionary constraint on these exons is consistent with previous observations on alternatively spliced exons and is a signature of enrichment of functional elements, such as splicing enhancer or silencer elements [18, 21]. Accordingly, exon 5 contains binding sites for SRSF1 and hnRNP A1 [8-9] and exon 5 alternative splicing is regulated by multiple splicing factors (Fig. 5). Intriguingly, recent studies demonstrated that protein-coding regions with low synonymous mutation rate are significantly enriched in Intrinsically Disordered Regions (IDRs) [42-43]. IDRs are polypeptide chains lacking a stable tertiary structure and the relaxed protein structural constraint provide an advantage when a protein-coding region simultaneously contains additional regulatory sites such as splicing enhancer and repressor sequences [43]. In order to analyze the structural properties of IGF-1 domains, we used DisCons [35], a sequence analysis tool able to identify protein disorder segments in an evolutionary context. As expected, among 27 mammalian species analyzed, the IGF-1 core region was predicted to be structured for the most part (Fig. 7). Conversely all the E-peptides were predicted to be almost completely “constrained disorder” (Fig. 7). Namely, they were enriched in disorder-promoting residues and there was conservation of the property of disorder across mammals. These results do not depend on the disorder predictor algorithm because core results were qualitatively replicated using PONDR® VSL2 instead of IUPred (data not shown). Thus, during the course of mammalian

evolution, the structure of IGF-1 protein has been strongly conserved to maintain both the folded structure of IGF-1 core and its intrinsically disordered E-peptide tails.

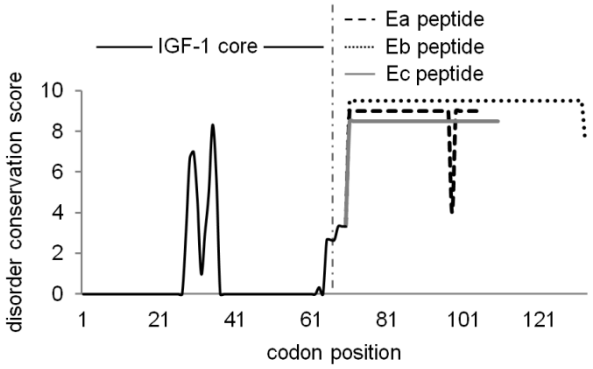


Figure 7. The sliding-window plot of the disorder conservation score of IGF-1 domains. The x axis represents the codon positions and the y axis shows disorder conservation profiles for IGF-1 core (black line), Ea (dashed line), Eb (dotted line) and Ec (gray line) peptides. The disorder conservation score for IGF-1 domains was calculated using the Disorder Conservation (DisCons) software with default parameters [35]. Low values correspond to order, high values to disorder.

4. Discussion

In the present work we provided evolutionary evidence supporting the emergence of a functional alternative exon 5 by exonization of a MIR-b element in mammalian *IGF-1* intron 4 (Fig. 1C). The exon 5 gain leads to the generation of two new IGF-1 E-peptides in the mammalian lineage: Eb and Ec. Our work indicates that the Ea peptide represents the ancestral E-peptide, common to all vertebrates, whereas Eb and Ec peptides are an evolutionary novelty appeared in Mammalia after the exonization of MIR-b transposon by exaptation [44]. The MIR element belongs to a family of transposable elements, which were actively propagating prior to mammalian radiation [45]. The MIR family, with a conservation rate between 60 and 70%, is one of the most diverged transposable element families identifiable in the human genome and only recent molecular approaches have improved their chances to be identified [27, 46]. For example, MIRs have a 95% chance of being missed by RepeatMasker if they are 50 bp in length, and 50% chance of being missed if 100 bp long [46]. In our study, we successfully identified and aligned the MIR-b element to 103 nucleotides of the human *IGF-1* exon 5 sequence only by using the

recently introduced “hmmer” search engine of RepeatMasker, which uses the hidden Markov model search tool nhmmer and the Dfam library [27-28].

Several genome-wide studies have focused on the exaptation and exonization of MIR elements and found over 100 exonized MIR elements in the human genome, located both in the UTRs and the CDSs of annotated genes [47-48]. These studies showed that exonization of retroposed sequences can occur at any time following their insertion [49-50]. Notably, both DNA sequence analyses performed in our study and mRNA analyses conducted in birds [13], fish [15] and reptiles [14] showed that *IGF-1* exon 5 is absent in these vertebrates. Therefore, the emergence of *IGF-1* exon 5 may have taken place around the mammalian diversification, which coincides with the timing of integration of the MIR elements [51].

The fact that MIR-b integrated in the *IGF-1* exon 5 in its antisense orientation and contributed the splice sites and the oligopyrimidine tract may have facilitated its exonization immediately after integration [50]. Moreover, the 3'ss of *IGF-1* exon 5 is located near the highly conserved core region of MIR-b, a feature that is present in many of the exonized-MIR described so far [38, 50] and that may have also contributed to its early exonization [50]. The conservation and the relative strength of the 3'ss of *IGF-1* exon 5, together with the maintenance of open reading frame in all main mammalian branches, suggest that during evolution the creation of a functional 3'ss inside the MIR-b integrated in the *IGF-1* intron 4 was essential for successful exonization of this element [52].

Exonization of MIR-b in the mammalian *IGF-1* gene also created a c5'ss that allows splicing of the IGF-1Ec variant in some species. Our analyses highlighted that, in contrast to the 3'ss, the mammalian c5'ss shows a considerable variation in term of the number of matching nucleotides to the vertebrate 5'ss consensus, with the most marked difference observed between humans and Muridae (Figs. 2A, and 2B and Supplementary Fig. S2). This led us to hypothesize that the spliceosome ability to recognize the c5'ss present in exon 5 may vary among mammalian species. Hence, the relative quantity of the IGF-1Eb and IGF-1Ec isoforms might also differ among mammals. Accordingly, we observed a differential expression pattern of IGF-1 isoforms across 9 different mammalian species analyzed (Fig. 3 and Supplementary Fig. S3). In particular, we found a marked difference in IGF-1 splicing between humans and Muridae, with a prevalence of IGF-1Eb in humans and IGF-1Ec in Muridae. In other mammalian species, including the primate macaque,

expression of the IGF-1Eb isoform was slightly higher or equal to the IGF-1Ec. Hence, the expression of IGF-1Eb is not restricted to humans, and this splice variant should no longer be considered human-specific [1, 39].

Divergence of alternative splicing represents one of the major driving forces to shape phenotypic differences across species [53-57]. Such changes could arise from divergence in cis-regulatory elements and/or trans-acting splicing factors [58]. Therefore, we next focused on the mechanism underlying the diversification of IGF-1 alternative splicing between human and mouse. For this purpose we constructed minigene-based systems of the alternatively spliced regions. Both minigenes recapitulated the splicing pattern observed in tissues, as the human and mouse minigenes produced very low levels of IGF-1Ec and IGF-1Eb isoforms, respectively (Fig. 4 and Supplementary Fig. S4). Nevertheless, our minigene splicing assays also showed some small differences with respect to mouse tissues. In particular, mouse minigene produces a small amount of the IGF-1Eb variant, whereas this isoform is minimally expressed in mouse tissues (Figs. 3A and 4B and Supplementary Fig. S4). Moreover, alternative splicing in the mouse cell lines is mostly shifted toward the IGF-1Ec isoform (Fig. 4B and Supplementary Fig. S4). We suppose that these differences might be due to alterations of the genomic and epigenetic structure of the IGF-1 gene with respect to the minigene and/or to the different promoter present in the minigenes, as all these features have been shown to modulate the outcome of pre-mRNA splicing [40-41]. Nevertheless, our splicing assays indicate that the IGF-1 minigenes are reliable tools for the analysis of the splicing regulation of the corresponding endogenous genes. By using these minigenes, we tested whether cis-acting elements and/or trans-acting factors were responsible for the different splicing patterns observed in vivo between mouse and human IGF-1 mRNA. Alternative splicing outcome is determined by competition between splice sites in different exons and is influenced by trans-acting splicing factors, which in turn influence splice site recognition by the spliceosome [59]. As a consequence, both changes in the sequence of cis-acting elements and in the expression levels of splicing factors can modulate usage of specific exons [58, 60]. Our results argue against changes in expression of specific splicing factors as a key determinant between the splicing differences across species. Indeed, we found that splicing of the mouse and human IGF-1 minigenes is mostly preserved across different cell type contexts (Fig. 4B and Supplementary Figure S4). Moreover, by modulating the expression of specific splicing factors it was not possible to promote splicing of the IGF-

1Ec variant from the human minigene (Fig. 5B, left panel and Supplementary Fig. S5C) or endogenous human gene (Fig. 5D, left panel and Supplementary Fig. S5B). Similarly, neither the overexpression (Fig. 5B, right panel) nor the siRNA-mediated depletion (Fig. 5D, right panel and Supplementary Figs. S5B and S5C) of splicing factors affected the expression of mouse IGF-1Eb isoform. We tested several hnRNPs and SR proteins, including SRSF1 and hnRNP A1 that were previously shown to regulate splicing of human IGF-1 [8-9]. Most of these trans-acting factors modulated the human IGF-1Eb/IGF-1Ea and mouse IGF-1Ec/IGF-1Ea ratio, but none of them promoted the variant typical of the other species. Thus, our results suggest that the mechanism regulating this process does not rely on differential expression of species-specific trans-acting factors.

An alternative possibility is the presence of cis-acting elements conferring species-specific features to the exon 5 region. This hypothesis is in line with a recent study showing that most vertebrate specie-specific splicing patterns are primarily cis-directed [54, 58]. Accordingly, our analysis revealed that the consensus bases within the c5'ss are highly divergent among mammalian species (Fig. 2A and Supplementary Fig. S2). This observation led us to hypothesize that variation in the c5'ss strength might represent a potential source of species-specific splicing pattern. In support of this hypothesis, strengthening the human c5'ss by replacing it with the mouse one (hMUT: TGvsGA) was sufficient to promote usage of this splice site and to yield IGF-1Ec variant (Figs. 6A and 6B). On the other hand, weakening the mouse c5'ss by swapping it with the human sequence (mMUT: GAvsTG) completely prevented its usage (Figs. 6A and 6C). Nevertheless, our results also indicate that other cis-acting elements likely contribute to the splicing pattern typically observed in tissues. First, we found that “murinization” of the human c5’ss only partially promoted splicing of the IGF-1Ec variant, but did not restore the IGF-1Ec/IGF-1Ea ratio observed with the mouse minigene and in mouse tissues (Figs. 3A and 6B). Furthermore, “humanization” of the mouse c5'ss (mMUT: GAvsTG) prevented splicing of the IGF-1Ec variant, but did not promote IGF-1Eb variant (Figs. 6A and 6C). By contrast, we observed the selection of another cryptic splice site (*5'ss) located 38 nt downstream of the canonical one (Figs. 6A and 6C). Notably, this cryptic splice site is not conserved in the human gene, thus explaining the generation of the unusual IGF-1 mRNA variant instead of IGF-1Eb from the mMUT (GAvsTG) minigene. Moreover, overexpression of SRSF1 did not promote splicing of the IGF-1Eb variant from the mMUT (GAvsTG) minigene (Supplementary Fig. S6A).

SRSF1-dependent splicing of human IGF-1Eb requires its binding to a purine-rich enhancer located in exon 5 [8-9]. Sequence alignment of human and mouse exon 5 revealed that the SRSF1 binding site is not conserved in the mouse (Supplementary Fig. S6B), possibly explaining the different behavior of SRSF1 in presence of the mouse minigene. These results strongly indicate that the strength of the c5'ss in exon 5 is a prerequisite for IGF-1 alternative splicing diversification between species, but also highlight that the co-evolution of additional exonic and intronic cis-regulatory elements contribute to such diversification [40-41]. Notably, our phylogenetic analysis of the c5'ss indicates that this splice site is relative strong not only in mouse but also in rat and all members of Cricetidae family (Fig. 2B and Supplementary Fig. S2). It is possible that strengthening of this c5'ss in a common ancestor of Muridae and Cricetidae, as well as the emergence of other splicing regulatory motifs, have determined the marked shift toward production of the IGF-1Ec isoform in these species and possibly a rodent-specific gain of functional properties. Interestingly, a recent study comparing the IGF-1 isoform expression in both mouse and human muscle samples at different ages, showed that the age-related change of IGF-1 splice variants is species-dependent and, unlike the mouse, only the human IGF-1Eb isoform was regulated during ageing in skeletal muscles [61]. Thus, the results obtained with murine models on IGF-1Ec/mechano growth factor regulation must be interpreted with caution and additional studies comparing IGF-1 expression levels across species are needed to clarify the functional role of the IGF-1 isoforms.

Compared with mature IGF-1 relatively little is known about the mechanism of action of the different E peptides [1, 10-12]. From an evolutionary point of view the unchanged persistence over long evolutionary periods of MIRb-derived IGF-1 exon 5 implies its functional relevance [52, 62]. Moreover, E-peptides are protein-coding regions in which synonymous mutation rates are extremely low compared to IGF-1 core (Table 1), indicating additional sequence constraints beyond those dictated by the structure and function of the proteins. These additional constraints probably stem from the demands of regulatory sites involved in transcript splicing (Fig. 5) and [8-9, 18], nevertheless the presence of other regulatory sites such as specific RNA secondary structures and microRNA targets [20-21] cannot be excluded. The systematic analysis of the role of synonymous variants and the comparative splicing evaluation of mammalian sequence

divergences [63-64] will help to characterize the functional roles of these regulatory elements.

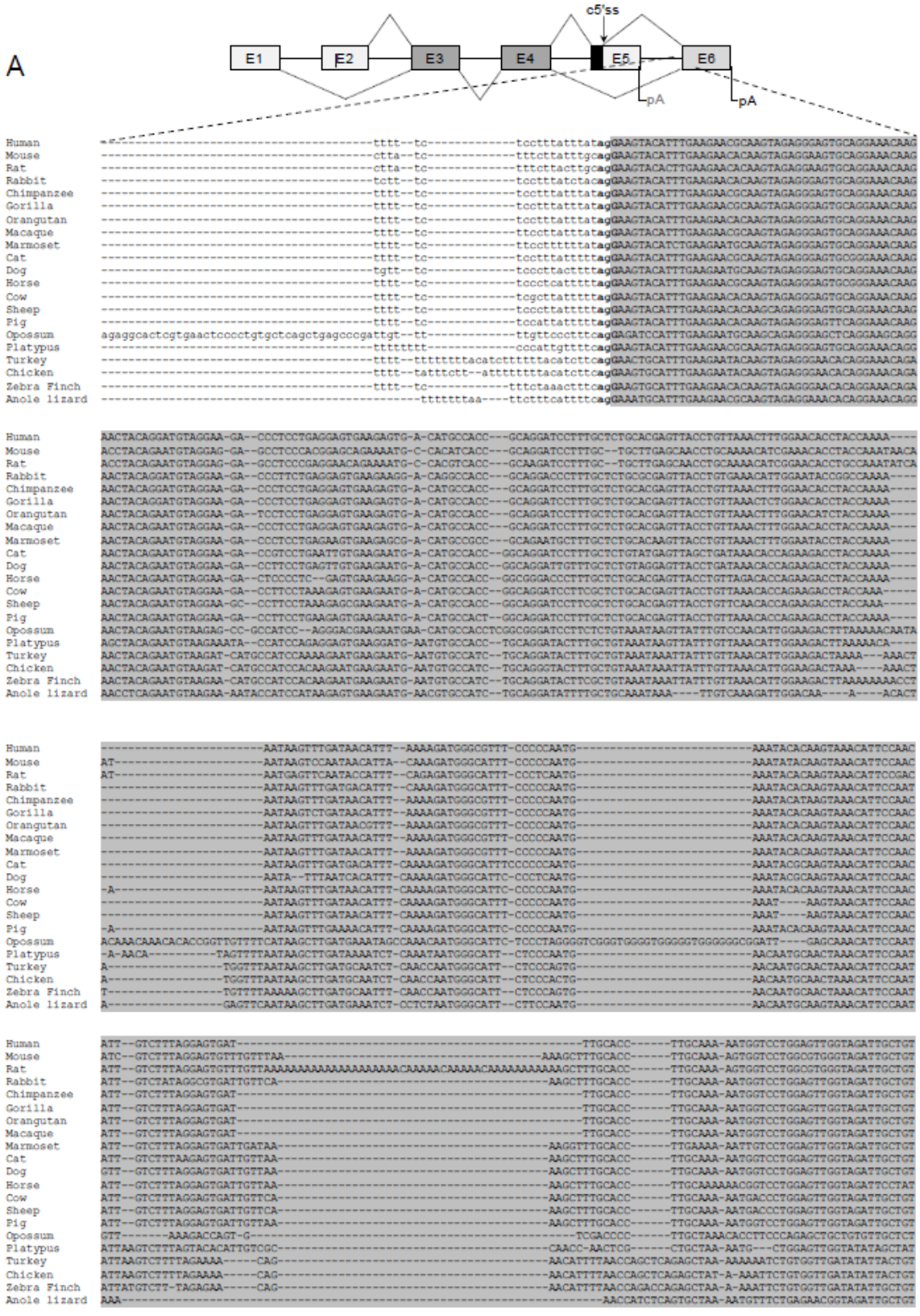
Analysis of structural properties of IGF-1 domains adds a new layer of complexity to the function of E-peptides. Indeed, our work represents the first evidence that E-peptides contain disorder-promoting amino acids and that there is substantial evolutionary pressure to keep the different E-peptides as intrinsically disordered regions (IDRs) (Fig. 7). There is a growing interest on IDRs since they are usually enriched in posttranslational modification sites and may exert a number of regulatory functions on their "host" protein [65-66]. Intriguingly, we and others recently demonstrated that the IGF-1 protein retaining C-terminal E peptides are the predominant forms produced intracellularly, instead of mature IGF-1, and are subjected to extensive post-translational modifications [67-68], further hinting to their functional relevance.

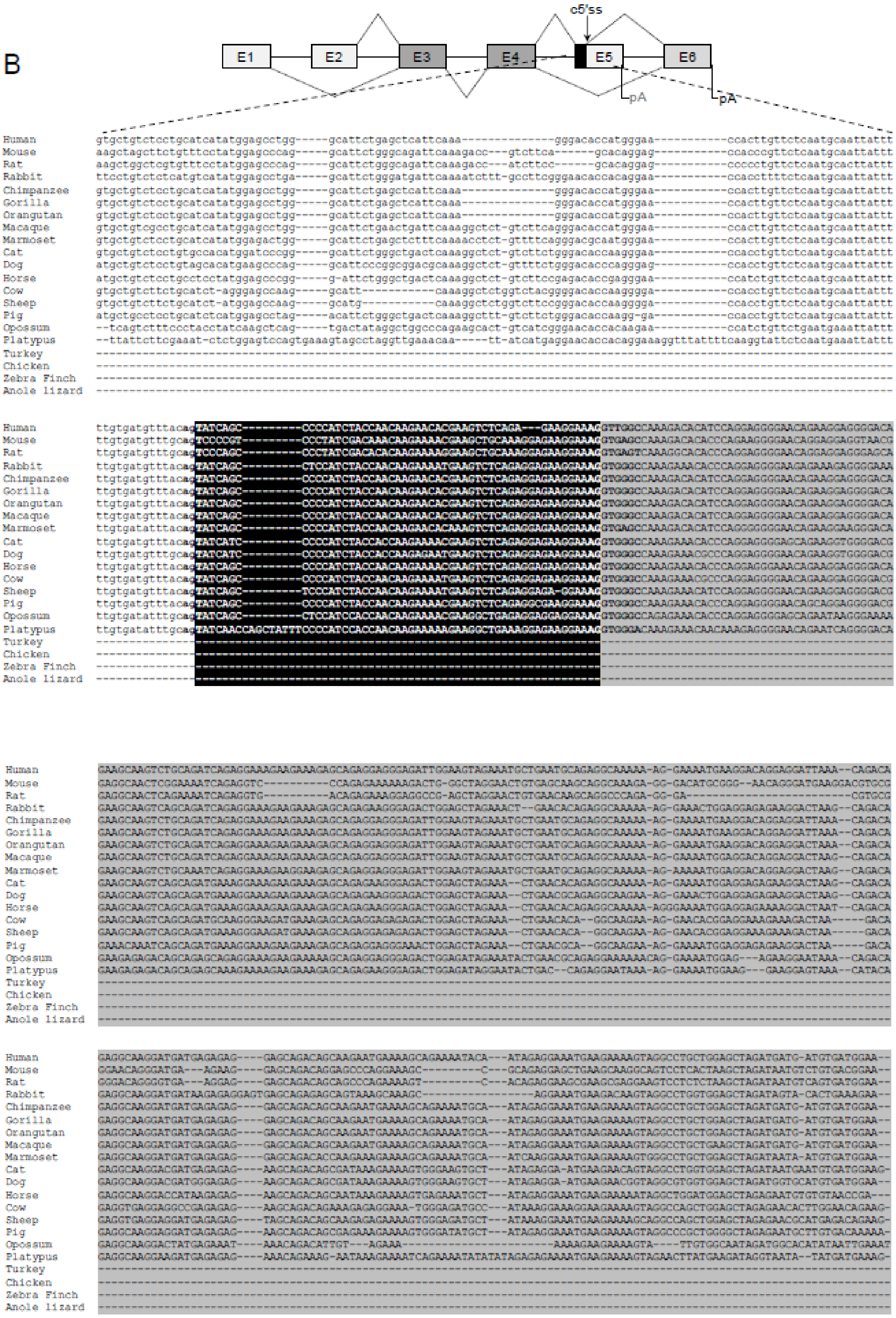
Exonization of previously non-coding sequences, together with creation of novel domain combinations, has been directly related to the increase of organismal complexity [66, 69-71]. Accordingly, we propose that MIR-b exonization during mammalian evolution determined the IGF-1 exon 5 gain and hence the addition of two new disordered tails to IGF-1: the Eb and Ec-tails. Thus, novel exon and E-peptide combinations may have created new layers of regulation to mature IGF-1 in mammalian species. Targeting these regulatory elements may represent a new strategy to control IGF-1 bioavailability in different physiological/pathological conditions, with particular attention to possible differences between species.

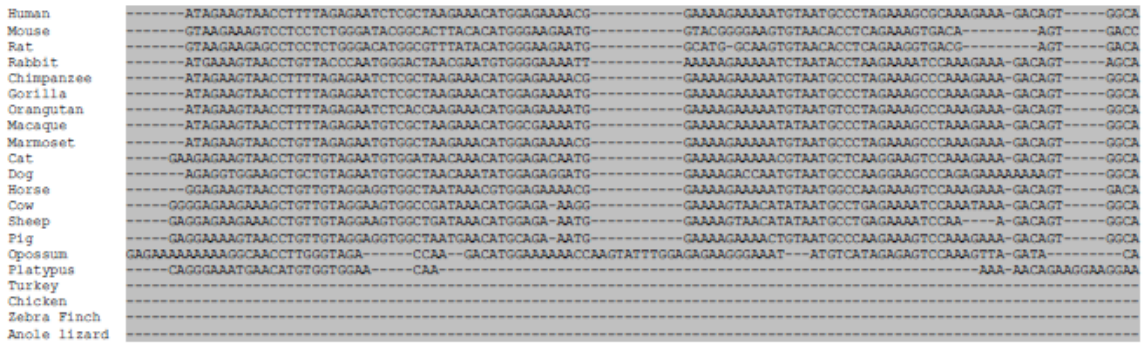
5. Acknowledgement

We wish to thank Dr. Franco Berrino of the Department of Predictive & Preventive Medicine, National Institute of Cancer, Milan and Dr. Giorgio Arnaldi, Department of Internal Medicine, Polytechnic University of Marche Region, Ancona, Italy for a critical reading of the manuscript. This research was supported by the Italian Ministry of Health, "Ricerca finalizzata 2009" (grant number: RF-2009-1532789) and "Ricerca Finalizzata 2011 (grant number: GR-2011-02348423); by the Associazione Italiana Ricerca sul Cancro (AIRC IG14581) and by Telethon (CGP14095)."

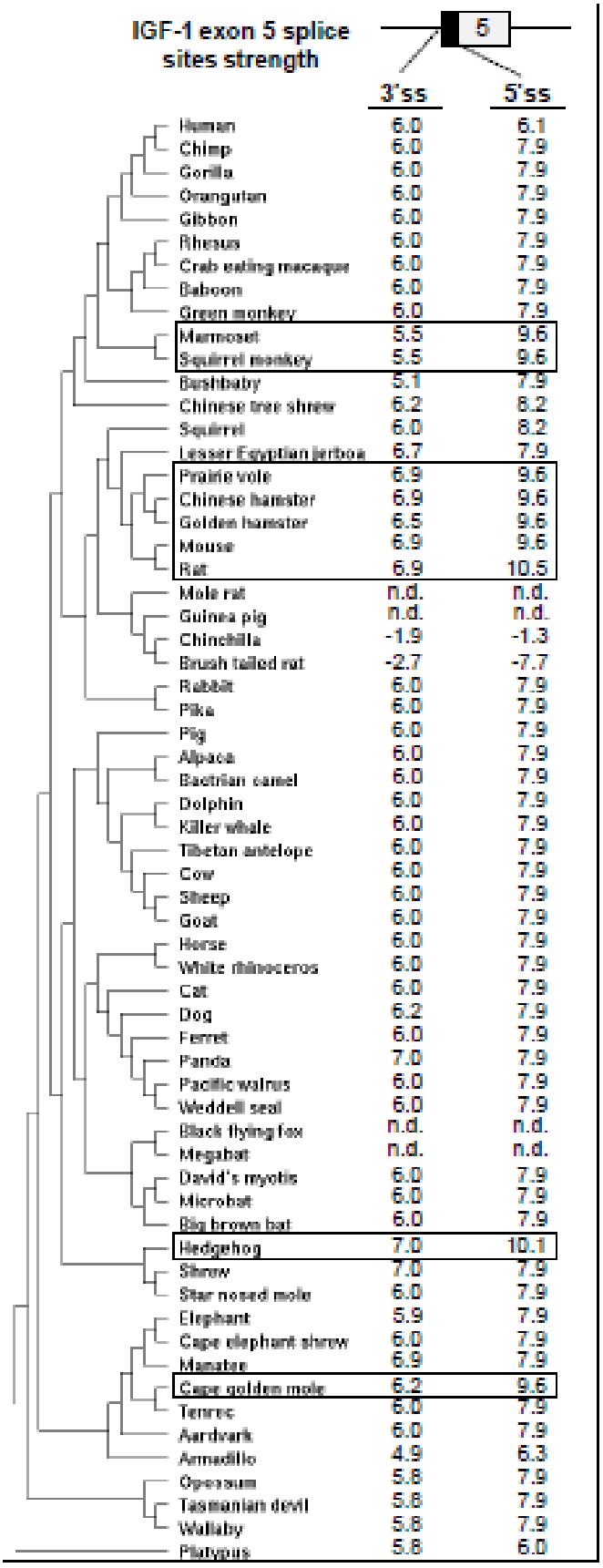
Supplementary data




B
c5'ss
E1
E2
E3
E4
E5
E6
pA
pA
Human
Mouse
Rat
Rabbit
Chimpanzee
Gorilla
Orangutan
Macaque
Marmoset
Cat
Dog
Horse
Cow
Sheep
Pig
Opossum
Platypus
Turkey
Chicken
Zebra Finch
Anole lizard



Supplementary Figure S1. Partial sequence alignment of IGF-1 genomic sequences among 23 amniote vertebrates retrieved from Ensembl database (23 amniote vertebrates Pecan). (A) The nucleotide sequences surrounding the alternatively spliced exon 6 were shown. The exon 6 sequences are in upper case letters and highlighted in gray, the intron sequences are in lower case letters. Bold letters indicates the 3'splice site of exon 6, dashed lines indicate sequence gaps. Human chromosomal location:GRCh38:12:102537063:102537363. (B) The nucleotide sequences surrounding the alternatively spliced exon 5 were shown. The exon 5 sequences are indicated in upper case letters and highlighted in black and gray: the black nucleotides designated a part of exon 5 common to IGF-1Eb and IGF-1Ec variants, while the grey region is included only in the IGF-1Eb isoform; the intron sequences are in lower case letters. The 3' splice site and the cryptic 5' splice site (c5'ss) of exon 5 are shown in bold. Dashed lines indicate sequence gaps. Human chromosomal location:GRCh38: 12:102458404:102458924. Ensembl Genomes: Human (*Homo sapiens*) GRCh38.p5; Mouse (*Mus musculus*) GRCm38.p4; Rat (*Rattus norvegicus*) Rnor_6.0; Rabbit (*Oryctolagus cuniculus*) OryCun2.0; Chimpanzee (*Pan troglodytes*) CHIMP2.1.4 ; Gorilla (*Gorilla gorilla gorilla*) gorGor3.1; Orangutan (*Pongo abelii*) PPYG2; Macaque (*Macaca mulatta*) MMUL 1.0 ; Marmoset (*Callithrix jacchus*) C_jacchus3.2.1; Cat (*Felis catus*) Felis_catus_6.2; Dog (*Canis lupus familiaris*) CanFam3.1; Horse (*Equus caballus*) Equ Cab 2; Cow (*Bos taurus*) UMD3.1; Sheep (*Ovis aries*) Oar_v3.1; Pig (*Sus scrofa*) Sscrofa10.2; Opossum (*Monodelphis domestica*) monDom5 ; Platypus (*Ornithorhynchus anatinus*) OANA5; Turkey (*Meleagris gallopavo*) Turkey_2.01; Chicken (*Gallus gallus*) Galgal4; Zebra Finch (*Taeniopygia guttata*) taeGut3.2.4; Anole lizard (*Anolis carolinensis*) AnoCar2.0.



Supplementary Figure S2. Analyses of the 3′ and the 5′ splice-site signals of the IGF-1 exon 5 among 58 mammalian genomic sequence datasets available at the UCSC genome browser. Evolutionary tree of 58 organisms and the reconstructed ancestral state was adapted from the UCSC Genome Browser. Strength of 3' and 5'ss in exon5 obtained using the Maximum Entropy scores is shown for each species [Yeo G. and Burge CB. Comput. Biol. (2004);11:377-394]. Species with relative strong exon 5 5'ss are boxed. n.d.= not determined.

A

> IGF-1Ea E3 E4 E6

```
Pig(350)        ------------------------------------------------------------        exon 3
Pig(283)        ------------------------------------------------------------
Rabbit(537)     CCTCCTCGCATCTCTTCTACCTGGCCCTCTGCTTGCTCACCTTCACCAGCTCGGCCACAG
Macaque(537)    CCTCCTCGCATCTCTTCTACCTGGCACTGTGCCTGCTCACCTTCACCAGCTCCGCCACGG
Human(627)      CCTCCTCGCATCTCTTCTACCTGGCGCTGTGCCTGCTCACCTTCACCAGCTCTGCCACGG
Human(534)      CCTCCTCGCATCTCTTCTACCTGGCGCTGTGCCTGCTCACCTTCACCAGCTCTGCCACGG
Mouse(391)      ------------------------------------------------------------
Rat(381)        ------------------------------------------------------------

Pig(350)        ------------------------------------------------------------        exon 3
Pig(283)        ------------------------------------------------------------
Rabbit(537)     CCGGACCGGAGACGCTCTGCGGTGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Macaque(537)    CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Human(627)      CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Human(534)      CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Mouse(391)      ------------------------------------------------------------
Rat(381)        ------------------------------------------------------------

Pig(350)        ------------------/----------------------------------------        exon 3/4
Pig(283)        ------------------/----------------------------------------
Rabbit(537)     ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGATACGGCTCCAGCAGTCGGAGGGCACCTC
Macaque(537)    ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCACCTC
Human(627)      ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCGCCTC
Human(534)      ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCGCCTC
Mouse(391)      ------------------/------------GCTATGGCTCCAGCATTCGGAGGGCACCTC
Rat(381)        ------------------/------------GCTATGGCTCCAGCATTCGGAGGGCACCAC

Pig(350)        ----------CGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT        exon 4
Pig(283)        ----------CGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
Rabbit(537)     AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
Macaque(537)    AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGAAGGCTGGAGATGT
Human(627)      AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTAAGGAGGCTGGAGATGT
Human(534)      AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTAAGGAGGCTGGAGATGT
Mouse(391)      AGACAGGCATTGTGGATGAGTGTTGCTTCCGGAGCTGTGATCTGAGGAGACTGGAGATGT
Rat(381)        AGACGGGCATTGTGGATGAGTGTTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
                ********** ********************** ** ** **********

Pig(350)        ACTGTGCACCCCTCAAGCCTGCCAAGTCGGCCCGCTCCGTCCGTGCCCAGCGCCACACGG        exon 4
Pig(283)        ACTGTGCACCCCTCAAGCCTGCCAAGTCGGCCCGCTCCGTCCGTGCCCAGCGCCACACGG
Rabbit(537)     ACTGTGCACCCCTCAAGCCGGCAAAGGCAGCCCGCTCCGTCCGTGCCCAGCGCCACACCG
Macaque(537)    ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCCCGCTCTGTCCGTGCCCAGCGCCATACCG
Human(627)      ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCTCGCTCTGTCCGTGCCCAGCGCCACACCG
Human(534)      ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCTCGCTCTGTCCGTGCCCAGCGCCACACCG
Mouse(391)      ACTGTGCCCCACTGAAGCCTACAAAAGCAGCCCGCTCTATCCGTGCCCAGCGCCACACTG
Rat(381)        ACTGTGCTCCGCTGAAGCCTACAAAGTCAGCTCGTTCCATCCGGGCCCAGCGCCACACTG
                * ** ** ** ** *****  * **  * ** ** **  **** ************ ** *

Pig(350)        ACATGCCCAAGGCTCAGAAG/GAAGTACATTTGAAGAACACAAGTAGAGGGAGTTCAGGAA      exon 4/6
Pig(283)        ACATGCCCAAGGCTCAGAAG/GAAGTACATTTGAAGAACACAAGTAGAGGGAGTTCAGGAA
Rabbit(537)     ACATGCCCAAGACTCAGAAG/GAAGTACATTTGAAGAACACAAGTAGAGGGAGTGCAGGAA
Macaque(537)    ACATGCCCAAGACCCAGAAG/GAAGTACATTTGAAGAACGCAAGTAGAGGGAGTGCAGGAA
Human(627)      ACATGCCCAAGACCCAGAAG/GAAGTACATTTGAAGAACGCAAGTAGAGGGAGTGCAGGAA
Human(534)      ACATGCCCAAGACCCAGAAG/GAAGTACATTTGAAGAACGCAAGTAGAGGGAGTGCAGGAA
Mouse(391)      ACATGCCCAAGACTCAGAAG/GAAGTACATTTGAAGAACACAAGTAGAGGAAGTGCAGGAA
Rat(381)        ACATGCCCAAGACTCAGAAG/GAAGTACACTTGAAGAACACAAGTAGAGGAAGTGCAGGAA
                *********** * ****** ******** ********** ********** *** ******

Pig(350)        ACAAGAACTACAGAATGTAGGAAGACCTTCCTGAAGAGTGAAGAATGACATGCCACTGGC        exon 6
Pig(283)        ACAAGAACTACAGAATGTAGGAAGACCTTCCTGAAGAGTGAAGAATGACATGCCACTGGC
Rabbit(537)     ACAAGAACTACAGGATGTAGGAAGACCCTTCTGAGGAGTGAAGAAGGACAGGCCACC-GC
Macaque(537)    ACAAGAACTACAGAATGTAGGAAGACCCTCCTGAGGAGTGAAGAGTGACATGCCACC-GC
Human(627)      ACAAGAACTACAGGATGTAGGAAGACCCTCCTGAGGAGTGAAGAGTGACATGCCACC-GC
Human(534)      ACAAGAACTACAGGATGTAGGAAGACCCTCCTGAGGAGTGAAGAGTGACATGCCACC-GC
Mouse(391)      ACAAGACCTACAGAATGTAGGAGGAGCCTCCCACGGAGCAGAAAATGCCACATCACC-GC
Rat(381)        ACAAGACCTACAGAATGTAGGAGGAGCCTCCCGAGGAACAGAAAATGCCACGTCACC-GC
                ****** ****** ******** ** * * *    **    * *  * **   ***  **
```

