Scientific Report

TRANSPARENT PROCESS OPEN ACCESS

EMBO reports

The RNA-binding protein Arrest (Bruno) regulates alternative splicing to enable myofibril maturation in *Drosophila* flight muscle

Maria L Spletter[1], Christiane Barz[1], Assa Yeroslaviz[1], Cornelia Schönbauer[1], Irene R S Ferreira[1], Mihail Sarov[2], Daniel Gerlach[3,†], Alexander Stark[3], Bianca H Habermann[1] & Frank Schnorrer[1,*]

Abstract

In *Drosophila*, fibrillar flight muscles (IFMs) enable flight, while tubular muscles mediate other body movements. Here, we use RNA-sequencing and isoform-specific reporters to show that *spalt major* (*salm*) determines fibrillar muscle physiology by regulating transcription and alternative splicing of a large set of sarcomeric proteins. We identify the RNA-binding protein Arrest (Aret, Bruno) as downstream of *salm*. Aret shuttles between the cytoplasm and nuclei and is essential for myofibril maturation and sarcomere growth of IFMs. Molecularly, Aret regulates IFM-specific splicing of various *salm*-dependent sarcomeric targets, including *Stretchin* and *wupA* (TnI), and thus maintains muscle fiber integrity. As Aret and its sarcomeric targets are evolutionarily conserved, similar principles may regulate mammalian muscle morphogenesis.

Keywords alternative splicing; Arrest; *Drosophila*; flight muscle; myofibril
Subject Categories Development & Differentiation; RNA Biology
DOI 10.15252/embr.201439791 | Received 27 October 2014 | Revised 25 November 2014 | Accepted 2 December 2014 | Published online 22 December 2014
EMBO Reports (2015) 16: 178–191

Introduction

Mammals possess various muscle types that exhibit particular physiological properties to fulfill their diverse functions. For example, the heart muscle beats continuously throughout the life of the animal, slow skeletal muscle fibers support endurance exercises, and fast skeletal muscles empower peak forces but fatigue quickly. The major physiological and biophysical differences between muscle types are largely determined by differences in the expression patterns of structural proteins that build the contractile structures—the myofibrils and sarcomeres. One prominent example is the transcriptional regulation of the various muscle myosin heavy chain genes in mammals, often used as the basis for muscle fiber-type classification [1]. In addition to differential transcription, alternative splicing adds another level of regulation by creating a plethora of additional protein isoforms. In particular, alternative splicing of the large sarcomeric proteins, such as titin, contributes to physiological diversity. Differential splicing between skeletal muscles and heart results in a short and stiff, heart-specific titin isoform that is implicated in the high passive stiffness of mammalian heart [2,3].

Drosophila is a valuable model to study the mechanisms that instruct and execute muscle fiber-type diversity. The adult fly houses two different types of body muscles: fibrillar indirect flight muscles (IFMs) and tubular body muscles. Tubular muscles are similar to mammalian body muscle; they contain laterally aligned sarcomeres and contract synchronously in response to motor neuron stimulation, which triggers calcium influx. By contrast, fibrillar IFMs contain individual non-aligned myofibrils and use an asynchronous contraction mechanism. In addition to calcium influx, this mechanism requires physical stretch stimulation as a trigger. Thus, IFMs, similar to mammalian heart, display a high passive stiffness likely caused by a specific sarcomeric protein composition. Together, these biophysical features of IFM myofibers achieve the very high contraction frequencies and large power output of IFMs, enabling insect flight [4–6].

We have shown previously that the Zn-finger transcription factor *spalt major* (*salm*) is required and sufficient for fibrillar IFM fate choice during pupal development. Loss of *salm* from IFMs switches these muscles to a tubular fate, whereas gain of *salm* in tubular muscles converts them to the fibrillar fate [7]. Salm executes this switch by the regulation of targets on both the transcriptional and splicing level. However, as the initial study of the *salm* mutant IFMs was performed by microarray analysis which provided limited coverage of the various gene isoforms [7], it remained unclear to what extent alternative splicing contributes to the muscle fiber-type switch. Furthermore, it was unknown which RNA-binding protein may instruct the IFM-specific splicing pattern.

Here, we provide a systematic analysis of the *salm*-regulated genes and gene isoforms in IFMs by mRNA-Seq and identify a core

1 Max Planck Institute of Biochemistry, Martinsried, Germany
2 Max Planck Institute of Cell Biology and Genetics, Dresden, Germany
3 Research Institute of Molecular Pathology (IMP), Vienna Biocenter (VBC), Vienna, Austria
*Corresponding author. Tel: +49 89 8578 2434; E-mail: schnorrer@biochem.mpg.de
†Present address: Boehringer Ingelheim RCV GmbH & Co KG, Vienna, Austria

set of more than 700 fibrillar-specific gene isoforms, many of which code for sarcomeric components. We show that the RNA-binding protein Arrest (Aret, Bruno) acts downstream of *salm* to regulate a large number of these genes by instructing their alternative splicing. These targets include Stretchin (Strn-Mlck), Sls/Kettin, and WupA, which are incorporated into the growing sarcomeres during myofiber maturation. Thus, Aret ensures the proper isoform composition of the sarcomeric module during flight muscle development, enabling the construction of muscles fast and powerful enough to enable insect flight.

Results

Wild-type IFMs have a fibrillar morphology of their myofibrils, and their nuclei are spaced regularly between the myofibril bundles (Fig 1A). By contrast, leg or jump muscles display a tubular fiber morphology with their nuclei located in the center of the tube (Fig 1B, C). Muscle-specific RNAi-mediated knockdown of *salm* (*salm-IR*) or conditional deletion of *salm* with *Mef2-GAL4* using a novel conditional *salm* allele that is flanked by 2 FRT insertions ($salm^{FRT}$) results in a complete tubular conversion of the *salm* mutant IFMs (Fig 1D and Supplementary Fig S1), which has been observed previously [7]. To systematically identify the *salm* targets underlying the morphological and physiological differences between fibrillar IFMs and tubular muscles, we dissected IFMs, leg muscle, and jump muscle from wild-type adults, as well as $salm^{FRT}$ and *salm-IR* IFMs, and performed mRNA-Seq on biological duplicates. Bioinformatic analysis using DESeq2 to detect differential gene expression [8] identified 362 genes with a $\log_2$-fold change greater than 2 ($\log_2FC > 2$) whose expression are significantly enriched in wild-type IFMs as compared to *salm-IR* IFMs (Fig 2A). 133 of these genes are also significantly enriched in wild-type IFMs as compared to leg and jump muscles (Fig 2A and Supplementary Table S1). Thus, these 133 genes are fibrillar muscle specific, and their expression depends on *salm* function.

Many muscle genes, in particular the complex sarcomeric genes, are present in multiple isoforms and differentially expressed between muscle types [9]. Our previous microarray data suggested that the regulation of some IFM-specific isoforms could be *salm* dependent [7]. To systematically identify all exons and their respective gene isoforms that are regulated by *salm*, we performed a DEXSeq analysis of our mRNA-Seq data [10]. We identified 794 exons from 577 genes with a $\log_2FC > 2$ that are significantly enriched in IFMs as compared to leg or jump muscles and are dependent on *salm* (Fig 2B and Supplementary Table S1). Together with the 133 genes regulated at the gene level, our analysis identified a total of 703 genes that are upregulated in a *salm*-dependent fashion in fibrillar versus tubular muscle. We define these 703 genes as core fibrillar muscle-specific genes or gene isoforms (Supplementary Table S1).

Interestingly, these 703 genes are highly enriched for cytoskeletal or mitochondrial components (Fig 2C, D). To investigate fibrillar versus tubular expression of the sarcomeric genes in more detail, we clustered the $\log_2FC$ values of all exons from sarcomeric genes that are significantly differentially expressed (P-value < 0.05), in total 319 exons from 53 sarcomeric genes (Fig 2E). Generally, we see two major sub-clusters of sarcomeric exon expression: 'Fibrillar exons' are upregulated in IFMs as compared to legs in a *salm*-dependent manner, while 'tubular exons' are upregulated in legs or *salm-IR* IFMs as compared to wild-type IFMs (Fig 2E). Often, the same gene has both fibrillar and tubular exons, indicating muscle-type-specific isoform expression (Fig 2E).

To support our RNA-Seq analysis data and to investigate the expression and localization of fibrillar and tubular muscle-specific genes or gene isoforms, we generated a number of genomic fosmid reporter transgenes [11] in which we inserted a GFP tag into the protein or protein isoform of interest by recombineering [12]. We find very prominent IFM-specific expression of the titinlike gene *Stretchin* (*Strn-Mlck*) isoform R. This fibrillar isoform is expressed from its own promoter and has a unique splicing pattern resulting in an early termination as compared to the tubular isoforms (Fig 1G). The fosmid reporter shows that Strn-Mlck-IsoR protein is indeed IFM specific and localizes to the myosin thick filament of the sarcomeres (Fig 1H–J). Both Strn-Mlck-IsoR RNA and protein expression are entirely dependent on *salm* (Fig 1G, K–M). Conversely, alternative splicing of the other titin homolog *sls/kettin* results in early termination of tubular muscle-specific short isoforms A/D that localize to the Z-disks of tubular muscle (Fig 1N–Q). These isoforms are gained in *salm-IR* IFMs (Fig 1R–T). Alternative splicing of the LIM domain protein Limpet (Lmpt) results in the short IFM-specific isoform K, which lacks LIM domains, and the long tubular muscle-specific isoforms, including isoforms B/C/J with 5 LIM domains, which localize to I-bands of tubular muscle. Again, the muscle-specific splicing pattern depends on *salm* (Supplementary Fig S2A–I). Additionally, we confirm that the previously characterized IFM-specific expression of Act88F [13] strongly depends on *salm* (Supplementary Fig S2J–N), while expression of the normally tubular muscle-specific Mlp84B [14] is gained in *salm-IR* IFMs (Supplementary Fig S2O–S). Together, these systematic data suggest that *salm* indeed determines fibrillar muscle morphology by controlling expression and alternative splicing of many differentially expressed sarcomeric genes.

