Published in final edited form as:

Mol Cell Endocrinol. 2019 May 15; 488: 52–69. doi:10.1016/j.mce.2019.03.004.

Characterizing the Complexity of Australian Marsupial *Insulin-like Growth Factor 1* Genes

Peter Rotwein

Department of Biomedical Sciences, Paul L. Foster School of Medicine, Texas Tech Health University Health Sciences Center, El Paso, Texas 79905

Abstract

Insulin-like growth factor 1 (IGF1) actions are essential for somatic growth and tissue repair. *IGF1* gene regulation is controlled by many inputs, with growth hormone playing a major role. In most mammals, the 6-exon *IGF1/Igf1* gene produces multiple transcripts via independent activity of its promoters plus alternative RNA splicing and differential polyadenylation. Here, by analyzing public genomic and RNA-sequencing repositories, I have characterized three Australian marsupial *IGF1* genes. Koala, Tasmanian devil, and wallaby *IGF1* are more complicated than other mammals, as they contain up to 11 exons, and encode multiple mRNAs and predicted protein precursors, including potentially novel isoforms. Moreover, just two of multiple growth hormone-stimulated transcriptional enhancers found in other *IGF1/Igf1* loci are detected in these species. These observations define Australian marsupial *IGF1* genes and demonstrate that comprehensive interrogation of genomic and RNA-sequencing resources is an effective strategy for characterizing genes and gene expression in otherwise experimentally intractable organisms.

Keywords

IGF1; genomics; gene expression; gene regulation; gene evolution

Introduction

Insulin-like growth factor 1 (IGF1), a small, secreted protein that is produced in most vertebrate species (Das and Dobens, 2015; Daughaday et al., 1985; Rotwein, 1999; Rotwein, 2018; Schwartz and Bronikowski, 2016; Sussenbach et al., 1991), plays central roles in pre- and postnatal somatic growth (LeRoith, 2008; Lupu et al., 2001; Powell-Braxton et al., 1993; Woods et al., 1996). IGF1 also is involved in control of intermediary metabolism, and

Corresponding author: Peter Rotwein, Department of Biomedical Sciences, Paul L. Foster School of Medicine, Texas Tech University Health Sciences Center, 5001 El Paso Drive, El Paso, Texas 79905. Telephone: 915-215-4910, Fax: 915-783-5223, peter.rotwein@ttuhsc.edu.

Author's Contributions

The author is responsible for the ideas leading to this manuscript, for all data collection and analysis, and for writing and editing the paper.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

in tissue repair, and has been implicated in certain cancers and in other disease pathogenesis in adult humans (Gems and Partridge, 2013; Pollak, 2012).

Growth hormone (GH) is a critical regulator of *IGF1* gene expression in mammals (Adamo, 1995; Ahluwalia et al., 2004; Bichell et al., 1992; Chia, 2014; Eleswarapu et al., 2008; Lupu et al., 2001; Mathews et al., 1986), consistent with its physiological role in controlling somatic growth. GH binds to its trans-membrane receptor and induces activity of the Jak2 protein tyrosine kinase (Waters and Brooks, 2015; Waters, 2016), which then causes phosphorylation of intracellular tyrosine residues on the GH receptor, and recruitment of several signal transduction proteins (Waters and Brooks, 2015; Waters, 2016). Among the latter is the Stat5b transcription factor (Woelfle et al., 2003), which upon its activation has been found to dimerize, transit to the nucleus, and bind to DNA at sites near GH-responsive genes, leading to their rapid transcriptional activation (Chia et al., 2010; Chia and Rotwein, 2010; Varco-Merth et al., 2012; Woelfle and Rotwein, 2004).

IGF1/Igf1 genes and IGF1 proteins appear to be conserved in mammals (Daughaday et al., 1985; Rotwein, 1999; Rotwein, 2017a; Rotwein, 2017b; Sussenbach et al., 1991). In these species, the single-copy *IGF1/Igf1* gene has been shown to be composed of six exons and five introns, and to share other structural features, including two promoters, each with a promoter-specific leader exon (Hall et al., 1992; Hoyt et al., 1988; Lowe et al., 1987; Shimatsu and Rotwein, 1987), and the presence of multiple putative Stat5b-binding transcription enhancers in locus chromatin (Chia et al., 2010; Rotwein, 2017a; Rotwein, 2017b). As a result of differential promoter use, alternative RNA splicing, and variable mRNA polyadenylation, six major classes of mRNAs along with many minor species are produced in the cell types in which *IGF1/Igf1* is expressed. These six predominant transcripts each encode a distinct IGF1 protein precursor, which differ in the NH₂-terminal parts of their signal peptides and in the COOH-terminal extension peptides or E domains, but contain the identical mature single-chain 70-amino acid IGF1 protein (Rotwein, 1999; Rotwein, 2017a; Rotwein, 2017b, Wallis, 2009).

Based on recent studies, it appears that the *Igf1* gene is generally simpler in overall structure and mRNA expression in non-mammalian vertebrates than in mammals (Rotwein, 2017a; Rotwein, 2018). In most terrestrial vertebrates analyzed, *Igf1* is composed of four exons and three introns, and has a single promoter that that resembles mammalian *IGF1/Igf1* promoter 1 (Rotwein, 2018). In contrast, in many ray-finned fish, the single-promoter *Igf1* gene contains five exons and four introns, with an additional small, divergent coding exon located 3' to conserved exon 3 (Rotwein, 2018). However, in both groups, there are species with 'extra' *IGF1* exons that vary from the predominant pattern, but none with more than a single promoter (Rotwein, 2018). Moreover, non-mammalian vertebrates lack the Stat5b-binding elements that control GH-activated *IGF1/Igf1* gene transcription in mammals (Rotwein, 2017a; Rotwein, 2018), and thus the mechanisms of *Igf1* gene regulation by GH in these former species are uncharacterized. IGF1 protein precursors in non-mammalian vertebrates also are simpler than in mammals, as there is only a single signal peptide, but there are one or two alternative COOH-terminal E domains, depending on the species (Rotwein, 2018).

The current investigations were initiated after examination of the recently characterized koala genome (Johnson et al., 2018) revealed a single-copy *IGF1* gene with features that appeared to be very different than those found in other mammals, with many more exons being predicted along with multiple classes of putative *IGF1* transcripts containing potentially divergent coding regions. To gain more insight into the organization and expression of the koala *IGF1* gene, the analyses in this manuscript leveraged the vast resources of public genome and gene expression databases, and used them to determine if these unusual features were the norm in koala, and in Tasmanian devil and wallaby, the two other Australian marsupials whose genomes have been sequenced. Collectively, these observations not only delineate the complexity of Australian marsupial *IGF1* genes but also demonstrate that interrogating large-scale public genomic and RNA-sequencing resources is an effective strategy for gaining a full understanding of genes, gene expression, and gene evolution in different organisms, including those in which direct experimental evaluation is difficult.

Materials and Methods

1.1 Database searches and analyses.

Australian marsupial *IGF1* genes were accessed initially in GenBank (<https://www.ncbi.nlm.nih.gov/genome>) for koala (*Phascolarctos cinereus*, genome assembly GCA_900166895.1), and in the Ensembl Genome Browser (www.ensembl.org) for Tasmanian devil (*Sarcophilus harrisii*, Devil_ref v7.0) and wallaby (*Notamacropus eugenii*, Meug_1.0). Queries of all three genomes were performed with *IGF1* gene DNA segments from the other marsupials (using exons, putative promoter regions, and possible Stat5b binding elements), and with rat *Igf1* or human *IGF1* locus gene DNA segments (*Rattus norvegicus*, genome assembly Rnor_6.0 or *Homo sapiens*, genome assembly GRCh38.p12), using BlastN under normal sensitivity (maximum e-value of 10; mis-match scores: 1,-3; gap penalties: opening 5, extension, 2; filtered low complexity regions, and repeat sequences masked). The Sequence Read Archive of the National Center for Biotechnology Information (SRA NCBI; www.ncbi.nlm.nih.gov/sra) was searched with *IGF1* gene fragments from different marsupial species in order to characterize 5' and 3' ends of the genes, novel exons, and differential promoter use, and to identify wallaby *IGF1* exon 3, as these regions could not be evaluated fully within whole genome assemblies. Queries were performed using the megablast option (optimized for highly similar sequences; maximum target sequences–10,000; expect threshold–10; word size–11; match/mismatch scores–2, -3; gap costs–existence 5, extension 2; low-complexity regions filtered). Protein sequences were obtained from the NCBI Consensus CDS Protein Set (<https://www.ncbi.nlm.nih.gov/CCDS/>) or the Uniprot browser (<http://www.uniprot.org/>), or were determined by translation of DNA sequences with assistance of SerialCloner1.3. Results are presented in text, Tables, and figures as percent identity over entire query regions, unless otherwise specified. Other genomes that were searched included the following mammals: dog (*Canis lupus familiaris*, genome assembly CanFam3.1), elephant (*Loxodonta Africana*, LoxAfr3), megabat (*Pteropus vampyrus*, pteVam1), mouse (*Mus musculus*, GRCm38.p6), opossum (*Monodelphis domestica*, monDom5), pig (*Sus scrofa*, Sscrofa11.1), platypus (*Ornithorhynchus anatinus*, OANA5), and the following non-mammalian vertebrates: anole lizard (*Anolis carolinensis*,

AnoCar2.0), chicken (*Gallus gallus*, GRCg6a), Chinese soft shell turtle (*Pelodiscus sinensis*, PelSin_1.0), coelacanth (*Latimeria chalumnae*, LatCha1), frog (*Xenopus tropicalis*, JGI 4.2), lamprey (*Petromyzon marinus*, Pmarinus_7.0), spotted gar (*Lepisosteus oculatus*, LepOcu1), tilapia (*Oreochromis niloticus*, Orenil1.0), turkey (*Meleagris gallopavo*, Turkey_2.0.1), and zebrafish (*Danio rerio*, GRCz11).

1.2 Experimental strategy.

An initial assessment of Australian marsupial *IGF1* genes and transcripts within GenBank and Ensembl genome browsers showed that the assignments appeared to be incomplete or incorrect, and seemed very different from typical six-exon mammalian *IGF1/Igf1* genes (Rotwein, 2017a; Rotwein, 2017b) (see Fig.1). An iterative strategy was developed to characterize koala, Tasmanian devil, and wallaby *IGF1* genes as thoroughly as possible. Starting with exon assignments in GenBank or Ensembl (Fig.1), BlastN searches were conducted against each genome assembly for similar DNA segments. These queries were supplemented with searches using portions of rat *Igf1* or human *IGF1* genes, and were followed by secondary searches, using RNA sequencing libraries found in the SRA NCBI (Table 1) in order to characterize 5' and 3' ends of each marsupial *IGF1* gene, and to identify possible alternative RNA splicing or other features (DNA probes for screening are listed in Table 2). Finally, queries were performed with novel components of marsupial *IGF1* genes to determine if they were present in other mammalian or other vertebrate genomes.