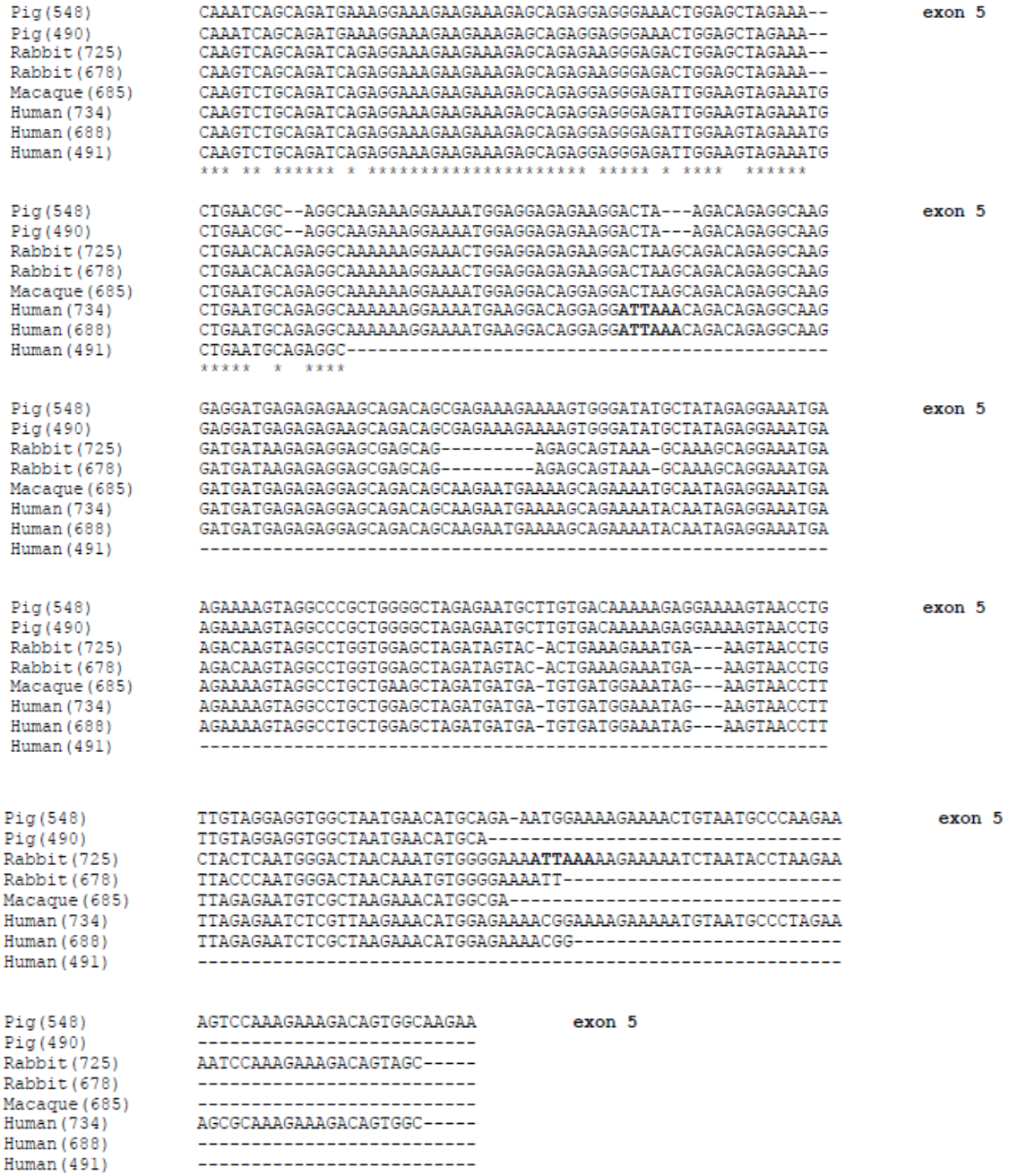
```
Pig(350)        AGGATCCTTTGCTCTGCACGAGTTACCTGTTAAACACCAGAAGACCTACCAAAA-----      exon 6
Pig(283)        AGGATCCTTTGCTCTGCACGAGTTACCTGTTAAACACCAGAAGACCTACCAAA-------
Rabbit(537)     AGGACCCTTTGCTCTGCGCGAGTTACCTGTGAAACATTGGAATACCGGCCAAAA------
Macaque(537)    AGGATCCTTTGCTCTGCACGAGTTACCTGTTAAACTTTGGAACACCTACCAAAA------
Human(627)      AGGATCCTTTGCTCTGCACGAGTTACCTGTTAAACTTTGGAACACCTACCAAAA------
Human(534)      AGGATCCTTTGCTCTGCACGAGTTACCTGTTAAACTTTGGAACACCTACCAAAA------
Mouse(391)      AGGATCCTTTG--CTGCTTGAGCAACCTGCAAAACATCGAAACACCTACCAAATAACAAT
Rat(381)        AAGATCCTTTG--CTGCTTGAGCAACCTGCAAAACATCGGAACACCTGCCAAATATCAAT
                * ** ******  ****  ***  *****  ****     ** ***  *****

Pig(350)        AATAAGTTTGAAAACATTTCAAAAGATGGGCATTCCCCCCAATGAAGTACACAAGTAAAC      exon 6
Pig(283)        ----------------------------------------------------------
Rabbit(537)     AATAAGTTTGATGACATTT-CAAAGATGGGCATTTCCCCCAATGAAATACACAAGTAAAC
Macaque(537)    AATAAGTTTGATAACATTT-AAAAGATGGGCGTTTCCCCCAATGAAATACACAAGTAAAC
Human(627)      AATAAGTTTGATAACATTT-AAAAGATGGGCGTTTCCCCCAATGAAATACACAAGTAAAC
Human(534)      AATAAGTTTGATAACATTT-AAAAGATGGGCGTTTCCCCCAATGAAATACACAAGTAAAC
Mouse(391)      AATAAGTCCAATAACATTA-CAAAGATGGGCATTTCCCCCAATGAAATATACAAGTAAAC
Rat(381)        AATGAGTTCAATACCATTT-CAGAGATGGGCATTTCCCTCAATGAAATACACAAG-----

Pig(350)        ATTCC-------------------------------------------------------      exon 6
Pig(283)        ----------------------------------------------------------
Rabbit(537)     ATTCC-------------------------------------------------------
Macaque(537)    ATTCC-------------------------------------------------------
Human(627)      ATTCCAACATTGTCTTTAGGAGTGATTTGCACCTTGCAAAAATGGTCCTGGAGTTGGTAG
Human(534)      AT----------------------------------------------------------
Mouse(391)      ATTCC-------------------------------------------------------
Rat(381)        ----------------------------------------------------------

Pig(350)        ----------------------------------        exon 6
Pig(283)        ----------------------------------
Rabbit(537)     ----------------------------------
Macaque(537)    ----------------------------------
Human(627)      ATTGCTGTTGATCTTTTATCAATAATGTTCTATAG
Human(534)      ----------------------------------
Mouse(391)      ----------------------------------
Rat(381)        ----------------------------------
```

B

> IGF-1Eb **E3 E4 E5**

```
Pig(548)        ----------------------------------------------------------      exon 3
Pig(490)        ----------------------------------------------------------
Rabbit(725)     CCTCCTCGCATCTCTTCTACCTGGCCCTCTGCTTGCTCACCTTCACCAGCTCGGCCACAG
Rabbit(678)     CCTCCTCGCATCTCTTCTACCTGGCCCTCTGCTTGCTCACCTTCACCAGCTCGGCCACAG
Macaque(685)    CCTCCTCGCATCTCTTCTACCTGGCACTGTGCCTGCTCACCTTCACCAGCTCCGCCACGG
Human(734)      CCTCCTCGCATCTCTTCTACCTGGCGCTGTGCCTGCTCACCTTCACCAGCTCTGCCACGG
Human(688)      CCTCCTCGCATCTCTTCTACCTGGCGCTGTGCCTGCTCACCTTCACCAGCTCTGCCACGG
Human(491)      CCTCCTCGCATCTCTTCTACCTGGCGCTGTGCCTGCTCACCTTCACCAGCTCTGCCACGG

Pig(548)        ----------------------------------------------------------      exon 3
Pig(490)        ----------------------------------------------------------
Rabbit(725)     CCGGACCGGAGACGCTCTGCGGTGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Rabbit(678)     CCGGACCGGAGACGCTCTGCGGTGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Macaque(685)    CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Human(734)      CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Human(688)      CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Human(491)      CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG

Pig(548)        ----------/----------------------------------------------      exon 3/4
Pig(490)        ----------/----------------------------------------------
Rabbit(725)     ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGATACGGCTCCAGCAGTCGGAGGGCACCTC
Rabbit(678)     ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGATACGGCTCCAGCAGTCGGAGGGCACCTC
Macaque(685)    ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCACCTC
Human(734)      ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCGCCTC
Human(688)      ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCGCCTC
Human(491)      ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCGCCTC
```

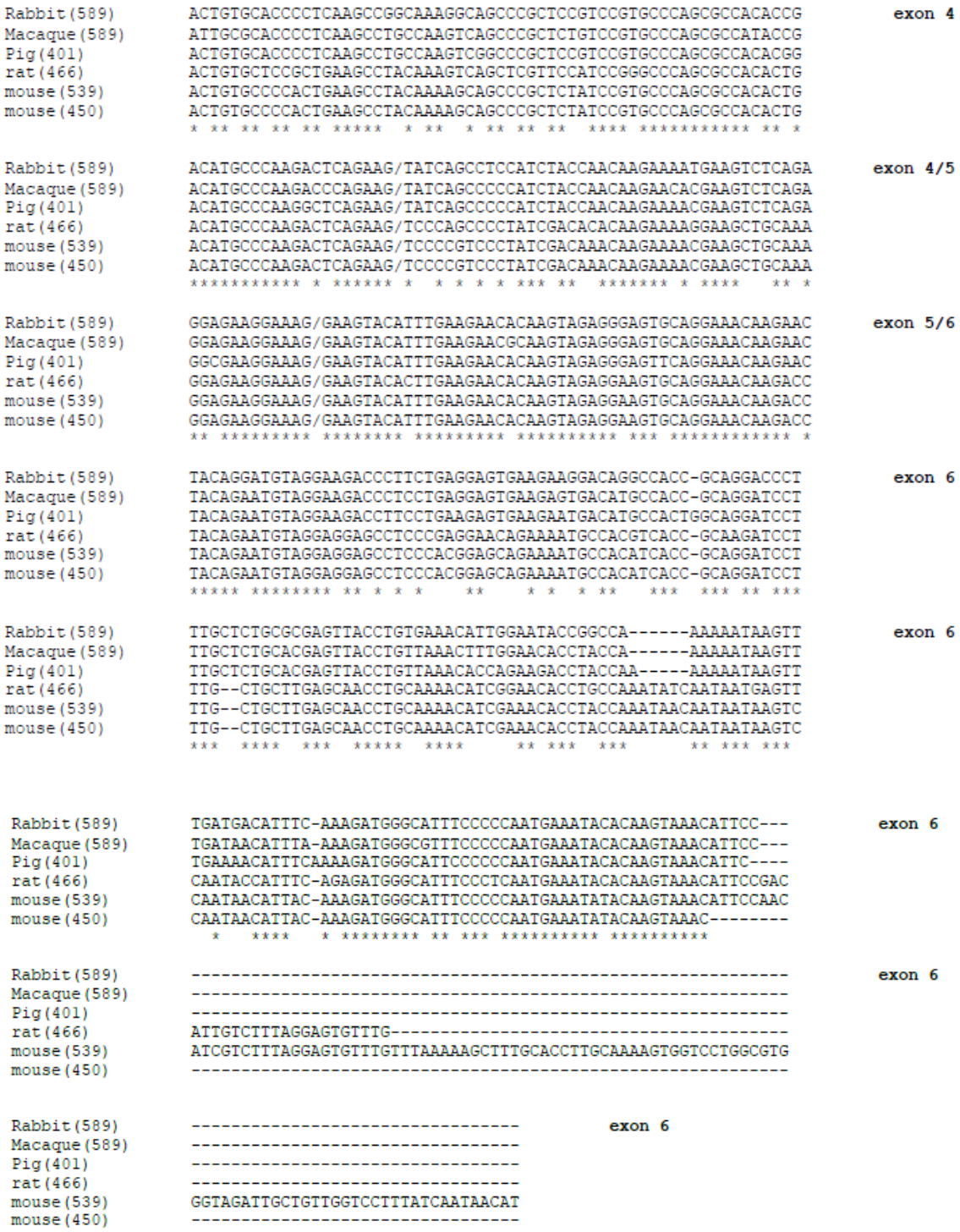
```
Pig(548)          ----------CGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT          exon 4
Pig(490)          ----------CGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
Rabbit(725)       AGACAGGCATCGTGGATGAGTGCTGCCTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
Rabbit(678)       AGACAGGCATCGTGGATGAGTGCTGCCTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
Macaque(685)      AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGAAGGCTGGAGATGT
Human(734)        AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTAAGGAGGCTGGAGATGT
Human(688)        AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTAAGGAGGCTGGAGATGT
Human(491)        AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTAAGGAGGCTGGAGATGT
                            *************** *************** ** *************

Pig(548)          ACTGTGCACCCCTCAAGCCTGCCAAGTCGGCCCGCTCCGTCCGTGCCCAGCGCCACACGG          exon 4
Pig(490)          ACTGTGCACCCCTCAAGCCTGCCAAGTCGGCCCGCTCCGTCCGTGCCCAGCGCCACACGG
Rabbit(725)       ACTGTGCACCCCTCAAGCCGGCAAAGGCAGCCCGCTCCGTCCGTGCCCAGCGCCACACCG
Rabbit(678)       ACTGTGCACCCCTCAAGCCGGCAAAGGCAGCCCGCTCCGTCCGTGCCCAGCGCCACACCG
Macaque(685)      ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCCCGCTCTGTCCGTGCCCAGCGCCATACCG
Human(734)        ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCTCGCTCTGTCCGTGCCCAGCGCCACACCG
Human(688)        ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCTCGCTCTGTCCGTGCCCAGCGCCACACCG
Human(491)        ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCTCGCTCTGTCCGTGCCCAGCGCCACACCG
                  * ** ************** ** *** * ** ***** ****************** ** *

Pig(548)          ACATGCCCAAGGCTCAGAAG/TATCAGCCCCCATCTACCAACAAGAAAACGAAGTCTCAGA      exon 4/5
Pig(490)          ACATGCCCAAGGCTCAGAAG/TATCAGCCCCCATCTACCAACAAGAAAACGAAGTCTCAGA
Rabbit(725)       ACATGCCCAAGACTCAGAAG/TATCAGCCTCCATCTACCAACAAGAAAATGAAGTCTCAGA
Rabbit(678)       ACATGCCCAAGACTCAGAAG/TATCAGCCTCCATCTACCAACAAGAAAATGAAGTCTCAGA
Macaque(685)      ACATGCCCAAGACCCAGAAG/TATCAGCCCCCATCTACCAACAAGAACACGAAGTCTCAGA
Human(734)        ACATGCCCAAGACCCAGAAG/TATCAGCCCCCATCTACCAACAAGAACACGAAGTTTCAGA
Human(688)        ACATGCCCAAGACCCAGAAG/TATCAGCCCCCATCTACCAACAAGAACACGAAGTTTCAGA
Human(491)        ACATGCCCAAGACCCAGAAG/TATCAGCCCCCATCTACCAACAAGAACACGAAGTTTCAGA
                  *********** * ****** ******** ****************** * ***** *****

Pig(548)          GGCGAAGGAAAGGTGGGCCAAAGAAACACCCAGGAGGGGAACAGCAGGAGGGGACGGAAA        exon 5
Pig(490)          GGCGAAGGAAAGGTGGGCCAAAGAAACACCCAGGAGGGGAACAGCAGGAGGGGACGGAAA
Rabbit(725)       GGAGAAGGAAAGGTGGGCCAAAGAAACACCCAGGAGGGGAAGAGAAAGAGGGGAAAGAAG
Rabbit(678)       GGAGAAGGAAAGGTGGGCCAAAGAAACACCCAGGAGGGGAAGAGAAAGAGGGGAAAGAAG
Macaque(685)      GGAGAAGGAAAGGTGGGCCAAAGACACATCCAGGAGGGGAACAGAAGGAGGGGACAGAAG
Human(734)        ---GAAGGAAAGGTTGGCCAAAGACACATCCAGGAGGGGAACAGAAGGAGGGGACAGAAG
Human(688)        ---GAAGGAAAGGTTGGCCAAAGACACATCCAGGAGGGGAACAGAAGGAGGGGACAGAAG
Human(491)        ---GAAGGAAAGGTTGGCCAAAGACACATCCAGGAGGGGAACAGAAGGAGGGGACAGAAG
                     *********** ********* *** ************* ** * *******  ***
```


Pig(548) CAAATCAGCAGATGAAAGGAAAGAAGAAAAGAGCAGAGGAGGGAAACTGGAGCTAGAAA-- exon 5
Pig(490) CAAATCAGCAGATGAAAGGAAAGAAGAAAAGAGCAGAGGAGGGAAACTGGAGCTAGAAA--
Rabbit(725) CAAGTCAGCAGATCAGAGGAAAGAAGAAAAGAGCAGAGAAGGGAGACTGGAGCTAGAAA--
Rabbit(678) CAAGTCAGCAGATCAGAGGAAAGAAGAAAAGAGCAGAGAAGGGAGACTGGAGCTAGAAA--
Macaque(685) CAAGTCTGCAGATCAGAGGAAAGAAGAAAAGAGCAGAGGAGGGAGATTGGAAGTAGAAATG
Human(734) CAAGTCTGCAGATCAGAGGAAAGAAGAAAAGAGCAGAGGAGGGAGATTGGAAGTAGAAATG
Human(688) CAAGTCTGCAGATCAGAGGAAAGAAGAAAAGAGCAGAGGAGGGAGATTGGAAGTAGAAATG
Human(491) CAAGTCTGCAGATCAGAGGAAAGAAGAAAAGAGCAGAGGAGGGAGATTGGAAGTAGAAATG
Pig(548) CTGAACGC--AGGCAAGAAAGGAAAATGGAGGAGAGAAGGACTA---AGACAGAGGCAAG exon 5
Pig(490) CTGAACGC--AGGCAAGAAAGGAAAATGGAGGAGAGAAGGACTA---AGACAGAGGCAAG
Rabbit(725) CTGAACACAGAGGCAAAAAAGGAAACTGGAGGAGAGAAGGACTAAGCAGACAGAGGCAAG
Rabbit(678) CTGAACACAGAGGCAAAAAAGGAAACTGGAGGAGAGAAGGACTAAGCAGACAGAGGCAAG
Macaque(685) CTGAATGCAGAGGCAAAAAAGGAAAATGGAGGACAGGAGGACTAAGCAGACAGAGGCAAG
Human(734) CTGAATGCAGAGGCAAAAAAGGAAAATGAAGGACAGGAGGATTAAACAGACAGAGGCAAG
Human(688) CTGAATGCAGAGGCAAAAAAGGAAAATGAAGGACAGGAGGATTAAACAGACAGAGGCAAG
Human(491) CTGAATGCAGAGGC----------
Pig(548) GAGGATGAGAGAGAAGCAGACAGCGAGAAAGAAAAGTGGGATATGCTATAGAGGAAATGA exon 5
Pig(490) GAGGATGAGAGAGAAGCAGACAGCGAGAAAGAAAAGTGGGATATGCTATAGAGGAAATGA
Rabbit(725) GATGATAAGAGAGGAGCGAGCAG---------AGAGCAGTAAA-GCAAAGCAGGAAATGA
Rabbit(678) GATGATAAGAGAGGAGCGAGCAG---------AGAGCAGTAAA-GCAAAGCAGGAAATGA
Macaque(685) GATGATGAGAGAGGAGCAGACAGCAAGAATGAAAAGCAGAAAATGCAATAGAGGAAATGA
Human(734) GATGATGAGAGAGGAGCAGACAGCAAGAATGAAAAGCAGAAAATACAATAGAGGAAATGA
Human(688) GATGATGAGAGAGGAGCAGACAGCAAGAATGAAAAGCAGAAAATACAATAGAGGAAATGA
Human(491) ----------
Pig(548) AGAAAAGTAGGCCCGCTGGGGCTAGAGAATGCTTGTGACAAAAAGAGGAAAAGTAACCTG exon 5
Pig(490) AGAAAAGTAGGCCCGCTGGGGCTAGAGAATGCTTGTGACAAAAAGAGGAAAAGTAACCTG
Rabbit(725) AGACAAGTAGGCCTGGTGGAGCTAGATAGTAC-ACTGAAAGAAATGA---AAGTAACCTG
Rabbit(678) AGACAAGTAGGCCTGGTGGAGCTAGATAGTAC-ACTGAAAGAAATGA---AAGTAACCTG
Macaque(685) AGAAAAGTAGGCCTGCTGAAGCTAGATGATGA-TGTGATGGAAATAG---AAGTAACCTT
Human(734) AGAAAAGTAGGCCTGCTGGAGCTAGATGATGA-TGTGATGGAAATAG---AAGTAACCTT
Human(688) AGAAAAGTAGGCCTGCTGGAGCTAGATGATGA-TGTGATGGAAATAG---AAGTAACCTT
Human(491) ----------
Pig(548) TTGTAGGAGGTGGCTAATGAACATGCAGA-AATGGAAAAGAAAACTGTAATGCCCAAGAA exon 5
Pig(490) TTGTAGGAGGTGGCTAATGAACATGCA----------
Rabbit(725) CTACTCAATGGGACTAACAAATGTGGGGAAATTAAAAAGAAAAATCTAATACCTAAGAA
Rabbit(678) TTACCCAATGGGACTAACAAATGTGGGGAAAATT----------
Macaque(685) TTAGAGAATGTCGCTAAGAAACATGGCGA----------
Human(734) TTAGAGAATCTCGTTAAGAAACATGGAGAAAACGGAAAAGAAAAATGTAATGCCCTAGAA
Human(688) TTAGAGAATCTCGCTAAGAAACATGGAGAAAACGG----------
Human(491) ----------
Pig(548) AGTCCAAAGAAAGACAGTGGCAAGAA exon 5
Pig(490) ----------
Rabbit(725) AATCCAAAGAAAGACAGTAGC-----
Rabbit(678) ----------
Macaque(685) ----------
Human(734) AGCGCAAAGAAAGACAGTGGC-----
Human(688) ----------
Human(491) ----------

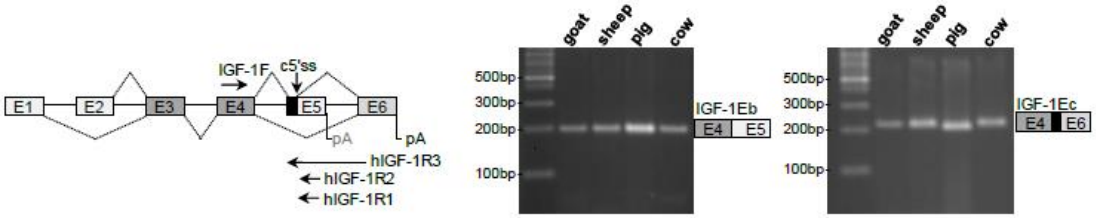
C

>IGF-1Ec E3 E4 E6

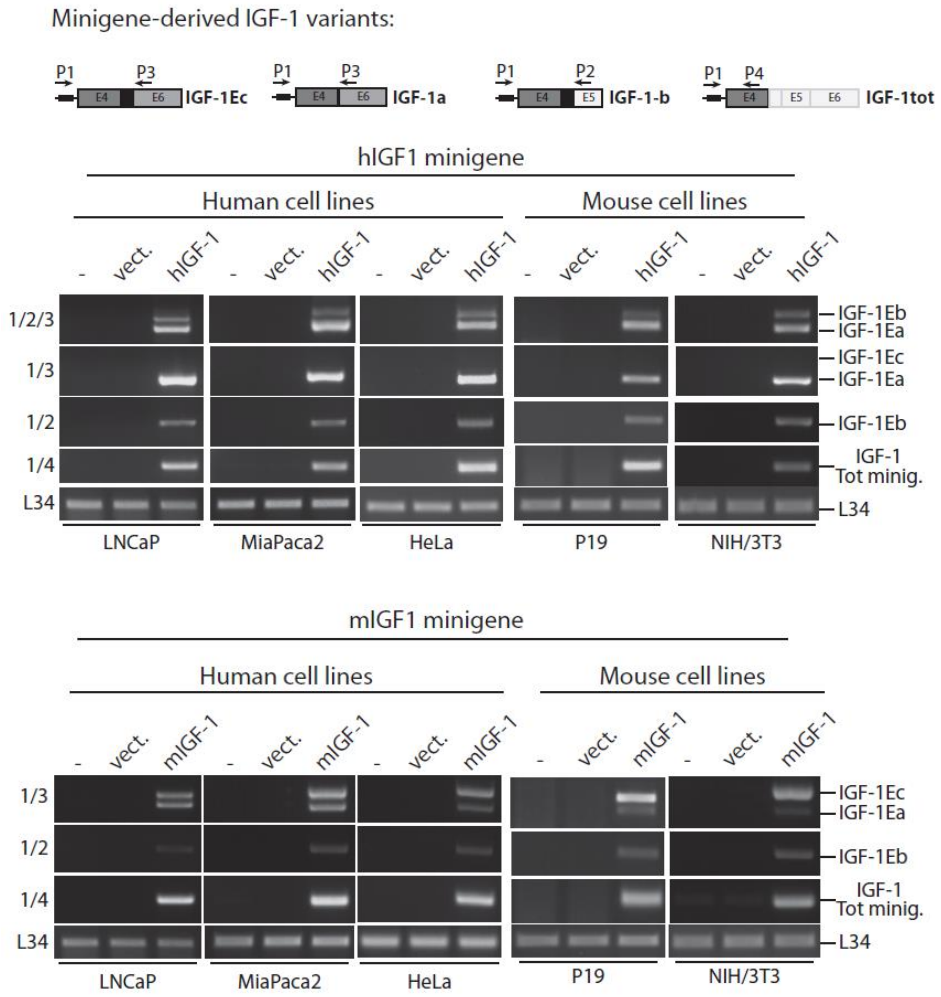
```
Rabbit(589)       CCTCCTCGCATCTCTTCTACCTGGCCCTCTGCTTGCTCACCTTCACCAGCTCGGCCACAG        exon 3
Macaque(589)      CCTCCTCGCATCTCTTCTACCTGGCACTGTGCCTGCTCACCTTCACCAGCTCCGCCACGG
Pig(401)          ------------------------------------------------------------
rat(466)          ------------------------------------------------------------
mouse(539)        ------------------------------------------------------------
mouse(450)        ------------------------------------------------------------

Rabbit(589)       CCGGACCGGAGACGCTCTGCGGTGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG        exon 3
Macaque(589)      CTGGACCGGAGACGCTCTGCGGGGCTGAGCTGGTGGATGCTCTTCAGTTCGTGTGTGGAG
Pig(401)          ------------------------------------------------------------
rat(466)          ------------------------------------------------------------
mouse(539)        ------------------------------------------------------------
mouse(450)        ------------------------------------------------------------

Rabbit(589)       ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGATACGGCTCCAGCAGTCGGAGGGCACCTC      exon 3/4
Macaque(589)      ACAGGGGCTTTTATTTCA/ACAAGCCCACAGGGTATGGCTCCAGCAGTCGGAGGGCACCTC
Pig(401)          ------------------/------------------------------------------
rat(466)          ------------------/------------GCTATGGCTCCAGCATTCGGAGGGCACCAC
mouse(539)        ------------------/------------GCTATGGCTCCAGCATTCGGAGGGCACCTC
mouse(450)        ------------------/------------GCTATGGCTCCAGCATTCGGAGGGCACCTC

Rabbit(589)       AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT        exon 4
Macaque(589)      AGACAGGCATCGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGAAGGCTGGAGATGT
Pig(401)          ----------CGTGGATGAGTGCTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
rat(466)          AGACGGGCATTGTGGATGAGTGTTGCTTCCGGAGCTGTGATCTGAGGAGGCTGGAGATGT
mouse(539)        AGACAGGCATTGTGGATGAGTGTTGCTTCCGGAGCTGTGATCTGAGGAGACTGGAGATGT
mouse(450)        AGACAGGCATTGTGGATGAGTGTTGCTTCCGGAGCTGTGATCTGAGGAGACTGGAGATGT
                  *********** ************************ ** **********
```


Rabbit(589) ACTGTGCACCCCTCAAGCCGGCAAAGGCAGCCCGCTCCGTCCGTGCCCAGCGCCACACCG exon 4
Macaque(589) ATTGCGCACCCCTCAAGCCTGCCAAGTCAGCCCGCTCTGTCCGTGCCCAGCGCCATACCG
Pig(401) ACTGTGCACCCCTCAAGCCTGCCAAGTCGGGCCCGCTCCGTCCGTGCCCAGCGCCACACGG
rat(466) ACTGTGCTCCGCTGAAGCCTACAAAGTCAGCTCGTTCCATCCGGGCCCAGCGCCACACTG
mouse(539) ACTGTGCCCCACTGAAGCCTACAAAAGCAGCCCGCTCTATCCGTGCCCAGCGCCACACTG
mouse(450) ACTGTGCCCCACTGAAGCCTACAAAAGCAGCCCGCTCTATCCGTGCCCAGCGCCACACTG
Rabbit(589) ACATGCCCAAGACTCAGAAG/TATCAGCCTCCATCTACCAACAAGAAAATGAAGTCTCAGA exon 4/5
Macaque(589) ACATGCCCAAGACCCAGAAG/TATCAGCCCCCATCTACCAACAAGAACACGAAGTCTCAGA
Pig(401) ACATGCCCAAGGCTCAGAAG/TATCAGCCCCCATCTACCAACAAGAAAACGAAGTCTCAGA
rat(466) ACATGCCCAAGACTCAGAAG/TCCCAGCCCCTATCGACACACAAGAAAAGGAAGCTGCAAA
mouse(539) ACATGCCCAAGACTCAGAAG/TCCCCGTCCCTATCGACAAACAAGAAAACGAAGCTGCAAA
mouse(450) ACATGCCCAAGACTCAGAAG/TCCCCGTCCCTATCGACAAACAAGAAAACGAAGCTGCAAA
Rabbit(589) GGAGAAGGAAAG/GAAGTACATTTGAAGAACACAAGTAGAGGGAGTGCAGGAAACAAGAAC exon 5/6
Macaque(589) GGAGAAGGAAAG/GAAGTACATTTGAAGAACGCAAGTAGAGGGAGTGCAGGAAACAAGAAC
Pig(401) GGCGAAGGAAAG/GAAGTACATTTGAAGAACACAAGTAGAGGGAGTTCAGGAAACAAGAAC
rat(466) GGAGAAGGAAAG/GAAGTACACTTGAAGAACACAAGTAGAGGAAGTGCAGGAAACAAGACC
mouse(539) GGAGAAGGAAAG/GAAGTACATTTGAAGAACACAAGTAGAGGAAGTGCAGGAAACAAGACC
mouse(450) GGAGAAGGAAAG/GAAGTACATTTGAAGAACACAAGTAGAGGAAGTGCAGGAAACAAGACC
Rabbit(589) TACAGGATGTAGGAAGACCCTTCTGAGGAGTGAAGAAGGACAGGCCACC-GCAGGACCCT exon 6
Macaque(589) TACAGAATGTAGGAAGACCCTCCTGAGGAGTGAAGAGTGACATGCCACC-GCAGGATCCT
Pig(401) TACAGAATGTAGGAAGACCTTCCTGAAGAGTGAAGAATGACATGCCACTGGCAGGATCCT
rat(466) TACAGAATGTAGGAGGAGCCTCCCGAGGAACAGAAAATGCCACGTCACC-GCAAGATCCT
mouse(539) TACAGAATGTAGGAGGAGCCTCCCACGGAGCAGAAAATGCCACATCACC-GCAGGATCCT
mouse(450) TACAGAATGTAGGAGGAGCCTCCCACGGAGCAGAAAATGCCACATCACC-GCAGGATCCT
Rabbit(589) TTGCTCTGCGCGAGTTACCTGTGAAACATTGGAATACCGGCCA------AAAAATAAGTT exon 6
Macaque(589) TTGCTCTGCACGAGTTACCTGTTAAACTTTGGAACACCTACCA------AAAAATAAGTT
Pig(401) TTGCTCTGCACGAGTTACCTGTTAAACACCAGAAGACCTACCAA-----AAAAATAAGTT
rat(466) TTG--CTGCTTGAGCAACCTGCAAAACATCGGAACACCTGCCAAATATCAATAATGAGTT
mouse(539) TTG--CTGCTTGAGCAACCTGCAAAACATCGAAACACCTACCAAATAACAATAATAAGTC
mouse(450) TTG--CTGCTTGAGCAACCTGCAAAACATCGAAACACCTACCAAATAACAATAATAAGTC
Rabbit(589) TGATGACATTTC-AAAGATGGGCATTTCCCCCAATGAAATACACAAGTAAACATTCC--- exon 6
Macaque(589) TGATAACATTTA-AAAGATGGGCGTTTCCCCCAATGAAATACACAAGTAAACATTCC---
Pig(401) TGAAAACATTTCAAAAGATGGGCATTCCCCCCAATGAAATACACAAGTAAACATTC----
rat(466) CAATACCATTTC-AGAGATGGGCATTTCCCTCAATGAAATACACAAGTAAACATTCCGAC
mouse(539) CAATAACATTAC-AAAGATGGGCATTTCCCCCAATGAAATATACAAGTAAACATTCCAAC
mouse(450) CAATAACATTAC-AAAGATGGGCATTTCCCCCAATGAAATATACAAGTAAAC--------
Rabbit(589) ---------- exon 6
Macaque(589) ----------
Pig(401) ----------
rat(466) ATTGTCTTTAGGAGTGTTTG----------
mouse(539) ATCGTCTTTAGGAGTGTTTGTTTAAAAAGCTTTGCACCTTGCAAAAGTGGTCCTGGCGTG
mouse(450) ----------
Rabbit(589) ---------- exon 6
Macaque(589) ----------
Pig(401) ----------
rat(466) ----------
mouse(539) GGTAGATTGCTGTTGGTCCTTTATCAATAACAT
mouse(450) ----------