Figure 1. Expression of Muscle-type-specific RNA and protein isoform depends on *salm*.

A–F Wild-type fibrillar IFMs (A) and tubular leg (B) and jump muscles (TDT) (C). Knock down of *salm* leads to tubular conversion of the IFMs (D), leg and jump muscles are unaffected (E, F).

G–T mRNA-Seq read counts showing expression of *Strn-Mlck* (G–M) and *kettin* isoforms (N–T) from wild-type IFMs, leg muscle and jump muscle as well as *salm-IR* or $salm^{FRT}$ conditional mutant IFMs (G, N), and by genomic GFP-tagged isoform markers in *Strn-Mlck* (H–M) and *kettin* (O–T). Note that the IFM-specific expression of Strn-Mlck-IsoR depends on *salm* (compare H and K), whereas the tubular muscle-specific Kettin-IsoA/D is gained in IFMs upon loss of *salm* (compare O and R). Insertion of the GFP tag is indicated by green arrows in G and N. Hairpin sequence of *Strn-Mlck-IR* and the MiMIC insertion site are marked in G. Note the fibrillar-specific exons marked with green boxes (G, N) and the tubular-specific exons in *Strn-Mlck* marked by a red box (G).

Data information: Black arrows indicate direction of transcription. Scale bars are 5 μm, and all images were cropped to the same size.

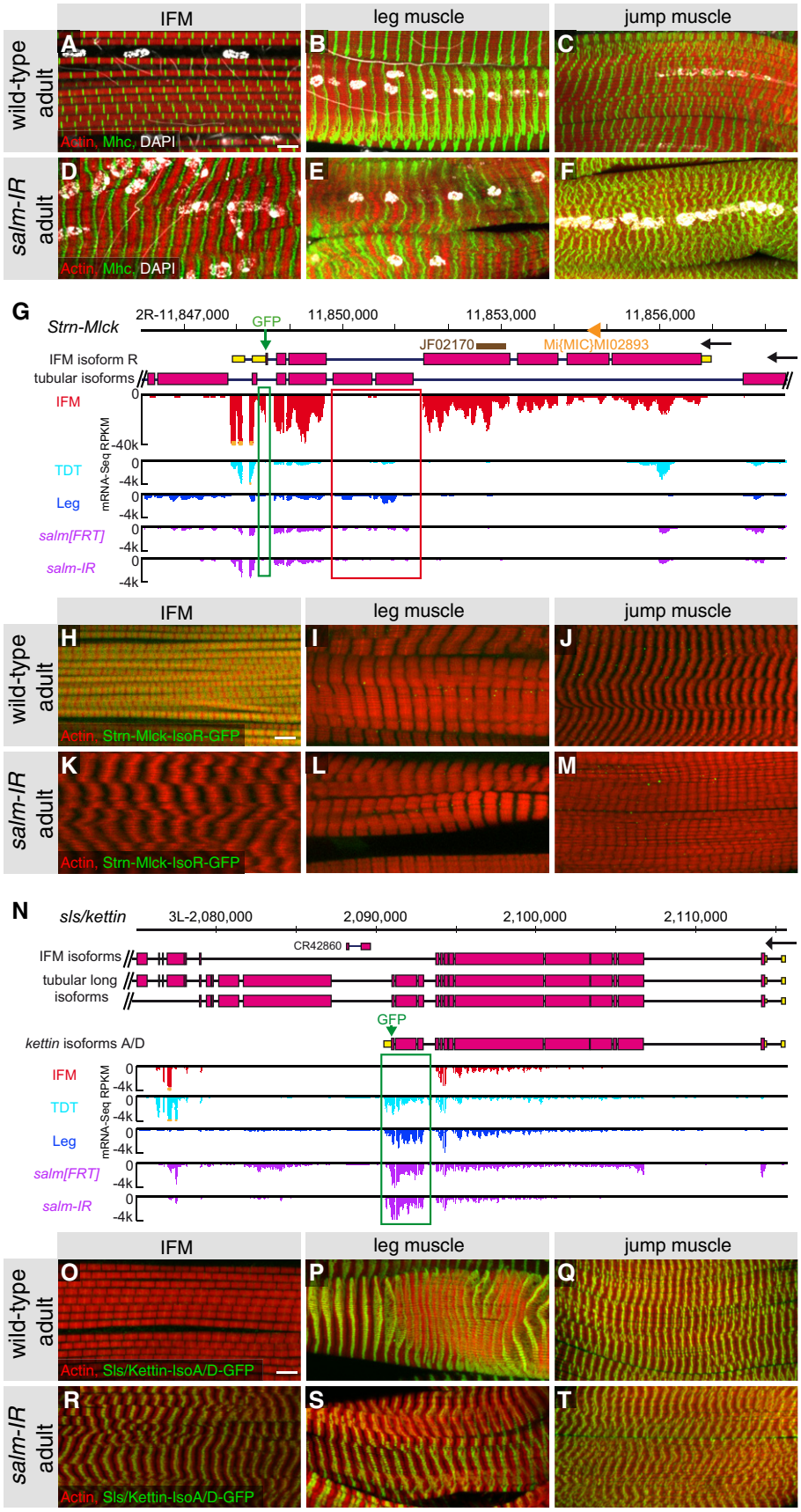

Figure 1.

To mechanistically investigate how *salm* instructs the IFM-specific splicing pattern of these identified sarcomeric genes, we looked for RNA-binding proteins that are regulated by *salm*. Our earlier work had identified the RNA-binding protein Arrest (Aret, Bruno), which contains 3 conserved RNA recognition motif (RRM) domains [15], as *salm* dependent [7]. Interestingly, genomewide muscle-specific RNAi data had shown that knockdown of *aret* using *Mef2-GAL4* can result in a flightless phenotype [16], making Aret a prime candidate to mediate IFM-specific splicing downstream of *salm*. Using developmental mRNA-Seq analysis of isolated IFMs, we found that *aret* mRNA is expressed highly in developing IFMs at 30 and 72 h APF and maintained at lower levels in adult IFMs but not in tubular leg or jump muscles (Fig 3A). Interestingly, Aret-specific antibodies detect Aret protein in the nuclei of adult IFMs, but not tubular muscles (Fig 3B–D). This IFM-specific expression pattern is lost in *salm-IR* IFMs (Fig 3E), suggesting that Aret indeed acts downstream of *salm* in IFMs.

To functionally investigate the role of *aret* in IFMs, we knocked down *aret* with *Mef2-GAL4* and a number of available hairpins from TRiP and VDRC. We found four partially non-overlapping hairpins, GD41568, KK107459, TRiP38983, and TRiP44483 (Fig 3A) that lead to viable adults flies that are entirely flightless (Fig 4A). Additionally, we investigated trans-heterozygous combinations of *aret* loss of function alleles $aret^{PA}/aret^{PD}$, $aret^{PA}/aret^{QB}$, and $aret^{PD}/aret^{QB}$, which were initially identified as female sterile due to developmental arrest of the germ line [17] and were later used to demonstrate that Aret is important to prevent premature *osk* mRNA translation during RNA transport [18]. All of these *aret* allelic combinations were indeed viable, female-sterile and entirely flightless (Fig 4A), demonstrating that *aret* is essential for IFM formation or function, but does not have an essential role in tubular muscle, as this would result in developmental lethality.

To investigate the IFM phenotype in detail, we stained young, day 1 adult hemithoraces of wild-type, *aret-IR,* and $aret^{PD}/aret^{QB}$ mutants with phalloidin and found that IFM fibers begin to thin and rupture close to their thoracic attachment sites (Fig 4B–D). Additionally, the sarcomeres of the *aret-IR* or *aret* mutant myofibrils appear too short and are sometimes entirely lost in day 1 adults (Fig 4E–G). Cross sections reveal myofibrils that are variable in diameter and often hollow after *aret* loss, in contrast to dense, regular myofibrils in wild-type (Supplementary Fig S3). Interestingly, a few days after eclosion, generally all IFM fibers of *aret-IR* or mutants are ruptured and the myofibrils entirely lose their sarcomeric organization (Fig 4H–M), suggesting a gradual IFM fiber degeneration during the first few days of adult life. These *aret-IR* or *aret* mutant flies remain viable and their tubular leg muscles do not display any obvious phenotypes (Fig 4N–P), again suggesting that Aret is only required in fibrillar IFMs.

The *aret* phenotype in young adult flies prompted us to investigate the developmental role of Aret in IFMs. We followed the development of the dorsal–longitudinal IFMs, which form by fusion of myoblasts to larval template muscles during early stages of pupal development [19]. We find that Aret protein is localized to substructures of the large larval nuclei in the muscle templates, but not in the nuclei of the fusing adult myoblasts at 14 h after puparium formation (APF) (Fig 5A). Aret expression remains low in the forming myotubes at 17 h APF, but becomes readily detectable from 24 h APF onwards. From 24 to 60 h APF, we find that Aret is often tightly associated with the nuclei or nuclear membrane and some Aret is present within the nuclei; however, the majority of Aret appears dispersed throughout the IFM cytoplasm (Fig 5C–F, I). Interestingly, this pattern drastically changes by 72 h APF when most Aret is shuttled into the nuclei, where it remains until adulthood (Fig 5G–H, J). Together, these localization patterns are consistent with a role of Aret in the nucleus; however, before 72 h APF, it may also have a function in the cytoplasm.