Results

1.1 Australian marsupial *IGF1* genes in genome databases.

In most mammals, *IGF1/Igf1* is a six-exon, five-intron single-copy gene, in which dual promoters and alternative splicing contribute to many distinct transcripts (Rotwein, 2017a; Rotwein, 2017b). At first glance, *IGF1* gene organization in Australian marsupials was different, as there appeared to be 10 *IGF1* exons in koala, and seven exons in wallaby (unlike most other species, wallaby *IGF1* is labeled as *PTHR11454_SF4*). In addition, exon 2 in wallaby was not readable within its genome assembly in Ensembl, and exons 5 and 7 were predicted to contain 4 and 10 nucleotides, respectively, which is unusually small for eukaryotic exons (Hawkins, 1988) (Fig 1; and see: http://www.ensembl.org/Notamacropus_eugenii/Transcript/Exons?db=core;g=ENSMEUG00000007436;r=GeneScaffold_265:16210-41182;t=ENSMEUT0000007453). Moreover, even though Tasmanian devil *IGF1* appeared to encode six exons and five introns, its gene organization also differed markedly from other mammals. For example, exons 3 and 4 were found to be 48 and 22 nucleotides in length, respectively, and had not been noted previously in other mammalian *IGF1/Igf1* genes, although small exons had been detected in *Igf1* genes in some fish (Rotwein, 2018) (Fig 1; and see: http://www.ensembl.org/Sarcophilus_harrisii/Transcript/Exons?db=core;g=ENSSHAG00000008195;r=GL861602.1:393966-499063;t=ENSSHAT00000009554).

Because of these potentially unusual features, and because Tasmanian devil and wallaby *IGF1* genes were incompletely characterized, as no 3' UTR had been annotated in the

former, and neither 5' nor 3' UTRs had been identified in the latter, a strategy was developed to define these *IGF1* genes more thoroughly. BlastN searches were initially conducted against each genome assembly using parts of other marsupial *IGF1* genes or portions of rat *Igf1* or human *IGF1* genes as queries, and were followed by searches of RNA sequencing libraries found in the SRA NCBI (Table 1) in order to characterize 5' and 3' ends of each marsupial *IGF1* gene, to identify and quantify possible differential promoter use and/or alternative RNA splicing, and to define other components of the gene and locus.

1.2 Koala *IGF1*.

In the recently characterized koala genome (Johnson et al., 2018), *IGF1* was mapped to a single scaffold (see: <https://www.ncbi.nlm.nih.gov/gene/?term=koala+AND+IGF1>). In this project, exons were defined using automated computational gene prediction algorithms, supplemented by some supporting evidence from RNA sequencing alignments (see: https://www.ncbi.nlm.nih.gov/nucore/NW_018343965.1#feature_NW_018343965.1). Several of these exon assignments seemed at odds with data on *IGF1/Igf1* genes from other mammals, particularly with regard to four 'extra' exons identified in koala (exons 4, 7-9; Fig. 1). Moreover, as no experimentally defined koala *IGF1* cDNAs were present in the NCBI nucleotide database, it was not possible to independently confirm these exon assignments from readily available experimental data. Thus, the iterative strategy of identifying exons and mapping other gene features was employed as described above.

The first result of these studies was the identification of 11 koala *IGF1* exons, with the 'extra' exon beyond those described in the genome assembly data being the homolog of the 22-base pair Tasmanian devil 'exon 4', as pictured in Fig. 1. All but one of these 11 exons was expressed in the liver-derived RNA-sequencing library, SRX501262, although at variable levels (Fig. 2A).

To define potential beginnings of koala *IGF1*, adjacent 60 base pair DNA segments found within and 5' to presumptive exon 1 were used to query SRX501262. Results showed that this exon was approximately 344 base pairs in length, and included a ~281-base pair 5' UTR and a 63-base pair coding region (Fig. 2B, left panel). As with other mammalian *IGF1/Igf1* genes (Rotwein, 2017a; Rotwein, 2017b), no potential TATA box, which helps to position RNA polymerase II at the start of transcription (Albright and Tjian, 2000), could be identified 5' to the beginning of the exon. Comparative mapping further showed that koala exon 1 was very similar to rat *Igf1* exon 1 (91% nucleotide identity, supplemental Fig. 1), as was the adjacent presumptive proximal promoter region (88% identity, supplemental Fig. 1). An analogous experimental approach was used to map exon 2, which in other mammals was shown to be the unique leader exon for promoter 2 (Rotwein, 2017a; Rotwein, 2017b). Exon 2 was found to be ~108 nucleotides in length and also lacked a TATA box, but unlike most other mammalian *IGF1/Igf1* genes did not encode an open reading frame (Rotwein, 2017a; Rotwein, 2017b) (Fig. 2B, right panel). Moreover, exon 2 was not well conserved with homologous rat *Igf1* exon (32% identity, supplemental Fig. 1).

In other mammals, a short version of *IGF1/Igf1* exon 5 undergoes alternative RNA splicing into exon 6, maintaining an open reading frame in the resultant transcript, and a longer un-spliced version contains its own 3' UTR and polyadenylation site (Rotwein, 2017a; Rotwein,

2017b). To determine if koala *IGF1* has an analogous organization, expression of the homolog of exon 5 (exon 7, Fig. 2C, left panel) was characterized using a strategy that was similar to the mapping of the 5' ends of exons 1 and 2. This led to the identification of spliced mRNAs, and to un-spliced transcripts with two different 3' UTRs, with total exon lengths of 217 and 300 base pairs, respectively (Table 3). Both 3' UTRs included near their 3' ends a presumptive poly A recognition sequence (Proudfoot, 2011) and tandem poly A addition sites (Fig. 2C, left panel). Except for the coding region at the proximal 5' end of the exon (78 % identity), DNA sequence similarity with the corresponding portion of rat *Igf1* exon was modest (41%).

A similar approach was used to map the 3' end of the analog of *IGF1/Igf1* exon 6 (exon 11 in koala), which in other mammalian species contains a 3' UTR of over several thousand base pairs in length (Rotwein, 2017a; Rotwein, 2017b). A single 3' end was mapped for this exon, which included a poly A recognition sequence of 'AATAAA', and a pair of adjacent poly A addition sites (Fig. 2C, right panel), and the exon was found to have a total length of 7438 base pairs (Table 3). Again, except for the coding region (83% identity over 63 base pairs), similarity with rat *Igf1* exon 6 was low (31%).

In other mammals, the unique leader exon directed by each promoter (exon 1 and 2, respectively) splices into common exon 3 (Rotwein, 2017a; Rotwein, 2017b). Thus, *IGF1/Igf1* transcripts directed by each promoter can be analyzed individually, and can be quantified and calibrated against each other (Rotwein, 2019). Potential differential promoter use in koala *IGF1* was assessed by constructing two 60-base pair DNA probes that spanned the junctions of exons 1 and 3, or exons 2 and 3, respectively (Table 2), and using them to interrogate SRX501262. Both classes of transcripts were expressed, with twice the number of sequences containing exons 1 and 3 (and directed by promoter 1) as exons 2 and 3 (from promoter 2, Fig. 2D). Thus, based on these results, two *IGF1* promoters appear to be active in koala liver, although promoter 2 appeared to be more active in koala than in other mammals (Rotwein, 2017a; Rotwein, 2017b).

A series of investigations was performed next to assess alternative splicing in koala *IGF1*. Combinations of distinct DNA probes were generated that spanned the junctions of many different potential exon pairs (Table 2), and were used to interrogate SRX501262. Many types of transcripts were expressed, although the number of sequence reads varied dramatically among the different groups (Fig. 3A-D). The collective results of these studies were that the 11-exon koala *IGF1* gene (Fig. 3E) encoded ten different major classes of koala *IGF1* mRNAs. This includes six that had been mapped in other mammalian species, and that contain the homologs of 'standard' mammalian *IGF1/Igf1* exons 1-6 (types A-C, Fig. 3F), and four transcripts that seem to be unique to koala (types D and E, Fig. 3F).

1.3 Tasmanian devil *IGF1*.

As noted above, although the Tasmanian devil *IGF1* gene described in Ensembl was stated to encode six exons and five introns, no 3' UTR had been defined for the last exon, and two exons were substantially smaller than detected in any other mammalian *IGF1/Igf1* genes (Rotwein, 2017a; Rotwein, 2017b), 48 and 22 nucleotide pairs, respectively (Fig. 1). To characterize the Tasmanian devil *IGF1* gene, the same experimental strategy was employed

as was used to define koala *IGF1* (Tables 1 and 2, Figs. 4 and 5). Transcripts that included 11 different Tasmanian devil *IGF1* exons were expressed in the liver RNA sequencing library, ERX1403324 (Table 1), with three being homologous to koala exons 4, 8, and 10 (these were Tasmanian devil exons 4, 7, and 10, respectively, Fig. 4A). In addition, mapping studies showed that Tasmanian devil exons 1 and 2 were 464 and 168 base pairs in length, respectively (Fig. 4B, Table 3). Comparative analysis demonstrated that Tasmanian devil exon 1 was very similar to rat *Igf1* exon 1, while exon 2 was not as well conserved (86% and 49% nucleotide identity with rat exons 1 and 2, respectively, supplemental Fig. 1). At the 3' end of Tasmanian devil *IGF1*, the un-spliced version of exon 8 (the homolog of exon 5 in other mammals, and exon 7 in koala), and exon 11 (similar to exon 6 in other mammals, and exon 11 in koala), were shown to encode 217 and 7508 nucleotide pairs, respectively (Fig. 4C, Table 3). Except for their coding regions (76% and 80% identity, respectively), these exons were dissimilar to the corresponding parts of rat *Igf1*, but were conserved with koala *IGF1* (Table 4).

Potential differential promoter use was assessed in Tasmanian devil by constructing two 60-base pair DNA probes that spanned the junctions of exons 1 and 3, or exons 2 and 3, respectively (Table 2), and using them to interrogate ERX1403324. Both classes of transcripts were expressed, with promoter 2-directed mRNAs being ~5-fold more abundant than those controlled by promoter 1 (Fig. 4D). Thus, as in koala, two *IGF1* promoters appear to be active in the liver in Tasmanian devil. The reasons why Tasmanian devil promoter 2 is more active than in other mammals in the liver (Rotwein, 2017a; Rotwein, 2017b) are not apparent, since both promoters regulate transcripts are identical except for their 5' ends, and encode the same IGF1 protein precursors.