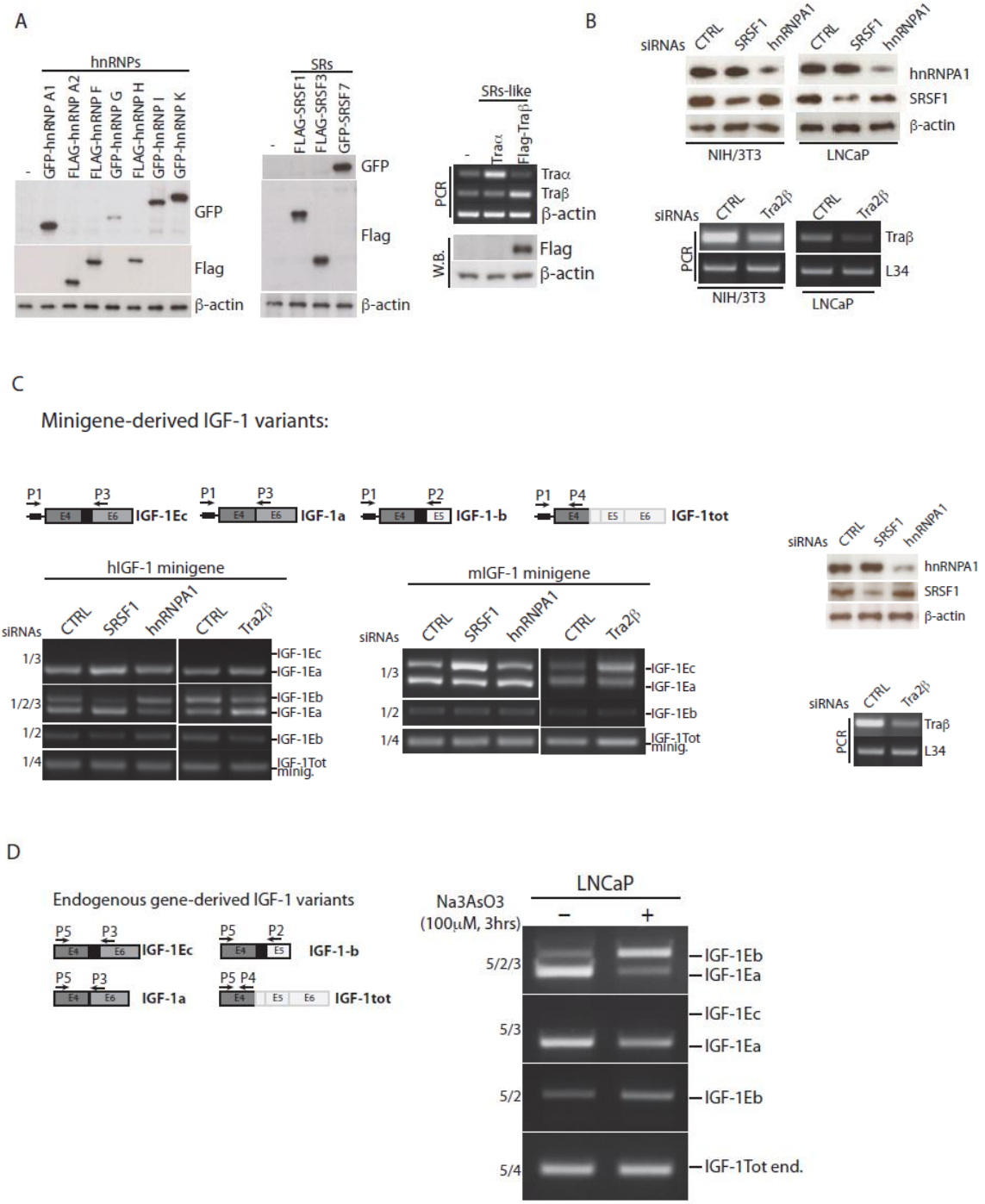
D



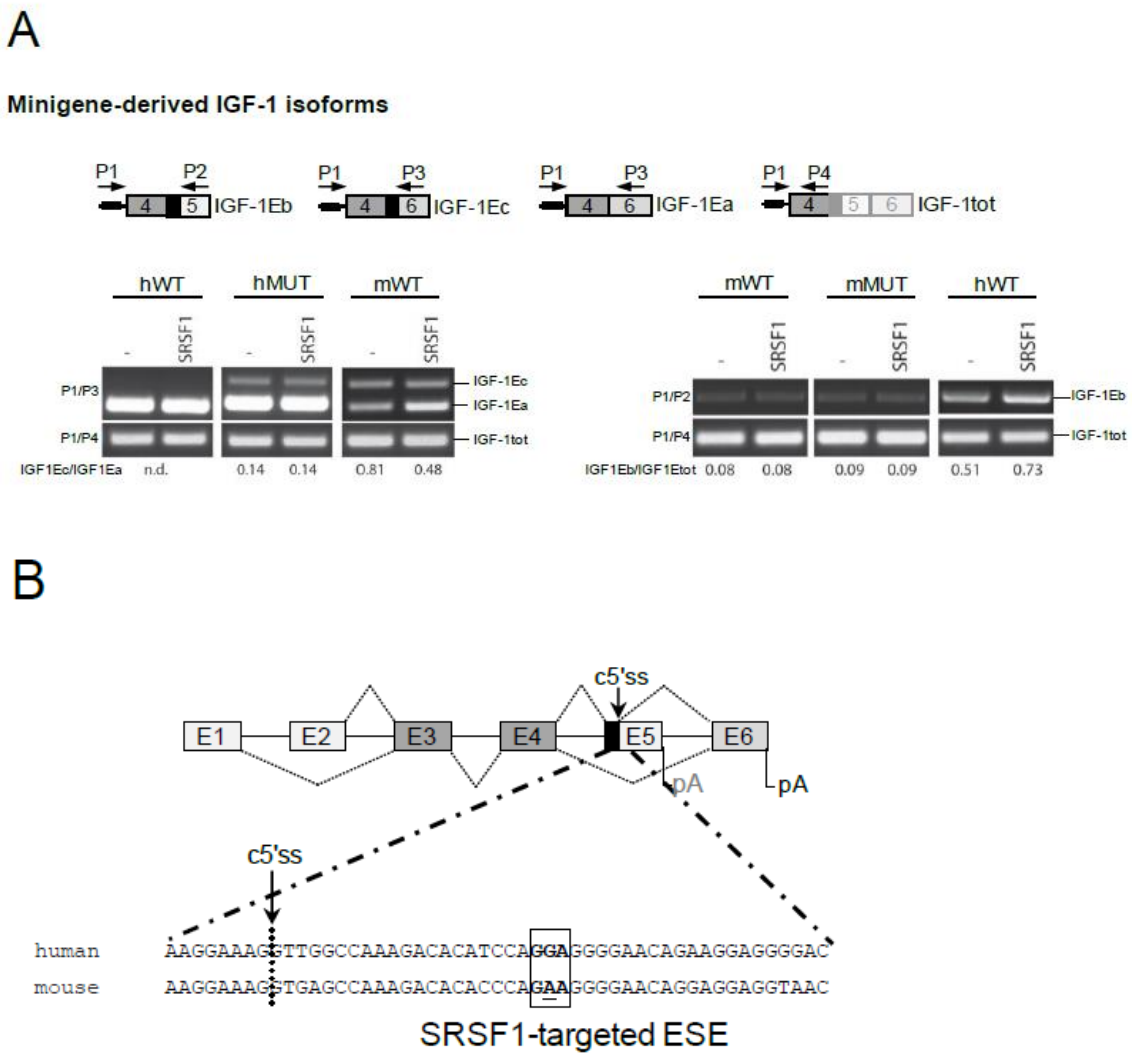
Supplementary Figure S3. Species-Specific differences in expression pattern of the IGF-1 isoforms among mammals. (A, B and C) 3' RACE-PCR sequences obtained from liver tissues of mammalian species (see Figure 3B). 3' RACE-PCR products were sequenced and the nucleotide sequences corresponding to IGF-1Ea (A), IGF-1-1Eb (B) and IGF-1Ec (C) were aligned using Clustal W [Thompson J.D. et al. Curr Protoc Bioinformatics Chapter 2 (2002) Unit 23]. Poly(A) Signal Miner (http://dnafsminer.bic.nus.edu.sg/PolyA.html) was used to predict poly(A) signal and, if detected in the sequences, highlighted in bold. PCR product sizes (nucleotides) for each sequence are indicated in parentheses. The nucleotide sequences matched to the following NCBI Reference Sequence (RefSeq): IGF-1Ea (Pig 350 and 283: NM_214256; Rabbit 537: XM_008256720/ XM_008256719; Macaque 537: NM_001260726; Human 627 and 534: NM_000618/ NM_001111284; Mouse 391: NM_001314010/ NM_001111275/ NM_001111276; Rat 381: NM_178866/ NM_001082479); IGF-1Eb (Pig 548 and 490: XM_005664196/ XM_005664195; Rabbit 725 and 678: XM_008256717/ XM_008256716; Macaque 685: NC_027903; Human 734, 688 and 491: NM_001111285); IGF-1Ec (Rabbit 589: XM_008256718; Macaque 589: XM_015152532/ XM_015152534; Pig 401: XM_005664197; Rat 466: NM_001082477/ NM_001082478; Mouse 539 and 450: NM_010512/ NM_001111274). (D) Amplification of IGF-1Eb and IGF-1Ec isoforms in liver tissue of goat, sheep, pig and cow with isoform-specific PCR strategy. Map of *IGF-1* gene with the relative position of primers used to separately amplify IGF-1 isoforms is shown on the left. The IGF-1F forward primer is common to all mammals (5'-CGTGGATGAGTGCTGCTTC-3'); the IGF-1R1 reverse primer is specific for IGF-1Eb amplification in goat, sheep and cow (5'-TCCTTCTGTTCCCCTCCTGG-3'); the IGF-1R2 reverse primer is specific for pig IGF-1Eb isoform (5'-CCCTCCTGGGTGTTTCTTTG-3'); the IGF-1R3 reverse primer is specific for IGF-1Ec amplification in all mammals (5'-CTTCAAATGTACTTCCTTTCC-3'). The conventional PCR was carried out as described in Material and Methods. The PCR products were loaded on 4.0% agarose gel and DNA fragments were eluted from gel, subcloned and sequenced. The nucleotide sequences matched to the following NCBI Reference Sequence (RefSeq): IGF-1Eb (Goat: XM_013963970/ XM_013963971/ XM_005680531/ XM_005680532/ XM_005680533/ XM_005680534/ XM_005680535; Sheep: NC_019460; Cow; XM_015471061/ XM_015471062/ XM_015471063/ XM_015471064; Pig: XM_005664196/XM_005664195); IGF-1Ec (Goat: NC_022297; Sheep: NC_019460; Cow: XM_005206490; Pig: XM_005664197).



Supplementary Figure S4. The expression pattern of human and mouse IGF-1 minigenes in different human and mouse cell lines. RT-PCR analysis of splicing assay of human (hIGF-1; *upper panel*) and mouse (mIGF-1, *lower panel*) IGF-1 minigenes performed in the indicated human and mouse cell lines. Untransfected cells (-) or cells transfected with the empty vector (vect.) were used as a PCR control. L34 was used as loading control. A diagram of primers position used for RT-PCR analysis is also shown.



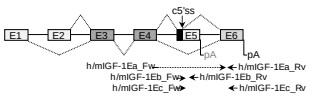
Supplementary Figure S5. Cis-acting elements contribute to species-specific variation of IGF-1 alternative splicing. (A,B) Representative Western-blot and RT-PCR analysis showing the expression level of hnRNP, SR and SR-like proteins in the splicing assay experiments. β-actin was used as loading control. C) RT-PCR analysis of splicing assay of human (hIGF-1) and mouse (mIGF-1) IGF-1 minigenes performed in HEK293T cells silenced with the indicated siRNAs. A representative western-blot showing the expression level of the silenced proteins is also shown. β-actin was used as loading control. D) RT-PCR analysis of splicing assay of endogenous human *IGF-1* gene performed in LNCaP treated for 3hrs with 100μM Sodium Arsenite (Na_3AsO_3).

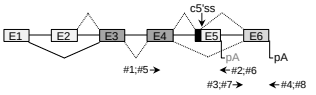


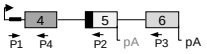
Supplementary Figure S6. (A) RT-PCR analysis of splicing assay performed in HEK293T cell line of mouse and human wild-type (WT), hMUT (TG*vs*GA) (*left panel*) and mMUT (GA*vs*TG) (*right panel*) IGF-1 minigenes in presence or not of SRSF1. The ratio between IGF-1Ec/IGF-1Ea (*left panel*) and IGF-1Eb/IGF-1tot (*right panel*) is also indicated. n.d.= not determined. A diagram of primers position used for RT-PCR analysis is also shown. (B) Representation of SRSF1 binding site (black box) in IGF-1 exon 5 identified by Smith PJ. et al. [Endocrinology (2002) 143:146-154] and Cléry A. et al. [PNAS (2013) 110:E2802-E2811]. Canonical cryptic 5' splice site (c5'ss) is also indicated.

Supplementary Table S1. List of the oligonucleotides used in this study.

#	Oligonucleotide name	sequence 5'→ 3'	Experiment	Primer annealing positions
/	hIGF1-Ea_Fw	GACATGCCCAAGACCCAGAAGGA	Real time RT-PCR (human and macaque)	
/	hIGF1-Ea_Rv	CGGTGGCATGTCACTCTTCACTC	Real time RT-PCR (human and macaque)	
/	hIGF1-Eb_Fw	CTACCAACAAGAACACGAAGTCTCA	Real time RT-PCR (human and macaque)	
/	hIGF1-Eb_Rv	TCTGATCTGCAGACTTGCTTCTG	Real time RT-PCR (human and macaque)	
/	hIGF1-Ec_Fw	GCCCCCATCTACCAACAAGAACAC	Real time RT-PCR (human and macaque)	
/	hIGF1-Ec_Rv	TCCCTCTACTTGCGTTCTTCAAA	Real time RT-PCR (human and macaque)	
/	mIGF1-Ea_Fw	GCCCAAGACTCAGAAGGAAGTAC	Real time RT-PCR (mouse)	
/	mIGF1-Ea_Rv	CGGTGATGTGGCATTTTCTG	Real time RT-PCR (mouse)	
/	mIGF1-Eb_Fw	CACACTGACATGCCCAAGAC	Real time RT-PCR (mouse)	
/	mIGF1-Eb_Rv	AGTTGCCTCCGTTACCTCCT	Real time RT-PCR (mouse)	
/	mIGF1-Ec_Fw	GCTGCAAAGGAGAAGGAAAG	Real time RT-PCR (mouse)	
/	mIGF1-Ec_Rv	CGGTGATGTGGCATTTTCTG	Real time RT-PCR (mouse)	
1	mIGF1-KPN1_Fw (P5)	ACGGTACCACAAGCCCACAGG	Mouse minigene (5'end)/splicing assay of endog. gene	
2	mIGF1-SalI_Rv	ACGTCGACCCCGATGTTCTTTGG	Mouse minigene (5'end)	
3	mIGF1-SalI_Fw	ACGTCGACCCCGATGTTCTTTGG	Mouse minigene (3'end)	
4	mIGF1-NotI_Rv	ACGCGGCCGCTTGCACCCTCCTGGAAA	Mouse minigene (3'end)	
5	hIGF1-EcoRI_Fw (P5)	ACGAATTCACAAGCCCACAGGGTA	Human minigene (5'end)/splicing assay of endog. gene	
6	hIGF1-SalI_Rv	ACGTCGACCCTAAACCCTGGACT	Human minigene (5'end)	
7	hIGF1-SalI_Fw	ACGTCGACTCTAGAGGCCAGAAGGCTGGATT	Human minigene (3'end)	
8	hIGF1-NotI_Rv	ACGCGGCCGCTGTTGCACCCTTACAG	Human minigene (3'end)	
9	mIGF1-GAvsTG_Fw	AAGGAAAGGTTGGCCAAAGACA	Mouse GAvsTG minigene (5'end)	
10	hIGF1-TGvsGA_Fw	AAGGAAAGGTGAGCCAAAGACA	Human TGvsGA minigene (5'end)	
11	PCI_Fw (P1)	GGTGTCCACTCCCAGTTCAA	Splicing assay	
12	m/hIGF1-Ex6_Rv (P3)	CACTTCCTCTACTTGTGTTCTTCAAATGTAC	Splicing assay	
13	mIGF1-Ex5_Rv (P2)	TTTCCGAGTTGCCTCCGTTA	Splicing assay	
14	hIGF1-Ex5_Rv (P2)	TTCCCCTCCTGGATGTGTCT	Splicing assay	
15	mIGF1-Ex4tot_Rv (P4)	GATCCAGCTCCGGAAGCAA	Splicing assay	
16	hIGF1-Ex4tot_Rv (P4)	TTGAGGGGTGCGCAATACAT	Splicing assay	
17	Tra2α_Fw	AAGGAGCCGAAGCCATTCTC	Expression level evaluation	
18	Tra2α_Rv	GCCAAACACTCCAAGGCAAG	Expression level evaluation	
19	Tra2β_Fw	CCAGCAGTCTAGGCGTTCAA	Expression level evaluation	
20	Tra2β_Rv	TGATCCTACGCCCATCAAGC	Expression level evaluation	
21	β-actin_Fw	ATGATGATATCGCCGCGCTC	Expression level evaluation	
22	β-actin_Rv	AGGGTGAGGATGCCTCTCTT	Expression level evaluation	
23	Control siRNA	AGACGAACAAGUCACCGAC-[dT] [dT]	Protein depletion experiments	
24	SRSF1 siRNA	CCAAGGACAUUGAGGACGT-[dT] [dT]	Protein depletion experiments	
25	hnRNP A1 siRNA	CAGCUGAGGAAGCTCTTCA-[dT] [dT]	Protein depletion experiments	
26	TRA2β siRNA # 1	AAGCUAAAGAACGTGCCAA-[dT] [dT]	Protein depletion experiments	
27	TRA2β siRNA # 2	GCCGAUGUGUCUAUUGUAU-[dT] [dT]	Protein depletion experiments	
28	TRA2β siRNA # 3	CUGUUGUCUUGGAGUAUUU-[dT] [dT]	Protein depletion experiments	







Supplementary file S1. IGF-1 isoform nucleotide sequences of the 27 mammalian species used for IGF-1 evolutionary analysis. The "CDS FASTA alignment from multiple alignments" data, derived from the "multiz100way" alignment data prepared from 100 vertebrate genomes [Blanchette M. et al. Genome Res (2004), 14, 708-715], were downloaded using the Table Browser tool of the UCSC Genome Browser. Sequences were subsequently realigned using MUSCLE [Edgar R.C. Nucleic Acids Res (2004), 32, 1792-1797] and protein-coding sequences from 27 mammalian species were extracted from these alignment datasets. The phylogenetic tree in Newick format was constructed using the IGF-1 core alignment of

6. References

[1] E.R. Barton, The ABCs of IGF-I isoforms: impact on muscle hypertrophy and implications for repair, Appl Physiol Nutr Metab 31 (2006) 791-797.

[2] L. Temmerman, E. Slonimsky, N. Rosenthal, Class 2 IGF-1 isoforms are dispensable for viability, growth and maintenance of IGF-1 serum levels, Growth Horm IGF Res 20 (2010) 255-263.

[3] A.M. Oberbauer, The Regulation of IGF-1 Gene Transcription and Splicing during Development and Aging, Front Endocrinol (Lausanne) 4 (2013) 39.

[4] S.L. Chew, P. Lavender, A.J. Clark, R.J. Ross, An alternatively spliced human insulin-like growth factor-I transcript with hepatic tissue expression that diverts away from the mitogenic IBE1 peptide, Endocrinology 136 (1995) 1939-1944.

[5] C.T. Roberts, Jr., S.R. Lasky, W.L. Lowe, Jr., W.T. Seaman, D. LeRoith, Molecular cloning of rat insulin-like growth factor I complementary deoxyribonucleic acids: differential messenger ribonucleic acid processing and regulation by growth hormone in extrahepatic tissues, Mol Endocrinol 1 (1987) 243-248.

[6] S. Yang, M. Alnaqeeb, H. Simpson, G. Goldspink, Cloning and characterization of an IGF-1 isoform expressed in skeletal muscle subjected to stretch, J Muscle Res Cell Motil 17 (1996) 487-495.

[7] M. Hill, G. Goldspink, Expression and splicing of the insulin-like growth factor gene in rodent muscle is associated with muscle satellite (stem) cell activation following local tissue damage, J Physiol 549 (2003) 409-418.

[8] P.J. Smith, E.L. Spurrell, J. Coakley, C.J. Hinds, R.J. Ross, A.R. Krainer, S.L. Chew, An exonic splicing enhancer in human IGF-I pre-mRNA mediates recognition of alternative exon 5 by the serine-arginine protein splicing factor-2/alternative splicing factor, Endocrinology 143 (2002) 146-154.

[9] A. Clery, R. Sinha, O. Anczukow, A. Corrionero, A. Moursy, G.M. Daubner, J. Valcarcel, A.R. Krainer, F.H. Allain, Isolated pseudo-RNA-recognition motifs of SR proteins can regulate splicing using a noncanonical mode of RNA recognition, Proc Natl Acad Sci U S A 110 (2013) E2802-2811.

[10] P. Rotwein, Editorial: the fall of mechanogrowth factor?, Mol Endocrinol 28 (2014) 155-156.

[11] A. Armakolas, M. Kaparelou, A. Dimakakos, E. Papageorgiou, N. Armakolas, A. Antonopoulos, C. Petraki, M. Lekarakou, P. Lelovas, M. Stathaki, C. Psarros, I. Donta, P.S. Galanos, P. Msaouel, V.G. Gorgoulis, M. Koutsilieris, Oncogenic Role of the Ec Peptide of the IGF-1Ec Isoform in Prostate Cancer, Mol Med 21 (2015) 167-179.

[12] R.W. Matheny, Jr., B.C. Nindl, M.L. Adamo, Minireview: Mechano-growth factor: a putative product of IGF-I gene expression involved in tissue repair and regeneration, Endocrinology 151 (2010) 865-875.

[13] Y. Kajimoto, P. Rotwein, Structure of the chicken insulin-like growth factor I gene reveals conserved promoter elements, J Biol Chem 266 (1991) 9724-9731.

[14] A.M. Sparkman, T.S. Schwartz, J.A. Madden, S.E. Boyken, N.B. Ford, J.M. Serb, A.M. Bronikowski, Rates of molecular evolution vary in vertebrates for insulin-like growth factor-1 (IGF-1), a pleiotropic locus that regulates life history traits, Gen Comp Endocrinol 178 (2012) 164-173.

[15] D.M. Tiago, V. Laize, M.L. Cancela, Alternatively spliced transcripts of Sparus aurata insulin-like growth factor 1 are differentially expressed in adult tissues and during early development, Gen Comp Endocrinol 157 (2008) 107-115.

[16] M. Reinecke, C. Collet, The phylogeny of the insulin-like growth factors, Int Rev Cytol 183 (1998) 1-94.

[17] M. Wallis, New insulin-like growth factor (IGF)-precursor sequences from mammalian genomes: the molecular evolution of IGFs and associated peptides in primates, Growth Horm IGF Res 19 (2009) 12-23.

[18] Y. Xing, C. Lee, Evidence of functional selection pressure for alternative splicing events that accelerate evolution of protein subsequences, Proc Natl Acad Sci U S A 102 (2005) 13526-13531.

[19] M.F. Lin, P. Kheradpour, S. Washietl, B.J. Parker, J.S. Pedersen, M. Kellis, Locating protein-coding sequences under selection for additional, overlapping functions in 29 mammalian genomes, Genome Res 21 (2011) 1916-1928.

[20] J.V. Chamary, J.L. Parmley, L.D. Hurst, Hearing silence: non-neutral evolution at synonymous sites in mammals, Nat Rev Genet 7 (2006) 98-108.

[21] Y. Xing, C. Lee, Can RNA selection pressure distort the measurement of Ka/Ks?, Gene 370 (2006) 1-5.

[22] M. Blanchette, W.J. Kent, C. Riemer, L. Elnitski, A.F. Smit, K.M. Roskin, R. Baertsch, K. Rosenbloom, H. Clawson, E.D. Green, D. Haussler, W. Miller, Aligning multiple genomic sequences with the threaded blockset aligner, Genome Res 14 (2004) 708-715.

[23] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res 32 (2004) 1792-1797.

[24] J.D. Thompson, T.J. Gibson, D.G. Higgins, Multiple sequence alignment using ClustalW and ClustalX, Curr Protoc Bioinformatics Chapter 2 (2002) Unit 2 3.

[25] G. Yeo, C.B. Burge, Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals, J Comput Biol 11 (2004) 377-394.

[26] J. Jurka, Repbase update: a database and an electronic journal of repetitive elements, Trends Genet 16 (2000) 418-420.

[27] T.J. Wheeler, J. Clements, S.R. Eddy, R. Hubley, T.A. Jones, J. Jurka, A.F. Smit, R.D. Finn, Dfam: a database of repetitive DNA based on profile hidden Markov models, Nucleic Acids Res 41 (2013) D70-82.

[28] T.J. Wheeler, S.R. Eddy, nhmmer: DNA homology search with profile HMMs, Bioinformatics 29 (2013) 2487-2489.

[29] G. Annibalini, M. Guescini, D. Agostini, R.D. Matteis, P. Sestili, P. Tibollo, M. Mantuano, C. Martinelli, V. Stocchi, The expression analysis of mouse interleukin-6 splice variants argued against their biological relevance, BMB Rep 45 (2012) 32-37.

[30] S. Barik, Site-directed mutagenesis in vitro by megaprimer PCR, Methods Mol Biol 57 (1996) 203-215.

[31] P. Bielli, R. Busa, S.M. Di Stasi, M.J. Munoz, F. Botti, A.R. Kornblihtt, C. Sette, The transcription factor FBI-1 inhibits SAM68-mediated BCL-X alternative splicing and apoptosis, EMBO Rep 15 (2014) 419-427.

[32] S.L. Pond, S.D. Frost, S.V. Muse, HyPhy: hypothesis testing using phylogenies, Bioinformatics 21 (2005) 676-679.

[33] S.V. Muse, Estimating synonymous and nonsynonymous substitution rates, Mol Biol Evol 13 (1996) 105-114.

[34] R. Szklarczyk, J. Heringa, S.K. Pond, A. Nekrutenko, Rapid asymmetric evolution of a dual-coding tumor suppressor INK4a/ARF locus contradicts its function, Proc Natl Acad Sci U S A 104 (2007) 12807-12812.

[35] M. Varadi, M. Guharoy, F. Zsolyomi, P. Tompa, DisCons: a novel tool to quantify and classify evolutionary conservation of intrinsic protein disorder, BMC Bioinformatics 16 (2015) 153.

[36] Z. Dosztanyi, V. Csizmok, P. Tompa, I. Simon, IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content, Bioinformatics 21 (2005) 3433-3434.

[37] R. Sorek, The birth of new exons: mechanisms and evolutionary consequences, RNA 13 (2007) 1603-1608.

[38] I. Vorechovsky, Transposable elements in disease-associated cryptic exons, Hum Genet 127 (2010) 135-154.

[39] A.M. Oberbauer, J.M. Belanger, G. Rincon, A. Canovas, A. Islas-Trejo, R. Gularte-Merida, M.G. Thomas, J.F. Medrano, Bovine and murine tissue expression of insulin like growth factor-I, Gene 535 (2014) 101-105.

[40] S. Naftelberg, I.E. Schor, G. Ast, A.R. Kornblihtt, Regulation of alternative splicing through coupling with transcription and chromatin structure, Annu Rev Biochem 84 (2015) 165-198.

[41] N.N. Singh, M.N. Lawler, E.W. Ottesen, D. Upreti, J.R. Kaczynski, R.N. Singh, An intronic structure enabled by a long-distance interaction serves as a novel target for splicing correction in spinal muscular atrophy, Nucleic Acids Res 41 (2013) 8144-8165.

[42] E. Kovacs, P. Tompa, K. Liliom, L. Kalmar, Dual coding in alternative reading frames correlates with intrinsic protein disorder, Proc Natl Acad Sci U S A 107 (2010) 5429-5434.

[43] M. Macossay-Castillo, S. Kosol, P. Tompa, R. Pancsa, Synonymous constraint elements show a tendency to encode intrinsically disordered protein segments, PLoS Comput Biol 10 (2014) e1003607.

[44] J. Brosius, S.J. Gould, On "genomenclature": a comprehensive (and respectful) taxonomy for pseudogenes and other "junk DNA", Proc Natl Acad Sci U S A 89 (1992) 10706-10710.

[45] L. Lin, P. Jiang, S. Shen, S. Sato, B.L. Davidson, Y. Xing, Large-scale analysis of exonized mammalian-wide interspersed repeats in primate genomes, Hum Mol Genet 18 (2009) 2204-2214.

[46] A.P. de Koning, W. Gu, T.A. Castoe, M.A. Batzer, D.D. Pollock, Repetitive elements may comprise over two-thirds of the human genome, PLoS Genet 7 (2011) e1002384.

[47] B. Mersch, N. Sela, G. Ast, S. Suhai, A. Hotz-Wagenblatt, SERpredict: detection of tissue- or tumor-specific isoforms generated through exonization of transposable elements, BMC Genet 8 (2007) 78.

[48] N. Sela, B. Mersch, N. Gal-Mark, G. Lev-Maor, A. Hotz-Wagenblatt, G. Ast, Comparative analysis of transposed element insertion within human and mouse genomes reveals Alu's unique role in shaping the human transcriptome, Genome Biol 8 (2007) R127.

[49] R. Baertsch, M. Diekhans, W.J. Kent, D. Haussler, J. Brosius, Retrocopy contributions to the evolution of the human genome, BMC Genomics 9 (2008) 466.

[50] M. Krull, M. Petrusma, W. Makalowski, J. Brosius, J. Schmitz, Functional persistence of exonized mammalian-wide interspersed repeat elements (MIRs), Genome Res 17 (2007) 1139-1145.

[51] F.S. de Souza, L.F. Franchini, M. Rubinstein, Exaptation of transposable elements into novel cis-regulatory elements: is the evidence always strong?, Mol Biol Evol 30 (2013) 1239-1251.

[52] J. Schmitz, J. Brosius, Exonization of transposed elements: A challenge and opportunity for evolution, Biochimie 93 (2011) 1928-1934.

[53] H. Keren, G. Lev-Maor, G. Ast, Alternative splicing and evolution: diversification, exon definition and function, Nat Rev Genet 11 (2010) 345-355.

[54] N.L. Barbosa-Morais, M. Irimia, Q. Pan, H.Y. Xiong, S. Gueroussov, L.J. Lee, V. Slobodeniuc, C. Kutter, S. Watt, R. Colak, T. Kim, C.M. Misquitta-Ali, M.D. Wilson, P.M. Kim, D.T. Odom, B.J. Frey, B.J. Blencowe, The evolutionary landscape of alternative splicing in vertebrate species, Science 338 (2012) 1587-1593.

[55] J. Merkin, C. Russell, P. Chen, C.B. Burge, Evolutionary dynamics of gene and isoform regulation in Mammalian tissues, Science 338 (2012) 1593-1599.

[56] T. Lappalainen, M. Sammeth, M.R. Friedlander, P.A. t Hoen, J. Monlong, M.A. Rivas, M. Gonzalez-Porta, N. Kurbatova, T. Griebel, P.G. Ferreira, M. Barann, T. Wieland, L. Greger, M. van Iterson, J. Almlof, P. Ribeca, I. Pulyakhina, D. Esser, T. Giger, A. Tikhonov, M. Sultan, G. Bertier, D.G. MacArthur, M. Lek, E. Lizano, H.P. Buermans, I. Padioleau, T. Schwarzmayr, O. Karlberg, H. Ongen, H. Kilpinen, S. Beltran, M. Gut, K. Kahlem, V. Amstislavskiy, O. Stegle, M. Pirinen, S.B. Montgomery, P. Donnelly, M.I. McCarthy, P. Flicek, T.M. Strom, H. Lehrach, S. Schreiber, R. Sudbrak, A. Carracedo, S.E. Antonarakis, R. Hasler, A.C. Syvanen, G.J. van Ommen, A. Brazma, T. Meitinger, P. Rosenstiel, R. Guigo, I.G. Gut, X. Estivill, E.T. Dermitzakis, Transcriptome and genome sequencing uncovers functional variation in humans, Nature 501 (2013) 506-511.

[57] A. Necsulea, H. Kaessmann, Evolutionary dynamics of coding and non-coding transcriptomes, Nat Rev Genet 15 (2014) 734-748.

[58] Q. Gao, W. Sun, M. Ballegeer, C. Libert, W. Chen, Predominant contribution of cis-regulatory divergence in the evolution of mouse alternative splicing, Mol Syst Biol 11 (2015) 816.

[59] D.L. Black, Mechanisms of alternative pre-messenger RNA splicing, Annu Rev Biochem 72 (2003) 291-336.

[60] M. Chen, J.L. Manley, Mechanisms of alternative splicing regulation: insights from molecular and genomics approaches, Nat Rev Mol Cell Biol 10 (2009) 741-754.

[61] M. Sandri, L. Barberi, A.Y. Bijlsma, B. Blaauw, K.A. Dyar, G. Milan, C. Mammucari, C.G. Meskers, G. Pallafacchina, A. Paoli, D. Pion, M. Roceri, V. Romanello, A.L. Serrano, L. Toniolo, L. Larsson, A.B. Maier, P. Munoz-Canoves, A. Musaro, M. Pende, C. Reggiani, R. Rizzuto, S. Schiaffino, Signalling pathways regulating muscle mass in ageing skeletal muscle: the role of the IGF1-Akt-mTOR-FoxO pathway, Biogerontology 14 (2013) 303-323.

[62] V. Gotea, W. Makalowski, Do transposable elements really contribute to proteomes?, Trends Genet 22 (2006) 260-267.

[63] F. Pagani, M. Raponi, F.E. Baralle, Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution, Proc Natl Acad Sci U S A 102 (2005) 6368-6372.

[64] Z.E. Sauna, C. Kimchi-Sarfaty, Understanding the contribution of synonymous mutations to human disease, Nat Rev Genet 12 (2011) 683-691.

[65] M. Buljan, G. Chalancon, S. Eustermann, G.P. Wagner, M. Fuxreiter, A. Bateman, M.M. Babu, Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks, Mol Cell 46 (2012) 871-883.

[66] G. Thieulin-Pardo, L. Avilan, M. Kojadinovic, B. Gontero, Fairy "tails": flexibility and function of intrinsically disordered extensions in the photosynthetic world, Front Mol Biosci 2 (2015) 23.

[67] J. Durzynska, A. Philippou, B.K. Brisson, M. Nguyen-McCarty, E.R. Barton, The pro-forms of insulin-like growth factor I (IGF-I) are predominant in skeletal muscle and alter IGF-I receptor activation, Endocrinology 154 (2013) 1215-1224.

[68] M. De Santi, G. Annibalini, E. Barbieri, A. Villarini, L. Vallorani, S. Contarelli, F. Berrino, V. Stocchi, G. Brandi, Human IGF1 pro-forms induce breast cancer cell proliferation via the IGF1 receptor, Cell Oncol (Dordr) (2015).

[69] M. Buljan, A. Frankish, A. Bateman, Quantifying the mechanisms of domain gain in animal proteins, Genome Biol 11 (2010) R74.

[70] P.R. Romero, S. Zaidi, Y.Y. Fang, V.N. Uversky, P. Radivojac, C.J. Oldfield, M.S. Cortese, M. Sickmeier, T. LeGall, Z. Obradovic, A.K. Dunker, Alternative splicing in concert with protein intrinsic disorder enables increased functional diversity in multicellular organisms, Proc Natl Acad Sci U S A 103 (2006) 8390-8395.

[71] V.N. Uversky, The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini, FEBS Lett 587 (2013) 1891-1901.

CHAPTER 2

HUMAN IGF1 PRO-FORMS INDUCE BREAST CANCER CELL PROLIFERATION VIA THE IGF1 RECEPTOR

Mauro De Santi[1], Giosuè Annibalini[2], Elena Barbieri[2], Anna Villarini[3], Luciana Vallorani[2], Serena Contarelli[2], Franco Berrino[3], Vilberto Stocchi[2] and Giorgio Brandi[1]

[1]Department of Biomolecular Sciences, Hygiene Unit, University of Urbino Carlo Bo, Urbino (PU), Italy.
[2]Department of Biomolecular Sciences, Exercise and Health Sciences Unit, University of Urbino Carlo Bo, Urbino (PU), Italy.
[3]Epidemiology & Prevention Unit, Department of Preventive & Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

Published in Cellular Oncology (Dordr), 39(2):149-159, 2016.

Human IGF1 pro-forms induce breast cancer cell proliferation via the IGF1 receptor

Mauro De Santi[1], Giosuè Annibalini[2], Elena Barbieri[2], Anna Villarini[3], Luciana Vallorani[2], Serena Contarelli[2], Franco Berrino[3], Vilberto Stocchi[2] and Giorgio Brandi[1]

Authors' affiliation

[1]Department of Biomolecular Sciences, Hygiene Unit, University of Urbino Carlo Bo, Urbino (PU), Italy.

[2]Department of Biomolecular Sciences, Exercise and Health Sciences Unit, University of Urbino Carlo Bo, Urbino (PU), Italy.

[3]Epidemiology & Prevention Unit, Department of Preventive & Predictive Medicine, Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy.

Corresponding Author:

Mauro De Santi, PhD

Department of Biomolecular Sciences, Hygiene Unit

University of Urbino Carlo Bo

Via S. Chiara, 27 - 61029 Urbino (PU) Italy

Phone: +39 0722 303545

Fax: +39 0722 303541

email: mauro.desanti@uniurb.it

Abstract

Background IGF1 is a key regulator of tissue growth and development and has been implicated in the initiation and progression of various cancers, including breast cancer. Through IGF1 mRNA splicing different precursor pro-peptides, *i.e.*, the IGF1Ea, IGF1Eb and IGF1Ec pro-forms, are formed whose biological roles in the pathogenesis of breast cancer have not been established yet. The objective of this study was to assess the biological activity of the IGF1 pro-forms in human breast cancer-derived cells.

Methods The different IGF1 pro-forms were generated through transient transfection of HEK293 cells with the respective vector constructs. The resulting conditioned media were applied *in vitro* to MCF7, T47D and ZR751 breast cancer-derived cell cultures. The *recombinant* human IGF1 pro-forms were also tested for their binding affinity to an anti-IGF1 specific antibody by immunoprecipitation. To determine whether the IGF1 pro-forms induce cell proliferation, mature IGF1 was neutralised in HEK293-derived conditioned media.

Results We found that the IGF1 pro-forms were the only forms that were produced intracellularly, whereas both mature IGF1 and the IGF1 pro-forms were detected extracellularly. We also found that E peptides can impair the IGF1 pro-form binding affinity for the anti-IGF1 antibody and, thus, hamper an accurate measurement of the IGF1 pro-forms. Additionally, we found that the IGF1 antibody can completely inhibit IGF1-induced breast cancer cell proliferation and IGF1 receptor (IGF1R) phosphorylation, wheras the same antibody was found to only partially inhibit the biological activity of the pro-forms. Moreover, we found that the IGF1 pro-form activities can completely be inhibited by neutralising the IGF1R. Finally, we compared the bioactivity of the IGF1 pro-forms to that of mature IGF1, and found that the IGF1 pro-forms were less capable of phosphorylating the IGF1R in the breast cancer-derived cells tested.

Conclusions Our data indicate that IGF1 pro-forms can induce breast cancer cell proliferation via the IGF1R, independent from the mature IGF1 form. These results underline the importance of an accurate assessment of the presence of IGF1 pro-forms within the breast cancer microenvironment.

Keywords: IGF1 pro-forms · breast cancer · IGF1 receptor

1. Introduction

Insulin-like growth factor-1 (IGF1) plays an important role in normal tissue growth and development. In addition, several studies have shown associations between circulating IGF1 levels and the risk to develop breast cancer[1-3]. Since the IGF1 receptor (IGF1R) is over-expressed in about 90% of the breast cancer cases and since IGF1R levels are higher in breast cancer cells than in normal breast tissues [4], targeting the IGF1 system appears to be an attractive therapeutic option.

IGF1 is synthesized as a precursor protein requiring proteolysis at both the N- and C-termini to produce mature IGF1 [5, 6]. The full-length IGF1 precursor, pre-pro-IGF1, contains an N-terminal signal peptide, a 70 amino acid mature IGF1 peptide and a C-terminal E-peptide extension [7]. The signal peptide is cleaved off during translation in the endoplasmic reticulum, resulting in pro-IGF1. The E-peptide can subsequently be cleaved off from pro-IGF1 by proprotein convertases like furin, resulting in mature IGF1. The uncleaved pro-IGF1 has, however, also been detected *in vitro* in conditioned media and *in vivo* in sera [8-13].

The complexity of the IGF1 system is further enhanced by alternative splicing of the IGF1 mRNA, thereby producing multiple IGF1 isoforms (IGF1 pro-forms) that, while bearing the same mature IGF1 sequence, contain different N- and C-terminal extensions [5]. In humans, the alternative splicing that occurs at the 3' end of the IGF1 mRNA gives rise to three possible IGF1 pro-forms with different C-terminal extensions, called the Ea, Eb and Ec peptides (Fig. 1A). Another level of complexity results from glycosylation of the IGF1Ea pro-form, as the human Ea-peptide of IGF1 contains an N-linked glycosylation site at Asn92 [6].