As adult myofibrils of *aret* mutant IFMs are too short, we investigated when this phenotype arises during IFM development. Distinct myofibrils are detectable from about 32 h APF onwards [20], with readily scorable sarcomeres present at 48 h APF in wild-type IFMs (Fig 6A, B). *aret-IR* myofibrils appear a bit more irregular at 32 h APF but form properly by 48 h, housing sarcomeres of comparable length to wild-type (48 h APF wild-type length: 1.92 μm, SD = 0.23 μm; *aret-IR*: 2.04 μm, SD = 0.24 μm, Fig 6E, F, I). Wild-type sarcomeres begin to grow, reaching 2.75 μm (SD = 0.10 μm) at 72 h APF and 3.30 μm (SD = 0.16 μm) at 90 h APF. Interestingly, *aret-IR* sarcomeres fail to grow, instead even shorten, resulting in 1.87 μm (SD = 0.31 μm) long sarcomeres at 72 h APF and 1.76 μm (SD = 0.17 μm) long ones at 90 h APF (Fig 6C, D, G–I). This suggests that Aret is required for myofiber maturation and sarcomere growth happening after 48 h APF, potentially correlating with its increased nuclear localization during later stages of IFM morphogenesis.

To mechanistically investigate the molecular cause of the myofibril and sarcomere maturation defect, we aimed to identify targets of Aret by performing developmental mRNA-Seq from isolated wild-type and *aret-IR* IFMs. We first focused on the 362 genes that we had shown above to be regulated by *salm* in IFMs. We find that IFM-specific expression of only 51 (14%) of these also depends on *aret* function, demonstrating that *aret* regulates a small subset of the *salm*

Figure 2. Systematic identification of *salm*-dependent fibrillar and tubular muscle-specific exons.

A Venn diagram comparing significantly differentially expressed (P-value < 0.05, DESeq2) genes whose $\log_2$-fold changes are greater than 2 ($\log_2FC > 2$). Note that expression of 133 IFM-specific genes is *salm* dependent.

B Venn diagram comparing significantly differentially expressed (P-value < 0.05, DEXSeq) exons with $\log_2FC > 2$. Note that expression of most IFM-specific exons is *salm* dependent (n = 794). These exons combined with the genes in (A) define the core group of fibrillar-specific genes.

C REVIGO treemap of GO component analysis of the 703 core fibrillar genes versus all expressed genes showing enrichment for muscle, mitochondrial, and cytoskeletal terms.

D Tree of selected GOrilla GO component terms highlighting enrichment of muscle structural components.

E Hierarchical clustering of $\log_2FC$ of all 319 exons from 53 sarcomeric genes that are significantly differentially expressed (P-value < 0.05, DEXSeq) comparing their expression in IFMs to entire legs, jump muscles, *salm-IR* IFMs or *aret-IR* IFMs. Exons cluster into 'fibrillar exons' (shown in reds) which are IFM specific and mostly *salm* dependent and 'tubular exons' (shown in blues). Black dots on right demark location of individual sarcomeric gene exons in the heatmap. Note that individual sarcomeric genes have both tubular- and fibrillar-specific exon expression.

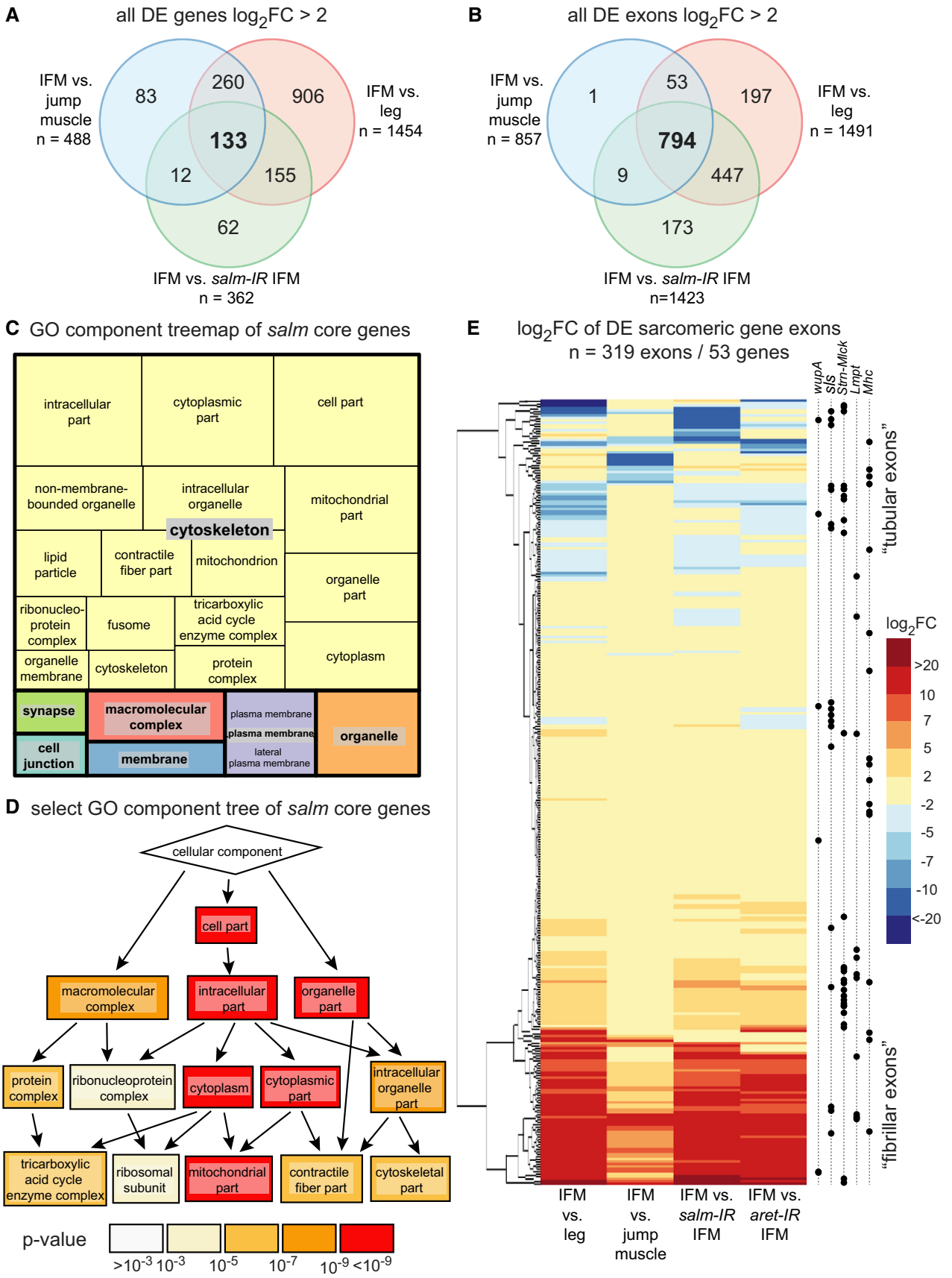

Figure 2.

targets at the transcriptional level of the entire gene unit (Fig 7A, Supplementary Table S2). Strikingly, we find that expression of 1119 of the 1423 (79%) *salm*-dependent exons also requires *aret* function in IFMs (Fig 7B). The $\log_2$FC values are also highly correlated (Pearson's coefficient = 0.7669482, Spearman's coefficient = 0.8383704) when comparing all exons significantly differentially expressed (*P*-value < 0.05) between *aret-IR* and *salm-IR* IFMs (Fig 7E, Supplementary Table S2). Overall, our analysis identifies only 24 genes, but 747 exons, which are upregulated in IFMs compared to leg and jump muscles and co-dependent on *salm* and *aret* (Fig 7C, D). These data strongly suggest that *salm* induces Aret expression, which then instructs the IFM-specific splicing pattern.

Many of the *salm* and *aret* co-regulated exons belong to sarcomeric genes, highlighting their key importance in building functionally different muscle types. Aret regulated exons cluster into 'fibrillar' and 'tubular' classes as observed for Salm-regulated exons,