Alternative splicing involving the 11 exons in Tasmanian devil *IGF1* was analyzed next by constructing a series of DNA probes spanning the junctions of different exon pairs (Table 2), and using them to interrogate ERX1403324 (Fig. 5A-D). Collectively, the results of all of these analyses define an 11-exon Tasmanian devil *IGF1* gene that includes exons similar to novel koala *IGF1* exons 4, 5 and 10 (Tasmanian devil exons 4, 6 and 10, respectively), plus a unique Tasmanian devil exon 5 of 48 nucleotide pairs (Fig. 5E). However, in contrast to what was observed in koala, only six predominant transcripts were detected in liver RNA in Tasmanian devil, and all were highly similar to those observed in other mammals (Fig. 5F).

1.4 Wallaby *IGF1*.

As stated above, the wallaby *IGF1* gene described in Ensembl was incomplete, as it lacked any 5' or 3' UTRs. It also included a series of small exons of 75, 4, 67, and 10 base pairs, respectively (exons 4 – 7), separated by 'introns' of 2, 11, and 1 nucleotide pairs that were not present in other mammalian *IGF1/Igf1* genes (Rotwein, 2017a; Rotwein, 2017b) (Fig. 1). In addition, presumably because of poor genome quality, no DNA sequence was available for apparent wallaby *IGF1* exon 2, which would be the analog of exon 3 in other species. Thus, the same approaches were employed to more thoroughly characterize wallaby *IGF1* as were used to map other Australian marsupial *IGF1* genes, along with additional analyses designed to identify the 'missing' exon 2 (Figs. 6 and 7). As with Tasmanian devil, homologs of koala *IGF1* exons 4, 8 and 10 were found in the wallaby genome, and a total of

8 of the 9 wallaby *IGF1* exons were expressed in transcripts in the liver RNA sequencing library, DRX012248 (Table 1, Fig. 6A). The 5' ends of wallaby *IGF1* exon 1 and exon 2 also were characterized using DRX012248, and were determined to be 427 and 228 base pairs in length, respectively (Fig. 6B, Table 3), although mapping of the former was potentially incomplete as genomic DNA sequences were absent from the wallaby genome at this location in the further 5' direction (see: [http://www.ensembl.org/Notamacropus_eugenii/Transcript/Exons?](http://www.ensembl.org/Notamacropus_eugenii/Transcript/Exons?db=core;g=ENSMEUG00000007436;r=GeneScaffold_265:16210-41182;t=ENSMEUT00000007453)

[db=core;g=ENSMEUG00000007436;r=GeneScaffold_265:16210-41182;t=ENSMEUT00000007453](http://www.ensembl.org/Notamacropus_eugenii/Transcript/Exons?db=core;g=ENSMEUG00000007436;r=GeneScaffold_265:16210-41182;t=ENSMEUT00000007453)). As with the other Australian marsupials studied here, wallaby *IGF1* exon 1 also was conserved with rat *Igf1* exon 1 (89% nucleotide identity, supplemental Fig. 1), but exon 2 was not (48% identity over 158 base pairs, supplemental Fig. 1). The 'missing' DNA sequence from wallaby *IGF1* 'exon 2' noted above was defined by screening DRX012248 with koala DNA, and was identical in length (157 bp) and highly similar to exon 3 from the other two marsupials (Fig. 6C, Tables 3 and 4) and from non-marsupial mammals (Rotwein, 2017a; Rotwein, 2017b). At the 3' end of wallaby *IGF1*, un-spliced exon 6 (homolog of exon 5 in other mammals), and exon 9 (corresponding to exon 6 in other mammals), were shown to encode 495 and 7468 nucleotide pairs, respectively (Fig. 6D, Tables 3 and 4). Moreover, wallaby *IGF1* exon 6 was found to include all of the DNA that had been defined in Ensembl as individual exons 4 – 7 and the interspersed introns (see above and Fig. 1). However, mapping of 3' end of this exon must be considered provisional, as no presumptive poly A recognition sequence could be identified (Fig. 6D, left panel). Additionally, for exon 9, koala DNA sequences were used as probes, since the wallaby genome in this region consists of poor quality DNA (see: [http://www.ensembl.org/Notamacropus_eugenii/Transcript/Exons?](http://www.ensembl.org/Notamacropus_eugenii/Transcript/Exons?db=core;g=ENSMEUG00000007436;r=GeneScaffold_265:16210-41182;t=ENSMEUT00000007453)

[db=core;g=ENSMEUG00000007436;r=GeneScaffold_265:16210-41182;t=ENSMEUT00000007453](http://www.ensembl.org/Notamacropus_eugenii/Transcript/Exons?db=core;g=ENSMEUG00000007436;r=GeneScaffold_265:16210-41182;t=ENSMEUT00000007453)). Both exons 6 and 9 were not conserved with the corresponding rat *Igf1* exons except for their 5' coding regions (49% identity with rat exon 5 and 80% with rat exon 6), but were similar to the koala homologs, exons 7 and 11, respectively (Table 4).

Differential *IGF1* promoter use was analyzed in wallaby by constructing two 60-base pair DNA probes that spanned the junctions of exons 1 and 3, or exons 2 and 3, respectively (Table 2), and using them to examine DRX012248. Both classes of transcripts were expressed, with promoter 1-directed mRNAs being twice as abundant than those controlled by promoter 2 (Fig. 6E). Thus, as in koala and Tasmanian devil, both *IGF1* promoters appear to be active in the liver in wallaby. Alternative splicing involving different wallaby *IGF1* exons was analyzed by constructing DNA probes spanning the junctions of different exon pairs (Table 2), and using them to examine DRX012248 (Fig. 7A-C). Several types of transcripts were expressed, although the number of sequence reads varied widely among different groups (Fig. 7A-C). Taken together, these results define a 9-exon wallaby *IGF1* gene that contained the homologs of koala exons 4, 8, and 10 (exons 4, 7, and 8 in wallaby, respectively, Fig. 7D), but did not encode a homolog of koala exon 9 or unique Tasmanian devil exon 5. Moreover, only six major classes of koala *IGF1* mRNAs were highly expressed in liver RNA, and these matched the types of transcripts that had been mapped in other mammalian species (Rotwein, 2017a; Rotwein, 2017b) (Fig. 7E).

1.5 *IGF1* gene expression in marsupial organs.

Several RNA-sequencing libraries were available in the SRA NCBI data resource for analysis of tissue-specific *IGF1* gene expression in koala and wallaby, but not for Tasmanian devil (Table 1). *IGF1* mRNA was found in all of the organs examined at variable levels (Fig. 8A, B). Except for koala kidney, in which mRNAs controlled by both promoters were at similar mRNA levels, transcripts directed by promoter 1 were many times more abundant than mRNAs activated by promoter 2. Transcript levels for a constitutively expressed control gene, *MRPS17*, also varied over more than a 10-fold range, particularly in koala (Fig. 8A, B).

1.6 Conservation of *IGF1* protein sequences among marsupials.

In many other mammals, the mature IGF1 protein of 70 amino acids is contained within six different protein precursors that differ at their NH₂- and COOH-termini (Rotwein, 2017a; Rotwein, 2017b). The mature IGF1 molecule itself is divided into four domains, termed B, C, A, and D, with the first three being similar to the B and A chains of insulin, and the C-peptide of pro-insulin (Blundell and Humbel, 1980), and all four regions being analogous to the domains of IGF2 (Blundell and Humbel, 1980). A 70-amino acid IGF1 protein was found in Australian marsupials and was identical in the three species studied here, and was 89% and 91% identical to rat (8 differences) and human IGF1 (6 differences), respectively (Fig. 9, Table 5).

In other mammals, there are two different IGF1 signal peptides. A common COOH-terminal 27-residue segment is encoded by exon 3, and unique NH₂-terminal fragments by either exon 1 (21 amino acids) or exon 2 (5 amino acids) (Rotwein, 2017a). In koala, Tasmanian devil, and wallaby, a 48-residue signal peptide was identified that derived from exons 1 and 3, and was highly conserved, as only a single difference was found in wallaby (Ile³⁰ to Leu, Fig. 9B and C), although there were 13 differences with the rat and 9 with the human signal sequence (Table 5). It is not known how these alterations in marsupials might affect any aspects of signal peptide function, as there are no experimental studies on IGF1 biosynthesis in these species. The *IGF1* gene in the marsupials examined here lacked a methionine codon or open reading frame within exon 2, which had been observed previously in a few other mammals (Rotwein, 2017a). However, the same open reading frame would be present in *IGF1* transcripts containing exon 2 as for those with exon 1 because of another in-frame methionine codon near the beginning of the signal peptide segment encoded in exon 3 (Fig. 9C).

The common E region (16 amino acids) and E_A peptide (19 residues) at the COOH-terminus of the IGF1 protein precursor were nearly identical to one another in the three Australian marsupials, as only a single substitution was found in the latter domain in Tasmanian devil (Ser¹¹ to Asn, Fig 9D, right panel). The 16 amino acids in the E_A domain also were conserved with the corresponding part of the human IGF1A precursor (94% identity, Table 5). The predicted E_B and E_C peptides were less similar among Australian marsupials than other parts of IGF1 precursor proteins (Fig. 9E and F), and were less conserved with these parts of human IGF1 (Table 5), as has been observed previously with other mammals (Rotwein, 2017a; Rotwein, 2017b). The marsupial E_B peptide also varied in length from 77

to 95 amino acids, and E_C from 42 to 56 residues (Fig. 9E, Table 5). Both segments were enriched in basic amino acids (Fig. 9E and F), as has been found in other mammals too (Rotwein, 2017a). Of note, it has been postulated that the diversification observed in this part of IGF1 might be secondary to insertion of a transposon from the interspersed repetitive-b family into the genome within this segment of *IGF1* (the equivalent of exon 5) in a shared mammalian ancestor (Annibalini et al., 2016).

In koala *IGF1*, novel classes of mRNAs (Fig. 3F) contain open reading frames that diverge from the E_A, E_B, or E_C peptides described above, and encode proteins termed here as E_D, with 105 predicted amino acids, and E_E, with 34 residues (Fig. 9G and H). At present there are no biological data that support the existence of either of these IGF1 precursors. Moreover, as transcripts encoding open reading frames similar to those in koala *IGF1* were not found in either Tasmanian devil or wallaby mRNAs (Figs. 5F and 7E), the existence of these novel classes of IGF1 progenitor proteins seems at present to be conjectural.

1.7 Insights into *IGF1* gene regulation in marsupials.

The most important physiological activator of *IGF1/Igf1* in other mammals is GH (Rosenfeld and Hwa, 2009). GH stimulates *Igf1* gene expression in rats through a series of transcriptional enhancers found throughout the *Igf1* locus that contain inducible binding elements for the Stat5b transcription factor (Chia et al., 2010; Varco-Merth et al., 2012). In addition, these elements have been shown to bind both Stat5b and transcriptional co-factors, including p300, RNA polymerase II, and the mediator complex *in vivo* in rat liver, and to undergo reversible histone modifications (Chia et al., 2010). In cell culture experiments, the same elements were able to physically interact with *Igf1* promoters in a GH-dependent way (Alzhanov et al., 2015), and to respond positively to GH treatment in promoter-reporter gene expression experiments (Varco-Merth et al., 2012). Five of these segments were conserved in analogous regions of the human *IGF1* locus (Mukherjee et al., 2016), and also have been found to function as GH-activated transcriptional regulators in cell culture (Mukherjee et al., 2016).