Recent studies in humans have shown that the IGF1 splice variants can be differentially transcribed in response to varying conditions and pathologies, such as skeletal muscle damage [14, 15], endometriosis [16], or prostate [17], cervix [18] and colorectal cancer [19]. Moreover, although it is generally assumed that IGF1 exerts its biological actions predominantly through the mature peptide, different biological activities have been reported for the different IGF1 pro-forms and/or for their E-peptides, either exogenously administrated or over-expressed in various *in vitro* models [6, 14, 17, 20, 21].

Even though circulating IGF1 levels are affected by physical activity and diet [22], the biological significance of the IGF1 pro-forms is currently unknown. Also, the

physiological and molecular mechanisms that regulate their expression and their circulating levels are unclear [6]. Despite the fact that the regenerative properties of the IGF1Ea pro-form in cardiac and skeletal muscles has extensively been documented [21, 29, 30], little is known about the role of the various IGF1 pro-forms in cancer.

Here we report the biological activity of IGF1 pro-forms on human breast cancer-derived cell lines. We analysed the intracellular and extracellular expression patterns of the IGF1 pro-forms in transfected HEK293 cells and conditioned media. MCF7, T47D and ZR751 breast cancer-derived cells were grown in conditioned media to assess whether the IGF1 pro-forms induce cell proliferation and/or IGF1R phosphorylation. We further evaluated the bioactivities of the IGF1 pro-forms compared to the mature IGF1 form, in terms of cellular proliferation and IGF1R, AKT or ERK1/2 phosphorylation.

2. Materials and Methods

2.1 Cell cultures

The MCF7, T47D, ZR751 and HEK293 cell lines were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA). The cells were cultured in DMEM (MCF7 and HEK293) or RPMI-1640 (T47D and ZR751) media supplemented with 10% fetal bovine serum, 10 mg/L insulin (MCF7 and T47D), 2 mmol/L L-glutamine, 1x MEM Non-essential Amino Acid Solution, 0.1 mg/ml streptomycin and 0.1 U/L penicillin (growth media). Cells were maintained in a humidified incubator (5 % CO_2) at 37 °C during at maximum fifteen passages.

For the experiments, the breast cancer-derived cells were starved overnight in red phenol-free DMEM or RPMI-1640 media without FBS, after which the media were replaced by the same media with or without hormones. All cell culture materials were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2 MTS cell proliferation assay

Triplicate samples of 5×10^3 MCF7, T47D and ZR751 cells in 96-well plates were treated for 4 days with mature IGF1 or IGF1 pro-forms. Cell viabilities were evaluated using a CellTiter 96® Aqueous Non-Radioactive Cell Proliferation Assay (Promega, Madison, WI, USA) based on the ability of viable cells to convert soluble tetrazolium salt (MTS)

into a formazan product, as reported before [23]. The results are expressed as the relative number of viable cells in treated samples relative to controls (untreated cells).

2.3 Plasmid constructs

Plasmid constructs containing sequences encoding human prepro-IGF1Ea, prepro-IGF1Eb and prepro-IGF1Ec were kindly provided by Dr. Joanne Tonkin and Dr. Tommaso Nastasi, European Molecular Biology Laboratory (EMBL), Monterotondo (Rome, Italy). Each plasmid contained DNA encoding the class 1 IGF1 48-amino acid signal peptide, the mature 70-amino acid IGF1 peptide, the first 16 amino acids (aa) of the COOH-terminal peptide, and C-terminal sequences encoding either the Ea (19 aa), the Eb (61 aa) or the Ec (24 aa) peptide.

2.4 Cell transfection assays

HEK293 cells were cultured in DMEM without antibiotics at a density of $1x10^6$/well in 6 well plates. After overnight incubation, the cells were transfected using a TransIT®-LT1 Transfection Reagent (Mirus Bio, Madison, WI, USA) according the manufacturer's instructions. Briefly, 2.5 µg plasmid DNA was added to 250 µl growth medium without FBS and antibiotics and gently mixed, after which 7.5 µl TransIT®-LT1 Reagent was added. After a 30 min incubation at room temperature, the mixture was added drop-wise to the cells. After a 5 hour incubation, the culture medium was replaced by red phenol-free DMEM without FBS. Next, the cells were incubated for another 24 hours, after which supernatants were collected, clarified by 1,000 rpm centrifugation for 5 min, and directly used or stored at -80° C for further experiments. Transfected HEK293 cells were lysed for Western blotting or real-time PCR analyses. To increase IGF1 pro-form production, a furin convertase inhibitor chloromethylketone (CMK) (Enzo Life Sciences Inc., Farmingdale, NY, USA) was added at a 2.5 µmol/L final concentration during transfection. For E peptide cleavage, supernatants without CMK were treated with 10 nmol/L recombinant furin (R&D Systems Ltd, Minneapolis, MN, USA) overnight at room temperature with gently shaking [6].

2.5 ELISA assay

For the quantitative determination of human IGF1 concentrations in transfected HEK293 cell culture supernatants, a commercially available ELISA kit was used according to the

manufacturer's instructions (Quantikine® ELISA DG100, R&D Systems). Data were acquired in duplicate using a microplate reader (Multiskan EX, Thermo Fisher Scientific, Waltham, MA, USA) at 450 nm, after which the results were averaged.

2.6 Western blot analysis

MCF7, T47D, ZR751 and HEK293 cells were processed for Western blot analysis as previously reported [24]. Briefly, cells were lysed for 20 minutes on ice with 20 mmol/L HEPES (pH 7.9), 25 % v/v glycerol, 0.42 mol/L NaCl, 0.2 mmol/L EDTA, 1.5 mmol/L $MgCl_2$, 0.5 % v/v Nonidet P-40, 1 mmol /L DTT, 1 mmol/L Naf, 1 mmol/L Na_3VO_4, and 1× complete protease inhibitor cocktail (Roche Diagnostics Ltd, Mannheim, Germany). The cell lysates were frozen and thawed twice and clarified by centrifugation at 12,000 rpm for 10 minutes at 4°C. The proteins from the HEK293 cell supernatants were concentrated using an Amicon Ultra 3K centrifugal filter unit (Merck Millipore, Billerica, MA, USA). Total cell lysates and concentrated supernatants were fractionated by SDS-PAGE and electroblotted onto nitrocellulose membranes (0.2 μm pore size) (Bio-Rad Laboratories Inc., Hercules, CA, USA). The resulting blots were probed with the following primary antibodies: anti-phospho-IGF1 Receptor β (3024), anti-IGF1 Receptor β (3027), anti-phospho-p44/42 (ERK1/2) (9101), anti-p44/42 (ERK1/2) (9102), anti-phospho-Akt (Ser473) (9271) and anti-Akt (9272), all purchased from Cell Signalling Technology (Beverly, MA, USA) and anti-IGF1 (I8773) purchased from Sigma-Aldrich. Protein bands were detected using a horseradish peroxidase-conjugated secondary antibody (Bio-Rad Laboratories Inc). The blots were treated with enhanced chemiluminescence reagents (ECL Kit, Amersham Bioscience, Arlington Heights, IL, USA), and the immunoreactive bands were detected and quantified using a Chemi-Doc System (Bio-Rad Laboratories Inc) equipped with Quantity One software.

2.7 RNA extraction, cDNA synthesis and qRT-PCR

Total RNA was extracted and purified using an Omega Bio-Tek E.Z.N.A.™ Total RNA kit (Omega Bio-Tek, Norcross, GA, USA) according to the manufacturer's instructions. After digestion with DNase I (Qiagen, Hilden, Germany), cDNA was synthesized from 1 μg of total RNA using Omniscript RT (Qiagen) and random hexamers. Subsequently, quantitative real-time PCR was performed using an Applied Biosystems StepOnePlus™ Real Time PCR System in conjunction with a TaqMan® Universal PCR Master Mix No

AmpErase® UNG and commercially available 6-carboxyfluorescein (FAM)-labeled TaqMan primers for human IGF1 (Hs01547656_m1) and GAPDH (Hs03929097_g1), respectively (Applied Biosystems, Foster City, CA, USA). mRNA expression data were generated using the $2^{-\Delta CT}$ method. The real-time PCR conditions were: 95°C for 10 min followed by 40 cycles of two-steps at 95°C for 15 sec and 60°C for 1 min. The specificity of the amplification products was confirmed by thermal denaturation plots and by separation in 4% agarose gels.

2.8 Immunoprecipitation assay

To prepare magnetic beads for immunoprecipitation, Dynabeads® Protein G (Life Technologies, Monza, Italy) were washed twice with PBS/0.1% Tween-20 and incubated with 5 μg of IGF1 antibody (Sigma) for 1 hour at room temperature with end-over-end rotation. The bead-antibody complexes were washed with PBS/0.1% Tween-20 after which the IGF1 monoclonal antibody was covalently bound to the beads using BS^3 as cross-linkers according to the manufacturer's instructions (Thermo Scientific, Milano, Italy). Subsequently, the beads were washed three times with PBS/0.1% Tween-20 to remove non-covalently bound antibodies and incubated with 1 ml of tissue culture supernatant for 1 hour at room temperature with end-over-end rotation. Finally, the beads were washed three times with washing buffer and the bound proteins were eluted by heating the beads for 10 minutes at 70°C in 20 μl elution buffer and 10 μl SDS-PAGE sample buffer.

2.9 IGF1 and IGF1R neutralisation

To neutralise IGF1 activity, culture medium containing IGF1 or IGF1 pro-forms was incubated with 3 μg/ml anti-IGF1 antibody (Sigma) for 1 hour at 37°C. Next, MCF7 and ZR751 cells were cultured in IGF1-neutralised media to evaluate their effects on cell proliferation and IGF1R phosphorylation. To neutralise IGF1R activity, cells were pre-incubated with 5 μg/ml of anti-IGF1R antibody (R&D Systems) for 1 hour at 37°C and treated with IGF1 or IGF1 pro-forms to evaluate their effects on cell proliferation and IGF1R phosphorylation.

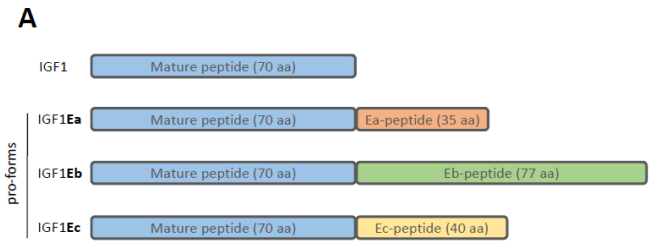
2.10 Statistical analyses

Statistical analyses were performed using one-way or two-way ANOVA as appropriate, followed by Bonferroni's multiple comparison post hoc tests (GraphPad Software, Inc., La Jolla, CA, USA).

3. Results

3.1 Expression of IGF1 and its pro-forms in HEK293 transfected cells

IGF1 pro-forms were generated through transient transfection of HEK293 cells with specific plasmid vector constructs for each pro-form. The resulting cell lysates and supernatants were analyzed by Western blotting using an antibody directed against the mature region of IGF1. The amount of IGF1 and its pro-forms in the supernatants were quantified by ELISA and concentrated using filter columns, after which 50 ng was loaded on gel. By doing so, we found that the IGF1 pro-forms were the only forms produced intra-cellularly by the transfected HEK293 cells, whereas both mature IGF1 and the IGF1 pro-forms were detected extra-cellularly (Fig. 1b). Notably, both glycosylated and non-glycosilated IGF1Ea were detected in the cell lysates, whereas only the glycosylated IGF1Ea pro-form (gly-IGF1Ea) was secreted. Moreover, we found that both IGF1Eb and IGF1Ec showed additional higher molecular weight bands, suggesting that also these pro-forms are subject to extensive post-translational modification.



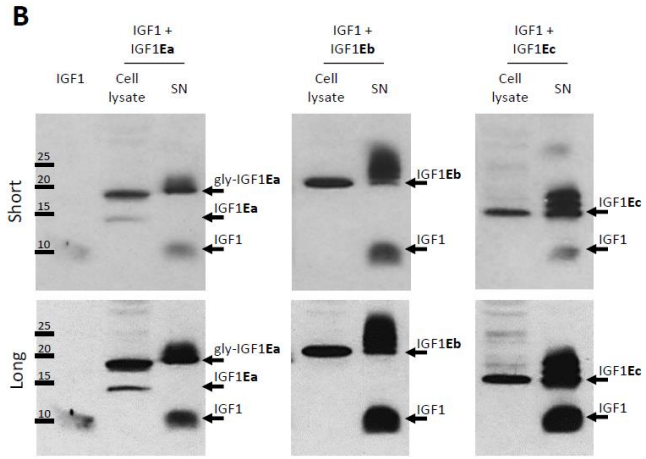


Fig. 1. Human IGF1 pro-forms. (**a**) Schematic presentation of mature IGF1 and IGF1 pro-forms. (**b**) Western blot analysis of HEK293 cells transfected with specific constructs. Cell lysates and supernatants (SN) were analysed 24 hours post-transfection using an anti-IGF1 antibody. Images are representative of three replicates giving similar results.

3.2 E peptides impair an accurate quantification of the IGF1 pool

As previously reported [13], the quantity of the gly-IGF1Ea pro-form may be underestimated in non-denaturating conditions such as ELISA assays, suggesting that the E peptide could impair the binding affinity between IGF1 and anti-IGF1 antibodies. In order to evaluate whether the ELISA assay provides an accurate measure of the IGF pool, HEK293 cells were transfected with IGF1Ea, IGF1Eb and IGF1Ec expression vectors with or without the furin inhibitor CMK. In doing so, the same IGF1 mRNA expression efficiencies were obtained in CMK treated and untreated cells (Supplementary Fig. S1). Next, the supernatants were analysed by both ELISA and Western blotting after filter column concentration for IGF1 quantification. As shown in Fig. 2a, Western blot analysis of conditioned media from the CMK treated or untreated HEK293 cells did not show any significant variation in the total IGF1 pool (*i.e.*, mature IGF1 and pro-forms). IGF1 quantification by ELISA of gly-IGF1Ea enriched medium did not show any change after CMK treatment, whereas a significant reduction of IGF1 was observed, after CMK treatment, in IGF1Eb and IGF1Ec enriched media (Fig. 2b). Therefore, we conclude that the E peptides in the IGF1 pro-forms may impair their affinity to the anti-IGF1 antibody under non-denaturating conditions and, hence, hamper the accuracy of the ELISA assay.

To further confirm this notion, conditioned media obtained from HEK293 cells transfected with the IGF1 pro-form vectors were immunoprecipitated with Dynabeads

coupled to an anti-IGF1 antibody. The proteins bound to the bead-antibody complex were subsequently recovered and analysed by Western blotting. As shown in Fig. 2c, the gly-IGF1Ea pro-form was, at least partially, recognized by the anti-IGF1 antibody, whereas the IGF1Eb and IGF1Ec pro-forms were only weakly recognized by the anti-IGF1 antibody in the immunoprecipitates. These results confirm that E peptides, in the IGF1 pro-forms, hamper an accurate measurement of the IGF1 pool.

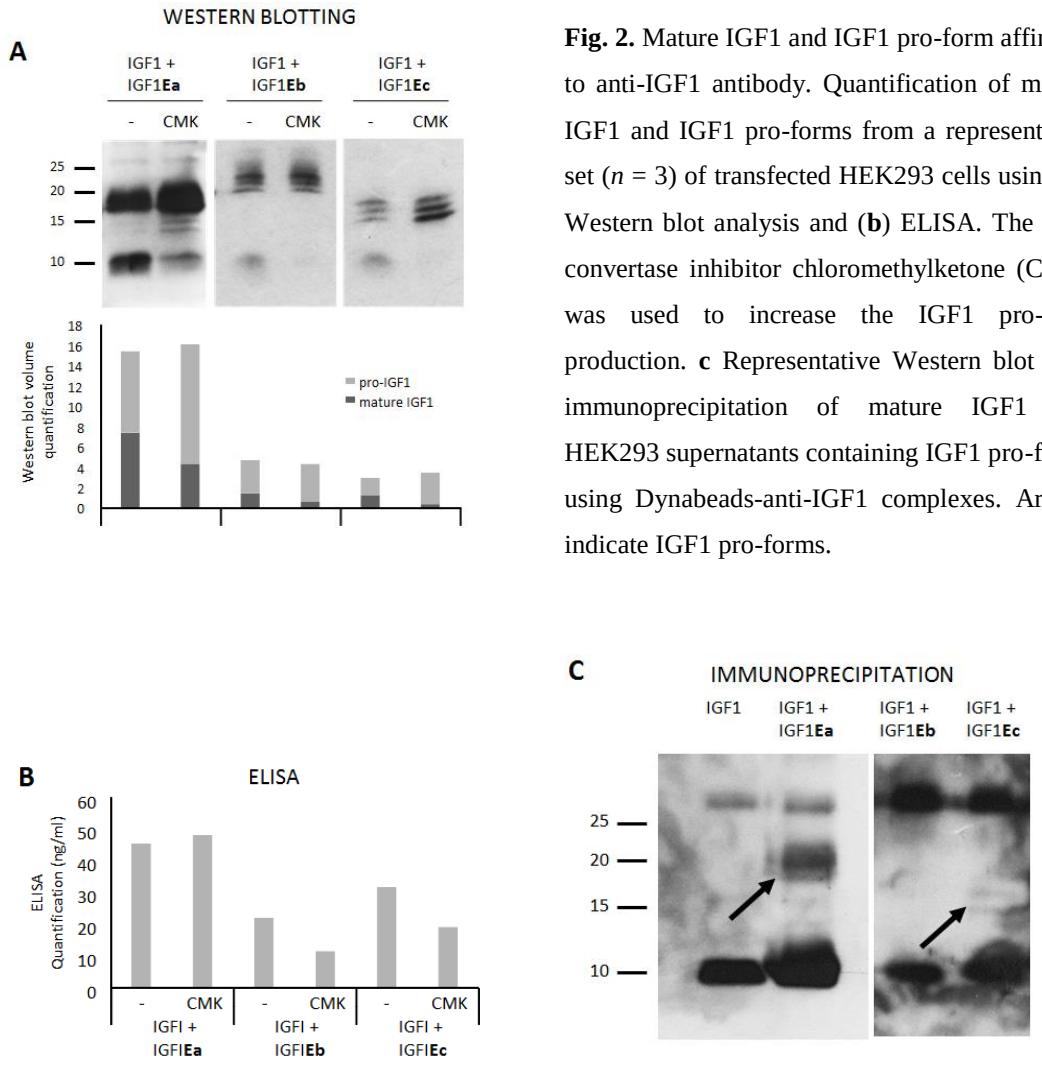


Fig. 2. Mature IGF1 and IGF1 pro-form affinities to anti-IGF1 antibody. Quantification of mature IGF1 and IGF1 pro-forms from a representative set (n = 3) of transfected HEK293 cells using (**a**) Western blot analysis and (**b**) ELISA. The furin convertase inhibitor chloromethylketone (CMK) was used to increase the IGF1 pro-form production. **c** Representative Western blot after immunoprecipitation of mature IGF1 and HEK293 supernatants containing IGF1 pro-forms using Dynabeads-anti-IGF1 complexes. Arrows indicate IGF1 pro-forms.

3.3 Biological activity of IGF1Ea, IGF1Eb and IGF1Ec enriched media

Next, the activity of each IGF1 pool in the MCF7 and ZR751 human breast cancer-derived cells was evaluated in terms of cell proliferation and IGF1R phosphorylation. Cell proliferation was evaluated using a MTS cell proliferation assay, an indirect assay

that evaluates the cellular metabolic activity. First, MCF7 cells were grown in the presence of increasing concentrations of mature IGF1. After 4 days of culture, the cellular proliferation was evaluated by both the MTS assay and by cell counting. As shown in supplementary Fig. 2S, both methods yielded similar results. Subsequently, the cells were cultured in IGF1 pro-form-enriched media, previously normalised to 10 ng/ml using an ELISA assay. Since it was not possible to accurately quantify the total IGF1 pool (*i.e.*, mature IGF1 and pro-forms) in IGF1 pro-form-enriched media using an ELISA assay (see above), we were unable to directly compare the effects between each IGF1 pool. We found, however, that the IGF1 pro-form-enriched media significantly induced both MCF7 and ZR751 cell proliferation compared to the control (unstimulated) cells (Fig. 3a and b). It was not possible to evaluate the proliferation response in T47D cells due to their poor growth in serum free medium (not shown). Supernatants of un-transfected HEK293 cells or HEK293 cells transfected with an empty vector did not affect cell proliferation (not shown). Next, the anti-IGF1 antibody was used to neutralize the activity of mature IGF1. As shown in Fig. 3a and b, we found that the anti-IGF1 antibody completely inhibited the IGF1-induced cell proliferation, whereas the same antibody only partially inhibited MCF7 (Fig. 3a) and ZR751 (Fig. 3b) cell proliferation induced by HEK293 supernatants containing the IGF1 pro-forms. Moreover, we found that the anti-IGF1 antibody markedly inhibited the IGF1R phosphorylation induced by mature IGF1, but not the phosphorylation induced by the IGF1 pro-forms (Fig. 3c and d). These results suggest that the IGF1 pro-forms can induce breast cancer cell proliferation and IGF1R phosphorylation. The activity of the IGF1 pro-forms is IGF1R dependent. In fact, by inhibiting IGF1R activation with an anti-IGF1R antibody, neither cell proliferation nor IGF1R phosphorylation, induced by either mature IGF1 or the IGF1 pro-forms, were detected (Fig. 3).

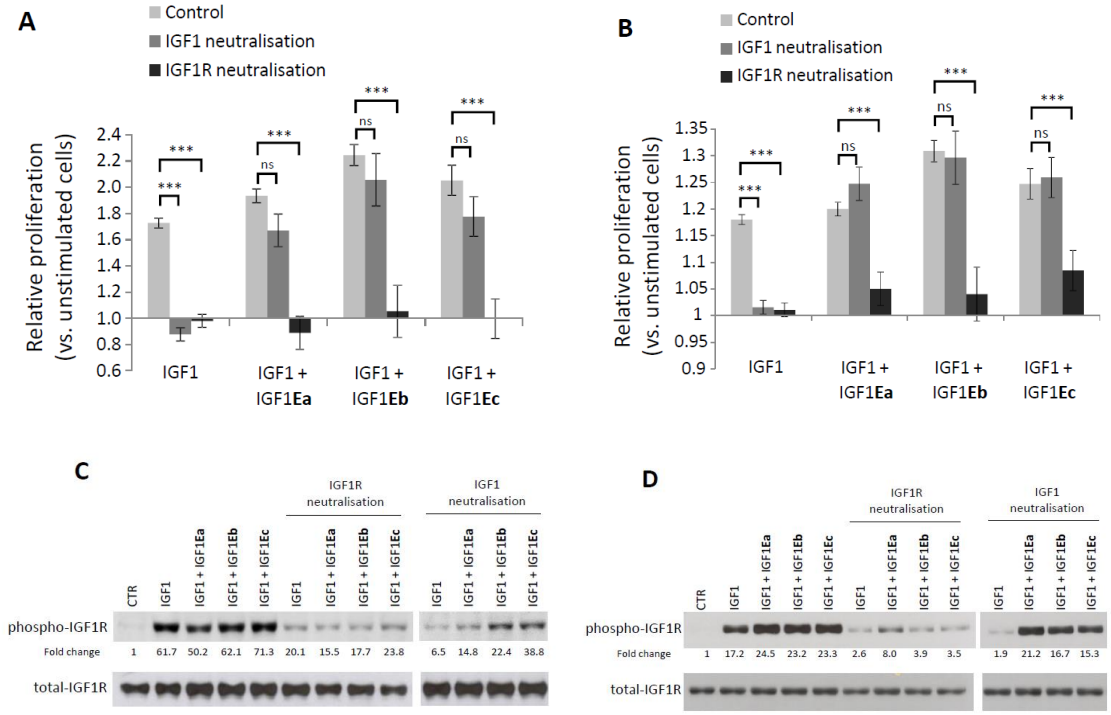


Fig. 3. IGF1 pro-forms induce cell proliferation via the IGF1R. (**a**) MCF7 and (**b**) ZR751 cells were cultured 4 days with mature IGF1 (10 ng/ml) or HEK293 supernatants containing IGF1 pro-forms (means ± SEM; *n* = 3). Cell proliferation was evaluated by MTS assay. Data are expressed as relative proliferation *vs.* unstimulated cells. *** Significantly different, $P < 0.001$; ns: not significantly different; 1-way ANOVA followed by Bonferroni's multiple comparison test. Representative Western blot (*n* = 3) of phospho-IGF-R levels in (**c**) MCF7 and (**d**) ZR751 cells stimulated for 10 minutes with mature IGF1 or HEK293 supernatants containing IGF1 pro-forms. IGF1R was used as a loading control. Densitometry values for specific proteins relative to unstimulated cells (set as one-fold) are included below the lanes. An anti-IGF1 antibody was used to neutralise the biological activity of IGF1. An anti-IGF1R antibody was used to inhibit IGF1R phosphorylation/activation.

3.4 Biological activity of mature IGF1 versus the IGF1 pro-forms

To evaluate the activity of the IGF1 pro-forms compared to mature IGF1, supernatants containing different ratios of mature IGF1 and pro-forms were produced. Recombinant furine was used to induce E peptide cleavage and to increase the amount of mature IGF1. The furin convertase inhibitor CMK was used to inhibit E peptide cleavage and to increase the IGF1 pro-form amounts during transfection. The supernatants were concentrated using filter columns and analysed by Western blotting using an anti-IGF1 antibody. We found that CMK markedly increased the IGF1 pro-form amounts, while in

the furine-treated supernatants the IGF1 pro-forms were not detectable (Fig. 4a, 4b and 4c).

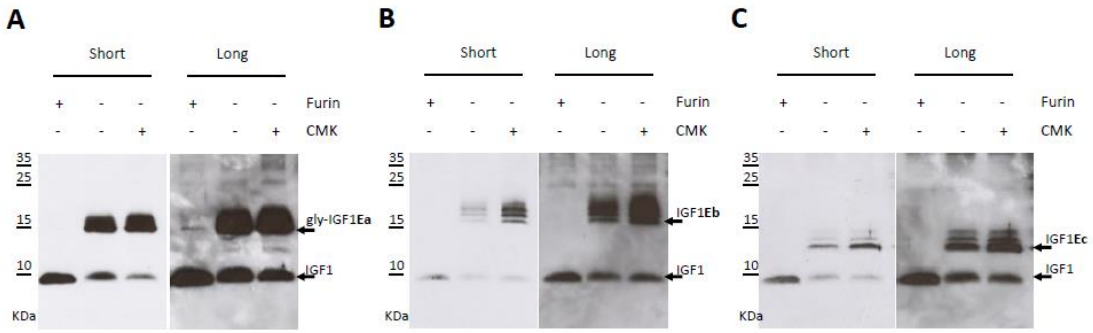


Fig. 4. IGF1 pro-form production and E peptides cleavage. Representative Western blots are shown for supernatants of HEK293 cells transfected with specific constructs for the (**a**) IGF1Ea, (**b**) IGF1Eb and (**c**) IGF1Ec pro-forms. Mature IGF1 and IGF1 pro-forms were detected using an anti-IGF1 antibody. The furin convertase inhibitor CMK was used to increase IGF1 pro-form production. Recombinant furin was used to induce E peptides cleavage.

Next, MCF7 cells were cultured in two-fold diluted conditioned media (from 1:4 to 1:32) containing different ratios of mature IGF1 and the IGF1 pro-forms, after which cellular proliferation and IGF1R, AKT and ERK1/2 phosphorylation were evaluated at the indicated time points. No significant differences in MCF7 cell proliferation were detected (supplementary Fig. S3a-b-c), but by increasing the amount of the gly-IGF1Ea pro-form in the cell culture medium, IGF1R phosphorylation was found to be markedly reduced (Fig. 5a), suggesting a minor binding affinity of gly-IGF1Ea for the IGF1R. While increasing the amount of gly-IGF1Ea also reduced AKT phosphorylation in MCF7 cells, no effect on ERK1/2 phosphorylation was observed (Fig. 5a). Decreased levels of IGF1R phosphorylation were also observed in MCF7 cells cultured with higher amounts of IGF1Ec and, partially, IGF1Eb, whereas no differences in AKT and ERK1/2 phosphorylation were observed (Fig. 5b and 5c).

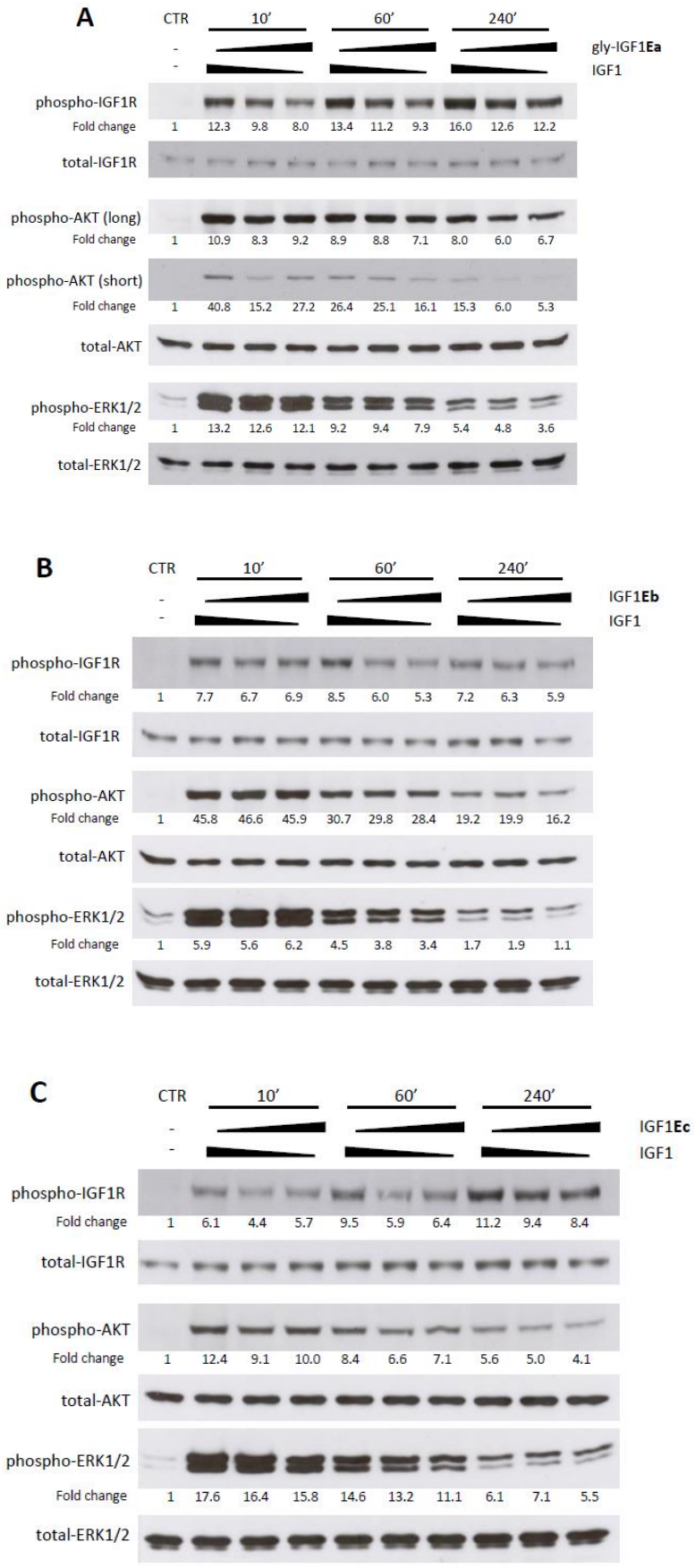


Fig. 5. Phosphorylation of IGF1R, AKT and ERK1/2 in MCF7 cells. Representative Western blot (*n* = 3) showing phospho-IGFR, phospho-AKT and phospho-ERK1/2 levels in MCF7 cells stimulated for the indicated times with HEK293 supernatants containing mature IGF1 and the (**a**) IGF1Ea, (**b**) IGF1Eb and (**c**) IGF1Ec pro-forms. IGF1R, AKT and ERK1/2 were used as loading controls. Densitometry values for specific proteins relative to unstimulated cells (set as one-fold) are included below the lanes.

The activity of the IGF1 pro-forms compared to mature IGF1 was also evaluated in T47D and ZR751 breast cancer-derived cells (Fig. 6). In conformity with the results obtained with MCF7 cells, we found that the glycosylated IGF1Ea and IGF1Ec pro-forms were less potent in phosphorylating IGF1R in both the T47D (Fig. 6a) and ZR751 (Fig. 6b) cells. Additionally, we found that furin and its inhibitor CMK did not alter the phosphorylation status of IGF1R, AKT and ERK1/2 when induced by mature IGF1 (supplementary Fig. S3d).

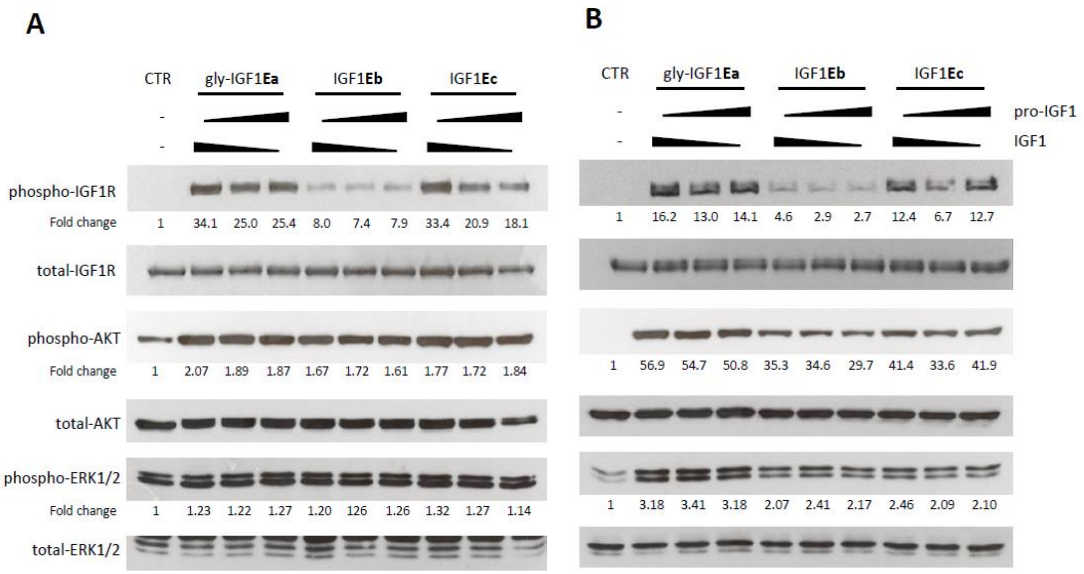


Fig. 6. Phosphorylation of IGF1R, AKT and ERK1/2 in (**a**) T47D and (**b**) ZR751 cells. Representative Western blot (n = 3) showing phospho-IGF1R, phospho-AKT and phospho-ERK1/2 levels in cells stimulated for 60 minutes with HEK293 supernatants containing mature IGF1 and the IGF1Ea, IGF1Eb and IGF1Ec pro-forms. IGF1R, AKT and ERK1/2 were used as loading controls. Densitometry values for specific proteins relative to unstimulated cells (set as one-fold) are included below the lanes.

4. Discussion

The IGF1 pathway plays a well-documented role in the development and/or progression of breast carcinomas [2]. IGF1 mRNA splicing events generate different precursor IGF1 polypeptides, namely the IGF1Ea, IGF1Eb and IGF1Ec pro-forms in humans, that share the mature peptide, but differ by the structure of their extension peptides, or E-peptides, in the C-terminus [5, 6]. The IGF1 pro-forms also undergo posttranslational modifications, such as glycosylation and proteolytic processing by proprotein convertases such as furin [6]. Convertase-mediated cleavage generally occurs intra-cellularly [25], but

it has also been reported that there are potential proprotein convertases that may process pro-IGF1 extra-cellularly, resulting in the secretion of unprocessed IGF1 pro-forms [10, 13]. Our data confirm this latter notion by revealing that the IGF1 pro-forms are the predominant forms inside the transfected HEK293 cells, and that they are also abundantly secreted in the cell culture media. Our results also showed that the non-glycosylated IGF1Ea form was detectable in the cell lysates only, whereas only the glycosylated form was secreted. The Ea-peptide of human IGF1 is a unique E peptide that contains an N-linked glycosylation site, and it has been hypothesized that its glycosylation may play a role in IGF1 biological activity modulation, such as bioavailability [26]. Interestingly, our data revealed that both IGF1Eb and IGF1Ec are subject to posttranslational modifications. These modifications still require detailed characterization.

As reported by Durzyńska *et al.* [13], ELISA measurements are more sensitive to mature IGF1 than to the IGF1 pro-forms, suggesting that the presence of the E-peptide may impair the ability of the anti-IGF1 antibody to recognize the native protein. Our results support this hypothesis, showing that the anti-IGF1 antibody has a higher affinity for mature IGF1 compared to the IGF1 pro-forms, especially IGF1Eb and IGF1Ec. Moreover, according to the literature [13], the ELISA quantification appears to be impaired in supernatants with large amounts of pro-forms after CMK treatment during the transfections, even though our Western blotting results did not show a decrease in total IGF1. Thus, it is difficult to compare the bioactivities of the different pro-forms, as the E-peptide in pro-IGF1 hamper the ability to accurately measure and subsequently normalize the IGF1 content under non-denaturing conditions.