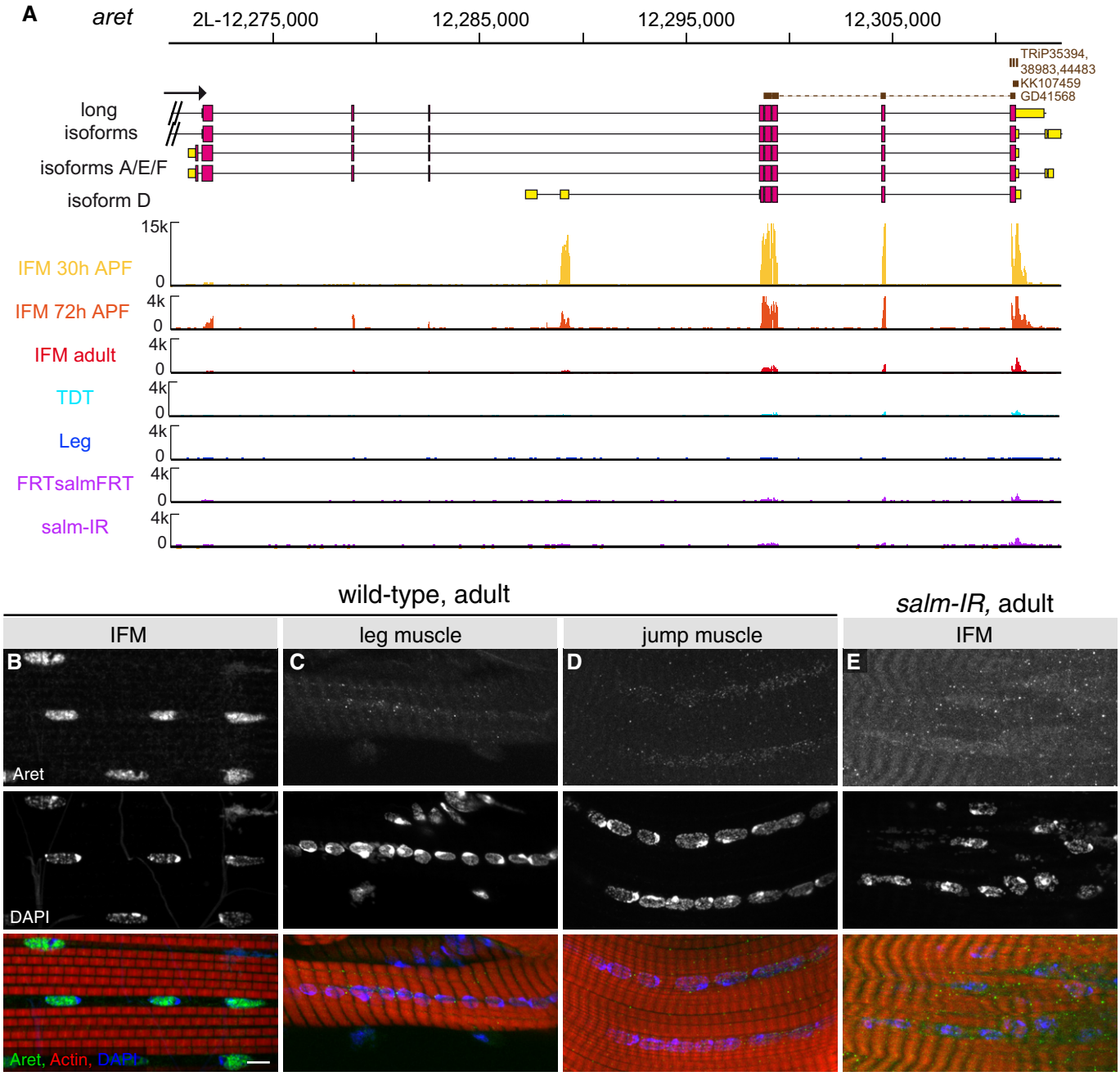

Figure 3. IFM-specific expression and nuclear localization of Aret.

A Developmental mRNA-SEQ analysis of *aret* expression in IFMs, leg muscle and jump muscle. *aret* mRNA is strongly expressed at 30 h APF and expression continues in adult IFMs, but not in leg or jump muscles. *aretIR* hairpin sequences are indicated.

B–E Aret protein localizes to IFM nuclei (B) but not to leg muscle (C) or jump muscle nuclei (D) in adults. Both *aret* mRNA and protein expression depend on *salm* (A, E). Scale bar are 5 μm.

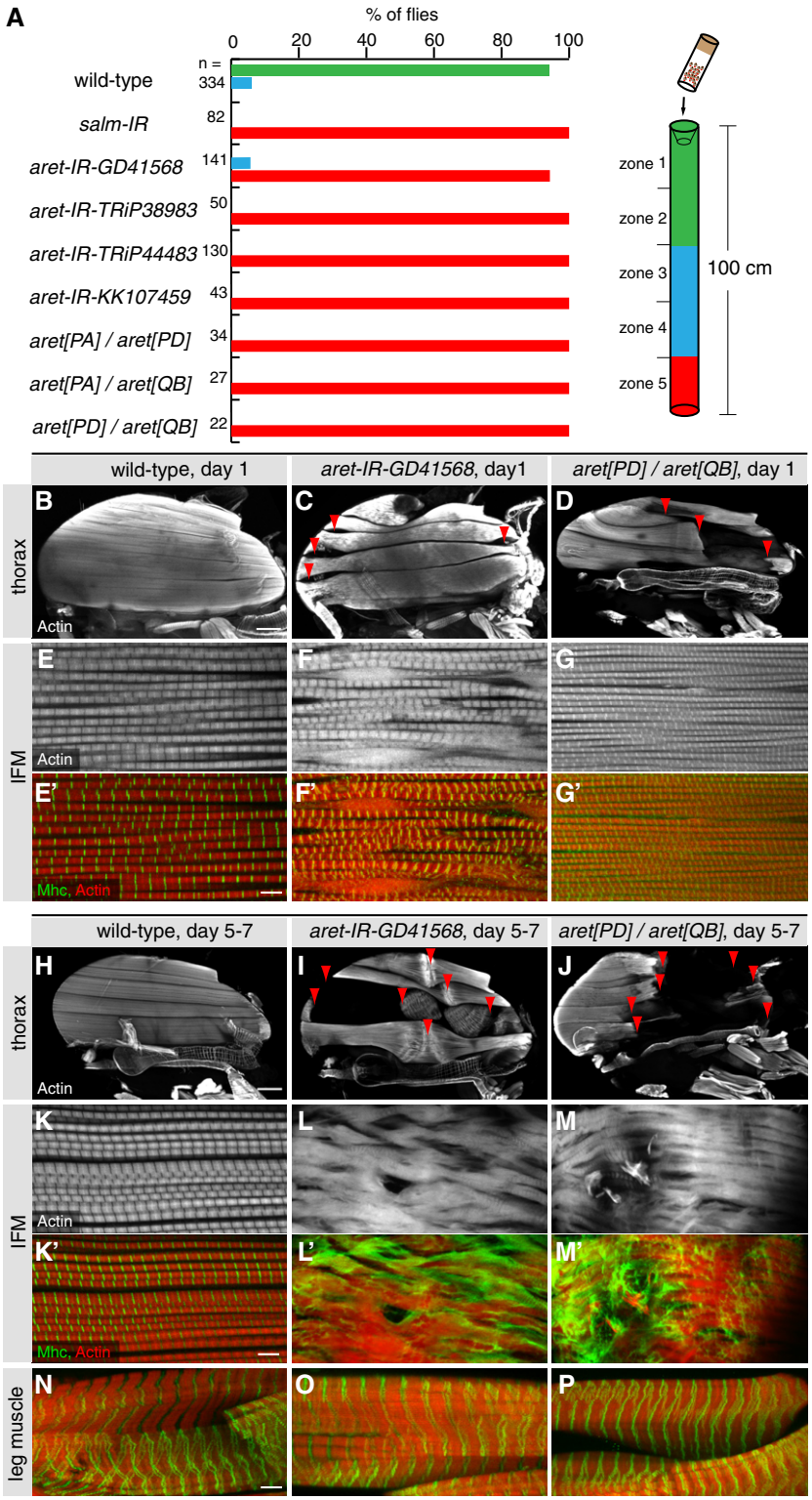

Figure 4. *aret* is required for flight muscle function and fiber integrity.

A Flight tests of various *aret*-specific RNAi hairpins and trans-heterozygous *aret* mutant combinations.

B–G Hemithoraces (B–D) and IFM myofibrils (E–G) from young (day 1) wild-type (B, E), *aret-IR* (C, F), or *aret* mutants (D, G). Note the thinner or ruptured IFM fibers upon *aret* removal (C, D, red arrow heads) and the shorter or entirely disrupted sarcomeres (F, G compare to E). Scale bars are 100 μm in (B–D).

H–P Hemithoraces (H–J) and IFM myofibrils (K–M) of aged adults (5–7 days) from wild-type, *aret-IR*, or *aret* mutants. Note the severe IFM fiber disruption (I, J, red arrow heads) and the severe myofibril and sarcomere defects upon *aret* removal (L, M compared to K). Leg muscles are unaffected (N–P). Scale bars are 100 μm (H–J), and 5 μm (E–G) and (K–P).

highlighting the disruption of the normal fibrillar splicing program in *aret-IR* IFMs (Fig 2E). The $\log_2$FC values are even more tightly correlated for sarcomeric protein exons (red dots in Fig 7E, Pearson's coefficient = 0.8365097; Spearman's coefficient = 0.8603421) than when comparing all exons significantly differentially expressed (*P*-value < 0.05) between *aret-IR* and *salm-IR* IFMs (Fig 7E). We observe both loss and gain of exon expression in *aret-IR* IFMs, although expression of more exons is lost, indicating that Aret can both promote fibrillar exon inclusion and inhibit the use of tubular exons (Fig 7E).