Only two of these elements could be mapped to Australian marsupial *IGF1* gene loci, the analogs of rat (R) elements R8-9 and R60-61 (Table 6, Fig. 10). Comparison of the DNA sequences of these elements revealed a high degree of similarity among koala, Tasmanian devil, and wallaby, ranging from 91% to 96% DNA sequence identity, including conservation of the three identified 9-nucleotide pair canonical Stat5b binding sites, and identical DNA spacing between the paired elements (Table 6). There was less similarity with the corresponding rat *Igf1* elements (75% and 78%, respectively, for R8-9 and R60-61) although the DNA sequences of the two Stat5b binding sites were identical in the homolog of R8-9 and the distance between the paired sites was very similar between rats and marsupials (226 base pairs in the former vs. 224 in the latter, Table 6) (Rotwein, 2017a). In contrast, in Australian marsupials only a single Stat5b binding site could be detected in the equivalent of R60-61 (Table 6).

1.8 Novel features of marsupial *IGF1* genes are found in other mammals.

The genomes of nine different mammals and ten non-mammalian vertebrates were interrogated for homologs of the novel *IGF1* exons identified here (four in koala and two in Tasmanian devil (Fig. 10)). No similar DNA sequences were detected in human, mouse, platypus, or rat for any of these six exons, or in anole lizard, chicken, Chinese soft shell turtle, coelacanth, frog, lamprey, spotted gar, tilapia, turkey, or zebrafish. However, an analog of koala *IGF1* exon 4 was characterized in opossum (90% identity over 268 of 332 base pairs), which as in koala was located near the 3' end of the large intron between traditional *IGF1* exons 3 and 4 (supplemental Fig. 2). Gene expression of mRNAs containing this possible exon was minimal in opossum liver (1 sequence read vs. 730 for traditional *IGF1* exon 3; data not shown). In dog, elephant, megabat, opossum, and pig genomes, DNA sequences of 36 to 62 base pairs were 92% to 95% identical with a segment of 125 base pair koala *IGF1* exon 8. These sequences also were found within the *IGF1* locus of each of these species, and mapped between 5 and 7.5 kb 3' to traditional *IGF1* exon 5 (data not shown).

Discussion

IGF1, a conserved 70-amino acid single-chain secreted protein, is an essential regulator of normal somatic growth during childhood in humans, and in juveniles of other mammals, in which it functions as a mediator of GH actions (Le Roith et al., 2001; LeRoith, 2008; Lupu et al., 2001; Powell-Braxton et al., 1993; Woods et al., 1996) and as an integrator for other inputs that affect overall health and thus growth potential (Baron et al., 2015). IGF1 also plays a role in metabolic regulation and in tissue repair throughout life (Berryman et al., 2008; Gallagher and LeRoith, 2011; Gems and Partridge, 2013; Pollak, 2012). IGF1 is produced after post-translational processing from precursor proteins. In at least 20 mammalian species from 13 different orders, these precursors of mature IGF1 are derived from the translation of six major classes of mRNAs that result from transcription from two distinct promoters, alternative RNA splicing, and differential RNA polyadenylation (Rotwein, 2017a).

The focus of this report is on the molecular genetics of *IGF1* in Australian marsupials, species whose ancestors diverged from other mammalian progenitors nearly 150 million years ago (Bininda-Emonds et al., 2007; Mitchell et al., 2014). Australian marsupial *IGF1* genes differ substantially from those in most other mammals. In the latter species, *IGF1/Igf1* is a six-exon gene (Rotwein, 2017a). In contrast, *IGF1* has either nine or 11 exons in the three Australian marsupials studied here, and encodes individual exons that appear either to be confined to just these three species, to be found uniquely in koala or in Tasmanian devil genomes (Figs. 2–5, 10), or to be shared with a few other marsupials, such as opossum (Supplemental Fig. 2). Despite these substantive and distinctive differences, which include the presence of additional unique *IGF1* transcripts in koala that predict novel IGF1 precursor proteins (Figs. 2, 9F, G), there are a number of features in Australian marsupial *IGF1* genes that are shared with other mammals. As outlined above, in the latter, each of two promoters directs distinct leader exons that define major groups of transcripts. These mRNAs in turn are translated into six different IGF1 protein precursors that vary in their NH₂- and COOH-

termini but contain the identical mature IGF1 (Rotwein, 2017a). In koala, Tasmanian devil, and wallaby, similar groups of mRNAs are produced from two promoters, that along with their leader exons variably share features with a representative mammalian *IGF1/Igf1* gene, rat *Igf1* (supplemental Fig.1). Six major Australian marsupial *IGF1* mRNAs also encode six different types of IGF1 precursors with distinct NH₂- and COOH-terminal segments that resemble the IGF1 progenitors in other mammals (Figs. 3, 5, 7, 9). Moreover, the 70-amino acid IGF1 molecule, which is identical in these three Australian marsupials, is 91% identical to human IGF1, with three of six amino acid differences being located in the more variable C-domain, which acts primarily as a linker between the B and A segments of the protein (Werner et al., 2008) (A³⁸, P³⁹, Q⁴⁰ in human to L, H, H in marsupials), and one occurring in the D-region (L⁶⁴ to I; Fig 9A), which also varies among other mammalian species (Rotwein, 2017a).

Although very little is known about importance of GH actions in Australian marsupial physiology, with published data being limited to a few observational studies in wallaby (Hetz et al., 2015; Hetz et al., 2016; Menzies et al., 2012), remarkably two of the putative GH-stimulated Stat5b binding enhancers found in *IGF1/Igf1* loci in other mammals (Rotwein, 2017a; Rotwein, 2017b) are present in koala, Tasmanian devil and wallaby. These elements are located in similar genomic positions, and share DNA sequence similarity with rat *Igf1* Stat5b binding domains, R8-9 and R60-61 (Fig. 10, Table 6). Moreover, these data expand what had been recognized previously in Tasmanian devil and wallaby, and in opossum (Supplemental Fig. 2), in which only a single element had been detected in prior analyses (Rotwein, 2017a). Of note, however, the opossum equivalent of R60-61 lacks 9-nucleotide canonical Stat5b binding sites in an otherwise conserved DNA segment with koala (Supplemental Fig. 2). Perhaps these two GH-regulated Stat5b binding enhancers found in Australian marsupials comprise part of a core signature in mammalian *IGF1/Igf1* loci, and these common elements may represent the most conserved and thus most essential components of a GH-activated signaling pathway to stimulate *IGF1/Igf1* gene transcription. Clearly additional studies will be needed to explore this hypothesis. Ideally such investigations would involve direct animal-based experiments involving transcriptional assays in conjunction with assessment of Stat5b binding to chromatin, as has been performed in rats (Chia et al., 2010). However, it is now possible to test GH-activated transcription in chimeric cell lines containing *IGF1/Igf1* loci (Alzhanov et al., 2015; Mukherjee et al., 2016), and thus the functional properties of these potential marsupial regulatory elements also could be examined in such cultured cells as an initial exploration of this idea.

A key observation that was an impetus for this study is that data in GenBank on koala, and especially in Ensembl on the Tasmanian devil and wallaby *IGF1* genes are both inaccurate and incomplete, most likely because exon assignments had not been determined experimentally, but rather had been defined through deployment of gene prediction algorithms. It is now clear that much of these data are wrong, for example the four small 'exons' of respectively 75, 4, 67, and 10 nucleotide pairs in wallaby *IGF1* ('exons' 4 – 7) that were apparently separated by miniature 'introns' of 2, 11, and 1 base pairs (Fig. 1) were incorrectly assigned. In fact, this entire region of the wallaby genome has been found to be part of *IGF1* exon 6 (Fig. 6D). More critically, the results described in this paper provide a

comprehensive demonstration of how public genomic and RNA-sequencing repositories can be used to facilitate in-depth analysis of gene structure and/or regulation in organisms that are otherwise intractable to direct experimental manipulation, not only correcting inaccurate data, but also raising new hypotheses with the potential to enhance our understanding of gene evolution during speciation.

Finally, the complicated roles of IGF1 in both physiology and disease are potentially represented in its complex gene structure and patterns of gene and protein expression. Conservation of aspects of the *IGF1* gene and the IGF1 peptide between Australian marsupials and other mammals argues that some transcriptional and other regulatory pathways, as well as IGF1 functions, have been maintained since the onset of mammalian speciation (Venditti et al., 2011; Venditti and Pagel, 2010), but that others, including novel exons and mRNA species, may be more closely aligned to uniquely marsupial physiology. In theory these ideas are testable in experimental systems, and could lead to new insights into the comparative biology of IGF1 regulation or action.

Conclusions.

These studies show that Australian marsupial *IGF1* genes have diverged from those of other mammals in terms of their organization and patterns of expression while still retaining features that were present in common ancestors. The results also illustrate how comprehensive interrogation of large-scale public genomic and RNA-sequencing resources can be an effective strategy for fully characterizing individual genes and assessing their activity in otherwise experimentally intractable organisms.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

I thank my colleagues for comments on the manuscript.

Funding

The ideas that led to this paper were supported initially by National Institutes of Health research grant, R01 DK069703 (to PR).