Since it was unclear whether pro-IGF1 is bioactive or simply an inactive precursor or source for mature IGF1 [7], we cultured MCF7 and ZR751 cells in IGF1-neutralised conditioned media. As expected, the anti-IGF1 antibody was able to completely neutralise the activity of mature IGF1 in terms of the induction of cell proliferation and IGF1R phosphorylation. On the contrary, the anti-IGF1 antibody was found to be ineffective in inhibiting the proliferation and IGF1R phosphorylation in cells cultured in conditioned media containing the IGF1 pro-forms (Fig. 7). These results suggest that the IGF1 pro-forms are able to induce breast cancer cell proliferation. *In vitro* studies have suggested that the E-peptides of the human IGF1 precursors may act as independent growth factors, inducing mitosis independently from IGF1R [6]. In contrast, we found that by neutralising the IGF1R, the induction of cell proliferation by mature IGF1 or the

IGF1 pro-forms was completely inhibited, suggesting that IGF1 pro-forms induce cell proliferation via IGF1R activation (Fig. 7).

Despite the vast amount of evidence regarding the biological activity of E peptides, little is known about the biological activities of the IGF1 pro-forms [6, 13, 21, 27]. Here, we evaluated the biological activities of the IGF1 pro-forms compared to those of mature IGF1. To this end, we generated a set of conditioned media containing different ratios of mature IGF1 and IGF2 pro-forms by using the proprotein convertase furin to induce pro-form cleavage and to increase the mature IGF1 amounts and, on the other hand, by using the convertase inhibitor CMK during transfection to inhibit pro-form cleavage and to increase the IGF1 pro-form amounts. We found that after culturing MCF7, T47D and ZR751 cells with increasing amounts of the GF1 pro-forms, phosphorylation of the IGF1R markedly decreased. This result correlates with recent data showing that glycosylated pro-IGF1Ea is less efficient in receptor activation than pro-IGF1 and mature IGF1 [13]. Despite the finding that the pro-forms decreased the activation of the IGF1R, no significant differences were observed in cellular proliferation and ERK1/2 phosphorylation compared to mature IGF1. Interestingly, AKT phosphorylation in MCF7 cells seems to be affected by gly-IGF1Ea. It has previously been suggested that IGF1Ea may activate alternative IGF1R downstream pathways [6], as the canonical PI3K/AKT/mTOR signaling pathway was not induced in transgenic mice over-expressing IGF1Ea [28, 29]. The effect of gly-IGF1Ea on AKT phosphorylation in breast cancer cells requires, however, independent confirmation.

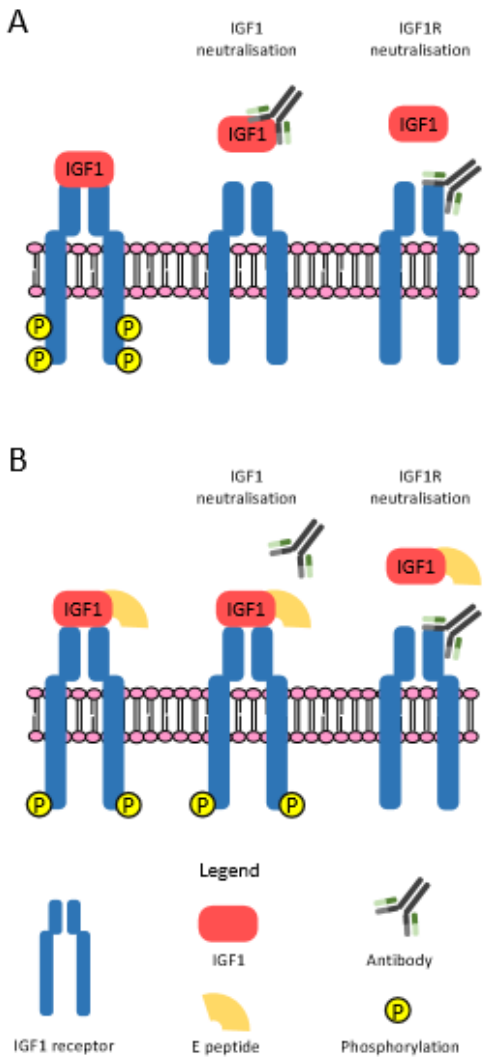


Fig. 7. Schematic presentation of IGF1 pro-form biological activity in breast cancer cells. (a) IGF1 induces IGF1R phosphorylation that is completely inhibited by neutralizing IGF1 or IGF1R. (b) E peptide decreases IGF1R phosphorylation induced by IGF1. IGF1 neutralisation is ineffective in inhibiting IGF1R phosphorylation that, on the other hand, is completely inhibited by IGF1R neutralization.

In conclusion, we found that IGF1 pro-forms can induce breast cancer cell proliferation via IGF1R phosphorylation. There are other data supporting a role of the IGF1 pro-forms in cancer, such as prostate [17], cervical [18] and colorectal cancer [19]. As yet, however, the biological activity of IGF1 variants in breast cancer development has not been established, and no analytical methods are available to correctly detect and quantify the IGF1 pro-forms. In fact, the available methods rely on the use of antibodies that primarily recognise the mature IGF1 peptide, thereby underestimating the pro-forms. The low specificity of anti-IGF1 antibodies to the pro-forms could also have implications for the design of breast cancer therapies, since current targeted strategies include anti-IGF1 antibodies to neutralise the IGF system [31]. The low affinities of pro-IGF1s for IGF1R, together with the poor prognosis associated with high IGF1R expression, make the search

for regulatory mechanism(s) and potentially specific bioactivities of the various IGF1 peptides an area of particular interest, and further studies will focus on the identification of the pro-IGF1s as candidate prognostic factors.

5. Acknowledgements

The authors would like to thank Dr. Joanne Tonkin and Dr. Tommaso Nastasi for kindly providing the IGF1 pro-form specific constructs for the cell transfections.

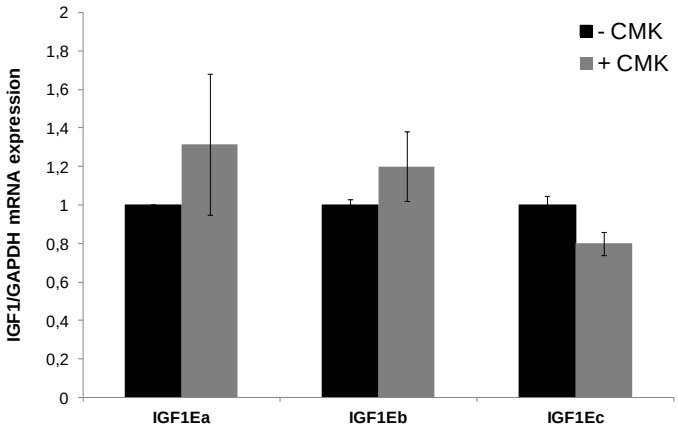
Funding

This study was supported by the RF-2009-1532789 Ministry of Health Project-Italy.

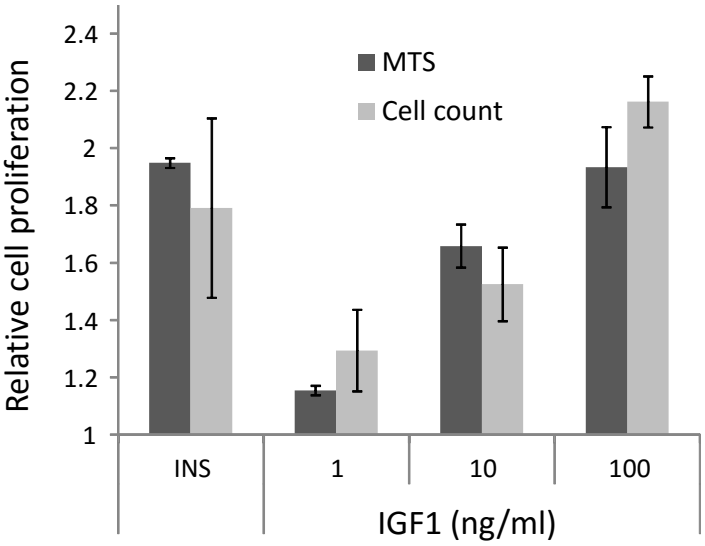
Declaration of interest

The authors declare that there is no conflict of interest that could be perceived as prejudicing the impartiality of the research reported.

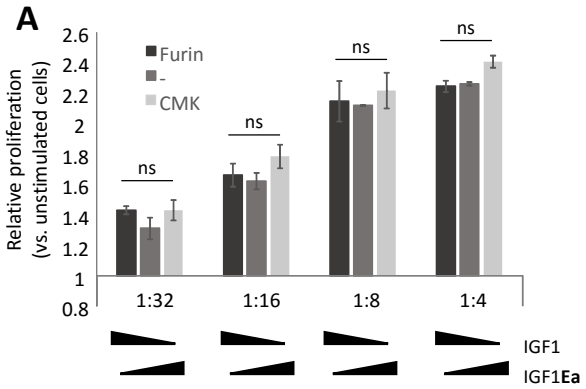
Supplementary data

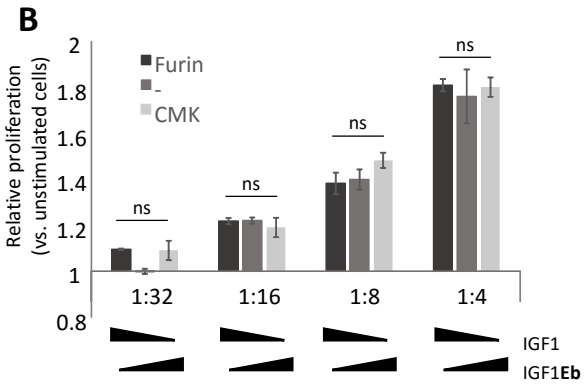


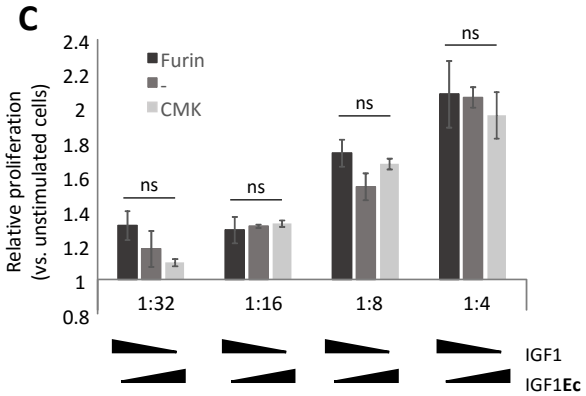
Supplementary Fig. S1. Relative expression of IGF1 in transfected HEK293 cells with CMK relative to control (no CMK). CMK did not significantly alters the IGF1 expression.



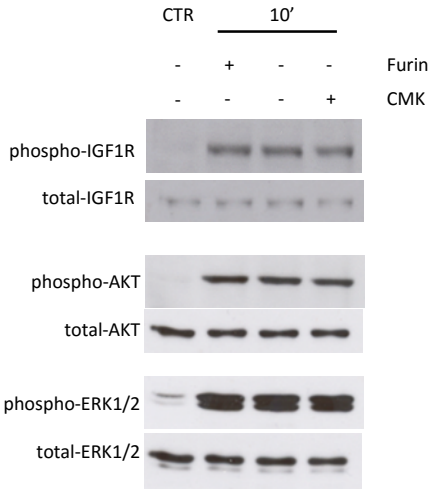
Supplementary Fig. S2. MCF7 cells were grown with increasing concentration of mature IGF1, and after 4 days of culture, cell proliferation was evaluated with both MTS assay and cell count (means ± SEM. n=3). Both methods given similar results. Insulin (INS, 10 µg/ml) was used as positive control.


A
Relative proliferation (vs. unstimulated cells)
2.6
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
Furin
-
CMK
ns
ns
ns
ns
1:32
1:16
1:8
1:4
IGF1
IGF1Ea


B
Relative proliferation (vs. unstimulated cells)
2
1.8
1.6
1.4
1.2
1
0.8
Furin
-
CMK
ns
ns
ns
ns
1:32
1:16
1:8
1:4
IGF1
IGF1Eb


C
Relative proliferation (vs. unstimulated cells)
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
Furin
-
CMK
ns
ns
ns
ns
1:32
1:16
1:8
1:4
IGF1
IGF1Ec

D



Supplementary Fig. S3. Proliferative activity of mature IGF1 *vs.* IGF1 pro-forms: relative cell proliferation MCF7 cultured 4 day with HEK293 supernatants containing mature IGF1 and (A) IGF1Ea, (B) IGF1Eb and (C) IGF1Ec pro-forms (means ± SEM; *n*=3). HEK293 supernatants were two-fold diluted from 1:4 to 1:32. Black triangles indicate changes in mature IGF1 and IGF1 pro-forms amounts. Furin convertase inhibitor CMK and recombinant furin were used to change mature IGF1/IGF1 pro-form ratios. Cell proliferation was evaluated by MTS assay. Data are expressed as relative proliferation *vs.* unstimulated cells. ns: not significantly different; 2-way ANOVA.

(D) Western blot analysis of phospho-IGF-R, phospho-AKT and phospho-ERK1/2 levels in MCF7 cells stimulated for 10 minutes with mature IGF1 with or without CMK and furin. IGF1R, AKT and ERK1/2 were used as a loading controls.

6. References

1. S.J Weroha, P. Haluska, IGF System in Cancer. Endocrinol. Metab. Clin. North Am. **41**, 335–350 (2012)
2. M.V. Karamouzis, A.G. Papavassiliou, Targeting insulin-like growth factor in breast cancer therapeutics. Crit. Rev. Oncol. Hematol. **84**, 8–17 (2012)
3. Endogenous Hormones and Breast Cancer Collaborative Group, Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies. Lancet Oncol. **11**, 530-542 (2010)
4. T.O. Nielsen, H.N. Andrews, M. Cheang, J.E. Kucab, F.D. Hsu, J. Ragaz, C.B. Gilks, N. Makretsov, C.D. Bajdik, C. Brookes, L.M. Neckers, V. Evdokimova, D.G. Huntsman, S.E. Dunn, Expression of the insulin-like growth factor I receptor and urokinase plasminogen activator in breast cancer is associated with poor survival: potential for intervention with 17-allylamino geldanamycin. Cancer Res. **64**, 286–291 (2004)
5. E.R. Barton, The ABCs of IGF1 isoforms: impact on muscle hypertrophy and implications for repair. Appl. Physiol. Nutr. Metab. **31,** 791-797 (2006)
6. A. Philippou, M. Maridaki, S. Pneumaticos, M. Koutsilieris, The complexity of the IGF1 gene splicing, posttranslational modification and bioactivity. Mol. Med. **7**, 202-214 (2014)
7. B.K. Brisson, E.R. Barton, New Modulators for IGF1 Activity within IGF1 Processing Products. Front. Endocrinol. **27**, 42 (2013)
8. D.R. Powell, P.D. Lee, D. Chang, F. Liu, R.L. Hintz, Antiserum developed for the E peptide region of insulin-like growth factor IA prohormone recognizes a serum protein by both immunoblot and radioimmunoassay. J. Clin. Endocrinol. Metab. **65**, 868–875 (1987)
9. C.A. Conover, B.K. Baker, L.K. Bale, J.T. Clarkson, F. Liu, R.L. Hintz, Human hepatoma cells synthesize and secrete insulin-like growth factor Ia prohormone under growth hormone control. Regul. Pept. **48**, 1-8 (1993)
10. C.A. Conover, B.K. Baker, R.L. Hintz, Cultured human fibroblasts secrete insulin-like growth factor IA prohormone. J. Clin. Endocrinol. Metab. **69**, 25-30 (1989)
11. H.E. Wilson, M. Westwood, A. White, P.E. Clayton, Monoclonal antibodies to the carboxy-terminal Ea sequence of pro-insulin-like growth factor-IA (proIGF1A) recognize proIGF1A secreted by IM9 B-lymphocytes. Growth Horm. IGF Res. **11**, 10-17 (2001)
12. E.R. Barton, S. Park, J.K. James, C.A. Makarewich, A. Philippou, D. Eletto, B. Brisson, O. Ostrovsky, Z. Li, Y. Argon, Deletion of muscle GRP94 impairs both muscle and body growth by inhibiting local IGF production. FASEB J. **26**, 3691-3702 (2012)

13. J. Durzynska, A. Philippou, B.K. Brisson, M. Nguyen-McCarty, E.R. Barton, The pro-forms of insulin-like growth factor I (IGF1) are predominant in skeletal muscle and alter IGF1 receptor activation. Endocrinol. **154**, 1215–1224 (2013)
14. A. Philippou, E. Papageorgiou, G. Bogdanis, A. Halapas, A. Sourla, M. Maridaki, N. Pissimissis, M. Koutsilieris, Expression of IGF1 isoforms after exercise-induced muscle damage in humans: characterization of the MGF E peptide actions in vitro. In Vivo **23**, 567-575 (2009)
15. B.R. McKay, C.E. O'Reilly, S.M. Phillips, M.A. Tarnopolsky, G. Parise, Co-expression of IGF1 family members with myogenic regulatory factors following acute damaging muscle-lengthening contractions in humans. J. Physiol. **586**, 5549-5560 (2008)
16. D.S. Milingos, A. Philippou, A. Armakolas, E. Papageorgiou, A. Sourla, A. Protopapas, A. Liapi, A. Antsaklis, M. Mastrominas, M. Koutsilieris, Insulinlike growth factor-1Ec (MGF) expression in eutopic and ectopic endometrium endometrium: characterization of the MGF E-peptide actions in vitro. Mol. Med. **17**, 21-28 (2010)
17. A. Armakolas, A. Philippou, Z. Panteleakou, A. Nezos, A. Sourla, C. Petraki, M. Koutsilieris, Preferential expression of IGF1Ec (MGF) transcript in cancerous tissues of human prostate: evidence for a novel and autonomous growth factor activity of MGF E peptide in human prostate cancer cells. Prostate **70**, 1233-1242 (2010)
18. M.M. Koczorowska, A. Kwasniewska, A. Gozdzicka-Jozefiak, IGF1 mRNA isoform expression in the cervix of HPV-positive women with pre-cancerous and cancer lesions. Exp. Ther. Med. **2**, 149-156 (2011)
19. A. Kasprzak, W. Szaflarski, J. Szmeja, M. Andrzejewska, W. Przybyszewska, E. Kaczmarek, M. Koczorowska, T. Kościński, M. Zabel, M. Drews, Differential expression of IGF1 mRNA isoforms in colorectal carcinoma and normal colon tissue. Int. J. Oncol. **42**, 305-316 (2012)
20. Y.H. Kuo, T.T. Chen, Novel activities of pro-IGF1 E peptides: regulation of morphological differentiation and anchorage-independent growth in human neuroblastoma cells. Exp. Cell Res. **280**, 75-89 (2002)
21. B. Poudel, D. Bilbao, P. Sarathchandra, R. Germack, N. Rosenthal, M.P. Santini, Increased cardiogenesis in P19-GFP teratocarcinoma cells expressing the propeptide IGF1Ea. Biochem. Biophys. Res. Comm. **416**, 293-299 (2011)
22. T. Norat, L. Dossus, S. Rinaldi, K. Overvad, H. Grønbaek, A. Tjønneland A, A. Olsen, F. Clavel-Chapelon, M.C. Boutron-Ruault, H. Boeing, P.H. Lahmann, J. Linseisen, G. Nagel, A. Trichopoulou, D. Trichopoulos, V. Kalapothaki, S. Sieri, D. Palli, S. Panico, R. Tumino, C. Sacerdote, H.B. Bueno-de-Mesquita, P.H. Peeters, C.H. van Gils, A. Agudo, P. Amiano, E. Ardanoz, C. Martinez, R. Quirós, M.J. Tormo, S. Bingham, T.J. Key, N.E. Allen, P.

Ferrari, N. Slimani, E. Riboli, R. Kaaks, Diet, serum insulin-like growth factor-I and IGF-binding protein-3 in European women. Eur. J. Clin. Nutr. **61**, 91-98 (2007)

23. L. Galluzzi, M. De Santi, R. Crinelli, C. De Marco, N. Zaffaroni, A. Duranti, G. Brandi, M. Magnani, Induction of endoplasmic reticulum stress response by the indole-3-carbinol cyclic tetrameric derivative CTet in human breast cancer cell lines. PLoS One **7**, e43249 (2012)
24. M. De Santi, L. Galluzzi, S. Lucarini, M.F. Paoletti, A. Fraternale, A. Duranti, C. De Marco, M. Fanelli, N. Zaffaroni, G. Brandi, M. Magnani, The indole-3-carbinol cyclic tetrameric derivative CTet inhibits cell proliferation via overexpression of p21/CDKN1A in both estrogen receptor-positive and triple negative breast cancer cell lines. Breast Cancer Res. **13**, R33 (2011)
25. S.J. Duguay, W.M. Milewski, B.D. Young, K. Nakayama, D.F. Steiner, Processing of wild-type and mutant proinsulin-like growth factor-IA by subtilisin-related proprotein convertases. J. Biol. Chem. **272**, 6663-6670 (1997)
26. M.S. Hede, E. Salimova, A. Piszczek, E. Perlas, N. Winn, T. Nastasi, N. Rosenthal, E-peptides control bioavailability of IGF1. PLoS One **7**, e51152 (2012)
27. E.R. Barton, Viral expression of insulin like growth factor-I isoforms promotes different responses in skeletal muscle. J. Appl. Physiol. **100**, 1778-1784 (2006)
28. M. Vinciguerra, A. Musarò, N. Rosenthal, Regulation of muscle atrophy in aging and disease. Adv. Exp. Med. Biol. **694**, 211-233 (2010)
29. M.P. Santini, L. Tsao, L. Monassier, C. Theodoropoulos, J. Carter, E. Lara-Pezzi, E. Slonimsky, E. Salimova, P. Delafontaine, Y.H. Song, M. Bergmann, C. Freund, K. Suzuki, N. Rosenthal, Enhancing repair of the mammalian heart. Circ. Res. **100**, 1732-1740 (2007)
30. M.P. Santini, J. Lexow, G. Borsellino, E. Slonimski, E. Zarrinpashneh, T. Poggioli, N. Rosenthal, IGF1Ea induces vessel formation after injury and mediates bone marrow and heart cross-talk through the expression of specific cytokines. Biochem. Biophys. Res. Comm. **410**, 201-207 (2011)
31. Y. Feng, Z. Zhu, X. Xiao, V. Choudhry, J.C. Barrett, D.S. Dimitrov, Novel human monoclonal antibodies to insulin-like growth factor (IGF)-II that potently inhibit the IGF receptor type I signal transduction function. Mol. Cancer Ther. **5**, 114-120 (2006)

CHAPTER 3

REGULATION OF IGF-1 STABILITY, LOCALIZATION AND SECRETION BY INTRINSICALLY DISORDERED E-DOMAIN TAILS

Giosuè Annibalini[1], Mauro De Santi[1], Serena Contarelli[1], Roberta Saltarelli[1], Michele Guescini[1], Luciana Vallorani[1], Giorgio Brandi[1], Vilberto Stocchi[1], Elena Barbieri[1]^

[1]Department of Biomolecular Sciences, University of Urbino Carlo Bo, 61029 Urbino, Italy.
^IIM, Interuniversity Institute of Myology.

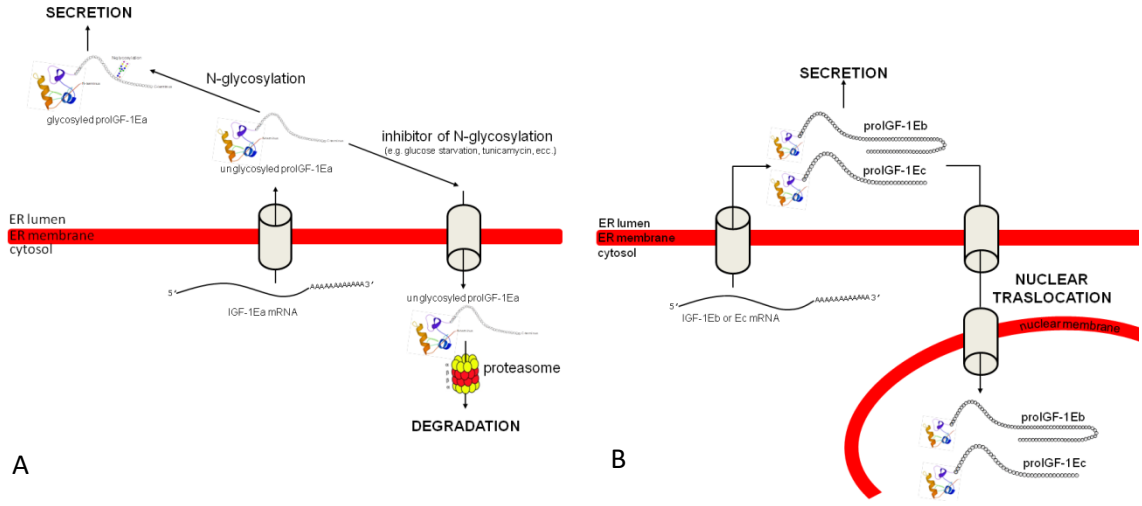
In preparation, 2017.

Regulation of IGF-1 stability, localization and secretion by intrinsically disordered E-domain tails

Giosuè Annibalini[1], Mauro De Santi[1], Serena Contarelli[1], Roberta Saltarelli[1], Michele Guescini[1], Luciana Vallorani[1], Giorgio Brandi[1], Vilberto Stocchi[1], Elena Barbieri[1]^

[1]Department of Biomolecular Sciences, University of Urbino Carlo Bo, 61029 Urbino, Italy.
^IIM, Interuniversity Institute of Myology.

Graphical Abstract



Proposed mechanism of actions for disordered E-domain on pro-IGF-1Ea (A) and pro-IGF-1Eb and pro-IGF-1Ec secretion (B). ER lumen in the upper panel, ER membrane in red and the cytosol in the lower panel. (A) The pro-IGF-1Ea nascent peptide is synthesized and released into the lumen of the ER. Here, the newly created polypeptide is subjected to glycosylation and hence secreted. N-linked glycosylation inhibitors determined the exit of unglycosylated pro-IGF-1Ea from the ER and its degradation by the cytosolic proteasome. (B) The pro-IGF-1Eb and pro-IGF-1Ec nascent peptides are synthesized and released into the ER. Then the occurrence of Eb- or Ec-domain regulates the subcellular localizations of pro-IGF-1Eb and pro-IGF-1Ec, promoting their partial nuclear accumulation.

1. Introduction

Insulin-like growth factor-1 (IGF-1) is a growth factor with multiple roles in various aspects of normal and pathological growth and differentiation [1-2]. The translation of the *igf-1* gene gives rise to an immature IGF-1 peptide, which has a signal peptide at the 5' end of the gene, a core region and a C-terminal E-domain extension. The passage of the polypeptide into the endoplasmic reticulum (ER) removes the signal peptide while the nascent IGF-1 pro-hormone (pro-IGF-1) is emerging, retaining the E-domain. Conversion of pro-IGF-1 to mature peptide requires the endoproteolytic cleavage of the E-domain by proprotein convertases, such as furin, which processes proproteins at highly conserved, unique pentabasic motif [3].

Due to alternative splicing of terminal exon 5 of the *igf-1* gene, three distinct pro-IGF-1s might exist: pro-IGF-1Ea, pro-IGF-1Eb and pro-IGF-1Ec [3-4]. These pro-hormones have the same IGF-1 mature sequence of 70 amino acids (aa) but different E-domains. In particular, the human Ea-domain is composed of 35 aa; the first 16 aa of Ea-domain are common in all E-domains, while 19 aa are unique to this isoform. The human Ea-domain contains a potential N-glycosylation site, N92, which follows the consensus sequence motif for N-glycosylation, NX(S/T) (where X can be any amino acid except proline). Accordingly, two bands corresponding to unglycosylated pro-IGF-1Ea (11.7 kDa) and glycosylated pro-IGF-1Ea (~17-22 kDa) were found in normal and IGF-1-overexpressing cells [5-6]. The human Eb and Ec-domains contain the 16 common aa and 61 and 24 additional isoform-specific aa respectively, with a predicted molecular weight of 16.5 kDa for pro-IGF-1Eb and 12.5 kDa for pro-IGF-1Ec. The human Eb and Ec-domains lack potential N-linked glycosylation consensus sequences [3].

Current evidence suggests that pro-IGF-1s are not a simple inactive precursor of mature IGF-1, but are stable intermediates of posttranslational processing. For example, normal, unstimulated mammalian tissues mainly produced the glycosylated pro-IGF-1Ea [5]. Moreover, in mouse skeletal muscle, viral delivery of IGF-1Ea or IGF-1Eb, but not mature IGF-1, increase muscle mass, suggesting that E-domains may be necessary to promote functional hypertrophy [7].

Bioinformatic analysis of pro-IGF-1 structures showed that the E-domains are putative intrinsically disordered regions (IDRs) [4, 8]. Moreover, we recently demonstrated that the structure of pro-IGF-1s has been strongly conserved to maintain both the folded structure of mature IGF-1 and its intrinsically disordered E-domain tails [4]. IDRs are

regions within proteins that exhibit high flexibility and may lack a secondary or tertiary structure [9]. IDRs may facilitate the regulation of protein function through various mechanisms. For example, owing to their conformational flexibility, IDRs have a high propensity to undergo posttranslational modifications, such as acetylation, glycosylation, methylation, or phosphorylation [10]. IDRs might also control protein half-life by efficiently engaging proteins to the proteasome [11-12]. Moreover, studies have identified IDRs as enriched in the alternatively spliced protein segments, indicating that protein isoforms may display functional diversity due to the alteration of tissue-specific and species-specific modules within these regions [13].

In this study, we analyzed the structural propriety of mature IGF-1 and pro-IGF-1Es, and we investigated the role of alternative E-domains on IGF-1 stability, localization and secretion.

2. Materials and Methods

2.1 Tissue sampling and cell cultures

Tissue sampling and cell cultures were carried out by standard methods as previously described [4]. Briefly, human and animal tissues, about 30 mg, were immediately submerged in liquid nitrogen and stored at -80°C prior to protein extraction. The HEK293, K562, HeLa, LoVo and MCF7 cell lines were obtained from the American Type Culture Collection (ATCC, Rockville, MD, USA). The cells lines were cultured in DMEM (HEK293, MCF7, LoVo) or RPMI-1640 (K562 and HeLa) media supplemented with 10% fetal bovine serum, 2 mmol/L L-glutamine, 1x MEM Non-essential Amino Acid Solution, 0.1 mg/ml streptomycin and 0.1 U/L penicillin. Cells were maintained in a humidified incubator (5% CO_2) at 37 °C. All cell culture materials were purchased from Sigma-Aldrich (St. Louis, MO, USA).

2.2 Protein extraction and Western blotting analysis

Tissues and cells were processed for Western blot analysis as previously reported [6]. Briefly, protein extracts were prepared by homogenizing with a Polytron homogenizer (KINEMATICA AG, Switzerland) in lysis buffer containing: 20 mmol/L HEPES (pH 7.9), 25% v/v glycerol, 0.42 mol/L NaCl, 0.2 mmol/L EDTA, 1.5 mmol/L MgCl2, 0.5%

v/v Nonidet P-40, 1 mmol /L DTT, 1 mmol/L Naf, 1 mmol/L Na_3VO_4, and 1× complete protease inhibitor cocktail (Roche Diagnostics Ltd, Mannheim, Germany). The lysates were frozen and thawed twice and clarified by centrifugation at 12,000 rpm for 10 minutes at 4°C. Total protein concentrations were determined using the Bradford colorimetric assay. Equal amount of total proteins were fractionated by SDS-PAGE on a 15% polyacrylamide gel and then transferred on PVDF membranes (GE Healthcare). The membranes were incubated overnight at 4°C with the primary antibodies directed towards: IGF-1 (1:2,000; no. 500P11), purchased from Peprotech (Rocky Hill, New Jersey, USA), pro-IGF-1s (1:2000; no. PA5-19382) purchased from Invitrogen (Carlsbad, California, USA), β-tubulin (1:2000; no. 2146), Lamin A/C (1:2000; no. 4777), Calnexin (1:2000; no. 2679) phospho-IGF-1 Receptor β (1:2000; no. 3024), IGF-1 Receptor β (1:2000; no. 3027), phospho-p44/42 (ERK1/2) (1:2000; no. 9101), p44/42 (ERK1/2) (1:2000; no. 9102), phospho-Akt (Ser473) (1:2000; no. 9271) and Akt (1:2000; no. 9272) were purchased from Cell Signaling Technology (Beverly, MA, USA). Membranes were washed and incubated with appropriate secondary HRP-conjugated antibodies (Bio-Rad Laboratories Inc). Protein bands were visualized using Clarity Western ECL Substrate (Bio-Rad) and the immunoreactive bands were quantified using Fluor-S MAX System (Bio-Rad Laboratories Inc) equipped with Quantity One software. β-tubulin was used for normalization.

2.3 Cell culture transfection assays

Transient cell transfection was carried out with TransIT-X2® Transfection Reagent (Mirus Bio, Madison, WI, USA) with plasmid constructs containing sequences encoding pro-IGF-1s as previously described [6]. Each plasmid contained DNA encoding the class 1 IGF-1 48-amino acid signal peptide, the mature 70-amino acid IGF-1 peptide, the first 16 aa common in all E-domains, and C-terminal sequences encoding either the Ea (19 aa), the Eb (61 aa) or the Ec (24 aa) domain. Where indicated, to inhibits N-glycosylation, cells were growth in glucose-depleted medium or treated with 0.1 µg/ml of Tunicamycin (Tun) (Sigma-Aldrich, St. Louis, MO, USA) for 24h. The supernatants of HEK293-transfected cells growth in glucose depleted media or Tun were also used to treated the MCF7 cells for 1h in order to evaluate their effects on IGF-1R, AKT and ERK1/2 phosphorylation. Afterwards the MCF7 cells were washed with PBS and lysed for western blotting analysis. Were indicated, the protein synthesis inhibitor

cycloheximide (CHX) (Sigma-Aldrich) and the proteasome inhibitor MG132 (Sigma-Aldrich) were added to HEK293 cells cultured with a final concentration of 25μg/ml and 10μM respectively in a time-course experiment. The cells were then collected at indicated time points, lysed and prepared to western blotting analysis. Deglycosylation of pro-IGF-1Ea was performed by incubation of pro-IGF-1Ea enriched media with 2500 U of PNGase F (NewEngland Biolabs) for 3 hours at 37°C, according to manufacturer's recommendations. Aliquot of pro-IGF-1Ea supernatant incubated with equal volume of PNGase assay reaction buffer without the enzyme PNGase F was used as control.

2.4 Limited proteolysis

Supernatants of IGF-1Ea-, IGF-1Eb- or IGF-1Ec-transfected HEK293 cells were concentrated using an Amicon Ultra 3K centrifugal filter unit (Merck Millipore, Billerica, MA, USA) and subjected to limited proteolysis. Enzymatic digestion was performed at 37°C adding bovine trypsin (Sigma-Aldrich) and proteinase K (Sigma-Aldrich) to protein extract at a 1:100 enzyme/substrate ratio (wt/wt). Reactions were removed over a time-course and the digested products were quenched with SDS sample buffer prior to SDS-PAGE and analyzed by Western blotting. The quantification of bands intensity were normalized to the "no trypsin" samples of each set.

2.5 Subcellular localization analysis

The cytosol, ER and nucleus isolations were performed as described in [14]. Briefly, cells were treated with permeabilization buffer for 5 minutes and were centrifuged at 3,000 g for 5 min to collect the cytosol fraction in the soluble fraction. The pellet was then washed, subjected to lysis buffer for 5 minutes and centrifuged again at 3,000 g for 5 min to collect the ER in the soluble fraction, whereas the pellet represent the nucleus fraction. Finally all the samples were clarified at 7,500 g for 10 minutes to remove cell debris and transferred to clean tubes for further analysis.

The HEK-293 cells subjected to immunofluorescence were seeded in 4-well chamber slide at a density of 5x104 cells/well, incubated overnight and transfected for IGF-1 isoforms expression as previously described. After overnight incubation, cells were fixed with 4% paraformaldehyde for 15 minutes, permeabilized with 0.2% TritonX-100, blocked with 5% of goat serum and incubated overnight at 4°C with the anti-mature IGF-1 antibody (Prepotech). Next, cells were incubated 1 hour with an anti-rabbit-PE

conjugated antibody, stained with DAPI, mounted with Fluoreshield (Sigma) and photographed with a fluorescence microscope (ZEISS AxioVert A.1).

2.6 Conservation of Ea-domain N-glycosylation site

BLASTP was used to align protein sequence of human Ea-domain (EVHLKNASRGSAGNKNYRM) against the entire UniProt database using E-threshold of 0.01 [15]. The obtained UniProt hits (250 sequences with a minimal sequence identity of ~60%) were aligned using MUSCLE [16] and WebLogo (http://weblogo.berkeley.edu/logo.cgi) was used to create relative frequency plots.