To support our bioinformatics data, we used our fosmid reporter lines and find that neither IFM-specific expression of Act88F nor IFM- or leg muscle-specific splicing of Lmpt depend on *aret* (Supplementary Fig S4A–N). Interestingly, both genes are already expressed at 30 h APF in IFMs and hence already present during early phases of myofibril formation. By contrast, the IFM-specific *Strn-Mlck-IsoR* is entirely lost in *aret-IR* IFMs (Supplementary Fig S5A–E). *Strn-Mlck-IsoR* mRNA is only expressed from 72 h onwards, correlating with myofibril maturation and the strong nuclear localization of Aret protein (Supplementary Fig S5A and G). Similarly, the tubular muscle-specific *sls/kettin-IsoA/D,* which is present at low levels at 30 h APF in IFMs but then entirely suppressed in IFMs from 72 h APF onwards, is strongly gained in *aret-IR* IFMs (Supplementary Fig S5F–J). This demonstrates that Aret is actively required to suppress splicing into the terminal exons of the short *sls/kettin* isoforms in developing IFMs. We also identified *wupA* (troponin I, TnI) as an Aret target and generated a fosmid reporter line for the tubular muscle-specific isoform (Supplementary Fig S5K–M). Interestingly, Aret is required for both splice suppression of a tubular muscle-specific exon and inclusion of a fibrillar muscle-specific exon from 30 h APF onwards (Supplementary Fig S5K). The *wupA* fosmid reporter confirms that the tubular *wupA* isoform is indeed gained in *aret-IR* IFMs (Supplementary Fig S5N, O). In addition to confirming these complex changes in alternative splicing, we could also confirm an identified Salm-dependent transcriptional change in *aret-IR* IFMs. Expression of the tubular muscle-specific Mlp84B is strongly gained in *aret-IR* IFMs (Supplementary Fig S4O–S). Since this gain only occurs after eclosion and not yet at 72 h APF, it is possibly promoted by an unknown transcription factor whose activity is regulated by Aret, potentially via alternative splicing. Statistically, we find that Aret indeed regulates a large number of exons specifically in adult IFMs (491), whereas only 129 exons are specifically regulated at 30 h APF during initiation of myofibrillogenesis and a smaller set of 52 exons are regulated at all analyzed developmental stages (Fig 7F). Together, these data demonstrate that a large subset of the fibrillar muscle-specific *salm* targets are regulated by Aret. This regulation happens mainly at the splicing level during later stages of flight muscle morphogenesis.

Mis-splicing of *wupA* (TnI) is implicated in muscle fiber degeneration caused by muscle hyper-contraction [21,22]. Interestingly, *aret-IR* IFMs also display splicing defects in *Mhc* and *up* (TnT), which are also implicated in muscle hyper-contraction and as a consequence can lead to muscle fiber loss [23] (Supplementary Fig S6). To test whether the *aret-IR* fiber degeneration phenotype is caused by uncontrolled myosin activity leading to muscle hyper-contraction, we crossed the IFM-specific *Mhc* null allele, Mhc^{10}, into the *aret-IR* background. At 90 h APF, the *aret-IR* IFM fiber morphology is comparable to wild-type; however, *aret-IR* fibers are torn during the first days after eclosion (Fig 8A–F). This fiber degeneration phenotype is entirely rescued by the additional removal of *Mhc* from IFMs, demonstrating that loss of Aret causes uncontrolled myosin activity and IFM fiber hyper-contraction in adults (Fig 8G–L).

A number of the 'hyper-contraction genes' regulate myosin activity. As the newly identified Strn-Mlck-IsoR protein is also strongly localized to the myosin filament (Fig 1H), we investigated its role in fiber contraction. We knocked down *Strn-Mlck-IsoR* by an isoform-specific hairpin with *Mef2-GAL4* (see Fig 1G) and found an IFM fiber degeneration phenotype after eclosion, remarkably similar to that of *aret-IR* (Fig 8M–O). This *Strn-Mlck-IsoR* RNAi phenotype was confirmed by a MiMIC insertion disrupting the IFM-specific isoform (Fig 8P–R and see Fig 1G). Together, this suggests that *Strn-Mlck-IsoR* is a major Aret target that regulates myosin activity and biophysical forces in adult IFMs.

Discussion

Functionally different muscle types are essential for normal life in higher animals. Most insects require fast oscillating indirect flight muscles to enable flight. In *Drosophila* and also in the beetle *Tribolium*, *salm* or its *Tribolium* homolog determines the fibrillar morphology of the IFMs [7]. Our systematic mRNA-Seq data revealed that in order to achieve fibrillar muscle morphogenesis, *salm* controls the expression of a large core set of fibrillar genes (more than 700). Many of these genes are present in distinct isoforms in fibrillar versus tubular muscles. These unique isoform combinations potentially determine the specific physiological and biophysical features of the different muscle types.

As many of the *salm* targets, in particular the complex sarcomeric genes, are regulated at the level of alternative splicing, *salm* needs to instruct a fibrillar muscle-specific splicing program. Our data suggest that this is largely achieved by IFM-specific expression of Aret (Supplementary Fig S7). Aret controls IFM-specific splicing of a very significant subset of sarcomeric genes, including the titin homolog Strn-Mlck, as well as repressing tubular-specific splicing in IFMs, such as tubular-specific events of the titin homolog sls/Kettin. Interestingly, a number of these splicing events occur between 48 h and 90 h APF during which the myofibrils, initially housing thin and short sarcomeres, mature to myofibrils with long and thick sarcomeres, which can contract in a stretch-sensitive manner. As Aret activity is essential for normal sarcomere growth and myofiber maturation, it is likely that incorporation of the Aret targets such as Strn-Mlck or IFM-specific WupA (TnI) instructs normal myofibril maturation (Supplementary Fig S7).

Splicing as well as alternative splicing occurs in the nucleus. Thus, a direct regulator of splicing should be located in the nucleus. Aret lacks an obvious nuclear localization signal and until 60 h APF is found largely in the cytoplasm of the developing IFMs. Nevertheless, some Aret is present within the nuclei, possibly in domains close to the nuclear membrane, where Aret could regulate fibrillar-type splicing of targets like *wupA*. Upon an unknown stimulus, most of the Aret protein translocates to the nuclei by 72 h APF, and now most of the Aret targets, including *Strn-Mlck* and *sls/kettin,* are spliced in fibrillar mode. Together, this enables correct sarcomere growth and myofibril maturation. In the mature IFMs, it prevents

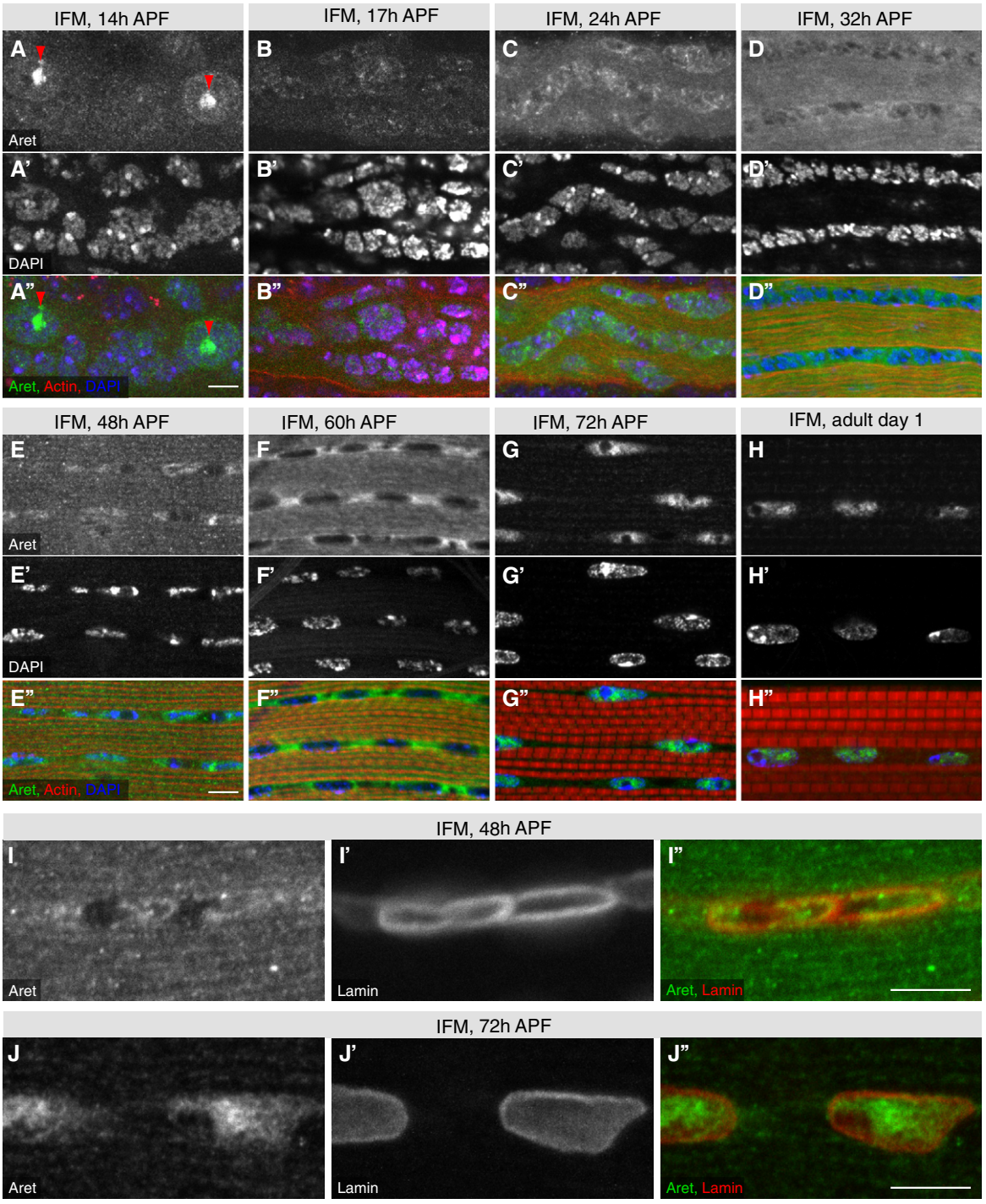

Figure 5. Aret shuttles between the cytoplasm and nucleus during IFM development.

A–H Aret protein is concentrated in sub-nuclear locations of the larval IFM template muscles at 14 h APF (red arrow heads in A). Aret protein is barely detectable at 17 h APF and its levels increase until 60 h APF, during which time Aret is located in close proximity to the nuclei and throughout the IFM cytoplasm (B–F). Aret protein is found in the nuclei at 72 h APF (G) and in adult IFMs (H). Scale bars are 5 μm.