References

- Adamo ML, 1995 Regulation of insulin-like growth factor I gene expression. *Diabetes Rev* 3, 2–27.
- Ahluwalia A, Clodfelter KH, Waxman DJ, 2004 Sexual dimorphism of rat liver gene expression: regulatory role of growth hormone revealed by deoxyribonucleic Acid microarray analysis. *Mol Endocrinol* 18, 747–760. [PubMed: 14684848]
- Albright SR, Tjian R, 2000 TAFs revisited: more data reveal new twists and confirm old ideas. *Gene* 242, 1–13. [PubMed: 10721692]
- Alzhanov D, Mukherjee A, Rotwein P, 2015 Identifying growth hormone-regulated enhancers in the *Igf1* locus. *Physiol Genomics* 47, 559–568. [PubMed: 26330488]
- 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, 2016 MIR retroposon exonization promotes evolutionary

variability and generates species-specific expression of IGF-1 splice variants. *Biochim Biophys Acta* 1859, 757–768. [PubMed: 27048986]

Baron J, Savendahl L, De Luca F, Dauber A, Phillip M, Wit JM, Nilsson O, 2015 Short and tall stature: a new paradigm emerges. *Nat Rev Endocrinol* 11, 735–746. [PubMed: 26437621]

Berryman DE, Christiansen JS, Johannsson G, Thorner MO, Kopchick JJ, 2008 Role of the GH/IGF-1 axis in lifespan and healthspan: lessons from animal models. *Growth Horm IGF Res* 18, 455–471. [PubMed: 18710818]

Bichell DP, Kikuchi K, Rotwein P, 1992 Growth hormone rapidly activates insulin-like growth factor I gene transcription in vivo. *Mol Endocrinol* 6, 1899–1908. [PubMed: 1480177]

Bininda-Emonds OR, Cardillo M, Jones KE, MacPhee RD, Beck RM, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A, 2007 The delayed rise of present-day mammals. *Nature* 446, 507–512. [PubMed: 17392779]

Blundell TL, Humbel RE, 1980 Hormone families: pancreatic hormones and homologous growth factors. *Nature* 287, 781–787. [PubMed: 6107857]

Chia DJ, 2014 Minireview: mechanisms of growth hormone-mediated gene regulation. *Mol Endocrinol* 28, 1012–1025. [PubMed: 24825400]

Chia DJ, Rotwein P, 2010 Defining the epigenetic actions of growth hormone: acute chromatin changes accompany GH-activated gene transcription. *Mol Endocrinol* 24, 2038–2049. [PubMed: 20702579]

Chia DJ, Varco-Merth B, Rotwein P, 2010 Dispersed chromosomal Stat5b-binding elements mediate growth hormone-activated insulin-like growth factor-I gene transcription. *J Biol Chem* 285, 17636–17647. [PubMed: 20378540]

Das R, Dobens LL, 2015 Conservation of gene and tissue networks regulating insulin signalling in flies and vertebrates. *Biochem Soc Trans* 43, 1057–1062. [PubMed: 26517923]

Daughaday WH, Kapadia M, Yanow CE, Fabrick K, Mariz IK, 1985 Insulin-like growth factors I and II of nonmammalian sera. *Gen Comp Endocrinol* 59, 316–325. [PubMed: 3926605]

Eleswarapu S, Gu Z, Jiang H, 2008 Growth hormone regulation of insulin-like growth factor-I gene expression may be mediated by multiple distal signal transducer and activator of transcription 5 binding sites. *Endocrinology* 149, 2230–2240. [PubMed: 18276757]

Gallagher EJ, LeRoith D, 2011 Minireview: IGF, Insulin, and Cancer. *Endocrinology* 152, 2546–2551. [PubMed: 21540285]

Gems D, Partridge L, 2013 Genetics of longevity in model organisms: debates and paradigm shifts. *Annu Rev Physiol* 75, 621–644. [PubMed: 23190075]

Hall LJ, Kajimoto Y, Bichell D, Kim SW, James PL, Counts D, Nixon LJ, Tobin G, Rotwein P, 1992 Functional analysis of the rat insulin-like growth factor I gene and identification of an IGF-I gene promoter. *DNA Cell Biol* 11, 301–313. [PubMed: 1605853]

Hawkins JD, 1988 A survey on intron and exon lengths. *Nucleic Acids Res* 16, 9893–9908. [PubMed: 3057449]

Hetz JA, Menzies BR, Shaw G, Rao A, Clarke IJ, Renfree MB, 2015 Growth axis maturation is linked to nutrition, growth and developmental rate. *Mol Cell Endocrinol* 411, 38–48. [PubMed: 25896544]

Hetz JA, Menzies BR, Shaw G, Stefanidis A, Cowley MA, Renfree MB, 2016 Effects of nutritional manipulation on body composition in the developing marsupial, *Macropus eugenii*. *Mol Cell Endocrinol* 428, 148–160. [PubMed: 27032712]

Hoyt EC, Van Wyk JJ, Lund PK, 1988 Tissue and development specific regulation of a complex family of rat insulin-like growth factor I messenger ribonucleic acids. *Mol Endocrinol* 2, 1077–1086. [PubMed: 3221875]

Johnson RN, O'Meally D, Chen Z, Etherington GJ, Ho SYW, Nash WJ, Grueber E, Cheng Y, Whittington CM, Dennison S, Peel E, Haerty W, O'Neill RJ, Colgan D, Russell TL, Alquezar-Planas DE, Attenbrow V, Bragg JG, Brandies PA, Chong AY, Deakin JE, Di Palma F, Duda Z, Eldridge MDB, Ewart KM, Hogg CJ, Frankham GJ, Georges A, Gillett AK, Govendir M, Greenwood AD, Hayakawa T, Helgen KM, Hobbs M, Holleley CE, Heider TN, Jones EA, King A, Madden D, Graves JAM, Morris KM, Neaves LE, Patel HR, Polkinghorne A, Renfree MB, Robin C, Salinas R, Tsangaras K, Waters PD, Waters SA, Wright B, Wilkins MR, Timms P, Belov K,

- 2018 Adaptation and conservation insights from the koala genome. *Nat Genet* 50, 1102–1111. [PubMed: 29967444]
- Kim SW, Lajara R, Rotwein P, 1991 Structure and function of a human insulin-like growth factor-I gene promoter. *Mol Endocrinol* 5, 1964–1972. [PubMed: 1791841]
- Le Roith D, Bondy C, Yakar S, Liu JL, Butler A, 2001 The somatomedin hypothesis: 2001. *Endocr Rev* 22, 53–74. [PubMed: 11159816]
- LeRoith D, 2008 Clinical relevance of systemic and local IGF-I: lessons from animal models. *Pediatr Endocrinol Rev* 5 Suppl 2, 739–743.
- Lowe WLJ, Roberts CTJ, Lasky SR, LeRoith D, 1987 Differential expression of alternative 5' untranslated regions in mRNAs encoding rat insulin-like growth factor I. *Proc Natl Acad Sci U S A* 84, 8946–8950. [PubMed: 3480521]
- Lupu F, Terwilliger JD, Lee K, Segre GV, Efstratiadis A, 2001 Roles of growth hormone and insulin-like growth factor 1 in mouse postnatal growth. *Dev Biol* 229, 141–162. [PubMed: 11133160]
- Mathews LS, Norstedt G, Palmiter RD, 1986 Regulation of insulin-like growth factor I gene expression by growth hormone. *Proc Natl Acad Sci U S A* 83, 9343–9347. [PubMed: 3467309]
- Menzies BR, Shaw G, Fletcher TP, Pask AJ, Renfree MB, 2012 Maturation of the growth axis in marsupials occurs gradually during post-natal life and over an equivalent developmental stage relative to eutherian species. *Mol Cell Endocrinol* 349, 189–194. [PubMed: 22056413]
- Mitchell KJ, Pratt RC, Watson LN, Gibb GC, Llamas B, Kasper M, Edson J, Hopwood B, Male D, Armstrong KN, Meyer M, Hofreiter M, Austin J, Donnellan SC, Lee MS, Phillips MJ, Cooper A, 2014 Molecular phylogeny, biogeography, and habitat preference evolution of marsupials. *Mol Biol Evol* 31, 2322–2330. [PubMed: 24881050]
- Mukherjee A, Alzhanov D, Rotwein P, 2016 Defining human insulin-like growth factor I gene regulation. *Am J Physiol Endocrinol Metab* 311, E519–29. [PubMed: 27406741]
- Pollak M, 2012 The insulin and insulin-like growth factor receptor family in neoplasia: an update. *Nat Rev Cancer* 12, 159–169. [PubMed: 22337149]
- Powell-Braxton L, Hollingshead P, Warburton C, Dowd M, Pitts-Meek S, Dalton D, Gillett N, Stewart TA, 1993 IGF-I is required for normal embryonic growth in mice. *Genes Dev* 7, 2609–2617. [PubMed: 8276243]
- Proudfoot NJ, 2011 Ending the message: poly(A) signals then and now. *Genes Dev* 25, 1770–1782. [PubMed: 21896654]
- Rosenfeld RG, Hwa V, 2009 The growth hormone cascade and its role in mammalian growth. *Horm Res* 71 Suppl 2, 36–40. [PubMed: 19407495]
- Rotwein P, 1986 Two insulin-like growth factor I messenger RNAs are expressed in human liver. *Proc Natl Acad Sci U S A* 83, 77–81. [PubMed: 3455760]
- Rotwein P, 1999 Molecular biology of IGF-I and IGF-II, in: Rosenfeld R, Roberts CJ (Eds.), *Humana Press*, Totowa, NJ, pp. 19–35.
- Rotwein P, 2017a Diversification of the insulin-like growth factor 1 gene in mammals. *PLoS One* 12, e0189642. [PubMed: 29240807]
- Rotwein P, 2017b Variation in the insulin-like growth factor 1 gene in primates. *Endocrinology* 158, 804–814. [PubMed: 28324014]
- Rotwein P, 2018 Insulin-like growth factor 1 gene variation in vertebrates. *Endocrinology* 159, 2288–2305. [PubMed: 29697760]
- Rotwein P, 2019 Quantifying promoter-specific Insulin-like Growth Factor 1 gene expression by interrogating public databases. *Physiol Rep* 7, e13970. [PubMed: 30604932]
- Schwartz TS, Bronikowski AM, 2016 Evolution and function of the insulin and insulin-like signaling network in ectothermic reptiles: some answers and more questions. *Integr Comp Biol* 56, 171–184. [PubMed: 27252221]
- Shimatsu A, Rotwein P, 1987 Mosaic evolution of the insulin-like growth factors. Organization, sequence, and expression of the rat insulin-like growth factor I gene. *J Biol Chem* 262, 7894–7900. [PubMed: 3034909]

- Sussenbach JS, Steenbergh PH, Jansen E, Holthuisen P, Meinsma D, van Dijk MA, Gloudemans T, 1991 Structural and regulatory aspects of the human genes encoding IGF-I and -II. *Adv Exp Med Biol* 293, 1–14.
- Varco-Merth B, Mirza K, Alzhanov DT, Chia DJ, Rotwein P, 2012 Biochemical characterization of diverse Stat5b-binding enhancers that mediate growth hormone-activated insulin-like growth factor-I gene transcription. *PLoS One* 7, e50278. [PubMed: 23185594]
- Venditti C, Meade A, Pagel M, 2011 Multiple routes to mammalian diversity. *Nature* 479, 393–396. [PubMed: 22012260]
- Venditti C, Pagel M, 2010 Speciation as an active force in promoting genetic evolution. *Trends Ecol Evol* 25, 14–20. [PubMed: 19720426]
- Wallis M, 2009 New insulin-like growth factor (IGF)-precursor sequences from mamalian genomes: the molecular evolution of IGFs and associated peptides in primates. *Growth Horm IGF Res* 19, 12–23 [PubMed: 18571449]
- Waters MJ, 2016 The growth hormone receptor. *Growth Horm IGF Res* 28, 6–10. [PubMed: 26059750]
- Waters MJ, Brooks AJ, 2015 JAK2 activation by growth hormone and other cytokines. *Biochem J* 466, 1–11. [PubMed: 25656053]
- Werner H, Weinstein D, Bentov I, 2008 Similarities and differences between insulin and IGF-I: structures, receptors, and signalling pathways. *Arch Physiol Biochem* 114, 17–22. [PubMed: 18465355]
- Woelfle J, Billiard J, Rotwein P, 2003 Acute control of insulin-like growth factor-1 gene transcription by growth hormone through STAT5B. *J Biol Chem* 278, 22696–22702. [PubMed: 12682066]
- Woelfle J, Rotwein P, 2004 In vivo regulation of growth hormone-stimulated gene transcription by STAT5b. *Am J Physiol Endocrinol Metab* 286, E393–401. [PubMed: 14761873]
- Woods KA, Camacho-Hubner C, Savage MO, Clark AJ, 1996 Intrauterine growth retardation and postnatal growth failure associated with deletion of the insulin-like growth factor I gene. *N Engl J Med* 335, 1363–1367. [PubMed: 8857020]