2.7 Statistical analysis

Data are represented as mean ± SEM af at least three independent experiments. Statistical analyses were performed using repeated measures ANOVA or one-way ANOVA as appropriate, followed by Bonferroni's multiple comparison post hoc tests. A p value <0.05 was considered statistically significant.

3. Results

3.1 Disorder propensity of Human IGF-1 isoforms

We used the D^2P^2 platform (http://d2p2.pro/) and limited proteolysis to predict intrinsically disordered regions of human IGF-1Ea (ENSP00000416811), IGF-1Eb (ENSP00000302665) and IGF-1Ec (ENSP00000376638) isoforms [17]. Figure 1 shows the plot generated by the D^2P^2 platform. This analysis showed that the mature IGF-1 is mostly ordered while all the predictors predict the E-domains as disordered. The Eb-domain also contains two predicted molecular recognition features (MoRFs) and two phosphorylation sites.

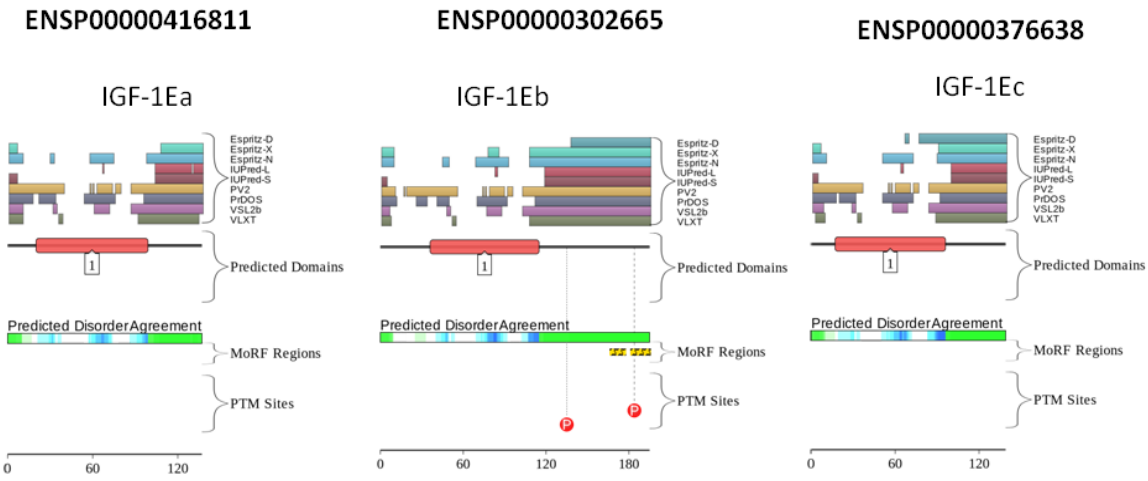


Figure 1. Evaluation of the intrinsic disorder propensity of human IGF-1 isoforms predicted using the D^2P^2 platform (http://d2p2.pro/). The red box correspond to the insulin-like domain; the green-and-white bar in the middle of the plot shows the predicted disorder agreement between nine predictors, with green parts corresponding to portions of sequence where at least 75% of the predictors agreed. Yellow bars show the location of the predicted disorder-based binding sites (molecular recognition features, MoRFs), whereas red circles at the bottom of the plot show the location of putative phosphorylation sites.

Subsequently, we used limited proteolysis to identify the regions of the polypeptide chain most prone to proteolysis and thus the sites of high flexibility or local unfolding [18]. The supernatant of HEK293 cells enriched of pro-IGF-1Ea, pro-IGF-1Eb, pro-IGF-1Ec and mature IGF-1 recombinant proteins were digested with trypsin, loaded on SDS-PAGE gels and probed with an anti-mature IGF-1 antibody. As shown in Figure 2, all pro-IGF-1s were sensitive to trypsin digestion while mature IGF-1 was significantly more resistant. Same results were obtained by proteinase K digestion of pro-IGF-1Ea (Supplementary Figure 1). These data show that pro-IGF-1s are composed of both protein structural domains, i.e. the mature IGF-1, and intrinsically disordered regions, i.e. the C-terminal E-domains.

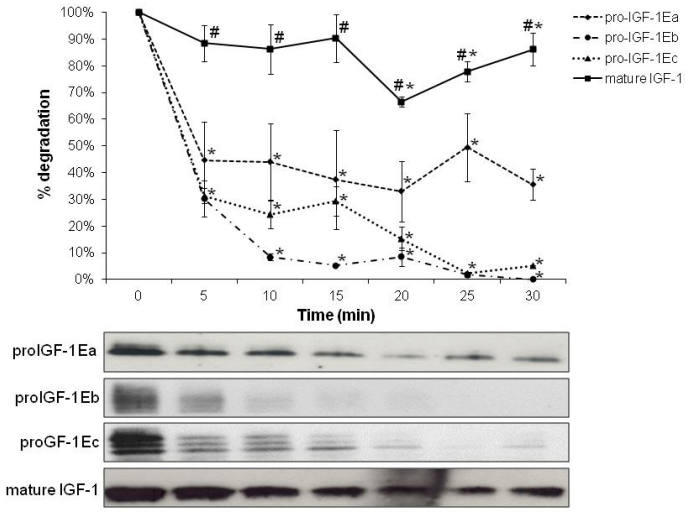


Figure 2. Limited proteolysis of IGF-1 isoforms following incubation with 0.2 µM trypsin at 37°C for different times. The IGF-1 isoforms were partially digested with trypsin. Reactions were removed over a time-course and the digested products were load on 12% SDS-PAGE and analysed by Western blotting with anti-IGF-1 antibody. Repeated measures ANOVA, # significantly different compared to pro-IGF-1Ea, pro-IGF-1Eb and pro-IGF-1Ec (p<0.05); * significantly different compared to 0 minute time point (p<0.05).

3.2 Intracellular IGF-1 are mainly expressed as pro-hormones, not mature IGF-1

Using RT-PCR analyses, we previously demonstrated that skeletal muscle, adipose tissues and liver of several mammalian species mainly expressed the IGF-1Ea isoform, which represents about 90% of IGF-1 transcripts [4]. The first goal of the present study was to examine the protein forms endogenously produced in these tissues. Immunoblotting of protein lysates using the anti-mature IGF-1 antibody showed a distinct ~17 kDa band, most likely representing glycosylated pro-IGF-1Ea, in all samples analyzed (Figure 3A). Notably, the band corresponding to mature IGF-1 (~7kDa) was not found in naïve tissues, in agreement with [5].

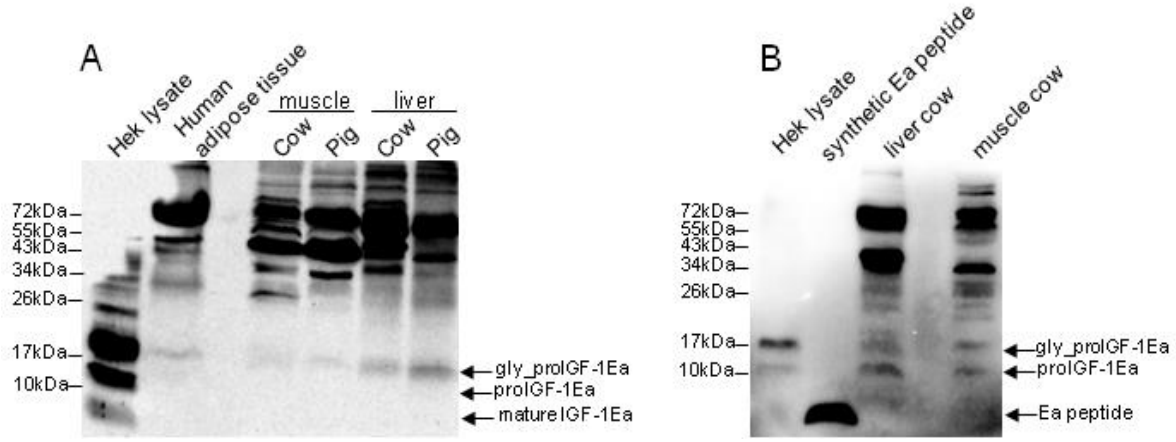


Figure 3. Immonoblotting of several mammalian tissues using an antibody directed against mature IGF-1 sequence (A) or common region of E-peptides (B). (A) Protein lysates (80 µg) were subjected to 12% SDS-

PAGE and immunoblotted with anti-mature IGF-1 antibody. A band at a molecular weight around 17 kDa, most likely representing glycosylated pro-IGF-1Ea, was detected in all the tissue samples tested. A cell lysate of HEK293T overexpressing IGF-1Ea was used has a positive control. The band corresponding to mature IGF-1 (~7kDa) was not found in tissues and is detectable in HEK293T overexpressing IGF-1Ea only after long exposure of the blots; (B) Immonoblotting of liver and muscle pig lysate using an antibody directed against E-domains (right). Two bands at a molecular weight around 12 kDa and 17 kDa were detected in all the tissue samples tested, most likely representing the unglycosilated and glycosylated pro-IGF-1Ea respectively. No band at a molecular weight of Ea peptide (around 4 kDa) was detected. A cell lysate of HEK293T overexpressing IGF-1Ea and synthetic human Ea peptide were used has a positive controls.

To further confirm the presence of pro-IGF-1s, we used an antibody directed against the common E-domain region of pro-IGF-1s (RSVRAQRHTD). The antibody specificity towards E-domain region was checked using HEK293 cells overexpressing IGF-1 isoforms (data not shown). As shown in Figure 3B, two bands of ~12kDa and ~17 kDa were detectable with the anti E-domain antibody, corresponding respectively to the molecular weight of unglycosylated and glycosylated pro-IGF-1Ea. As expected, no band corresponding to the molecular size of Ea-domain (~4kDa) was detected in the lysate of HEK293 cells overexpressing IGF-1Ea or tissues confirming that intracellularly, the E domain was not cleaved from the IGF-1 mature protein. Subsequently, we moved to a cell-based system to improve IGF-1 detection and to control the IGF-1 isoforms produced. We recently demonstrated that after over-expression of IGF-1 isoforms in HEK293 cells the pro-IGF-1s are the main forms produced intracellularly, and for pro-IGF-1Ea both unglycosylated (~12kDa) and glycosylated (~17kDa) forms were detected [6]. Hence the IGF-1 expression pattern of different tissues analyzed was recapitulated in our HEK293 cell-based transient gene expression system. Notably, the results were not cell-type dependent, since the same results were obtained by transfection of others cell lines including HeLa, K562 and LoVo (Supplementary Figure 2). HEK293 cell-based system was used for the next experiments due to their relatively high transfection efficiency. We next examined the effects of E-domains on pro-IGF-1s stability and secretion, starting with the predominant isoform produced in normal tissues, i.e. the pro-IGF-1Ea.

3.3 Glycosylation is necessary to stabilize pro-IGF-1Ea and regulate mature IGF-1 secretion

Multiple sequence alignment of vertebrate Ea domain showed that the N-glycosylation site of pro-IGF-1Ea had been conserved in mammals, birds, amphibians, and teleosts (Figure 4A).

This strong evolutionary conservation suggested that glycosylation plays an important role in the regulation of pro-IGF-1Ea.

Figure 4. WebLogo of Ea-domain sequences obtained from UniProt database. The relative frequency plots of amino acids of 250 E-domain sequences obtained from UniProt database is shown. The consensus sequence motif for N-glycosylation NX(S/T) (where X can be any amino acid except proline) is underline. The WebLogo was produced using the web server at http://weblogo.berkeley.edu/logo.cgi.

Subsequently, we wondered whether glucose withdrawal or direct inhibition of N-glycosylation by Tun might interfere with WT IGF-1Ea production. In this regard, we overexpressed pro-IGF-1Ea in HEK293 cells cultured in glucose-depleted medium or treated with Tun (Figure 5A). Notably, the band, corresponding to glycosylated pro-IGF-1Ea, completely disappeared in the absence of glucose or after treatment with Tun. Moreover, the analysis of culture media of IGF-1Ea-transfected HEK293 cells showed that the inhibition of glycosylation by glucose deprivation or Tun completely abrogated the glycosylated pro-IGF-1Ea secretion and markedly reduced the mature IGF-1 secretion (Figure 5B). Accordingly, the phosphorylation of IGF-1 receptor, ERK1/2 and AKT was inhibited when MCF-7 breast cancer cells were treated with conditioned media from IGF-1Ea-transfected HEK293 cells grown in glucose-depleted medium or treated with Tun (Figure 6). The marked reduction of pro-IGF-1Ea after glucose deprivation or Tun was not due to general suppression of transcription, as shown by IGF-1Ea mRNA quantification (Supplementary Figure 3A), or general protein synthesis inhibition, as shown by cotransfection of GFP (Supplementary Figure 3B).

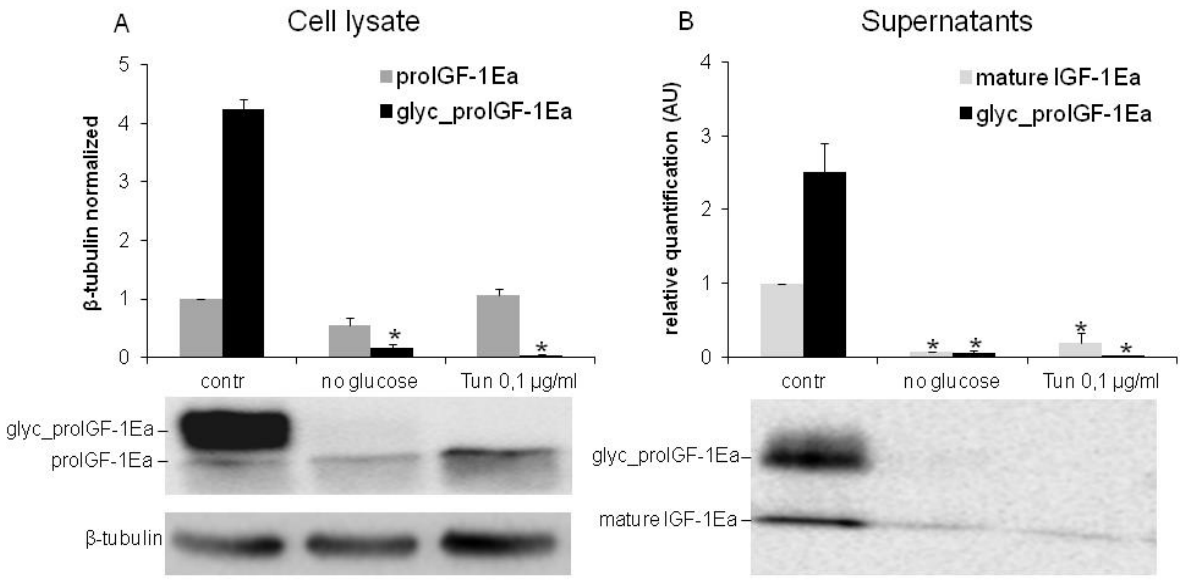


Figure 5. N-glycosilation is required for proper pro-IGF-1Ea secretion. IGF-1Ea was transiently expressed in HEK293 cells in glucose depleted medium or in the presence of 0.1 µg/ml of tunicamycin (Tun). After 24 h the cell lysates (A) and conditioned media (B) were analyzed by Western blot and relative expression level of glycosylatyed pro-IGF-1Ea, unglycosylated pro-IGF-1Ea and mature IGF-1 was calculated. The band at a molecular weight around 17 kDa, corresponding to glycosylated pro-IGF-1Ea, disappeared in absence of glucose or in presence of Tun in cell lysates (A) and in the secreted media (B). The band corresponding to mature IGF-1 (~7kDa) was markedly reduced in conditioned media after both treatments (B). Error bars represent the mean ± SEM; a one-way ANOVA was used to evaluate statistical significance (* $p < 0.05$). β-tubulin was used as a loading control for the cell lysates.

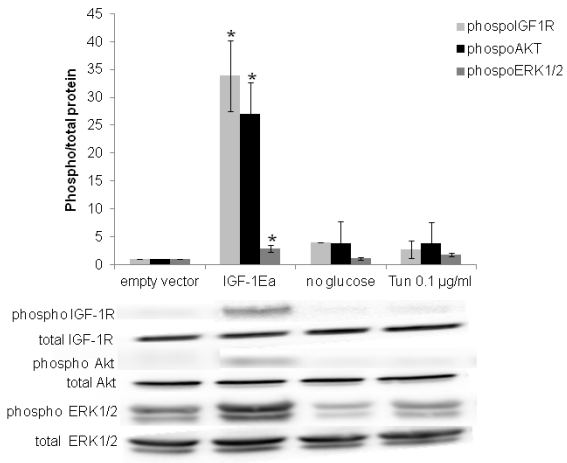


Figure 6. Phosphorylation of IGF-1R, AKT and ERK1/2 after treatment of MCF7 cells with conditioned media from IGF-1Ea-transfected HEK293 cells grown in glucose-depleted medium or treated with Tunicamycin (Tun). The phosphorylation of the IGFR pathway was markedly reduced after glycosylation inhibitor treatments. Error bars represent the mean ± SEM; a one-way ANOVA was used to evaluate statistical significance (* $p < 0.05$).

We then tested whether the modulation of pro-IGF-1Ea production by glucose was dose-dependent. Therefore, IGF-1Ea-transfected HEK293 cells were growth in medium with concentrations of glucose ranging from 0.65g/L (3.6mM) to 5g/L (27.8mM). As shown in figure 7 the expression level of pro-IGF-1Ea dose-dependently increased by glucose concentrations.

Collectively, these results document that the interference with Ea-domain glycosylation resulted in a dramatic decrease of intracellular pro-IGF-1Ea level and hence pro-IGF-1Ea and mature IGF-1 secretion.

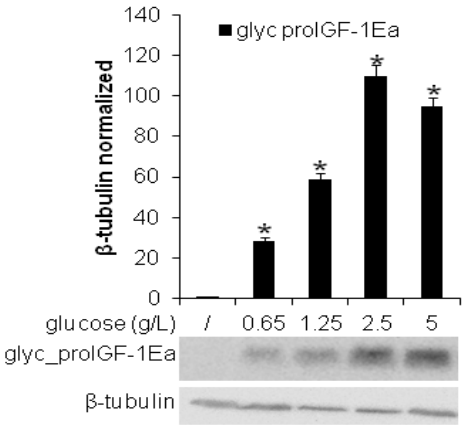


Figure 7. Influence of glucose concentration on pro-IGF-1Ea production. IGF-1Ea was transiently expressed in HEK293 cells grown in culture medium with different glucose concentration. After 24 h the cell lysates were analyzed by Western blot and relative expression level of glycosylatyed pro-IGF-1Ea was calculated. Error bars represent the mean ± SEM; a one-way ANOVA was used to evaluate statistical significance (* $p < 0.05$). β-tubulin was used as a loading control for the cell lysates.

3.4 The turnover of unglycosylated pro-IGF-1Ea is faster than glycosylated pro-IGF-1Ea and depends on proteasome activity

Inhibition of pro-IGF-1Ea glycosylation by glucose starvation or Tun treatment, did not determine a concomitant increase of unglycosylated pro-IGF-1Ea (Figure 5A). One possibility is that the unglycosylated pro-IGF-1Ea was rapidly degraded. Thus, we next sought to examine the role of glycosylation in the stability of pro-IGF-1Ea. In the presence of protein synthesis inhibitor cycloheximide (CHX), the turnover rate for unglycosylated pro-IGF-1Ea was faster than glycosylated pro-IGF-1Ea (Figure 8A). To test the involvement of 26S proteasome machinery on unglycosylated pro-IGF-1Ea degradation, we subsequently treated IGF-1Ea-transfected HEK293T cells with proteasome inhibitor MG132. As shown in Figure 8B, we found an increase of unglycosylated pro-IGF-1Ea while glycosylated pro-IGF-1Ea was only marginally affected by the proteasome inhibitor. These results demonstrated that unglycosylated pro-IGF-1Ea was unstable and degraded faster than glycosylated pro-IGF-1-Ea.

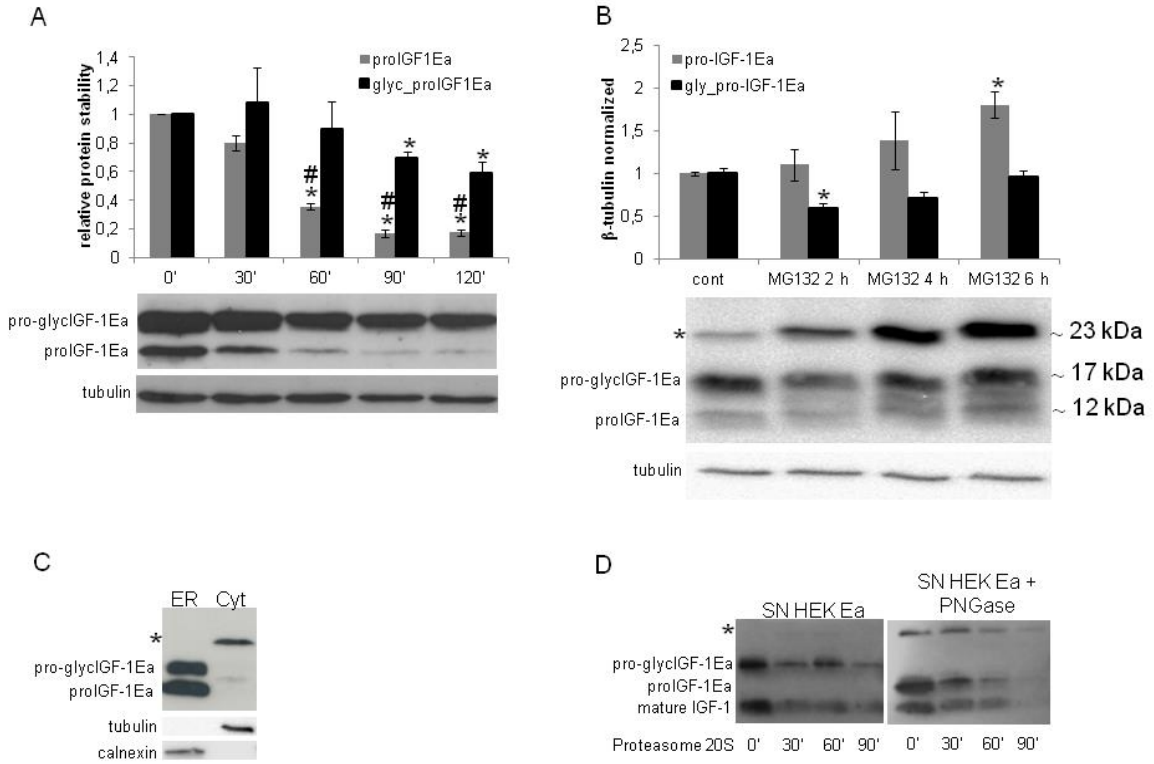


Figure 8. Analysis of pro-IGF-1Ea stability. IGF-1Ea was transiently expressed in HEK293 cells in the presence of 25μg/ml of protein synthesis inhibitor cycloheximide (CHX) (A) or 10μM of the proteasome inhibitor MG132 (B) in a time-course experiment. Cells were collected at different time points, and relative expression level of glycosylated and unglycosylated pro-IGF-1Ea was calculated. After MG132 treatment intracellular accumulation of a ~23kDa band was found (indicated with an asterisk in figure 8B), probably representing unglycosylated pro-IGF-1Ea dimer. Cytosol, endoplasmic reticulum (ER) and nucleus isolations of the conditioned media treatment with MG132 (C). Two bands at a molecular weight around 12 kDa and 17 kDa, representing the unglycosylated and glycosylated pro-IGF-1Ea respectively, were detected in the cytosol fraction, whereas the band at a molecular weight around 23 kDa was detected only in the cytosolic fraction. Deglycosylation of pro-IGF-1Ea enriched media using the *N*-Glycosidase F (PNGase F) (D). The PNGase treatment determined an accumulation of two bands: the unglycosylated pro-IGF-1Ea and the band at a molecular weight around 23 kDa probably corresponding to unglycosylated pro-IGF-Ea dimer. Repeated measures ANOVA, # significantly different compared to glyc_pro-IGF-1Ea (p<0.05); * significantly different compared to 0 minute time point (p<0.05). β-tubulin was used as a loading control for the cell lysates and the cytosol separation; Calnexin was used as a control for the ER separation.

Besides the increase of unglycosylated pro-IGF-1Ea, treatment with MG132 promoted a high accumulation of a band of approximately 23kDa (Figure 8B). We hypothesize that the 23kDa band corresponds to the dimer of unglycosylated pro-IGF-1Ea. In support of

this hypothesis we demonstrated that after MG132 treatment, the 23kDa band accumulated only in the cytoplasmic fraction and not in the ER fraction (Figure 8C). Moreover, enzymatic deglycosylation of the pro-IGF-1Ea enriched media determines the accumulation of both unglycosylated pro-IGF-1Ea and the putative 23kDa dimer (Figure 8D). These results demonstrated that addition of N-glycan on N92 of Ea-domain prevent the degradation of pro-IGF-1Ea by the proteasome, probably overcoming the folding limitation of Ea-domain and its tendency to self-aggregate.

3.5 The alternative Eb- and Ec-domains made the pro-IGF-Eb and pro-IGF-Ec isoforms insensitive to glucose starvation or Tun treatment and determined nuclear accumulation of pro-hormones

The pro-IGF-1Eb and pro-IGF-1Ec ran at the expected molecular weight of 16.5 kDa and 12.5 kDa in SDS-PAGE gels respectively, suggesting that these pro-hormones do not harbor any posttranslational modifications [6]. Accordingly, unlikely pro-IGF-1Ea, pro-IGF-1Eb do not contains any potential glycosylation sites, and migrated at the same molecular weight in standard or glucose depleted media or after treatment of HEK239 cells with Tun (Figure 9). Noteworthy, unlikely pro-IGF-1Ea, also the quantity of pro-IGF-1Eb were unaffected by glucose deprivation or Tun treatment (Figure 9). Same results were obtained with the Ec-domain (data not shown). Hence, modulation of pro-IGF-1s stability and secretion were differentially regulated by E-domain tail and the presence of the Eb or Ec-domain, instead of Ea-domain, completely abrogates the response of pro-IGF-1s to glycosylation status of the HEK293 cells.

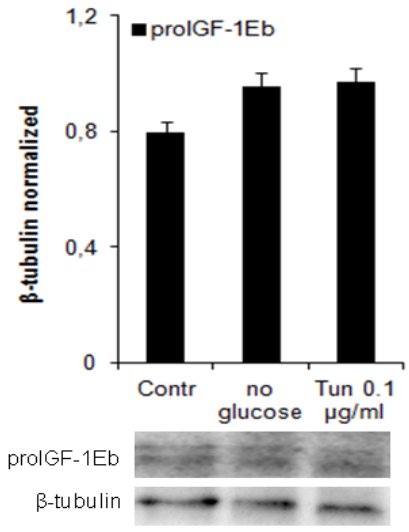


Figure 9. Effect of N-glycosylation inhibitors on IGF-1Eb production. IGF-1Eb was transiently expressed in HEK293 cells in glucose-depleted medium or in the presence with 0.1 µg /ml of Tunicamycin (Tun). After 24 h the cell lysates were analyzed by Western blot and relative expression level of IGF-1Eb was calculated. The bands of the two pro-hormones were unaffected by either of the treatment and migrated at the same level. β-tubulin was used as a loading control for the cell lysates.

Previous studies demonstrated that Eb- or Ec-

domains contain nuclear localization signal [19]. Accordingly, we found that pro-IGF-1Eb and pro-IGF-1Ec partially accumulated in the nucleus of HEK293-transfected cells (Figure 10). On the contrary, the pro-IGF-1Ea was mainly localized to the ER (Figure 10). These data suggested that pro-IGF-1Eb and pro-IGF-1Ec were less efficiently secreted. In addition, we previously demonstrated that lower amount of pro-IGF-1Eb, pro-IGF-1Ec and mature IGF-1 was detected in cell culture media of HEK293 cells overexpressing IGF-1Eb or IGF-1Ec isoforms compared to IGF-1Ea-transfected HEK293 cells [6]. These data showed that Eb- and Ec-domains control subcellular localization of pro-IGF-1s and hence the bioavailability of IGF-1.

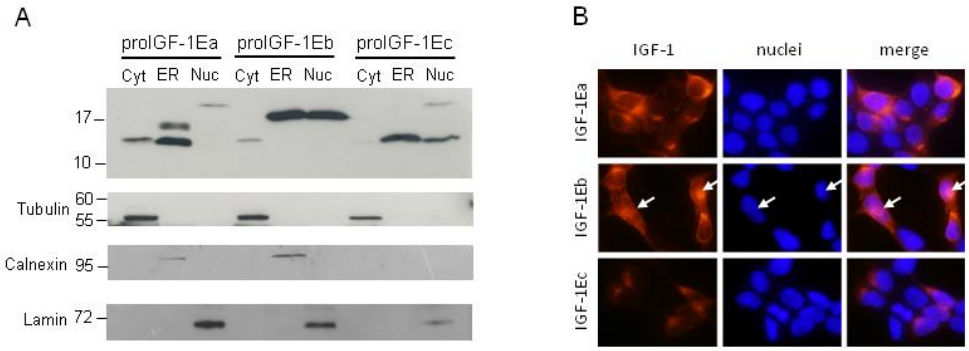


Figure 10. Subcellular localization of IGF-1 isoforms analyzed by cytosol, ER and nucleus isolations (A) or immunofluorescence (B). (A) pro-IGF-1Ea is present in the cytosol fraction (unglycosylated pro-IGF-1Ea) and the ER fraction (unglycosilated and glycosylated pro-IGF-1Ea), IGF-1Eb and IGF-1Ec are present in the ER and nucleus fraction. β-tubulin was used as a control for the cytosol separation, Calnexin as a control for the ER separation and Lamin as a control for the nucleus separation. (B) Immunofluorescence staining of pro-IGF-1Ea, pro-IGF-1Eb and pro-IGF-1Ec in transfected HEK293 cells. DAPI nuclear staining (blue).

4. Discussion

It is now well established that most peptide hormones and growth factors are initially synthesized as pro-hormones that are converted to active forms by endoproteolysis at specific sites [20]. Accordingly, the growth factor IGF-1 is produced as pro-hormone which contains a C-terminal domain, i.e. the E-domain, cleaved by furin convertase [3].
Evidence has been provided that only a small portion of pro-IGF-1 cleavage occurs intracellularly; hence most pro-IGF-1 might be converted to mature forms at the cell surface membrane or extracellularly [5-6]. The results of our study confirm these findings

showing that intracellular IGF-1 is mainly expressed as pro-IGF-1, not mature IGF-1, both *in vitro* and *in vivo* [Figure 3].

The amino acid composition of mature IGF-1 markedly differs compared to E-domains [4, 21]. In particular, the E-domains are enriched in disorder-promoting amino acids, and by bioinformatic analysis, we demonstrated that the E-domains are IDRs, which was also confirmed by limited proteolysis [Figure 1 and 2]. Therefore, the pro-IGF-1s consist of two different parts: the mature IGF-1, with a well-organized structure, and the flexible C-terminal E-domains. Notably, the disorder features of E-domains are conserved both across species and for all the three E-domains generated by alternative splicing [4].

Many studies showed that the conformational plasticity associated with intrinsic disorder provides IDRs with a complementary functional repertoire of ordered domains [10], especially if these IDRs are at the protein termini [22]. Therefore, we next focused on the functional role of C-terminal Ea-, Eb- and Ec-domains.

The Ea-domain contains a highly conserved N-glycosylation site (N92) [Figure 4], which is heavily glycosylated with sugar comprising over 30% of the total mass of the pro-IGF-1Ea both in human and mouse [Figure 3] [5-6]. Here we demonstrated that abrogation of N-glycosylation by glucose starvation or Tun treatment blocked the production of glycosylated pro-IGF-1Ea [Figure 5A]. Unexpectedly, the unglycosylated pro-IGF-1 did not accumulate in glucose-depleted media or after treatment with Tun. Hence, the secretion of pro-IGF-1Ea and mature IGF-1 was also substantially affected by glucose starvation or Tun treatment [Figure 5B]. More interestingly, conditioned media from IGF-1Ea-transfected cells grown in glucose-depleted medium or treated with Tun were unable to activate the IGF-1 receptor pathway in human breast cancer cell line MCF7 [Figure 6]. Noteworthy, the IGF-1Ea mRNA quantity and GFP expression were unaffected by N-linked glycosylation inhibitors, ruling out general inhibition of transcription or translation efficiencies [Supplementary Figure 3]. The effect of glucose on pro-IGF-1Ea production was dose-dependent, and maximum pro-IGF-Ea expression was observed in high glucose cultures (2.5 and 5 g/L) (Figure 7).

Subsequently, we demonstrated that unglycosylated pro-IGF-1 has a rapid turnover compared with glycosylated pro-IGF-1 [Figure 8A]. The proteasome inhibitors MG132 partially rescued the accumulation defect of unglycosylated pro-IGF-1Ea, but also markedly increased the production of a 23kDa band, probably representing unglycosylated pro-IGF-1Ea dimer [Figure 8B]. N-glycosylation is important for the

folding and trafficking of many glycoproteins [23]. Indeed glycosylation is a common mechanism to prevent the degradation of many secreted proteins [24]. Accordingly, our work indicates that addition of N-glycan on Ea-domain prevent the degradation of pro-IGF-1Ea by the proteasome, probably overcoming the folding limitation of unglycosylated pro-IGF-1Ea and its entry into the ER-associated degradation (ERAD) pathway.

In mammalian, in addition to Ea-domain, two alternative C-terminal tails can be produced by alternative splicing: the Eb- and Ec-domains [3-4]. The IGF-1Eb and IGF-1Ec isoforms are not so wide expressed as IGF-1Ea, although their level increased under specific conditions/stimuli (e.g. cancer [25-26] or exercise [27]) and are also expressed in a species-specific manner [4] suggesting isoform-specific functions. Accordingly, in the present study, we demonstrated that the behavior of pro-IGF-1Eb and pro-IGF-1Ec entirely differed compared to pro-IGF-1Ea. In particular, since both Eb- and Ec-domains lacking N-glycosylation sites, the intracellular and secreted level of pro-IGF-1Eb and pro-IGF-1Ec were unaffected by glucose starvation or Tun treatment [Figure 9]. Moreover, the Eb- and Ec-domains determined the partial nuclear localization of pro-IGF-1Eb and pro-IGF-1Ec [Figure 10]. Hence, the final location of pro-IGF-1Eb and pro-IGF-1Ec represent a balance of the competition between secretory and nuclear localization signals of Eb- and Ec-domains.

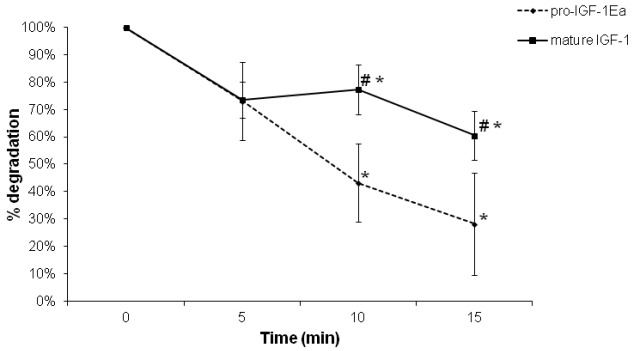
5. Conclusions

Despite the fact that IGF-1 has been implicated as a potential therapy in many pathological processes, its biogenesis is yet to be elucidated. Our data suggested that alternative E-domains act as flexible tails controlling pro-IGF-1s and mature IGF-1 bioavailability. In particular, we demonstrated that N-linked glycosylation regulates the stability and secretion of pro-IGF-1Ea, probably ensuring proper pro-hormone folding and favoring its passage through the secretory pathway.

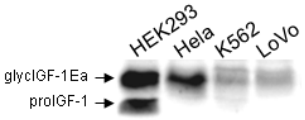
The alternative Eb-a nd Ec-domains lack N-terminal glycosylation sites hence pro-IGF-1Eb and pro-IGF-1Ec isoforms were insensitive to glycosylation status of the cells. Moreover, the Eb- and Ec-domains regulate the subcellular localizations of pro-IGF-1Eb and pro-IGF-1Ec, promoting their nuclear accumulation.

Thus, disordered E-domains play an important role in the structure, regulation and functioning of IGF-1.