I, J Co-stain of Aret and nuclear Lamin reveals some Aret in the nuclei at 48 h APF (I), whereas most of Aret is nuclear at 72 h APF (J). Scale bars are 5 μm.

muscle hyper-contraction and thus is essential for normal muscle fiber maintenance.

It is well established that Aret (Bruno) can regulate mRNA translation by binding to the 3′UTR of *osk* mRNA to prevent its premature translation during transport of the RNA from the nurse cells to the posterior pole of the oocyte in *Drosophila* [24,25]. A similar function for Aret in translational control of *grk* mRNA in the oocyte has also been suggested [26]. In both cases, Aret-dependent translational

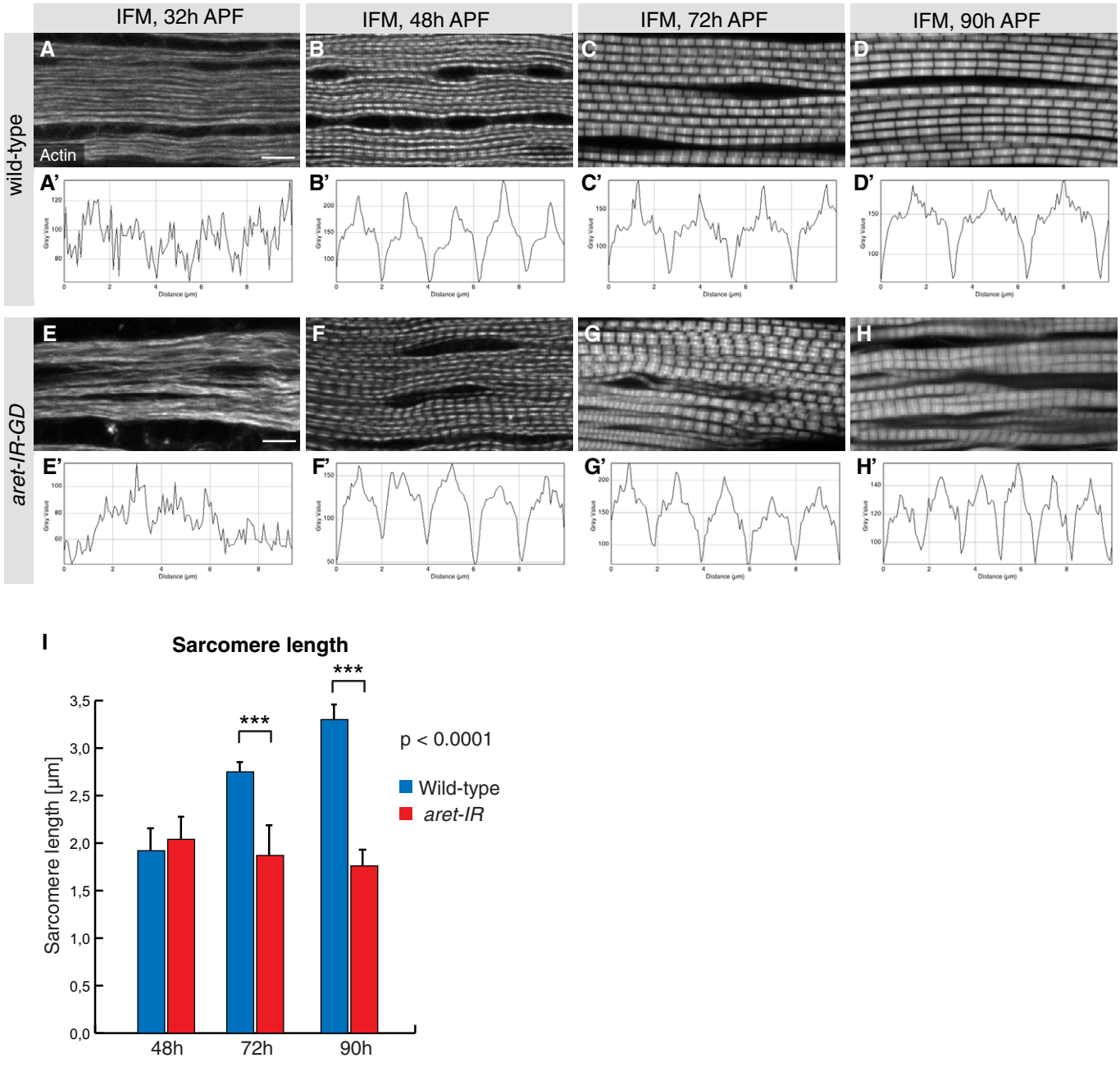

Figure 6. *aret* is essential for myofibril maturation and sarcomere elongation.

A–H Developing myofibrils and sarcomeres in wild-type (A–D) and *aret-IR* pupae (E–H) stained with phalloidin. Intensity plots 10 μm in length within one representative myofibril (A'–H'). Regular sarcomeres of about 2 μm at 48 h APF in wild-type (B) mature to about 3.3 μm at 90 h APF (D). *aret-IR* myofibrils and sarcomeres are present at approximately the normal length at 48 h APF (F) but fail to elongate until 90 h APF (G, H). Scale bars are 5 μm.

I Quantification of sarcomere length in wild-type (blue) and *aret-IR* (red), ***$P < 0.001$, unpaired Student's *t*-test. Standard deviation is shown.

repression occurs in the cytoplasm. However, it has been shown that upon a block of mRNA export during oogenesis or upon overexpression of Aret (Bruno) protein, Aret can be found in the nurse cell nuclei. This nuclear localization is more pronounced when the RNA-binding motifs were mutated [15]. This is consistent with our observations suggesting that Aret can shuttle between cytoplasm and nucleus not only in oocytes, but also in flight muscles. The cytoplasmic function of Aret in IFMs, if any, remains to be determined.

Proteins containing RNA recognition motif (RRM) domains are found frequently in the genome, with more than 250 examples in *Drosophila* [15,27]. Aret contains 3 RRMs, 2 N-terminal and 1 more C-terminal, an organization shared with the Elav family of proteins. *Drosophila* Elav is a well-established splicing factor that uses its RRMs to regulate mRNA splicing in neurons [28,29]. A role for Aret in regulating splicing in *Drosophila* was not known prior to this work. However, while this manuscript was in the final phase of

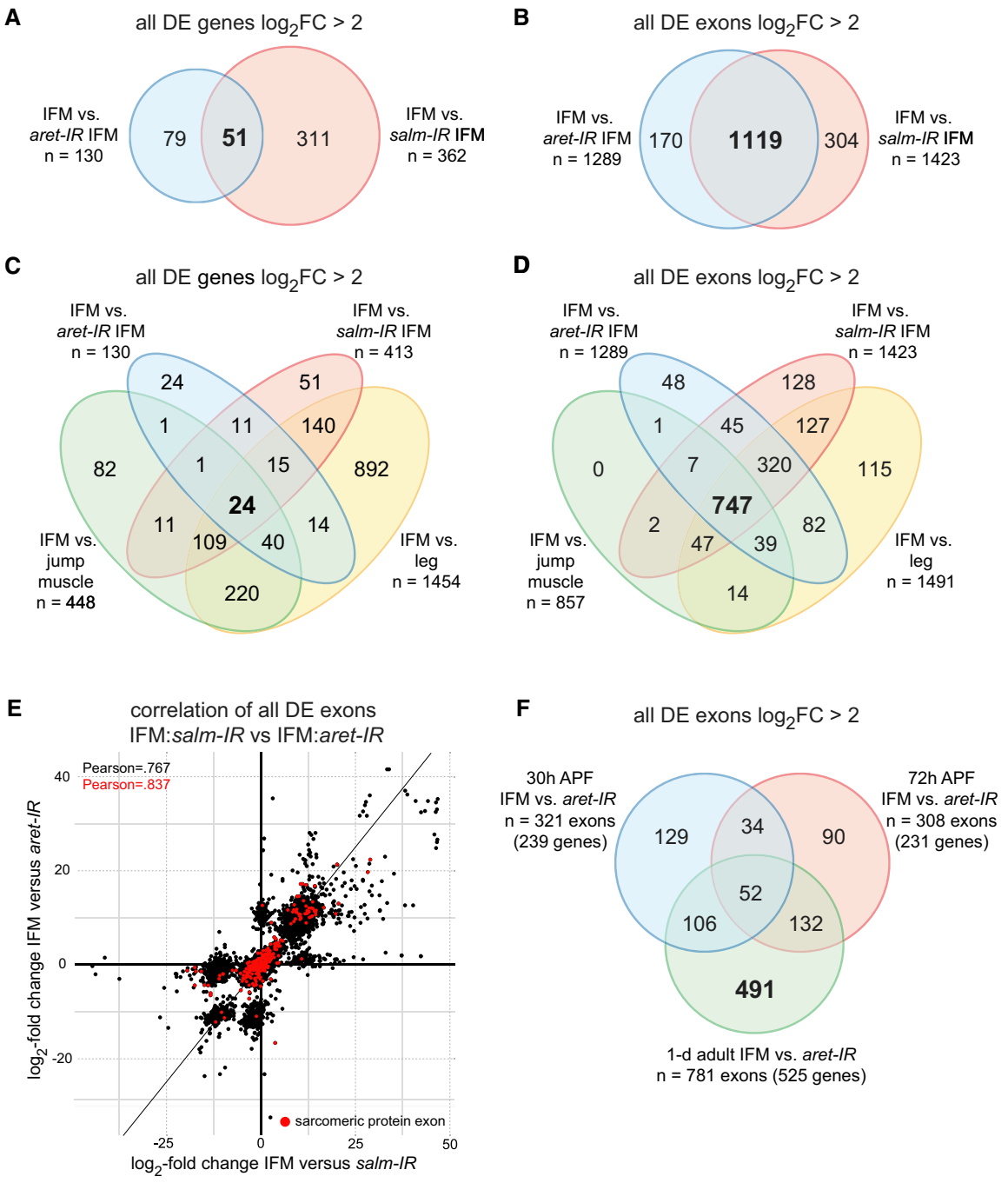

Figure 7. *aret* regulates fibrillar muscle-specific alternative splicing.