Highlights

- Australian marsupial *IGF1* genes are more complicated than *IGF1/Igf1* genes from other mammals.
- Australian marsupial *IGF1* genes encode unique exons that predict novel transcripts and IGF1 protein precursors.
- Australian marsupial *IGF1* genes retain the two-promoter structure of other mammalian *IGF1/Igf1* genes.
- Australian marsupial *IGF1* loci contain just two of the multiple growth hormone-activated and Stat5b-binding transcriptional enhancers found in other mammalian species.
- Interrogating large-scale public genomic and RNA-sequencing resources is an effective strategy for gaining a full understanding of genes, gene expression, and gene evolution in different organisms, including those in which direct experimental evaluation is difficult.

Figure 1. Comparison of marsupial *IGF1* genes from genome databases.

Maps of koala, Tasmanian devil, and wallaby *IGF1* genes are illustrated as they are presented in GenBank (koala), and in the Ensembl genome database (Tasmanian devil and wallaby). Exons are numbered and appear as boxes, and introns and flanking DNA are shown as horizontal lines. A scale bar is depicted.

Figure 2. Mapping the 5' and 3' ends of the koala *IGF1* gene and characterizing gene expression. **A.** Determination of koala *IGF1* gene expression in the liver for each exon, using data from the liver RNA-sequencing library, SRX501262, and individual genomic probes from each of the 11 putative exons (Table 2). **B.** Mapping the 5' ends of koala *IGF1* exon 1 (left panel) and exon 2 (right panel) by using gene expression data from SRX501262, and 60-nucleotide genomic segments a-g and a-e respectively as probes. **C.** Characterizing the putative 3' ends of koala *IGF1* exon 7 (left panel) and exon 11 (right panel) by using data from SRX501262 and 60-nucleotide genomic segments a-f and a-h, respectively as probes. Possible

polyadenylation signals are underlined in the DNA sequences below each graph, and the vertical arrows denote potential 3' ends of *IGF1* mRNAs (poly A addition sites). **D.** Comparing use of koala *IGF1* promoters 1 and 2 in liver by assessing the number of sequence reads in SRX501262, using the two 60-nucleotide DNA probes listed in Table 2 that measure transcripts containing exons 1 and 3 (promoter 1) or exons 2 and 3 (promoter 2).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 3. Mapping alternative RNA splicing in the koala *IGF1* gene.

A. Characterizing alternative splicing involving koala *IGF1* exon 3 by analyzing the number of sequence reads in SRX501262 using the two 60-nucleotide DNA probes listed in Table 2 that measure transcripts containing exons 3 and 4 or exons 3 and 6. **B.** Mapping alternative splicing involving koala *IGF1* exon 6 by analyzing the number of sequence reads in SRX501262 with the five 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 6 and 7, 6 and 8, 6 and 9, 6 and 10, and 6 and 11. **C.** Characterizing alternative splicing involving koala *IGF1* exon 7 by analyzing the number of sequence reads

in SRX501262 using the four 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 7 and 8, 7 and 9, 7 and 10, and 7 and 11. **D.** Mapping alternative splicing involving koala *IGF1* exons 8, 9, and 10 by analyzing the number of sequence reads in SRX501262 with the six 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 8 and 9, 8 and 10, 8 and 11, 9 and 10, 9 and 11, and 10 and 11. **E.** Structure of the koala *IGF1* gene based on the analyses in Fig. 2 and above. The higher power view shows promoters (P) 1 and 2, and exons 1, 2, and 3. Labeling is as in Fig. 1, with individual exons numbered. **F.** Schematic of the 10 major classes of koala *IGF1* mRNAs, illustrating their derivation from different combinations of *IGF1* exons. The colored regions represent coding segments (black for signal peptides, blue for mature IGF1, and orange, red, yellow and gray for different E regions), and the white areas noncoding portions of transcripts.

Figure 4. Mapping the 5' and 3' ends of the Tasmanian devil *IGF1* gene and characterizing gene expression.

A. Determination of Tasmanian devil *IGF1* gene expression in the liver for each exon, using data from the liver RNA-sequencing library, ERX1403324, and individual genomic probes from each of the 11 putative exons (Table 2). **B.** Mapping the 5' ends of Tasmanian devil *IGF1* exon 1 (left panel) and exon 2 (right panel) with gene expression data from ERX1403324, and 60-nucleotide genomic segments a-h and a-e respectively as probes. **C.** Characterizing the putative 3' ends of Tasmanian devil *IGF1* exon 8 (left panel) and exon 11

(right panel) using data from ERX1403324, and 60-nucleotide genomic segments a-e and a-g respectively as probes. Possible polyadenylation signals are underlined in the DNA sequences below the graphs, and the vertical arrows denote potential 3' ends of *IGF1* mRNAs (poly A addition sites). **D.** Comparing use of Tasmanian devil *IGF1* promoters 1 and 2 in liver by assessing the number of sequence reads in ERX1403324, using the two 60-nucleotide DNA probes listed in Table 2 that measure transcripts containing exons 1 and 3 (promoter 1) or exons 2 and 3 (promoter 2).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Figure 5. Mapping alternative RNA splicing in the Tasmanian devil *IGF1* gene.

A. Characterizing alternative splicing involving Tasmanian devil *IGF1* exon 3 by analyzing the number of sequence reads in ERX1403324 using the 60-nucleotide DNA probes listed in Table 2 that measure transcripts containing exons 3 and 4, 3 and 5, 3 and 6, and 3 and 7. **B.** Mapping alternative splicing involving Tasmanian devil *IGF1* exon 5 by analyzing the number of sequence reads in ERX1403324 with the six 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 5 and 6, 5 and 7, 5 and 8, 5 and 9, 5 and 10, and 5 and 11. **C.** Characterizing alternative splicing involving Tasmanian devil *IGF1* exon 7 by

analyzing the number of sequence reads in ERX1403324 using the 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 7 and 8, 7 and 9, 7 and 10, and 7 and 11. **D.** Mapping alternative splicing involving Tasmanian devil *IGF1* exon 8 by analyzing the number of sequence reads in ERX1403324 with the three 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 8 and 9, 8 and 10, or 8 and 11. **E.** Structure of the Tasmanian devil *IGF1* gene based on the analyses in Fig. 4 and above. The higher power view shows promoters (P) 1 and 2, and exons 1, 2, and 3. Labeling is as in Fig. 1, with individual exons numbered. **F.** Schematic of the six major classes of Tasmanian devil *IGF1* mRNAs, illustrating their derivation from different combinations of *IGF1* exons. The colored regions represent coding segments (black for signal peptides, blue for mature IGF1, and orange, red and yellow for different E regions), and the white areas noncoding portions of transcripts.

Figure 6. Mapping the 5' and 3' ends of the wallaby *IGF1* gene and characterizing gene expression.

A. Determination of wallaby *IGF1* gene expression in the liver for each exon, using data from the liver RNA-sequencing library, DRX012248, and individual genomic probes from each of the 9 putative exons (Table 2). **B.** Mapping the 5' ends of wallaby *IGF1* exon 1 (left panel) and exon 2 (right panel) with gene expression data from DRX012248, and 60-nucleotide genomic segments a-g and a-e respectively as probes. **C.** Mapping wallaby *IGF1* exon 3 by using the analogous koala exon to screen DRX012248. Illustrated is a schematic

of the exon, with the DNA sequence at its 5' and 3' termini, and the degree of identity with koala *IGF1* exon 3. **D.** Characterizing the putative 3' ends of wallaby *IGF1* exon 6 (left panel) and exon 9 (right panel) using data from DRX012248, and 60-nucleotide genomic segments a-f and a-h respectively as probes. A possible polyadenylation signal is underlined in the DNA sequence below the right graph, and the vertical arrows denote potential 3' ends of *IGF1* mRNAs (poly A addition sites). **E.** Comparing use of wallaby *IGF1* promoters 1 and 2 in liver by assessing the number of sequence reads in DRX012248, using the two 60-nucleotide DNA probes listed in Table 2 that measure transcripts containing exons 1 and 3 (promoter 1) or exons 2 and 3 (promoter 2).

Figure 7. Mapping alternative RNA splicing in the wallaby *IGF1* gene.

A. Mapping alternative splicing involving wallaby *IGF1* exon 5 by analyzing the number of sequence reads in DRX012248 with the four 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 5 and 6, 5 and 7, 5 and 8, and 5 and 9. **B.**

Characterizing alternative splicing involving wallaby *IGF1* exon 6 by analyzing the number of sequence reads in DRX012248 using the three 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 6 and 7, 6 and 8, and 6 and 9. **C.** Mapping alternative splicing involving wallaby *IGF1* exon 7 by analyzing the number of sequence reads in

DRX012248 with the two 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 7 and 8, and 7 and 9. **D.** Structure of the wallaby *IGF1* gene based on the analyses in Fig. 6 and above. The higher power view shows promoters (P) 1 and 2, and exons 1, 2, and 3. Labeling is as in Fig. 1, with individual exons numbered. **E.** Schematic of the six different major classes of wallaby *IGF1* mRNAs, illustrating their derivation from different combinations of *IGF1* exons. The colored regions represent coding segments (black for signal peptides, blue for mature IGF1, and orange, red and yellow for different E regions), and the white areas noncoding portions of transcripts.

Figure 8. Promoter-specific *IGF1* gene expression in different marsupial organs.

A. Upper graph: *IGF1* transcripts regulated by promoters 1 and 2 were quantified in koala adrenal gland, brain, heart, kidney, lung, spleen, and uterus by querying the RNA-sequencing libraries listed in Table 1 with the two 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 1 and 3 (promoter 1) or exons 2 and 3 (promoter 2).