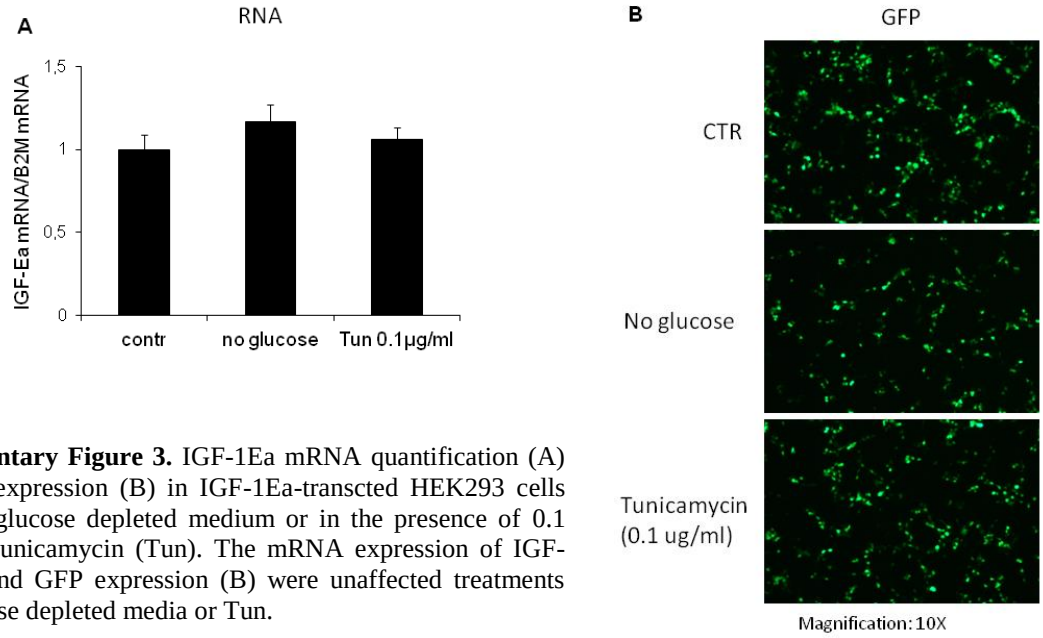
Supplementary data



Supplementary Figure 1. Limited proteolysis of pro-IGF-1Ea with proteinase K at 37°C for different times. pro-IGF-1Ea were partially digested with proteinase K for indicated times. The digested products were load on 12% SDS-PAGE and and analysed by Western blotting with anti-IGF-1 antibody. Repeated measures ANOVA, # significantly different compared to pro-IGF-1Ea (p<0.05); * significantly different compared to 0 minute time point (p<0.05).



Supplementary Figure 2. Immunoblotting using the anti-mature IGF-1 antibody to HEK293T, Hela, K562 and LoVo cells transfected with plasmid constructs containing sequences encoding for IGF-1Ea. A band at a molecular weight around 17 kDa, most likely representing glycosylated pro-IGF-1Ea, was detected in all mammalian cells transfected tested. Two bands at a molecular weight around 12 kDa and 17 kDa were detected in the HEK293T cell lysate, most likely representing the unglycosilated and glycosylated pro-IGF-1Ea respectively.



Supplementary Figure 3. IGF-1Ea mRNA quantification (A) and GFP expression (B) in IGF-1Ea-transcted HEK293 cells grown in glucose depleted medium or in the presence of 0.1 µg/ml of tunicamycin (Tun). The mRNA expression of IGF-1Ea (A) and GFP expression (B) were unaffected treatments with glucose depleted media or Tun.

6. References

1. Dyer, A.H., et al., *The role of Insulin-Like Growth Factor 1 (IGF-1) in brain development, maturation and neuroplasticity.* Neuroscience, 2016. **325**: p. 89-99.
2. Christopoulos, P.F., P. Msaouel, and M. Koutsilieris, *The role of the insulin-like growth factor-1 system in breast cancer.* Mol Cancer, 2015. **14**: p. 43.
3. Barton, E.R., *The ABCs of IGF-I isoforms: impact on muscle hypertrophy and implications for repair.* Appl Physiol Nutr Metab, 2006. **31**(6): p. 791-7.
4. Annibalini, G., et al., *MIR retroposon exonization promotes evolutionary variability and generates species-specific expression of IGF-1 splice variants.* Biochim Biophys Acta, 2016. **1859**(5): p. 757-68.
5. Durzynska, J., et al., *The pro-forms of insulin-like growth factor I (IGF-I) are predominant in skeletal muscle and alter IGF-I receptor activation.* Endocrinology, 2013. **154**(3): p. 1215-24.
6. De Santi, M., et al., *Human IGF1 pro-forms induce breast cancer cell proliferation via the IGF1 receptor.* Cell Oncol (Dordr), 2016. **39**(2): p. 149-59.
7. Barton, E.R., J. DeMeo, and H. Lei, *The insulin-like growth factor (IGF)-I E-peptides are required for isoform-specific gene expression and muscle hypertrophy after local IGF-I production.* J Appl Physiol (1985), 2010. **108**(5): p. 1069-76.
8. Redwan, E.M., M.H. Linjawi, and V.N. Uversky, *Looking at the carcinogenicity of human insulin analogues via the intrinsic disorder prism.* Sci Rep, 2016. **6**: p. 23320.
9. Lieutaud, P., et al., *How disordered is my protein and what is its disorder for? A guide through the "dark side" of the protein universe.* Intrinsically Disord Proteins, 2016. **4**(1): p. e1259708.
10. Dunker, A.K., et al., *Intrinsically disordered proteins and multicellular organisms.* Semin Cell Dev Biol, 2015. **37**: p. 44-55.
11. Ciechanover, A., *Intracellular protein degradation: from a vague idea through the lysosome and the ubiquitin-proteasome system and onto human diseases and drug targeting.* Neurodegener Dis, 2012. **10**(1-4): p. 7-22.
12. Pauwels, K., P. Lebrun, and P. Tompa, *To be disordered or not to be disordered: is that still a question for proteins in the cell?* Cell Mol Life Sci, 2017. **74**(17): p. 3185-3204.
13. Buljan, M., et al., *Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks.* Mol Cell, 2012. **46**(6): p. 871-83.
14. Jagannathan, S., C. Nwosu, and C.V. Nicchitta, *Analyzing mRNA localization to the endoplasmic reticulum via cell fractionation.* Methods Mol Biol, 2011. **714**: p. 301-21.
15. Magrane, M., *UniProt Knowledgebase: a hub of integrated protein data.* Database (Oxford), 2011. **2011**: p. bar009.
16. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput.* Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
17. Oates, M.E., et al., *D(2)P(2): database of disordered protein predictions.* Nucleic Acids Res, 2013. **41**(Database issue): p. D508-16.
18. Fontana, A., et al., *Identifying disordered regions in proteins by limited proteolysis.* Methods Mol Biol, 2012. **896**: p. 297-318.
19. Tan, D.S., A. Cook, and S.L. Chew, *Nucleolar localization of an isoform of the IGF-I precursor.* BMC Cell Biol, 2002. **3**: p. 17.

20. Chretien, M., *How the prohormone theory solved two important controversies in hormonal and neural Peptide biosynthesis.* Front Endocrinol (Lausanne), 2013. **4**: p. 148.
21. Hede, M.S., et al., *E-peptides control bioavailability of IGF-1.* PLoS One, 2012. **7**(12): p. e51152.
22. Uversky, V.N., *The most important thing is the tail: multitudinous functionalities of intrinsically disordered protein termini.* FEBS Lett, 2013. **587**(13): p. 1891-901.
23. Helenius, A. and M. Aebi, *Roles of N-linked glycans in the endoplasmic reticulum.* Annu Rev Biochem, 2004. **73**: p. 1019-49.
24. Skropeta, D., *The effect of individual N-glycans on enzyme activity.* Bioorg Med Chem, 2009. **17**(7): p. 2645-53.
25. Pickard, A., et al., *The IGF axis in HPV associated cancers.* Mutat Res Rev Mutat Res, 2017. **772**: p. 67-77.
26. Christopoulos, P.F., A. Philippou, and M. Koutsilieris, *Pattern of IGF-1 variants' expression in human cancer cell lines using a novel q-RT-PCR approach.* Anticancer Res, 2015. **35**(1): p. 107-15.
27. Vassilakos, G., A. Philippou, and M. Koutsilieris, *Identification of the IGF-1 processing product human Ec/rodent Eb peptide in various tissues: Evidence for its differential regulation after exercise-induced muscle damage in humans.* Growth Horm IGF Res, 2017. **32**: p. 22-28.

CHAPTER 4

THE INSULIN-LIKE GROWTH FACTOR POOL IN BREAST CANCER AND THE EFFECT OF EXERCISE AND LIFESTYLE

1. Introduction

Insulin-like growth factors (IGFs) are strong mitogens involved in regulating cell proliferation in breast cancer (BC) and targeting IGF-1R was one of the most investigated areas in anticancer drug development [1]. The human *igf-1* gene contains six exons that can be differentially spliced to create multiple transcript variants as deeply described in the first chapter. Exon 1 and exon 2 are mutually exclusive first exons and generate different signal peptides with local or circulating functions that have not yet been clarified [2]. At the 3' end of the gene, alternative splicing gives rise to three mRNAs, each encoding a unique E-terminal peptide [3]. We have previously demonstrated that these IGF-1 isoforms induce BC cells proliferation via IGF-1R as also described in the second chapter [4], and we revealed that they are enriched in regulatory elements at the mRNA and protein level further hinting for their functional relevance (See Chapter 1, [5]). Numerous studies have focused mostly on circulating IGF-1 showing a cross-talk between IGF-1, circulating estrogen and ER stimulation, leading to increased IGF-1 expression [6]. Nowadays, there are few data about the activation of IGF-1/IGF-1R signaling in tumor microenvironment [7, 8]. Moreover there are no data about the expression of IGF-1 isoforms in luminal BC and their association with clinical-histopathological features and lifestyle habits of BC patients.

Indeed, the host exposure to regular exercise significantly reduces the risk of development and recurrence of several cancers and might improve clinical outcomes following a diagnosis of a primary disease. However, the molecular mechanisms that establish the apparent anticancer effects of physical exercise on tissues that are distant from those directly implicated in the exercise response are poorly understood [8, 9].

Recently the remodeling of integrity and composition of the cancer microenvironment (TME) in response to specific signals derived from lifestyle factors such as exercise have

been described by Lee W. Jones and collaborator in an elegant review in Nature Reviews Cancer 17, 620–632 (2017) (Figure 1).

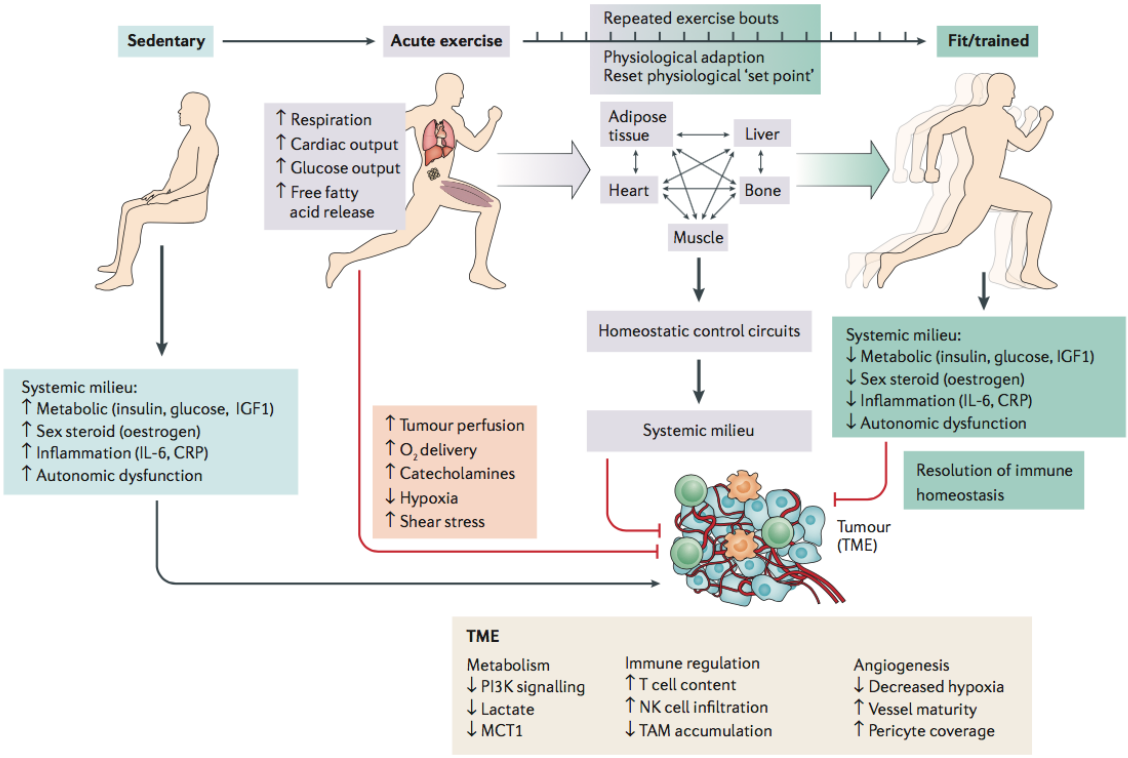


Figure. 1. Exercise-dependent regulation of the tumor microenvironment. Prolonged exposure to physical inactivity is associated with elevated circulating concentrations of numerous growth factors and hormones—a pro-tumorigenic milieu (blue boxes). By contrast, host exposure to acute bouts of exercise stimulates inter-organ signaling achieved by the secretion of hormones, cytokines and growth factors into the host systemic milieu from various tissues and organs (for example, skeletal muscle, heart, bone, liver and adipose tissue), which can subsequently regulate multiple highly integrated homeostatic control circuits that operate at the cellular, tissue and whole-organismal levels (purple boxes). Over time, chronic exercise-induced perturbation of inter-organ signaling promotes physiological adaptation across homeostatic control circuits (establishment of a higher homeostatic 'set point') that, in concert, stimulates the reprogramming of the systemic milieu, potentially characterized by alterations in the specific cell types and/or molecules (green boxes), potentially altering their availability and composition in 'distant' tumor microenvironments (TMEs). Exercise-induced alterations in the systemic milieu influence key regulatory mechanisms in the TME, such as angiogenesis, immune regulation and metabolism, thus having a cumulative antitumorigenic effect (ochre box). In addition to activating the secretion of numerous factors from skeletal muscle, during acute exercise, blood flow is redirected to the metabolically active skeletal muscle that paradoxically occurs in conjunction with increased tumor blood perfusion and reduced tumor hypoxia. As such, this represents an alternative mechanism of exercise regulation of the TME (red box). CRP, C-reactive protein; IGF-1, insulin-like growth factor 1; IL-6, interleukin 6; MCT1, monocarboxylate transporter 1; NK cell, natural killer cell; TAM, tumor-associated macrophage [8].

A recent grant of Italian Ministry of Health (RF-2009-1532789) allowed us to study and evaluate the biological activity of IGF-1 isoforms in BC cells showing an induction of cell proliferation [4]. This study underlines the importance of an accurate assessment of the presence of IGF-1 pro-forms within the BC microenvironment in order to select a patient's course of treatment, since these pro-forms promote BC progression independently of the mature IGF-1. The paper highlights the plasticity of cancer cells and the need for a different long-term approach in the battle against BC. In this regard and in collaboration with Dr. Franco Berrino and Dr. Anna Villarini from the National Cancer Institute in Milan, it was possible to analyze modification in lifestyle, including dietary change-associated with moderate physical activity for the BC prevention [10].

In particular, in this study data obtained from the Italian DIANA-5 clinical trial (randomized trial of diet, physical activity and BC recurrences: the DIANA-5 study in progress since 2008, funded by AIRC and Health Department), to assess whether Mediterranean diet and moderate physical activity influence circulating levels of unbound IGF-1, were analyzed. The DIANA-5 study focus on the effectiveness of a diet based on Mediterranean and macrobiotic recipes and principles, associated with moderate physical activity, in reducing additional BC events (prevention of BC recurrence) in women with early-stage invasive BC. The patients involved in this project are aged between 35 and 70 years diagnosed with BC over the past five years without BC recurrences or other cancers. The control group receives routine care for diet and physical activity based on international recommendations (WCRF/AICR 2007); scientists assisted the intervention group with regards to eating habits and to the acquisition of a more active lifestyle. The main aim of the study is to evaluate the impact of moderate intensity exercise intervention combined with a controlled diet on serum IGF-1 levels in the BC survivors of DIANA 5, as a response dependent on diet and/or dependent on physical activity and to assess whether the risk of BC recurrence may be related to specific IGF-1 isoform or the overall IGF-1s pool.

2. Materials and Methods

2.1 DIANA-5 study design

The overview of DIANA-5 study design is described in Villarini et al., 2012 [10]. Between 2008 and 2012, the study randomly assigned 1675 patients to an intensive diet and exercise intervention (841 patients) or to a control group (834 patients) to be followed-up after 5 years. General lifestyle recommendations for the prevention of cancer are given to both groups, and the intervention group is being offered a comprehensive lifestyle intervention, including cooking classes, conferences, common meals and exercise sessions. Adherence assessments occurred at baseline and at 12 months and are planned at 36 and 60 months. They include recall 24 h, anthropometric measures, body fat distribution assessed with impedance scale, one week registration of physical activity with a multisensor arm-band monitor, metabolic and endocrine blood parameters. Outcome BC events are assessed through self report at semi-annual meetings or telephone interview and are validated through medical record verification. At baseline and after 12 months, blood was drawn, separated in aliquots of red blood cells, buffy coat, plasma and serum, and stored at -80 °C. The collection of further blood samples is planned at 36 and 60 months.

2.2 Intervention

Before randomization into intervention or control group, all women received a leaflet illustrating the WCRF/AICR recommendations for cancer prevention (Table 1). After randomization, only the intervention group was offered a counselling program supported by cooking classes, physical activity sessions, conferences and print materials. The counselling program purposes to increase physical activity, control weight, promotes a healthy and low calorie diet.

The physical activity aims were primarily based on achieving and maintaining regular participation in a moderate intensity physical activity (approximately 3 to 5 metabolic equivalents) program of 210 min/week (30 min on the average per day) over at least 3 days/week, and decreasing sedentary behaviour by 30 min/day on at least 5 days/week. During the first 12 months, one group of exercise session per month is offered to improve program adoption. The focus is on maintaining moderate intensity activities, such as walking. In addiction women are encouraged to progress in more vigorous activity.

Participants have been provided with a pedometer for self-monitoring and compliance enhancement.

The dietary goals were based on reducing glycaemic and insulinemic response through moderate caloric restriction, reduce consumption of high glycaemic index foods, reduce protein intake, particular animal protein, and preferring consumption of satiating foods.

Table 1. WCRF/AICR 2007 recommendations

1. Be as lean as possible within the normal range of body weight
2. Be physically active as part of everyday life
3. Limit consumption of energy-dense food & avoid sugary drinks
4. Eat mostly food of plant origin, with a variety of non-starchy vegetables and fruit every day and unprocessed cereals and/or pulses within every meal
5. Limit intake of red meat & avoid processed meat
6. Limit alcoholic drinks
7. Limit consumption of salt & avoid mouldy cereals or pulses
8. Aim to meet nutritional needs through diet alone
9. Mothers to breastfeed; children to be breastfed
10. Cancer survivors: follow the recommendations for cancer prevention WCRF, World Cancer Research Fund; AICR, American Institute of Cancer Research.

2.3 Level of Physical Activity

In 563 DIANA-5 BC survivors was collected metabolic physical activity information using a metabolic holter (SenseWear® armband device). The SenseWear® armband was wear on the back of the upper arm for a full week to evaluate physical activity duration, total energy expenditure, active and resting energy expenditure in metabolic equivalents (METs), total number of steps, sleep duration and lying down. Data were analyzed through Software Cronolife. The patients of the intervention group were asked to maintain regular participation in a moderate intensity physical activity (approximately 3 to 5 METs) program of 210 minutes/week (30 min on average per day) over at least 3 days/week. The analyses were conducted with intervals of 12 months from the beginning of the intervention. To provide motivational feedbacks the patients will be provided with a pedometer to collect daily walking information. In addition, agreements with some public gyms, fitness centers, yoga or martial arts centers, and dance schools were draw up, to give the opportunity to the patients to attend (once a month) theoretical and practical lessons. Women wishing to take up vigorous sports were encouraged to do it.

During the study, to monitor the patients compliance, recall 24-hour of physical activity and nutrition were delivered to high risk women. For diary compilation the patients were contacted by telephone and requested to information regarding the previous day. Patients were follow-up for a longer period (5 years) to repeat the IGF-1 analysis.

2.4 Radioimmunoassay (RIA) for IGF-1

The radioimmunoassay was performed according to the protocol with the Insulin-like growth factor-1 reagent pack provided by Amersham (code IM 1721). The reaction mixture contained 100 µL of assay buffer, unlabeled IGF-1 (0.05–3.2 ng) or serum samples and 100 µL of diluted antibody 1:4000 with assay buffer. Samples were incubated for 30 min at room temperature and then 100 µL of 125I-IGF-1 was added and incubated for 48 h at 4° C. After the incubation 500 µL of second antibody reagent was added. The mixture was incubated for 10 min at room temperature and centrifuged at 25,000 g for 10 min. The radioactivity of the sediment was determined. A standard competition curve was established using 0.05–3.2 ng of unlabeled IGF-1 per tube. The radioactivity of control samples was subtracted from the radioactivity of the test samples to correct for non-specific binding.

2.5 Statistical Analysis

Statistical analysis approach to estimate the likely effect of the proposed lifestyle intervention, data from epidemiological studies of prognosis and lifestyle were used and described in [10]. All variables are summarized with descriptive statistics: total count, mean, standard deviation, median, gamma and IQR for continuous variables, and cumulative frequency (count and percentage) for categorical variables. The continuous variables will be subjected to verification of normal or gaussian distribution and the comparative statistics pre-post exercise will be performed with Friedman test if the outcome measures won't result normally distributed; on the other hand, the two-way repeated measures ANOVA test will be applied to measure the post-treatment outcome results variations. The categorical variables will be compared using the Chi-square test. Logistic regression equation will be used to search predictive factors of the slowing down of tumor cellular growth. Calculation of the statistical power has been performed according to the variation of IGF-1 measured after the lifestyle intervention and for this calculation we hypothesized that a comprehensive dietary intervention plus physical

activity will reduce BC recurrence rate by 33% or more, but we also considered a more conservative estimate of a 25% reduction.

3. Results

In order to determine whether nutrition and physical activity can affect circulating IGF-1 levels, radioimmunoassay was used to quantify the circulating IGF-1 levels in the serum of 600 DIANA-5 BC survivors.

In addition, the volunteers worn a metabolic holter for 6 consecutive days to monitor physical activity, assess energy expenditure, METs. METs and indicate of food consumption frequencies through a 24h recall. Compared to the 600 women, 563 were included in the statistical analysis since in 37 women the holter was not properly worn and physical activity data were incomplete. The anthropometric and clinical characteristics of the population in the study are given in Table 2.

Variables	Mean (± SD)
Age	52.16 (± 8.29)
Waist circumference (cm)	86.62 (± 12.06)
Weight (Kg)	68.86 (± 12.85)
Height (m)	160. 56 (± 6.44)
BMI (kg/m^2)	26.83 (± 5.12)
Diastolic pressure (mmHg)	82.34 (±11.51)
Systolic pressure (mmHg)	126.26 (±18.27)
Glycaemia (mg/dl)	95.44 (±18.77)
Total cholesterol (mg/dl)	209.16 (±36.88)
HDL cholesterol (mg/dl)	60.07 (±15.36)
LDL cholesterol (mg/dl)	126.33 (±36.63)
Triglycerides (mg/dl)	116.73 (± 85.08)
IGF-1 (ng/ml)	161.77 (± 61.41)
Education	
Primary school	23.68%
High school	43.80%
University degree	32.52%
Current smokers	
YES	11.01%
NO	88.99%

Table 2. Characteristics of DIANA-5 study population at baseline by this study group.

In the study group, the median of circulating IGF-1 levels was 151.32 ng/ml. The average of circulating IGF-1 levels in those with an IGF-1<151.32 ng/ml results to be 117.38 ng/ml, whereas those with IGF-1≥151.32 ng/ml are 206.01 ng/ml.

The data show a statistically significant relationship between the circulation levels of

IGF-1 and the consumption of milk and dairy foods, total consumption of predominantly protein foods (meats, eggs, milk, dairy products, yogurt, cheeses, fish, shellfish, crustaceans, legumes, soy) as well as the consumption of predominantly protein animal foods, the consumption of protein animal foods without meats and the total consumption of sugars added to foods and beverages. The significance for protein foods is not confirmed when they are excluded from the pattern of milk consumption and dairy products.

There is also a significant increase in the levels of IGF-1 in relation to the METs of physical activity and the daily minutes doing moderate physical activity, but there is no significant relation with for average walking steps in one day.

Additionally, IGF-1 levels resulted influenced by BMI (with a growing trend up to the 3rd quintile) and age (with a gradual and constant drop aging related).

Table 3 presents data obtained by evaluating the all population through the t-test and the Wilcoxon test in relation to the frequency of consumed foods and physical activity expressed as average of METs, minutes of moderate daily physical activity and the average of daily steps.

	IGF-1<151.32 ng/ml n.281	**IGF-1≥151.32 ng/ml n.282**	**P (t-test)**	**P (wilcoxon test)**
Milk and dairy products	1.34 (±0.97)	1.56 (±1.05)	**0.013**	**0.02**
Total protein foods	2.91 (±1.56)	3.26 (±1.8)	**0.01**	**0.01**
Animal protein	2.66 (±1.48)	2.97 (±1.73)	**0.02**	**0.04**
Animal protein foods without milk and dairy products	1.56 (±1.09)	1.70 (±1.3)	0.15	0.21
Vegetable protein	0.24 (±0.5)	0.29 (±0.5)	0.29	0.18
Protein animal foods without meats	1.85 (±1.21)	2.12 (±1.4)	**0.016**	**0.02**
Total added sugar foods	1.58 (±1.52)	1.82 (±1.51)	0.06	**0.03**
Physical activity (METs)	1.39 (± 0.23)	1.45 (± 0.21)	**0.002**	
Daily minutes doing moderate physical activity	79.81 (± 46.03)	93.50 (± 53.10)	**0.001**	
Daily walking steps	10342 (± 3555)	10815 (± 3326)	0.1	

Table 3. Frequency of consumed foods and physical activity in relation to IGF-1 levels.

In Table 4 is notable that less active or sedentary subjects (METs<1.4) have IGF-1 levels significantly lower than more active subjects, excluded by the number of daily average steps.

	Less active or sedentary subjects n.275	Active subjects n.288	P (t-test)
± 10,000 steps	158.03 (± 56.04)	165.14 (± 65.79)	0.17
Daily minutes doing moderate physical activity above and below the median (81.34)	153.096 (± 56.62)	169.5 (± 65.01)	**0.002**
METs above and below the median (1.4)	154.04 (± 55.64)	169.15 (± 65.7)	**0.003**

Table 4. IGF-1 levels in active and sedentary subjects.

The relationship between IGF-1 levels of age and BMI was assessed and in agreement with Pasanisi et al. 2011 [11], it emerges that IGF-1 levels are affected by BMI, with a growing trend up to the 3rd quintile and by aging, with a gradual and constant drop in the age (Figure 2).

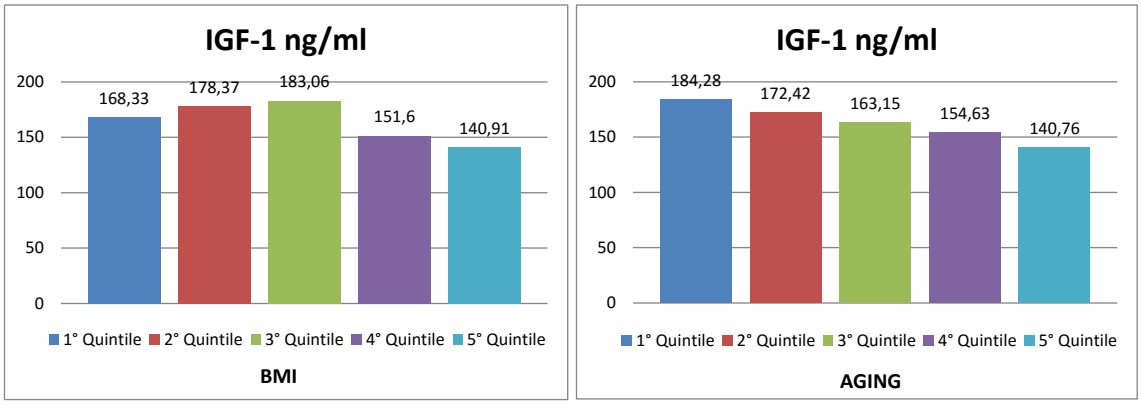


Figure 2. Relationship between IGF-1 levels and BMI and age.

In accord with these, the relationship between the circulation levels of IGF-1, age and BMI was evaluated. In women aged <52.1 years (Table 5), data show a statistically significant relationship with the consumption of predominantly protein foods (meats, salami, eggs, milk, dairy products, yoghurt, cheeses, fish, molluscs, crustaceans, legumes, soya) with the consumption of predominantly protein animals, and with the total consumption of sugars added to food and drink. In addition, a positive trend with the consumption of milk and dairy products was revealed. Significance for protein foods is not confirmed when they are excluded from the pattern of milk consumption and dairy products. There is, however, no significant increase in IGF-1 levels in relation to the

METs of physical activity, to the daily minutes of doing moderate physical activity and to the daily walking average.

	IGF-1<151.32 ng/ml n.115	IGF-1≥151.32 ng/ml n.167	P (t-test)	P (wilcoxon test)
Milk and dairy products	1.33 (±0.92)	1.55 (±1.05)	0.08	0.14
Total protein foods	2.88 (±1.57)	3.32 (±1.82)	**0.04**	**0.05**
Animal protein	2.59 (±1.49)	3.00 (±1.85)	**0.04**	0.09
Animal protein foods without milk and dairy products	1.55 (±1.09)	1.77 (±1.3)	0.17	0.11
Vegetable protein	0.29 (±0.59)	0.31 (±0.53)	0.74	**0.5**
Protein animal foods without meats	1.83 (±1.19)	2.09 (±1.5)	0.12	0.19
Total added sugar foods	0.92 (±0.94)	1.06 (±1.03)	**0.04**	**0.04**
Physical activity (METs)	1.43 (± 0.22)	1.47 (± 0.21)	0.13	
Daily minutes doing moderate physical activity	87.5 (± 49.01)	94.2 (± 57)	0.3	
Daily walking steps	10759 (±3572)	10893 (± 3477)	0.75	

Table 5. Relationship between IGF-1 levels and nutrient consumption in women aged<52.1 years.

Table 6 show the comparison with less active or sedentary subjects (METs <1.4) and the most active subjects, but no significant data emerge.

	Less active or sedentary subjects n.117	Active subjects n.165	P (t-test)
± 10,000 steps	169.63 (± 58.33)	176.24 (± 68.3)	0.39
Daily minutes doing moderate physical activity above and below the median (81.34)	169.62 (± 59.41)	176.51 (± 67.59)	0.36
METs above and below the median (1.4)	170.15 (± 57.78)	175.67 (± 68.13)	0.47

Table 6. IGF-1 levels in active and sedentary subjects in women aged<52.1 years.

In women with age>=52.1 years (Table 7) the data show a statistically significant relationship only to the consumption of predominantly protein without meat and a positive trend with the consumption of milk and dairy products is confirmed. Significance for protein foods is not confirmed when they are excluded from the pattern of milk consumption and dairy products. There is also a significant increase in the levels of IGF-1

in relation to the METs of physical activity and the daily minutes of doing moderate physical activity, but there is no significant significance for daily walk averaged.

	IGF-1<151.32 ng/ml n.166	IGF-1≥151.32 ng/ml n.115	P (t-test)	P (wilcoxon test)
Milk and dairy products	1.36 (±1.0)	1.57 (±0.98)	0.07	0.08
Total protein foods	2.93 (±1.55)	3.19 (±1.63)	0.18	0.17
Animal protein	2.72 (±1.47)	2.93 (±1.55)	0.23	0.22
Animal protein foods without milk and dairy products	1.57 (±1.02)	1.61 (±1.1)	0.73	0.89
Vegetable protein	0.21 (±0.43)	0.25 (±0.45)	0.44	0.38
Protein animal foods without meats	1.86 (±1.23)	2.15 (±1.25)	**0.05**	**0.04**
Total added sugar foods	0.9 (±0.96)	1.13 (±0.96)	0.1	0.1
Physical activity (METs)	1.36 (± 0.24)	1.41 (± 0.2)	**0.07**	
Daily minutes doing moderate physical activity	74.43 (± 43.12)	92.4 (± 47)	**0.001**	
Daily walking steps	10053 (±3524)	10702 (± 3150)	0.1	

Table 7. Relationship between IGF-1 levels and nutrient consumption in women age>=52.1 years.

It is noted that less active or sedentary subjects (METs <1.4) have IGF-1 levels significantly lower in relation to daily minutes doing moderate physical activity and to the METs, whereas this is excluded by the number of daily steps average (Table 8).

	Less active or sedentary subjects n.158	Active subjects n.123	P (t-test)
± 10,000 steps	148.27 (± 52.26)	152.08 (± 60.75)	0.57
Daily minutes doing moderate physical activity above and below the median (81.34)	140.86 (± 50.77)	161.18 (± 60.96)	**0.002**
METs above and below the median (1.4)	142.10 (± 50.98)	160.41 (± 61.49)	**0.006**

Table 8. IGF-1 levels in active and sedentary subjects in women age>=52.1 years.

In women with BMI<26.7 (Table 9), the data show a statistically significant relationship with the consumption of milk and dairy products, predominantly protein foods, with the consumption of predominantly protein animals, with the consumption of the total protein foods excluding the meat, and with the total consumption of sugars added to foods and

beverages. Significance for protein foods is not confirmed when they are excluded from the pattern of milk consumption and dairy products. There is, however, no significant increase in IGF-1 levels in relation to the METs of physical activity, to the daily minutes of doing moderate physical activity and to daily walking on average.

	IGF-1<151.32 ng/ml n.127	IGF-1≥151.32 ng/ml n.179	P (t-test)	P (wilcoxon test)
Milk and dairy products	1.25 (±0.96)	1.56 (±0.77)	**0.008**	**0.01**
Total protein foods	2.77 (±1.53)	3.3 (±1.78)	**0.008**	**0.006**
Animal protein	2.51 (±1.48)	2.99 (±1.68)	**0.01**	**0.01**
Animal protein foods without milk and dairy products	1.51 (±1.11)	1.73 (±1.27)	0.13	0.11
Vegetable protein	0.26 (±0.55)	0.3 (±0.53)	0.52	0.36
Protein animal foods without meats	1.86 (±1.24)	2.15 (±1.38)	0.06	**0.05**
Total added sugar foods	0.9 (±0.92)	1.17 (±1.03)	**0.02**	**0.02**
Physical activity (METs)	1.52 (± 0.21)	1.53 (± 0.18)	0.8	
Daily minutes doing moderate physical activity	94.37 (± 52.07)	102.30 (± 55.02)	0.17	
Daily walking steps	11260 (±3614)	11276 (± 3362)	0.98	

Table 9. Relationship between IGF-1 levels and nutrient consumption in women with BMI<26.7.

By comparing less active or sedentary subjects (METs <1.4) with the most active, no significant data emerged (Table 10).

	Less active or sedentary subjects n.83	Active subjects n.223	P (t-test)
± 10,000 steps	175.50 (± 58.99)	175.88 (± 71.41)	0.96
Daily minutes doing moderate physical activity above and below the median (81.34)	169.81 (± 60.0)	179.51 (± 70.6)	0.21
METs above and below the median (1.4)	176.44 (± 65.25)	175.47 (± 67.44)	0.91

Table 10. IGF-1 levels in active and sedentary subjects in women with BMI<26.7.

In women with BMI>=26.7 (Table 11), the data show no statistically significant relationship the circulating IGF-1 levels, with dietary consumption and physical activity.

	IGF-1<151.32 ng/ml n.154	IGF-1≥151.32 ng/ml n.103	P (t-test)	P (wilcoxon test)
Milk and dairy products	1.42 (±0.96)	1.54 (±1.1)	0.35	0.48
Total protein foods	3.02 (±1.57)	3.21 (±1.85)	0.3	0.57
Animal protein	2.79 (±1.46)	2.95 (±1.82)	0.45	0.73
Animal protein foods without milk and dairy products	1.60 (±1.08)	1.66 (±1.34)	0.66	0.87
Vegetable protein	0.22 (±0.46)	0.26 (±0.46)	0.55	0.43
Protein animal foods without meats	1.84 (±1.19)	2.06 (±1.43)	0.17	0.31
Total added sugar foods	0.91 (±0.98)	0.95 (±0.94)	0.77	0.63
Physical activity (METs)	1.28 (± 0.19)	1.30 (± 0.17)	0.32	
Daily minutes doing moderate physical activity	67.81 (± 36.39)	77.36 (± 45.5)	0.06	
Daily walking steps	9579 (±3327)	10014 (± 3120)	0.29	

Table 11. Relationship between IGF-1 levels and nutrient consumption in women with BMI>=26.7.

Even comparing less active or sedentary (METs <1.4) subjects with the most active no significant data emerges (Table 12).

	Less active or sedentary subjects n.192	Active subjects n.65	P (t-test)
± 10,000 steps	143.98 (± 49.42)	146.72 (± 65.7)	0.6
Daily minutes doing moderate physical activity above and below the median (81.34)	142.32 (± 51.12)	149.96 (± 50.0)	0.23
METs above and below the median (1.4)	144.35 (± 47.96)	147.47 (± 54.45	0.6

Table 12. IGF-1 levels in active and sedentary subjects in women with BMI>=26.7.

4. Discussions

The DIANA multicenter randomized controlled trials from 1 to 5 Diana Projects carried out in Milan analyzed if changes on lifestyle, such as a diet based on Mediterranean and macrobiotic principles associated with moderate physical activity, influence the incidence of BC. In particular, data deriving from healthy postmenopausal women (DIANA-1) [12, 13] and from BC patients (DIANA-2) [14] showed that highly satiating and insulin-lowering diet, based on traditional Mediterranean and macrobiotic recipes, significantly

decreases body weight, serum testosterone, and the bioavailability of both estrogens and IGF-1.