A Venn diagram comparing significantly differentially expressed (*P*-value < 0.05, DESeq2) genes with a $\log_2$FC > 2 between *aret-IR* and *salm-IR*. 51 genes are co-regulated by Salm and Aret.

B Venn diagram comparing significantly differentially expressed (*P*-value < 0.05, DEXSeq) exons with $\log_2$FC > 2 between *aret-IR* and *salm-IR*. Note that expression of 78.6% (1119/1423) of Salm-dependent exons is also Aret dependent.

C Venn diagram comparing significantly differentially expressed (*P*-value < 0.05, DESeq2) genes with a $\log_2$FC > 2 between IFM:*aret-IR*, IFM:*salm-IR*, IFM:leg, and IFM:jump muscle. Only 24 fibrillar-specific genes are co-regulated by Salm and Aret.

D Venn diagram comparing significantly differentially expressed (*P*-value < 0.05, DEXSeq) exons with $\log_2$FC > 2 between IFM:*aret-IR*, IFM:*salm-IR*, IFM:leg, and IFM:jump muscle. 747 fibrillar-specific exons co-depend on Aret and Salm.

E Correlation plot of the $\log_2$FC IFM:*salm-IR* versus IFM:*aret-IR*. All significantly differentially expressed exons (n = 5939, *P*-value < 0.05, DEXSeq) are plotted in black, while sarcomeric protein exons are plotted in red. Pearson's correlation coefficients for all exons (black) and the sarcomeric exons subset (red) are indicated. Note that many exons are co-regulated by both Salm and Aret and that Aret promotes both inclusion and exclusion of exons.

F Venn diagram comparing significantly differentially expressed (*P*-value < 0.05, DEXSeq) exons with $\log_2$FC > 2 between WT IFM:*aret-IR* IFM at 30 h APF, 72 h APF and in 1-d adults. Notably, 781 exons (491 uniquely) are regulated at the adult time point, while only ~300 exons are regulated at each developmental time point.

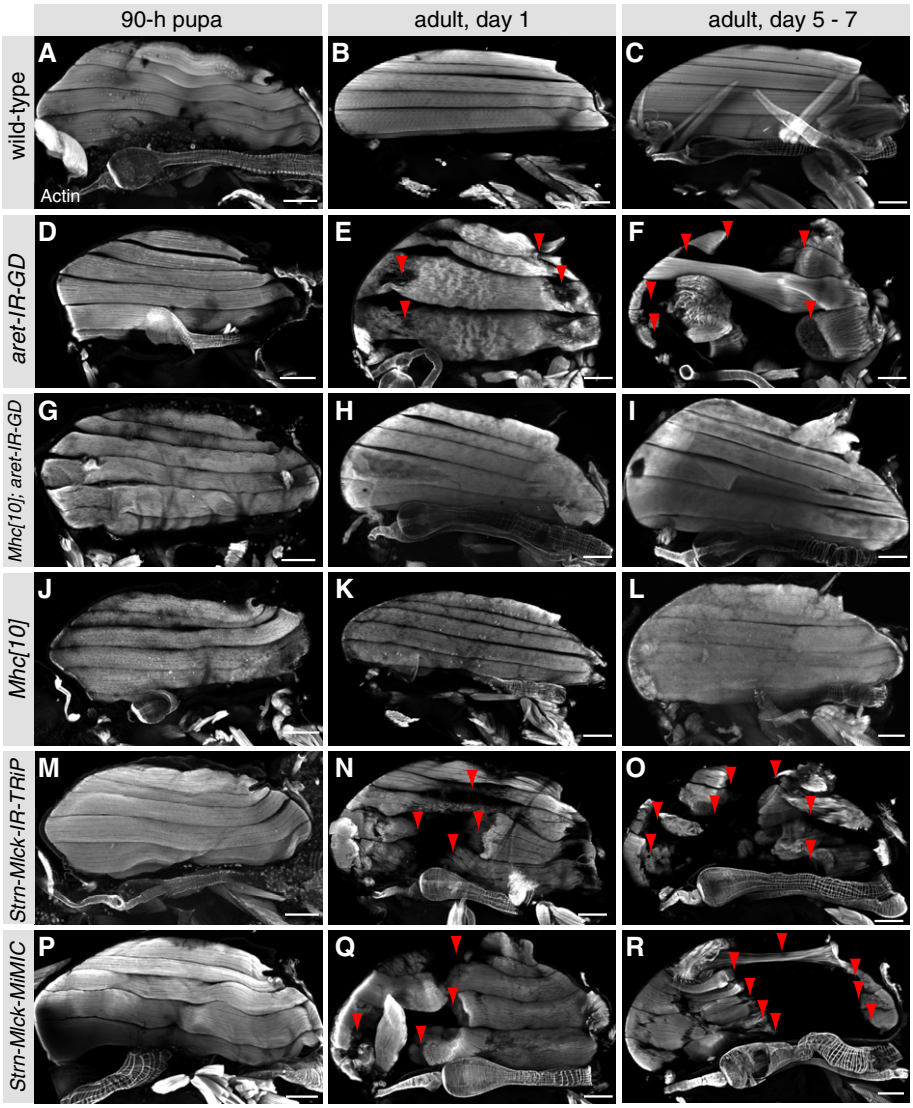

Figure 8. Age-dependent *aret-IR* fiber degeneration is caused by hyper-contraction.

A–L Hemithoraces of 90 h APF pupae (A, D, G, J), 1-day adults (B, E, H, K) and 5–7 days aged adults (C, F, I, L). Wild-type IFM fibers remain intact in aged adults (A–C), whereas *aret-IR* fibers are successively ruptured and lost (D–F, red arrow heads). This fiber loss is entirely suppressed upon removal of *Mhc* function from *aret-IR* IFMs using the *Mhc [10]* allele (G–L). Scale bars are 100 μm.

M–R Hemithoraces of IFM-specific *Strn-Mlck* knock down 90 h APF pupae (M), 1 day (N) and 5–7 days adults (O) and *Strn-Mlck MiMIC* $^{[MI02893]}$ 90 h APF (P), 1 day (Q) and 5–7 days adults (R). Note the progressive fiber degeneration upon aging (M–R, red arrow heads). Scale bars are 100 μm.

preparation, a parallel study showed that Aret regulates splicing of *sls*, *wupA*, and *ZASP52* in IFMs. Additionally, it can instruct the fibrillar splicing mode if expressed ectopically in tubular muscle or in S2 cells, suggesting that it regulates the splicing machinery directly [30].

In vertebrates, alternative splicing is also a prominent feature of different muscle types [31]. In particular in the heart, which shares some similarities with insect flight muscle, alternative splicing is very distinct to skeletal muscle and is one important mechanism to control the different physiological properties of both tissues. RBM20 regulates heart-specific splicing of titin by promoting exon skipping of the flexible PEVK exons in titin [32]. This is functionally important as human patients with a mutation in RBM20 suffer from hereditary cardiomyopathies [33]. A similar role for muscle-type splicing in heart and skeletal muscle was recently identified for RBM24 [34], highlighting the importance of muscle-type-specific splice regulation. While both RBM20 and RBM24 contain only a single RRM domain, the mammalian homologs of Aret called CELF 1–6 (CUGBP, Elav-like family) contain 3 RRMs with a similar spacing as in Aret. Interestingly, they have been implicated in regulating alternative splicing in various tissues including splicing of troponin T in the heart [35]. However, the presence of multiple genes makes genetic analysis difficult. This indicates that the mechanism of Aret-mediated alternative splicing is conserved to mammals, suggesting that insights gained in *Drosophila* will also be applicable to vertebrate muscle biology.

Materials and Methods

All fly work was performed at 27°C to enhance GAL4 activity. Immunostainings were performed using standard protocols [36]. All antibodies and fly stocks are listed in figure legends and in the Supplementary Information. Fosmids tagged with GFP were generated similarly as in previous studies [11,12] and will be published in detail elsewhere. All fosmids used in this study are listed in the Supplementary Information. Sarcomere length was quantified based on phalloidin staining in Fiji (Image J) and significance evaluated with unpaired Student's *t*-tests.

For the mRNA-Seq analysis, IFMs, jump muscles and entire legs marked with *Mef2-GAL4, UAS-GFP-Gma* were dissected at the indicated time points. RNA samples were prepared and processed based on a published protocol [37]. Briefly, total RNA was isolated with Tri-Pure reagent (Roche), mRNA selected over oligo-dT beads (Invitrogen), fragmented with peak length ~300 bp, reverse-transcribed with the Invitrogen SuperScript-III kit and dUTP labeled during second-strand synthesis. Libraries were prepared and sequenced according to standard Illumina protocols. RNA sequencing (RNA-seq) was performed at the CSF Next-Generation Sequencing Unit (http://csf.ac.at). Reads were filtered and trimmed using the FASTX Toolkit and cutadapt and mapped to the Ensembl BDGP5.25 genome assembly using Tophat v2.0. Reads were visualized on the UCSC server by normalizing to the largest library size (Supplementary Table S3). Libraries were evaluated with featureCounts v1.4.2, and differential expression analysis was performed on the gene level with DESeq2 and on the exon/isoform level with DEXSeq. Additional data processing was handled in R. GO analysis was performed with GOrilla [38] and REVIGO [39]. Additional details can be found in the Supplementary Information.