Lower graph: mRNA abundance of a control transcript from the *MRPS17* gene measured in the same organs as the upper graph. **B. Upper graph:** *IGF1* transcripts regulated by promoters 1 and 2 were quantified in wallaby adrenal gland, heart, lung, and spleen by querying the RNA-sequencing libraries listed in Table 1 with the two 60-nucleotide DNA probes in Table 2 that measure transcripts containing exons 1 and 3 (promoter 1) or exons 2 and 3 (promoter 2). **Lower graph:** mRNA abundance of a control transcript from the *MRPS17* gene from the same organs as the upper graph. For **A** and **B**, results are plotted as the number of sequence reads per organ.

Figure 9. Alignments of marsupial IGF1 protein segments.

A. Amino acid sequences of mature 70-residue IGF1 from koala, Tasmanian devil, and wallaby in single letter code. The sequences are identical for all three species. **B.** Amino acid sequences of the 48-residue IGF1 signal peptide. The single difference in the wallaby protein is indicated in red text. **C.** Amino acid sequences of the shorter 25-residue IGF1 signal peptide. The single difference in the wallaby protein is indicated in red text. **D.** Amino acid sequences of the common E peptide (left) and the E_A segment (right). The sequences are identical for all three marsupial species, except for a single substitution in Tasmanian devil, which is marked in red text. **E.** Amino acid sequences of the E_B peptide. Differences are shown in red text. **F.** Amino acid sequences of the E_C peptide. Differences are shown in red text. For **A** – **F**, see Table 5 for amino acid identities among marsupials and with human IGF1. **G.** Predicted amino acid sequence for koala E_D peptide, as determined from data found in gene expression library SRX501262. **H.** Predicted amino acid sequence for koala E_E peptide, as determined from data found in gene expression library SRX501262. For **G** and **H**, although similar sequences could be present for Tasmanian devil and wallaby, the specific transcripts potentially encoding these proteins were not identified in the gene expression libraries listed in Table 1.

Figure 10. Marsupial *IGF1* genes and loci.

Maps of koala, Tasmanian devil, and wallaby *IGF1* are shown. Exons are depicted as boxes, and introns and flanking DNA as horizontal lines. The enlargement above the koala map depicts the two presumptive *IGF1* promoters, P1 and P2, and exons 1 – 3. Exons 1-4 are found in all species. The other exons are color-coded to illustrate homologous relationships among different Australian marsupials. Exons in shades of gray in koala (exon 9) and Tasmanian devil (exon 5) are unique to each species. Dark red circles represent locations of sites with DNA sequence similarity to elements shown to bind Stat5b in a GH-inducible way in rat *Igf1* and human *IGF1* loci (Chia et al., 2010) (see Table 6). The percentage of nucleotide identity with different parts of koala *IGF1* is indicated for each exon and locus element (black for exons, dark red for putative Stat5b binding elements). Other abbreviations are as follows: K8-9, K60-61, nomenclature for potential koala Stat5b sites. A similar nomenclature has been adapted for Tasmanian devil (T) and wallaby (W). The dotted lines represent discontinuities in the DNA sequence in the wallaby genome assembly.

Table 1:

RNA-sequencing libraries screened for gene expression

Species	Organ	Experiment	Platform	Layout	Bases sequenced ($\times 10^9$)
Koala	Liver	SRX501262	Illumina	paired	22.9
	Adrenal gland	SRX501344	Illumina	paired	10.9
	Brain	SRX501343	Illumina	paired	12.3
	Heart	SRX501342	Illumina	paired	24.5
	Kidney	SRX501314	Illumina	paired	10.7
	Lung	SRX501320	Illumina	paired	11.7
	Spleen	SRX500280	Illumina	paired	11.8
	Uterus	SRX501302	Illumina	paired	10.3
Tasmanian devil	Liver	ERX1403324	Illumina	paired	22.8
Wallaby	Liver	DRX012248	Illumina	paired	1.7
	Adrenal gland	DRX012252	Illumina	paired	1.8
	Heart	DRX012250	Illumina	paired	2.0
	Lung	DRX012230	Illumina	paired	2.0
	Spleen	DRX012251	Illumina	paired	1.7

Table 2:

Probes for screening RNA-sequencing libraries

Species	Gene	Probe
Koala	<i>IGF1 exon 1</i>	GACAAATTCAACAGTCTTTCAACACAATTCTCTAAGTGCTGCCTTTGTGATTTCTTGAAG
	<i>IGF1 exon 2</i>	GCCACATAATACCCACCTTGACCTGCTGTAAAAGACCTGCAACCCAATTGTCACACCTCC
	<i>IGF1 exon 3</i>	TTGACCAGTTCAGCCACGGCAGGGCCCGAGACGCTCTGCGGAGCCGAGCTGGTGGATGCT
	<i>IGF1 exon 4</i>	CTACTTCTCTTCTCTGGTTGTGCCAAGATCTGCCTTCTCTTCAGCTTCCCGCAAATGC
	<i>IGF1 exon 5</i>	GTGAAGAGAATTCCACACTG
	<i>IGF1 exon 6</i>	AGTGCTGCTTCCGGAGCTGTGATCTGAGGCGGCTAGAGATGTACTGTGCACCAATTAAGC
	<i>IGF1 exon 7</i>	TATCAGCCCCATCCACCAACAAGAGAACGAAGGCTGAAAGGAGGAGGAAAGGTGGGCCA
	<i>IGF1 exon 8</i>	GGTGGGACAGTGGATCCAACCTGGTGGTGCCTGGGGAAGATTGGTTGGAGCAGTCTCCA
	<i>IGF1 exon 9</i>	CTAGAGCTGCAAGGGGACACTGAGGTGCTGAGACCTTCTAGGTTTGTGGATGGCAATGAA
	<i>IGF1 exon 10</i>	GTGAAGAAGCTGAGGCCAGAGATGTTTGCAAGATCATAAACATAGAGCTGGAAGAGGTC
	<i>IGF1 exon 11</i>	GAAATACATTTGAAGAACGCAAGCAGAGGAAGCTCAGGAAGCAGGAACCTACAGAATGTAA
	<i>IGF1 ex 1+3</i>	TCTAAGTGCTGCCTTTGTGATTTCTTGAAGGTGAAGATGCATACTGTGTCTCCATCCAT
	<i>IGF1 ex 2+3</i>	AAAGACCTGCAACCCAATTGTCACACCTCCGTGAAGATGCATACTGTGTCTCCATCCAT
	<i>IGF1 ex 3+4</i>	TTGTTTTCGGGAGAAAGAGGTTTCTACTTCACTACTTCTCTCTCTGTTGTGCCAAGAT
	<i>IGF1 ex 3+5</i>	TTGTTTTCGGGAGAAAGAGGTTTCTACTTCAGCAAGCCCACTGGATATGGCTCTAGCAGTC
	<i>IGF1 ex 6+7</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGTATCAGCCCCATCCACCAACAAGAGAACG
	<i>IGF1 ex 6+8</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGGTTGGGACAGTGGATCCAACCTGGTGGTGC
	<i>IGF1 ex 6+9</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGCTAGAGCTGCAAGGGGACACTGAGGTGCTG
	<i>IGF1 ex 6+10</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGGTGAAGAAGCTGAGGCCAGAGATGTTTGC
	<i>IGF1 ex 6+11</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGGAAATACATTTGAAGAACGCAAGCAGAGGA
	<i>IGF1 ex 7+8</i>	AGAGAACGAAGGCTGAAAGGAGGAGGAAAGGTTGGGACAGTGGATCCAACCTGGTGGTGC
	<i>IGF1 ex 7+9</i>	AGAGAACGAAGGCTGAAAGGAGGAGGAAAGCTAGAGCTGCAAGGGGACACTGAGGTGCTG
	<i>IGF1 ex 7+10</i>	AGAGAACGAAGGCTGAAAGGAGGAGGAAAGGTGAAGAAGCTGAGGCCAGAGATGTTTGC
	<i>IGF1 ex 7+11</i>	AGAGAACGAAGGCTGAAAGGAGGAGGAAAGGAAATACATTTGAAGAACGCAAGCAGAGGA
	<i>IGF1 ex 8+9</i>	TAGACATGGGGGATATTTCCGGGCTCCCCACTAGAGCTGCAAGGGGACACTGAGGTGCTG
	<i>IGF1 ex 8+10</i>	TAGACATGGGGGATATTTCCGGGCTCCCCAGTGAAGAAGCTGAGGCCAGAGATGTTTGC
	<i>IGF1 ex 8+11</i>	TAGACATGGGGGATATTTCCGGGCTCCCCAGAAATACATTTGAAGAACGCAAGCAGAGGA
	<i>IGF1 ex 9+10</i>	GGACTGCCCACCCAATACCCTTCATCCATGGTGAAGAAGCTGAGGCCAGAGATGTTTGC
	<i>IGF1 ex 9+11</i>	GGACTGCCCACCCAATACCCTTCATCCATGGAAATACATTTGAAGAACGCAAGCAGAGGA
	<i>IGF1 ex 10+11</i>	ATTAGATATTAAGACAGCAAGTGATGAGAAATACATTTGAAGAACGCAAGCAGAGGA
	<i>MRPS17</i>	GTTTCACGTGCTTGCTCGTGCTACCGCAATGCTTTGAGGAGCACAATGTCTCCAACCTG
Tasmanian devil	<i>IGF1 exon 1</i>	GACAAATTCAACAGTCTTTCAACCCAATTCTCTAAGTGCTGCCTTTGTGATTTCTTGAAG
	<i>IGF1 exon 2</i>	GCCACATAATACCCACCTTGACCTGCTGTAAAAGACCTGCAACCCAATTGTCACACCTCC
	<i>IGF1 exon 3</i>	TTGACCAGTTCGGCCACAGCGGGCCCGAGACACTCTGCGGAGCTGAGCTGGTGGATGCC
	<i>IGF1 exon 4</i>	CTACTTCTCTTCTTGGCTTGTGCCAAGATCTGCCTTCTCTCCAGCTTCCCGCAAATGA
	<i>IGF1 exon 5</i>	CTTAAATGTGAAAAAGGCTTCATGAGCGTCTTAACTCCCCGCCAAG