It is well documented that hormonal and micro-environmental factors can influence growth and progression of cancer disease [15] and exercise might affect both of them. Numerous studies have investigated the relation between lifestyle factors and BC [16, 17]. Sedentary lifestyle and alcohol intake have been documented to be associated with higher levels of estrogens and androgens [18, 19], insulin [20] and C-peptide [21]. Nutrition and in particular hyper-caloric diet, high protein intake, high glycemic index food and high-fat diet are major determinants of metabolic syndrome, insulin resistance and increased androgenic synthesis and activity. The relationship between a high-fat diet and BC has been attributed to overweight and body adiposity [22], which is associated with altered hormonal environment. On the contrary, moderate calorie restriction and high intake of satiating foods have been associated with low levels of circulating hormones, including thyroid, insulin, testosterone, and estrogen [23].

Post-menopausal overweight is associated with increased concentrations of total and free androgens, estrogens, insulin levels and decreased SHBG which all promote cancer development [24]. The role of insulin in the BC development has been demonstrated. Insulin stimulates the synthesis of androgens in the ovary, the expression of growth hormone receptors, increase levels of bioavailable estrogen via aromatase activity in adipose tissue and through reductions in concentrations of sex hormone-binding globulin (SHBG), a protein that binds to estrogen to decrease sex hormones bioavailability [25]. In addition, insulin inhibits the liver production of and IGFBP-1 and 2, thus increasing the bioavailability of IGF-1 and stimulating the mitogenic and antiapoptotic actions of IGF-1 in BC [26].

The importance of IGF-1 axis in the development and progression of BC has been demonstrated [27]. The over-expression of IGF-1R in BC has been reported and its high expression has been related to poorer survival [28]. However, the direct relationship between circulating IGF-1 as a result of dietary intake and its effect on IGF-1 axis signaling and cancer progression is not well establish.

The levels of circulating IGF-1 increasing by a diet rich in protein, in particular in milk protein [29]. It was postulated that high intake of satiating foods could reduce the risk of cancer development, in particular through phytonutrient, which may regulate cell growth and differentiation and induce apoptosis [30]. In addition, it was shown that moderate

calorie restriction could reduce cancer incidence in rodents due to its inhibitory effect on IGF-1 synthesis [31]. However, this was not revealed in humans because the caloric restriction was insufficient to decrease the IGF-1 concentration [32]. Thus the relation between diet intervention and BC risk in the modulation of circulating IGF-1 levels have yet to be conducted.

Physical activity, including both aerobic and resistance exercise, may affect the prevention and the progression of BC [33] and has been also proposed to promote psychological well-being during and following cancer treatment [34, 35]. Numerous studies have demonstrated that exercise can improve insulin sensitivity, reduces circulating insulin and sex hormones levels [36]. In particular in a randomized controlled trial in postmenopausal women of 12-month diet and exercise interventions showed that sex hormones levels, in particular estradiol levels decreased significantly with a combined diet and exercise intervention [37].

The role of physical exercise in the modulation of the IGF system at both cellular and systemic level, remains poorly understood. Numerous studies on BC lifestyle have investigated the relation between exercise and circulating IGF-1 levels, but with discordant results. Some studies have reported an increase or no difference or even a decrease in circulating IGF-1 levels associated with physical activity [38]. In particular the Yale Exercise and Survivorship trial showed significant reductions in IGF-1 and IGFBP-3 in postmenopausal women compared to non-active after a 6-month walking intervention [39]. Moreover, randomized trials in postmenopausal women based on chronic exercise intervention indicate that exercise confer similar effects in patients with primary BC [39, 40]. On the contrary, another study reported no significant changes in IGFBP-1 and IGFPB-3 after a 12-week intervention [41]. Although there is a lack of consensus in the literature on the effects of physical exercise in the modulation of circulating IGF-1 and BC development at both cellular and systemic levels, it is emerging the need of specification on exercise protocol management (for example acute *vs* chronic or aerobic *vs* resistance exercise), and also in terms of modalities, intensity, frequency and duration for a corrected determination of exercise dose–response.

The physiological muscle adaptations to exercise induce tissue-relevant biological processes (including metabolic, angiogenic and immune responses). In these light, the acute exercise activates a diverse network of transcription factors, kinases and co-regulatory proteins that culminate in gene expression changes that may induce local and

systemic catabolic or anabolic response depending on type of exercise and muscle recruited. In agreement, we demonstrated that a single isoinertial exercise session, which emphasize the eccentric overload, at the maximum strength induces muscle damage with an early muscular adaptation, showing a local decrease of all IGF-1 muscle transcript variants and a significant increase of circulating IGF-1 within the 24h post-exercise (Contarelli et al., In preparation). These results reveled different responses induced by acute exercise on cellular and circulating IGF-1. By contrast, chronic exercise, both aerobic and resistance, significantly increases glucose uptake in skeletal muscle *via* insulin-independent mechanisms, leading to decreased basal circulating levels of insulin, IGF-1 and glucose [42]. In this line, the physiological adaptations to exercise acute and chronic, lead early relevant biological processes in skeletal muscle tissue and induce systemic responses in the whole-body homeostasis, which might act on reprogramming 'distant' cancer tissue microenvironments.

Exercise-induced effects are involved on the maintenance of whole organism metabolic control and might alter homeostatic control signals that regulate nutrient and growth factor uptake and availability at both tissue and cancer cell level, causing downstream effects on metabolism and bioenergetics. In a model of mammary carcinogenesis, physical activity caused a delay in carcinogenesis [43], with a concomitant reductions of insulin, IGF-1 and leptin levels in circulation [44]. These decreases at the systemic level concurrent with alterations in metabolic intracellular signaling, indicated by increased activation of AMPK and reductions of activated AKT and mTOR [43].

5. Conclusions

The data show within DIANA-5 clinical trial a statistically significant relationship with the IGF-1 levels, the consumption of milk and dairy food-free diet, with total consumption of predominantly protein foods (meats, meats, eggs, milk, dairy products, yogurt, cheeses, fish, shellfish, crustaceans, legumes, soy) as well as with the consumption of predominantly protein animal foods, the consumption of protein animal foods without meats and the total consumption of sugars added to foods and beverages. The significance for protein foods is not confirmed, when they are excluded from the pattern of milk consumption and dairy products. There is also a significant increase in the levels of IGF-1 in relation to the METs of physical activity and the daily minutes doing

moderate physical activity, but there is no significant significance for daily average walking steps. The IGF-1 levels resulted influenced by BMI and age.

We propose that the transient increases in exercise factors during exercise appear to be mediating the positive effect of regular exercise on BC survivors. Interesting, based on recent evidence, exercise might induce either pro-tumorigenic or anti-tumorigenic effects, depending on the context. However, when the known effects are considered in their global view, we assume that exercise infer an overall cancer-suppressive effect in the most of oncology scenarios. If appropriately prescribed and dosed, exercise might regulate the host-tumor microenvironment interaction preventing cancer recurrence and improving overall survival.

6. References

1. Singh, P., J.M. Alex, and F. Bast, *Insulin receptor (IR) and insulin-like growth factor receptor 1 (IGF-1R) signaling systems: novel treatment strategies for cancer.* Medical oncology, 2014. **31**(1): p. 805.
2. Oberbauer, A., *The Regulation of IGF-1 Gene Transcription and Splicing during Development and Aging.* Front Endocrinol (Lausanne), 2013. **4**(39).
3. Durzyńska, J., et al., *Human Eb peptide: not just a by-product of pre-pro-IGF1b processing?* Hormone and Metabolic Research, 2013. **45**(06): p. 415-422.
4. De Santi, M., et al., *Human IGF1 pro-forms induce breast cancer cell proliferation via the IGF1 receptor.* Cellular Oncology, 2016. **39**(2): p. 149-159.
5. Annibalini, G., et al., *MIR retroposon exonization promotes evolutionary variability and generates species-specific expression of IGF-1 splice variants.* Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 2016. **1859**(5): p. 757-768.
6. Li, H., et al., *Genetic polymorphisms of insulin-like growth factor 1 and insulin-like growth factor binding protein 3, xenoestrogen, phytoestrogen, and premenopausal breast cancer.* Current oncology, 2016. **23**(1): p. e17.
7. Chong, K.Y., et al., *The prognostic significance of the insulin-like growth factor-1 ligand and receptor expression in breast cancer tissue.* Journal of surgical oncology, 2011. **104**(3): p. 228-235.
8. Koelwyn, G.J., et al., *Exercise-dependent regulation of the tumor microenvironment.* Nature Reviews Cancer, 2017. **17**(10): p. nrc. 2017.78.
9. Neufer, P.D., et al., *Understanding the cellular and molecular mechanisms of physical activity-induced health benefits.* Cell metabolism, 2015. **22**(1): p. 4-11.
10. Villarini, A., et al., *Lifestyle and breast cancer recurrences: the DIANA-5 trial.* Tumori, 2012. **98**(1): p. 1-18.
11. Pasanisi, P., et al., *Serum levels of IGF-I and BRCA penetrance: a case control study in breast cancer families.* Familial cancer, 2011. **10**(3): p. 521-528.
12. Berrino, F., et al., *Reducing bioavailable sex hormones through a comprehensive change in diet: the diet and androgens (DIANA) randomized trial.* Cancer Epidemiology and Prevention Biomarkers, 2001. **10**(1): p. 25-33.
13. Kaaks, R., et al., *Effects of dietary intervention on IGF-I and IGF-binding proteins, and related alterations in sex steroid metabolism: the Diet and Androgens (DIANA) Randomised Trial.* European journal of clinical nutrition, 2003. **57**(9): p. 1079-1088.
14. Berrino, F., et al., *Serum testosterone levels and breast cancer recurrence.* International journal of cancer, 2005. **113**(3): p. 499-502.
15. Kumar, V., et al., *Robbins and Cotran Pathologic Basis of Disease, Professional Edition E-Book*2014: Elsevier Health Sciences.
16. Dieli-Conwright, C.M., K. Lee, and J.L. Kiwata, *Reducing the risk of breast cancer recurrence: an evaluation of the effects and mechanisms of diet and exercise.* Current breast cancer reports, 2016. **8**(3): p. 139-150.
17. Patterson, R.E., et al., *Physical activity, diet, adiposity and female breast cancer prognosis: a review of the epidemiologic literature.* Maturitas, 2010. **66**(1): p. 5-15.
18. McTiernan, A., et al., *Relation of BMI and physical activity to sex hormones in postmenopausal women.* Obesity, 2006. **14**(9): p. 1662-1677.

19. Rinaldi, S., et al., *Relationship of alcohol intake and sex steroid concentrations in blood in pre-and post-menopausal women: the European Prospective Investigation into Cancer and Nutrition.* Cancer Causes & Control, 2006. **17**(8): p. 1033-1043.
20. Campbell, K.L. and A. McTiernan, *Exercise and biomarkers for cancer prevention studies.* The Journal of nutrition, 2007. **137**(1): p. 161S-169S.
21. Irwin, M.L., et al., *Relationship of obesity and physical activity with C-peptide, leptin, and insulin-like growth factors in breast cancer survivors.* Cancer Epidemiology and Prevention Biomarkers, 2005. **14**(12): p. 2881-2888.
22. De Pergola, G. and F. Silvestris, *Obesity as a major risk factor for cancer.* Journal of obesity, 2013. **2013**.
23. Holloszy, J.O. and L. Fontana, *Caloric restriction in humans.* Experimental gerontology, 2007. **42**(8): p. 709-712.
24. Cohen, D.H. and D. LeRoith, *Obesity, type 2 diabetes, and cancer: the insulin and IGF connection.* Endocr Relat Cancer, 2012. **19**(5): p. F27-45.
25. Rock, C.L., et al., *Favorable changes in serum estrogens and other biologic factors after weight loss in breast cancer survivors who are overweight or obese.* Clin Breast Cancer, 2013. **13**(3): p. 188-195.
26. Christopoulos, P.F., P. Msaouel, and M. Koutsilieris, *The role of the insulin-like growth factor-1 system in breast cancer.* Molecular cancer, 2015. **14**(1): p. 43.
27. Belardi, V., et al., *Insulin and IGFs in obesity-related breast cancer.* Journal of mammary gland biology and neoplasia, 2013. **18**(3-4): p. 277-289.
28. Law, J.H., et al., *Phosphorylated insulin-like growth factor-i/insulin receptor is present in all breast cancer subtypes and is related to poor survival.* Cancer research, 2008. **68**(24): p. 10238-10246.
29. Qin, L.-Q., K. He, and J.-Y. Xu, *Milk consumption and circulating insulin-like growth factor-I level: a systematic literature review.* International journal of food sciences and nutrition, 2009. **60**(sup7): p. 330-340.
30. Shu, L., et al., *Phytochemicals: cancer chemoprevention and suppression of tumor onset and metastasis.* Cancer and Metastasis Reviews, 2010. **29**(3): p. 483-502.
31. Berryman, D.E., et al., *Role of the GH/IGF-1 axis in lifespan and healthspan: lessons from animal models.* Growth Hormone & IGF Research, 2008. **18**(6): p. 455-471.
32. Fontana, L., S. Klein, and J.O. Holloszy, *Long-term low-protein, low-calorie diet and endurance exercise modulate metabolic factors associated with cancer risk.* The American journal of clinical nutrition, 2006. **84**(6): p. 1456-1462.
33. Baumann, F.T., et al., *Physical activity in breast cancer patients during medical treatment and in the aftercare-a review.* Breast Care, 2013. **8**(5): p. 330-334.
34. Meneses-Echavez, J.F., E. Gonzalez-Jimenez, and R. Ramírez-Vélez, *Supervised exercise reduces cancer-related fatigue: a systematic review.* Journal of physiotherapy, 2015. **61**(1): p. 3-9.
35. Meneses-Echávez, J.F., E. González-Jiménez, and R. Ramírez-Vélez, *Effects of supervised exercise on cancer-related fatigue in breast cancer survivors: a systematic review and meta-analysis.* BMC cancer, 2015. **15**(1): p. 77.
36. Kruk, J. and U. Czerniak, *Physical activity and its relation to cancer risk: updating the evidence.* Asian Pac J Cancer Prev, 2013. **14**(7): p. 3993-4003.
37. Campbell, K.L., et al., *Reduced-calorie dietary weight loss, exercise, and sex hormones in postmenopausal women: randomized controlled trial.* Journal of clinical oncology, 2012. **30**(19): p. 2314-2326.

38. Meneses-Echávez, J.F., et al., *The insulin-like growth factor system is modulated by exercise in breast cancer survivors: a systematic review and meta-analysis.* BMC cancer, 2016. **16**(1): p. 682.
39. Irwin, M.L., et al., *Randomized controlled trial of aerobic exercise on insulin and insulin-like growth factors in breast cancer survivors: the Yale Exercise and Survivorship study.* Cancer Epidemiology and Prevention Biomarkers, 2009. **18**(1): p. 306-313.
40. Fairey, A.S., et al., *Effects of exercise training on fasting insulin, insulin resistance, insulin-like growth factors, and insulin-like growth factor binding proteins in postmenopausal breast cancer survivors.* Cancer Epidemiology and Prevention Biomarkers, 2003. **12**(8): p. 721-727.
41. Sprod, L.K., et al., *Health-related quality of life and biomarkers in breast cancer survivors participating in tai chi chuan.* Journal of cancer survivorship, 2012. **6**(2): p. 146-154.
42. Stanford, K.I. and L.J. Goodyear, *Exercise and type 2 diabetes: molecular mechanisms regulating glucose uptake in skeletal muscle.* Advances in physiology education, 2014. **38**(4): p. 308-314.
43. Zhu, Z., et al., *Effect of nonmotorized wheel running on mammary carcinogenesis: circulating biomarkers, cellular processes, and molecular mechanisms in rats.* Cancer Epidemiology and Prevention Biomarkers, 2008. **17**(8): p. 1920-1929.
44. Thompson, H.J., et al., *Wheel Running Induced Changes in Plasma Biomarkers and the Carcinogenic Response in the 1-Methyl-1-Nitrosourea Induced Rat Model for Breast Cancer.* Cancer Prevention Research, 2010: p. canprevres. 0078.2010.

CONCLUSIONS

Breast Cancer (BC) is the most common type of cancer for women worldwide and is the second leading cause of cancer death in women. Numerous studies have linked high serum levels of IGF-1 and BC, although the mechanisms of action of this growth factor are not yet fully known.

In contrast to the IGF-1 circulating level little is known about the local role of IGF-1. Growing evidence has shown that tissue production of IGF-1 is controlled by complex transcriptional and post-translational modifications. Indeed, distinct IGF-1 isoforms may arise *via* the use of different promoters, alternative splicing, proteolytic processing and glycosylation events.

We demonstrated that *igf-1* isoforms are differentially expressed across tissues and species and using an *igf-1* minigene approach we showed that several cis-acting elements (e.g. 5' splice site strength and exonic splice enhancer) finely regulated the splicing pattern of *igf-1*. These results prompt us to get more insights into IGF-1 isoforms native regulation.

We highlighted that the IGF-1 protein retaining C-terminal E-domains (i.e. IGF-1 pro-hormones) are the predominant forms produced intracellularly, instead of mature IGF-1, and are subjected to extensive post-translational modifications, further hinting to their functional relevance.

The characterization of IGF-1 pro-hormone structures showed that the E-domains contain disorder-promoting amino acids and that there is substantial evolutionary pressure to keep the different E-domains as intrinsically disordered regions (IDRs). We also demonstrated that the E-domains are regulatory elements for the modulation of IGF-1 isoforms stability, intracellular localization and efficient secretion.

Moreover, the results of our experiments indicate that IGF-1 pro-hormones possess biological roles *per se*, i.e. independently from the mature IGF-1, and that pro-hormones can activate IGF-1 receptor signaling pathways and BC cell proliferation. This study underlines the importance of an accurate assessment of the IGF-1 pool to select a patient's course of treatment.

The level of circulating IGF-1 can be influenced both by diet and physical activity. In this regard, preliminary results obtained from Italian DIANA-5 clinical trial, show a significant relationship between IGF-1 level and protein foods as well as the total consumption of sugars added to foods and beverages. There is also a significant increase in the levels of IGF-1 in relation to physical activity and the daily minutes doing

moderate physical activity. Further analyses are still ongoing to detect and characterize the IGF-1 isoforms in serum and tissues and identify association with BC recurrence, metastasis or second tumor.

IGF-1 is a point of convergence for major signaling pathways involved in BC growth. This Thesis work highlights the importance of assessing the IGF-1 isoforms, i.e. splice variants, pro-hormones and mature IGF-1, and their association with IGF-1R signaling in BC.

A better comprehension of the complexity of IGF-1 isoforms regulation and function, alongside their accurate detection, quantification and occurrence at a local and systemic level, will be useful for the National Health Service in developing strategies for primary and secondary prevention of BC.

REFERENCES

Introduction

1. Luigi, L., N. Annalisa, and G. Francesco, *The IGF-I Signaling Pathway.* Current Pharmaceutical Design, 2007. **13**(7): p. 663-669.
2. Philippou, A., et al., *The complexity of the IGF1 gene splicing, posttranslational modification and bioactivity.* Mol Med, 2014. **20**: p. 202-14.
3. Le Roith, D., et al., *The Somatomedin Hypothesis: 2001.* Endocrine Reviews, 2001. **22**(1): p. 53-74.
4. Sjögren, K., et al., *A model for tissue-specific inducible insulin-like growth factor-I (IGF-I) inactivation to determine the physiological role of liver-derived IGF-I.* Endocrine, 2002. **19**(3): p. 249-256.
5. LeRoith, D., et al., *INsulin-like growth factors in health and disease.* Annals of Internal Medicine, 1992. **116**(10): p. 854-862.
6. Perrini, S., et al., *The GH/IGF1 axis and signaling pathways in the muscle and bone: mechanisms underlying age-related skeletal muscle wasting and osteoporosis.* Journal of Endocrinology, 2010. **205**(3): p. 201-210.
7. Leung, K.-C., et al., *Estrogen Regulation of Growth Hormone Action.* Endocrine Reviews, 2004. **25**(5): p. 693-721.
8. Isley, W.L., L.E. Underwood, and D.R. Clemmons, *Dietary components that regulate serum somatomedin-C concentrations in humans.* Journal of Clinical Investigation, 1983. **71**(2): p. 175-182.
9. Thissen, J.-P., J.-M. Ketelslegers, and L.E. Underwood, *Nutritional Regulation of the Insulin-Like Growth Factors*.* Endocrine Reviews, 1994. **15**(1): p. 80-101.
10. Siegfried, J.M., et al., *A mitogenic peptide amide encoded within the E peptide domain of the insulin-like growth factor IB prohormone.* Proceedings of the National Academy of Sciences, 1992. **89**(17): p. 8107-8111.
11. Wallis, M., *New insulin-like growth factor (IGF)-precursor sequences from mammalian genomes: the molecular evolution of IGFs and associated peptides in primates.* Growth Hormone and IGF Research, 2009. **19**(1): p. 12-23.
12. Barton, E.R., *The ABCs of IGF-I isoforms: impact on muscle hypertrophy and implications for repair.* Applied Physiology, Nutrition, and Metabolism, 2006. **31**(6): p. 791-797.
13. Adamo, M.L., et al., *Structure, expression, and regulation of the IGF-I gene*, in *Current Directions in Insulin-Like Growth Factor Research*1994, Springer. p. 1-11.
14. Temmerman, L., E. Slonimsky, and N. Rosenthal, *Class 2 IGF-1 isoforms are dispensable for viability, growth and maintenance of IGF-1 serum levels.* Growth Hormone & IGF Research, 2010. **20**(3): p. 255-263.
15. Hameed, M., et al., *Expression of IGF-I splice variants in young and old human skeletal muscle after high resistance exercise.* The Journal of physiology, 2003. **547**(1): p. 247-254.
16. Oberbauer, A., *The Regulation of IGF-1 Gene Transcription and Splicing during Development and Aging.* Front Endocrinol (Lausanne), 2013. **4**(39).
17. Duguay, S.J., J. Lai-Zhang, and D.F. Steiner, *Mutational analysis of the insulin-like growth factor I prohormone processing site.* Journal of Biological Chemistry, 1995. **270**(29): p. 17566-17574.

18. Rotwein, P., *Two insulin-like growth factor I messenger RNAs are expressed in human liver.* Proceedings of the National Academy of Sciences, 1986. **83**(1): p. 77-81.
19. Philippou, A., et al., *Expression of IGF-1 isoforms after exercise-induced muscle damage in humans: characterization of the MGF E peptide actions in vitro.* In vivo, 2009. **23**(4): p. 567-575.
20. Armakolas, A., et al., *Preferential expression of IGF-1Ec (MGF) transcript in cancerous tissues of human prostate: evidence for a novel and autonomous growth factor activity of MGF E peptide in human prostate cancer cells.* Prostate, 2010. **70**(11): p. 1233-42.
21. Milingos, D.S., et al., *Insulinlike growth factor-1Ec (MGF) expression in eutopic and ectopic endometrium: characterization of the MGF E-peptide actions in vitro.* Mol Med, 2011. **17**(1-2): p. 21-8.
22. Chew, S.L., et al., *An alternatively spliced human insulin-like growth factor-I transcript with hepatic tissue expression that diverts away from the mitogenic IBE1 peptide.* Endocrinology, 1995. **136**(5): p. 1939-1944.
23. Hill, M. and G. Goldspink, *Expression and Splicing of the Insulin-Like Growth Factor Gene in Rodent Muscle is Associated with Muscle Satellite (stem) Cell Activation following Local Tissue Damage.* The Journal of physiology, 2003. **549**(2): p. 409-418.
24. Duguay, S.J., et al., *Processing of wild-type and mutant proinsulin-like growth factor-IA by subtilisin-related proprotein convertases.* Journal of Biological Chemistry, 1997. **272**(10): p. 6663-6670.
25. Duckert, P., S. Brunak, and N. Blom, *Prediction of proprotein convertase cleavage sites.* Protein Engineering Design and Selection, 2004. **17**(1): p. 107-112.
26. Wilson, H., et al., *Monoclonal antibodies to the carboxy-terminal Ea sequence of pro-insulin-like growth factor-IA (proIGF-IA) recognize proIGF-IA secreted by IM9 B-lymphocytes.* Growth Hormone & IGF Research, 2001. **11**(1): p. 10-17.
27. Durzynska, J., et al., *The pro-forms of insulin-like growth factor I (IGF-I) are predominant in skeletal muscle and alter IGF-I receptor activation.* Endocrinology, 2013. **154**(3): p. 1215-24.
28. Yuasa, K., et al., *Subtilisin-like Proprotein Convertase PACE4 is Required for Skeletal Muscle Differentiation.* The Journal of Biochemistry, 2009. **146**(3): p. 407-415.
29. Conover, C.A., et al., *Human hepatoma cells synthesize and secrete insulin-like growth factor Ia prohormone under growth hormone control.* Regulatory peptides, 1993. **48**(1): p. 1-8.
30. Bach, M.A., et al., *Alternative splicing produces messenger RNAs encoding insulin-like growth factor-I prohormones that are differentially glycosylated in vitro.* Molecular Endocrinology, 1990. **4**(6): p. 899-904.
31. Imperiali, B., et al., *Chemistry and biology of asparagine-linked glycosylation.* Pure and applied chemistry, 1999. **71**(5): p. 777-787.
32. Brisson, B.K. and E.R. Barton, *New Modulators for IGF-I Activity within IGF-I Processing Products.* Front Endocrinol (Lausanne), 2013. **4**: p. 42.
33. Koczorowska, M.M., A. Kwasniewska, and A. Gozdzicka-Jozefiak, *IGF1 mRNA isoform expression in the cervix of HPV-positive women with pre-cancerous and cancer lesions.* Experimental and therapeutic medicine, 2011. **2**(1): p. 149-156.
34. Kasprzak, A., et al., *Differential expression of IGF-1 mRNA isoforms in colorectal carcinoma and normal colon tissue.* International journal of oncology, 2013. **42**(1): p. 305-316.

35. Barton, E.R., J. DeMeo, and H. Lei, *The insulin-like growth factor (IGF)-I E-peptides are required for isoform-specific gene expression and muscle hypertrophy after local IGF-I production.* Journal of applied physiology, 2010. **108**(5): p. 1069-1076.
36. Barton, E.R., *Viral expression of insulin-like growth factor-I isoforms promotes different responses in skeletal muscle.* Journal of applied physiology, 2006. **100**(6): p. 1778-1784.
37. Shimatsu, A. and P. Rotwein, *Mosaic evolution of the insulin-like growth factors. Organization, sequence, and expression of the rat insulin-like growth factor I gene.* Journal of Biological Chemistry, 1987. **262**(16): p. 7894-7900.
38. Tan, D.S., A. Cook, and S.L. Chew, *Nucleolar localization of an isoform of the IGF-I precursor.* BMC cell biology, 2002. **3**(1): p. 17.
39. Yang, S., et al., *Cloning and characterization of an IGF-1 isoform expressed in skeletal muscle subjected to stretch.* Journal of Muscle Research & Cell Motility, 1996. **17**(4): p. 487-495.
40. Goldspink, G., *Research on mechano growth factor: its potential for optimising physical training as well as misuse in doping.* British Journal of Sports Medicine, 2005. **39**(11): p. 787-788.
41. Yang, S.Y. and G. Goldspink, *Different roles of the IGF-I Ec peptide (MGF) and mature IGF-I in myoblast proliferation and differentiation.* FEBS Lett, 2002. **522**(1-3): p. 156-160.
42. Matheny, J.R.W., B.C. Nindl, and M.L. Adamo, *Minireview: Mechano-Growth Factor: A Putative Product of IGF-I Gene Expression Involved in Tissue Repair and Regeneration.* Endocrinology, 2010. **151**(3): p. 865-875.
43. Philippou, A., et al., *The role of the insulin-like growth factor 1 (IGF-1) in skeletal muscle physiology.* In vivo, 2007. **21**(1): p. 45-54.
44. Brisson, B.K. and E.R. Barton, *Insulin-like growth factor-I E-peptide activity is dependent on the IGF-I receptor.* PLoS One, 2012. **7**(9): p. e45588.
45. Gallagher, J.T., *Heparan sulfate: growth control with a restricted sequence menu.* Journal of Clinical Investigation, 2001. **108**(3): p. 357-361.
46. Hede, M.S., et al., *E-peptides control bioavailability of IGF-1.* PLoS One, 2012. **7**(12): p. e51152.
47. Baxter, R.C., *Insulin-like growth factor (IGF)-binding proteins: interactions with IGFs and intrinsic bioactivities.* American Journal of Physiology - Endocrinology And Metabolism, 2000. **278**(6): p. E967-E976.
48. Cohen, P., *Overview of the IGF-I System.* Hormone Research in Paediatrics, 2006. **65(suppl 1)**(Suppl. 1): p. 3-8.
49. Hwa, V., Y. Oh, and R.G. Rosenfeld, *The Insulin-Like Growth Factor-Binding Protein (IGFBP) Superfamily*.* Endocrine Reviews, 1999. **20**(6): p. 761-787.
50. Baxter, R.C., J.L. Martin, and V.A. Beniac, *High molecular weight insulin-like growth factor binding protein complex. Purification and properties of the acid-labile subunit from human serum.* Journal of Biological Chemistry, 1989. **264**(20): p. 11843-8.
51. Clemmons, D.R., *Insulin-like growth factor binding proteins and their role in controlling IGF actions.* Cytokine & Growth Factor Reviews, 1997. **8**(1): p. 45-62.
52. Clemmons, D.R., *Modifying IGF1 activity: an approach to treat endocrine disorders, atherosclerosis and cancer.* Nat Rev Drug Discov, 2007. **6**(10): p. 821-33.
53. Rodgers, B.D., E.H. Roalson, and C. Thompson, *Phylogenetic analysis of the insulin-like growth factor binding protein (IGFBP) and IGFBP-related protein gene families.* General and Comparative Endocrinology, 2008. **155**(1): p. 201-207.

54. Rajpathak, S.N., et al., *The role of insulin-like growth factor-I and its binding proteins in glucose homeostasis and type 2 diabetes.* Diabetes/metabolism research and reviews, 2009. **25**(1): p. 3-12.
55. Pollak, M.N., E.S. Schernhammer, and S.E. Hankinson, *Insulin-like growth factors and neoplasia.* Nature Reviews Cancer, 2004. **4**(7): p. 505-518.
56. Nakae, J., Y. Kido, and D. Accili, *Distinct and Overlapping Functions of Insulin and IGF-I Receptors.* Endocrine Reviews, 2001. **22**(6): p. 818-835.
57. De Meyts, P. and J. Whittaker, *Structural biology of insulin and IGF1 receptors: implications for drug design.* Nat Rev Drug Discov, 2002. **1**(10): p. 769-783.
58. Rommel, C., et al., *Mediation of IGF-1-induced skeletal myotube hypertrophy by PI (3) K/Akt/mTOR and PI (3) K/Akt/GSK3 pathways.* Nature cell biology, 2001. **3**(11): p. 1009-1013.
59. Bikle, D.D., et al., *Role of IGF-I Signaling in Muscle Bone Interactions.* Bone, 2015. **80**: p. 79-88.
60. Karamouzis, M.V. and A.G. Papavassiliou, *The IGF-1 network in lung carcinoma therapeutics.* Trends in molecular medicine, 2006. **12**(12): p. 595-602.
61. Durai, R., et al., *Biology of insulin-like growth factor binding protein-4 and its role in cancer.* International journal of oncology, 2006. **28**(6): p. 1317-1325.
62. Renehan, A.G., et al., *Insulin-like growth factor (IGF)-I, IGF binding protein-3, and cancer risk: systematic review and meta-regression analysis.* The Lancet, 2004. **363**(9418): p. 1346-1353.
63. Giovannucci, E., et al., *A prospective study of plasma insulin-like growth factor-1 and binding protein-3 and risk of colorectal neoplasia in women.* Cancer Epidemiology and Prevention Biomarkers, 2000. **9**(4): p. 345-349.
64. Fürstenberger, G. and H.-J. Senn, *Insulin-like growth factors and cancer.* The Lancet Oncology, 2002. **3**(5): p. 298-302.
65. Sachdev, D. and D. Yee, *The IGF system and breast cancer.* Endocr Relat Cancer, 2001. **8**(3): p. 197-209.
66. Dupont, J. and D. Le Roith, *Insulin-like growth factor 1 and oestradiol promote cell proliferation of MCF-7 breast cancer cells: new insights into their synergistic effects.* Molecular Pathology, 2001. **54**(3): p. 149-154.
67. O'Connor, R., C. Fennelly, and D. Krause, *Regulation of survival signals from the insulin-like growth factor-I receptor*, 2000, Portland Press Limited.
68. Treins, C., et al., *Regulation of hypoxia-inducible factor (HIF)-1 activity and expression of HIF hydroxylases in response to insulin-like growth factor I.* Molecular Endocrinology, 2005. **19**(5): p. 1304-1317.
69. Philippou, A., et al., *IGF1Ec expression in MG-63 human osteoblast-like osteosarcoma cells.* Anticancer Research, 2011. **31**(12): p. 4259-4265.
70. Ruan, W. and D.L. Kleinberg, *Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development.* Endocrinology, 1999. **140**(11): p. 5075-5081.
71. Macias, H. and L. Hinck, *Mammary gland development.* Wiley Interdisciplinary Reviews: Developmental Biology, 2012. **1**(4): p. 533-557.
72. Hormones, T.E. and B.C.C. Group, *Insulin-like growth factor 1 (IGF1), IGF binding protein 3 (IGFBP3), and breast cancer risk: pooled individual data analysis of 17 prospective studies.* The Lancet Oncology, 2010. **11**(6): p. 530-542.

73. Kaaks, R., et al., *Insulin-like growth factor I and risk of breast cancer by age and hormone receptor status—A prospective study within the EPIC cohort.* International journal of cancer, 2014. **134**(11): p. 2683-2690.
74. Pasanisi, P., et al., *Serum levels of IGF-I and BRCA penetrance: a case control study in breast cancer families.* Familial cancer, 2011. **10**(3): p. 521-528.
75. Pasanisi, P., et al., *Serum insulin-like growth factor-I and platelet-derived growth factor as biomarkers of breast cancer prognosis.* Cancer Epidemiology and Prevention Biomarkers, 2008. **17**(7): p. 1719-1722.
76. Duggan, C., et al., *Associations of insulin-like growth factor and insulin-like growth factor binding protein-3 with mortality in women with breast cancer.* International journal of cancer, 2013. **132**(5): p. 1191-1200.
77. Jones, R., et al., *Transgenic overexpression of IGF-IR disrupts mammary ductal morphogenesis and induces tumor formation.* Oncogene, 2007. **26**(11): p. 1636.
78. Nielsen, T.O., et al., *Expression of the insulin-like growth factor i receptor and urokinase plasminogen activator in breast cancer is associated with poor survival.* Cancer research, 2004. **64**(1): p. 286-291.
79. Taunk, N.K., et al., *Prognostic significance of IGF-1R expression in patients treated with breast-conserving surgery and radiation therapy.* Radiotherapy and Oncology, 2010. **96**(2): p. 204-208.
80. Zhu, C., et al., *PI3K/Akt and MAPK/ERK1/2 signaling pathways are involved in IGF-1-induced VEGF-C upregulation in breast cancer.* Journal of cancer research and clinical oncology, 2011. **137**(11): p. 1587.
81. Morgillo, F., et al., *Serum insulin-like growth factor 1 correlates with the risk of nodal metastasis in endocrine-positive breast cancer.* Current oncology, 2013. **20**(4): p. e283.
82. Mu, L., et al., *Favorable outcome associated with an IGF-1 ligand signature in breast cancer.* Breast Cancer Research and Treatment, 2012. **133**(1): p. 321-331.
83. Mu, L., et al., *Expressions of IGF-I, IGF-II and IGFBP-3 in breast cancer and their relations with disease characteristics*, 2007, AACR.

Chapter 1

Annibalini G, Bielli P, De Santi M, Agostini D, Guescini M, Sisti D, Contarelli S, Brandi G, Villarini A, Stocchi V, Sette C, Barbieri E. MIR retroposon exonization promotes evolutionary variability and generates species-specific expression of IGF-1 splice variants. Biochim Biophys Acta (BBA) - Gene Regulatory Mechanisms, 2016 May; 1859(5):757-68. doi: 10.1016/j.bbagrm.2016.03.014. Epub 2016 Apr 19.

Chapter 2

De Santi M, Annibalini G, Barbieri E, Villarini A, Vallorani L, Contarelli S, Berrino F, Stocchi V, Brandi G. Human IGF-1 pro-forms induce breast cancer cell

proliferation via the IGF-1 receptor. Cellular Oncology, (Dordr) 2016 Apr; 39(2):149-159. doi: 10.1007/s13402-015-0263-3. Epub 2015 Dec 23.

Chapter 3

Glycosylation stabilizes IGF-1Ea pro-hormone and regulates its secretion – In preparation.

Chapter 4

Italian Diana5 clinical trial: lifestyles diary food-free diet and/or Mediterranean principles and moderate physical activity influence circulating levels of unbound IGF-1 - In preparation.

Circulating IGF-1 and early muscle adaptive responses to an acute isoinertial exercise - In preparation.