Data availability

Three supplementary datasets are provided listing: (1) all genes that are significantly differentially expressed in the DESeq2 comparison of IFM, leg muscle, jump muscle, *salm-IR* IFM, and *aret-IR* IFM (Supplementary Raw Data S1); (2) all exons that are significantly differentially expressed in the DEXSeq comparison of IFM, leg muscle, jump muscle, *salm-IR* IFM, and *aret-IR* IFM (Supplementary Raw Data S2); (3) all exons that are significantly differentially expressed in the DEXSeq comparison of IFM to *aret-IR* IFM at 30 h APF, 72 h APF and 1-d adults (Supplementary Raw Data S3). mRNA-Seq data are publicly available from NCBI's Gene Expression Omnibus (GEO) under accession number GSE63707. Individual libraries are available from the Sequence Read Archive (SRA) under accession numbers GSM1555978–GSM1555995.

Supplementary information for this article is available online: http://embor.embopress.org

Acknowledgements

We thank Anne Ephrussi for generously sharing *aret* alleles and Aret antibodies and the Bloomington and VDRC stock centers for fly stocks. We are grateful to Reinhard Fässler for generous support and to Bettina Stender for excellent technical assistance. Our work was supported by the Max Planck Society, Humboldt, EMBO long-term (688-2011), and NIH-NRSA (5F32AR062477) post-doctoral fellowships (M.L.S.); a Career Development Award from the Human Frontier Science Program (F.S.), the EMBO Young Investigator Program (A.S., F.S.) and the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC Grant 310939.

Author contributions

MLS performed most of the experiments with important support by CB who also generated the *salm*^{FRT} allele. MLS, AY, DG, AS, and BHH performed the bioinformatic data analysis. CS contributed to *aret* and *Strn-Mlck* phenotypic analysis. IF and MS generated the GFP fosmid clones. FS conceived and supervised the project; MLS and FS made the figures and wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

1. Schiaffino S, Reggiani C (2011) Fiber types in mammalian skeletal muscles. *Physiol Rev* 91: 1447 – 1531
2. Linke WA, Kulke M, Li H, Fujita-Becker S, Neagoe C, Manstein DJ, Gautel M, Fernandez JM (2002) PEVK domain of titin: an entropic spring with actin-binding properties. *J Struct Biol* 137: 194 – 205
3. Guo W, Bharmal SJ, Esbona K, Greaser ML (2010) Titin diversity – alternative splicing gone wild. *J Biomed Biotechnol* 2010: 1 – 8
4. Lehmann F, Dickinson M (1997) The changes in power requirements and muscle efficiency during elevated force production in the fruit fly *Drosophila melanogaster*. *J Exp Biol* 200: 1133
5. Josephson R (2006) Comparative physiology of insect flight muscle. In *Nature's Versatile Engine: Insect Flight Muscle Inside and Out*, Vigoreaux J (ed.), pp 35 – 43. Georgetown, TX: Landes Bioscience.
6. Josephson RK, Malamud JG, Stokes DR (2000) Asynchronous muscle: a primer. *J Exp Biol* 203: 2713 – 2722
7. Schönbauer C, Distler J, Jährling N, Radolf M, Dodt H-U, Frasch M, Schnorrer F (2011) Spalt mediates an evolutionarily conserved switch to fibrillar muscle fate in insects. *Nature* 479: 406 – 409
8. Love MI, Huber W, Anders S (2014) Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. *Genome Biology* 15: 550
9. Venables JP, Tazi J, Juge F (2011) Regulated functional alternative splicing in Drosophila. *Nucleic Acids Res* 40: 1 – 10
10. Anders S, Reyes A, Huber W (2012) Detecting differential usage of exons from RNA-seq data. *Genome Res* 22: 2008 – 2017
11. Ejsmont R, Sarov M, Winkler S, Lipinski K, Tomancak P (2009) A toolkit for high-throughput, cross-species gene engineering in *Drosophila*. *Nat Methods* 6: 435 – 437
12. Sarov M, Murray JI, Schanze K, Pozniakovski A, Niu W, Angermann K, Hasse S, Rupprecht M, Vinis E, Tinney M *et al* (2012) A genome-scale resource for in vivo tag-based protein function exploration in *C. elegans*. *Cell* 150: 855 – 866
13. Barthmaier P, Fyrberg E (1995) Monitoring development and pathology of *Drosophila* indirect flight muscles using green fluorescent protein. *Dev Biol* 169: 770 – 774
14. Clark KA, Bland JM, Beckerle MC (2007) The *Drosophila* muscle LIM protein, Mlp84B, cooperates with D-titin to maintain muscle structural integrity. *J Cell Sci* 120: 2066 – 2077
15. Snee M, Benz D, Jen J, Macdonald PM (2008) Two distinct domains of Bruno bind specifically to the oskar mRNA. *RNA Biol* 5: 1 – 9
16. Schnorrer F, Schönbauer C, Langer CCH, Dietzl G, Novatchkova M, Schernhuber K, Fellner M, Azaryan A, Radolf M, Stark A *et al* (2010)

Systematic genetic analysis of muscle morphogenesis and function in *Drosophila*. *Nature* 464: 287 – 291

17. Schüpbach T, Wieschaus E (1991) Female sterile mutations on the second chromosome of *Drosophila melanogaster*. II. Mutations blocking oogenesis or altering egg morphology. *Genetics* 129: 1119 – 1136
18. Kim-Ha J, Kerr K, Macdonald PM (1995) Translational regulation of oskar mRNA by bruno, an ovarian RNA-binding protein, is essential. *Cell* 81: 403 – 412
19. Fernandes J, Bate M, VijayRaghavan K (1991) Development of the indirect flight muscles of *Drosophila*. *Development* 113: 67 – 77
20. Weitkunat M, Kaya-Copur A, Grill SW, Schnorrer F (2014) Tension and force-resistant attachment are essential for myofibrillogenesis in *Drosophila* flight muscle. *Curr Biol* 24: 705 – 716
21. Barbas JA, Galceran J, Torroja L, Prado A, Ferrús A (1993) Abnormal muscle development in the heldup3 mutant of *Drosophila melanogaster* is caused by a splicing defect affecting selected troponin I isoforms. *Mol Cell Biol* 13: 1433 – 1439
22. Nongthomba U, Clark S, Cummins M, Ansari M, Stark M, Sparrow JC (2004) Troponin I is required for myofibrillogenesis and sarcomere formation in *Drosophila* flight muscle. *J Cell Sci* 117: 1795 – 1805
23. Nongthomba U, Ansari M, Thimmaiya D, Stark M, Sparrow J (2007) Aberrant splicing of an alternative exon in the *Drosophila* troponin-T gene affects flight muscle development. *Genetics* 177: 295 – 306
24. Webster PJ, Liang L, Berg CA, Lasko P, Macdonald PM (1997) Translational repressor bruno plays multiple roles in development and is widely conserved. *Genes Dev* 11: 2510 – 2521
25. Chekulaeva M, Hentze MW, Ephrussi A (2006) Bruno acts as a dual repressor of oskar translation, promoting mRNA oligomerization and formation of silencing particles. *Cell* 124: 521 – 533
26. Filardo P, Ephrussi A (2003) Bruno regulates gurken during *Drosophila oogenesis*. *Mech Dev* 120: 289 – 297
27. Maris C, Dominguez C, Allain FHT (2005) The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression. *FEBS J* 272: 2118 – 2131
28. Soller M, White K (2003) ELAV inhibits 3′-end processing to promote neural splicing of ewg pre-mRNA. *Genes Dev* 17: 2526 – 2538
29. Soller M, White K (2005) ELAV multimerizes on conserved AU4-6 motifs important for ewg splicing regulation. *Mol Cell Biol* 25: 7580
30. Oas ST, Bryantsev AL, Cripps RM (2014) Arrest is a regulator of fiber-specific alternative splicing in the indirect flight muscles of *Drosophila*. *J Cell Biol* 206: 895 – 908
31. Spletter ML, Schnorrer F (2014) Transcriptional regulation and alternative splicing cooperate in muscle fiber-type specification in flies and mammals. *Exp Cell Res* 321: 90 – 98
32. Li S, Guo W, Dewey CN, Greaser ML (2013) Rbm20 regulates titin alternative splicing as a splicing repressor. *Nucleic Acids Res* 41: 2659 – 2672
33. Guo W, Schafer S, Greaser ML, Radke MH, Liss M, Govindarajan T, Maatz H, Schulz H, Li S, Parrish AM *et al* (2012) RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat Med* 18: 766 – 773
34. Yang J, Hung L-H, Licht T, Kostin S, Looso M, Khrameeva E, Bindereif A, Schneider A, Braun T (2014) RBM24 is a major regulator of muscle-specific alternative splicing. *Dev Cell* 31: 87 – 99
35. Barreau C, Paillard L, Méreau A, Osborne HB (2006) Mammalian CELF/Bruno-like RNA-binding proteins: molecular characteristics and biological functions. *Biochimie* 88: 515 – 525
36. Weitkunat M, Schnorrer F (2014) A guide to study *Drosophila* muscle biology. *Methods* 68: 2 – 14
37. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* 5: 621 – 628
38. Eden E, Lipson D, Yogev S, Yakhini Z (2007) Discovering motifs in ranked lists of DNA sequences. *PLoS Comput Biol* 3: e39
39. Supek F, Bošnjak M, Škunca N, Šmuc T (2011) REVIGO summarizes and visualizes long lists of gene ontology terms. *PLoS One* 6: e21800