Species	Gene	Probe
	<i>IGF1 exon 6</i>	GGTGAAGAGCATTTCCACACTG
	<i>IGF1 exon 7</i>	AATGCTGCTCCGGAGCTGTGATCTGAGACGGCTAGAGATGTACTGTGCGCCAATTAAGC
	<i>IGF1 exon 8</i>	TATCAGCCCCATCCACCAACAAGAGAACGAAGGCTGAGAGGAGGAGGAAAGGTGGGCCA
	<i>IGF1 exon 9</i>	TGGGACAATGGATCCAAATTGTGGTGGCTGGAGAAAATTTGATTGGAGCAGTCTCCTGTT
	<i>IGF1 exon 10</i>	GATGCACATTCTGTGGTATTAGATATTTTAAATAGCAAATGATAAAAGACTGGAATTT
	<i>IGF1 exon 11</i>	GAAATACATTGGAAGAACGCAAGCAGAGGGAAGCTCAGGAAGCAGGAACTACAGAATGTAA
	<i>IGF1 ex 1+3</i>	TCTAAGTGCTGCCTTTGTGATTTCTTGAAGGTGAAGATGCATACCGTGTCTCCATCCAT
	<i>IGF1 ex 2+3</i>	AAAGACCTGCAACCCAATTGTACACCTCCGTGAAGATGCATACCGTGTCTCCATCCAT
	<i>IGF1 ex 3+4</i>	TTGTTTGTGGAGAAAGAGGTTTCTACTTCACTACTTCTCTTCTTGGCTTGTGCCAAGAT
	<i>IGF1 ex 3+5</i>	TTGTTTGTGGAGAAAGAGGTTTCTACTTCACTTAAATGTGGAAGAGGCTTCATGAGCGT
	<i>IGF1 ex 3+6</i>	TTGTTTGTGGAGAAAGAGGTTTCTACTTCAGGTGAAGAGCATTTCACACTG
	<i>IGF1 ex 3+7</i>	TTGTTTGTGGAGAAAGAGGTTTCTACTTCATTGTTTGTGGAGAAAGAGGTTTCTACTT
	<i>IGF1 ex 5+6</i>	CTTCATGAGCGTCTTAACTTCCCCGCCAAGGGTGAAGAGCATTTCACACTG
	<i>IGF1 ex 5+7</i>	CTTCATGAGCGTCTTAACTTCCCCGCCAAGGCAAGCCCACTGGGTATGGCTCTAGCAGTC
	<i>IGF1 ex 5+8</i>	CTTCATGAGCGTCTTAACTTCCCCGCCAAGTATCAGCCCCATCCACCAACAAGAGAACG
	<i>IGF1 ex 5+9</i>	CTTCATGAGCGTCTTAACTTCCCCGCCAAGGATGGGACAATGGATCCAAATTGTGGTGGC
	<i>IGF1 ex 5+10</i>	CTTCATGAGCGTCTTAACTTCCCCGCCAAGTATAAGCATAGAGTTGAAAGAGGTCTAAGA
	<i>IGF1 ex 5+11</i>	CTTCATGAGCGTCTTAACTTCCCCGCCAAGGAAATACATTTGAAGAACGCAAGCAGAGGG
	<i>IGF1 ex 7+8</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGTATCAGCCCCATCCACCAACAAGAGAACG
	<i>IGF1 ex 7+9</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGGATGGGACAATGGATCCAAATTGTGGTGGC
	<i>IGF1 ex 7+10</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGTATAAGCATAGAGTTGAAAGAGGTCTAAGA
	<i>IGF1 ex 7+11</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGGAAATACATTTGAAGAACGCAAGCAGAGGG
	<i>IGF1 ex 8+9</i>	AGAGAACGAAGGCTGAGAGGAGGAGGAAAGGATGGGACAATGGATCCAAATTGTGGTGGC
	<i>IGF1 ex 8+10</i>	AGAGAACGAAGGCTGAGAGGAGGAGGAAAGTATAAGCATAGAGTTGAAAGAGGTCTAAG
	<i>IGF1 ex 8+11</i>	AGAGAACGAAGGCTGAGAGGAGGAGGAAAGGAAATACATTTGAAGAACGCAAGCAGAGGG
	<i>MRPS17</i>	GTTTCACGTGCTTGCCCGAGGCACTGGCAATGCTCTGAGGAGCACGATGTCTCCGATTG
Wallaby	<i>IGF1 exon 1</i>	GACAAATTCAACAGTCTTTCAACACAATTCTCTAAGTGCTGCCTTTGTGATTTCTTGAAG
	<i>IGF1 exon 2</i>	GCCACATAATACCCACCTGACCTGCTGTAAAAGACCTGCAACCAATTGTCACACCTCC
	<i>IGF1 exon 3</i>	TTGTGTTTGCTTACCTTGACCAGTTCAGCCACGGCGGGGCCGAGACGCTCTGCGGAGCC
	<i>IGF1 exon 4</i>	CTACTTCTCTGCTCTGGTTGTGCCAAGATCTGCCTTCTTCTCATCTTCCCGGCAAATGA
	<i>IGF1 exon 5</i>	AATGCTGCTTCCGGAGCTGTGATCTGAGGCGGCTAGAGATGTACTGTGCGCCAATTAAGC
	<i>IGF1 exon 6</i>	TATCAGCCCCATCCAGCAACAAGAGAACGAACGCTGAGAGGAGGAAAGGTGGGCCAGAG
	<i>IGF1 exon 7</i>	TGGGACAATGTATCCAACCTTTGTGCTGTCTGGGAAGATTTGGTTGAAGCAGTCTCCAGC
	<i>IGF1 exon 8</i>	CAGAATCATAACCTAGAGCTGCAGGAGATCTCTGAAGTCATCTGGTCTACCTCATCATC
	<i>IGF1 exon 9</i>	GAAATCCATTTGAAGAACGCAAGCAGAGGGAGCTCAGGAAGCAGGAACTACAGAATGTAAAG
	<i>IGF1 ex 1+3</i>	TCTAAGTGCTGCCTTTGTGATTTCTTGAAGGTGAAGATGCATACTGTGTCTCCCTCCAT
	<i>IGF1 ex 2+3</i>	AAAGACCTGCAACCCAATTGTACACCTCCGTGAAGATGCATACTGTGTCTCCCTCCAT
	<i>IGF1 ex 5+6</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGTATCAGCCCCATCCAGCAACAAGAGAACG
	<i>IGF1 ex 5+7</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGAAATGGGACAATGTATCCAACCTTTGTGCTGT

Species	Gene	Probe
	<i>IGF1 ex 5+8</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGCATTACTTTACAGAATGGACAGGATCACAG
	<i>IGF1 ex 5+9</i>	CGCCACACCGATATGCCCAAAGCCCAAAAGGAAATCCATTGAAGAACGCAAGCAGAGGG
	<i>IGF1 ex 6+7</i>	ACAAGAGAACGAACGCTGAGAGGAGGAAAGAATGGGACAATGTATCCAACCTTGTGCTGT
	<i>IGF1 ex 6+8</i>	ACAAGAGAACGAACGCTGAGAGGAGGAAAGCATTACTTTACAGAATGGACAGGATCACAG
	<i>IGF1 ex 6+9</i>	ACAAGAGAACGAACGCTGAGAGGAGGAAAGGAAATCCATTGAAGAACGCAAGCAGAGGG
	<i>IGF1 ex 7+8</i>	TCCGGGCTCCCCACTGCTTCTGGGACTTCACATTACTTTACAGAATGGACAGGATCA
	<i>IGF1 ex 7+9</i>	TCCGGGCTCCCCACTGCTTCTGGGACTTCAGAAATCCATTTGAAGAACGCAAGCAGAGGG
	<i>MRPS17</i>	GTTCACGTGCTTGCTCTTGGCACTGGCAAGGCTTTGAGGAGCACGATGTCTCCGACTG

Table 3:

Organization of marsupial *IGF1* genes^{*}

Species	Exon 1	Intron 1	Exon 2	Intron 2	Exon 3	Intron 3	Exon 4	Intron 4	Exon 5	Intron 5	Exon 6	Intron 6	Exon 7	Intron 7	Exon 8	Intron 8	Exon 9	Intron 9	Exon 10	Intron 10	Exon 11	Length
Koala	344	1809	108	3866	157	57430	332	16129	22	549	182	2673	217,300	4777	125	4837	95	3844	128	7709	7438	111702
Tas. devil	464	2203	168	3614	157	58333	332	19215	22 ^{**}	718	182	2867	217	4052	125	8811	none	none	128	5934	7508	115154
Wallaby	427 [#]	1916	228	1561	157	3992	339	13765	none	none	182	2832	495	5405	125	ndx	none	ndx	132	ndx	7468 [^]	unknown

^{*} Assignments are based on koala *IGF1* (see Fig. 3); length is in base pairs

[#] Full-length not mapped because of poor quality DNA sequence

^{**} There is an exon of 48 bp located 609 bp further 5' that is only found in Tasmanian devil (defined as exon 5, see Fig. 5E). The equivalents for Koala exons 5-8 are exons 6-9 in Tasmanian devil.

[^] mapped using Koala exon 11

ndx – not detected because different exons have not been mapped to the same contiguous DNA strand

Table 4:

Nucleotide identity with koala *IGF1* exons (%)

Species	Exon 1 (344 bp)*	Exon 2 (108 bp)#	Exon 3 (157 bp)	Exon 4 (352 bp)	Exon 5 (22 bp)	Exon 6 (182 bp)	Exon 7 (217, 300 bp)^	Exon 8 (125 bp)	Exon 9 (95 bp)	Exon 10 (128 bp)¶	Exon 11 (7438 bp)
Tas. devil	99	100	94	89	95	97	94	86	none	85	93
Wallaby	99	100	96	92	none	98	93	90	none	88	92**

* 464 bp in Tasmanian devil and 427 bp in wallaby.
168 bp in Tasmanian devil and 228 bp in wallaby.
^ 217 bp in Tasmanian devil and 495 bp in wallaby.
¶ only 61 bp can be aligned in Tasmanian devil and 52 bp in wallaby.
** only 3020 bp can be aligned because of poor quality wallaby DNA sequence

Table 5:

Amino acid identities with koala IGF1 (%)

Species	Signal peptide 1 (48 AA)	Signal peptide 2 (25 AA)	Mature IGF-I (70 AA)	Common E (16 AA)	E _A peptide (19 AA)	E _B peptide (95 AA)	E _C peptide (56 AA)
Tas. devil	100	100	100	100	95	95 [*]	89
Wallaby	98	96	100	100	100	85	90 [^]
Human	81	88	91	94	79	60 [¶]	79 [‡]

^{*} 77 AA,[^] 42 AA (see Fig 9);[¶] 60 AA;[‡] 24 AA

Table 6:Comparison of putative Stat5b binding elements in marsupial *IGF1* Loci

K8-9* (351 bp)	Koala	Tasmanian devil	Wallaby	K60-61* (376 bp)	Koala	Tasmanian devil	Wallaby
% Identity	100	95.6	96.4	% Identity	100	93.6	91.4
5' Site	TTCTAAGAA	TTCTAAGAA	TTCTAAGAA	5' Site	none	none	none
Inter-site length (bp)	224	224	224	Inter-site length (bp)	-	-	-
3' Site	TTCTTAGAA	TTCTTAGAA	TTCTTAGAA	3' Site	TTCTTAGAA	TTCTTAGAA	TTCTTAGAA
Location from exon 1	-80,598 bp	-89,949 bp	Unknown [#]	Location from exon 1	+80,738	+70,007	+12,572

*
K = koala[#]
Located on a different genomic scaffold