NineTeen Complex-subunit Salsa is required for efficient splicing of a subset of introns and dorsal–ventral patterning

OM SINGH RATHORE,[1] RUI D. SILVA,[1] MARIANA ASCENSÃO-FERREIRA,[2] RICARDO MATOS,[1] CÉLIA CARVALHO,[2] BRUNO MARQUES,[1] MARGARIDA N. TIAGO,[1] PEDRO PRUDÊNCIO,[1,2] RAQUEL P. ANDRADE,[1,3] JEAN-YVES ROIGNANT,[4,5] NUNO L. BARBOSA-MORAIS,[2] and RUI GONÇALO MARTINHO[1,2,6]

[1]Center for Biomedical Research (CBMR), Universidade do Algarve, Faro, 8005-139 Portugal
[2]Instituto de Medicina Molecular João Lobo Antunes, Faculdade de Medicina, Universidade de Lisboa, 1649-028 Lisboa, Portugal
[3]Department of Medicine and Biomedical Sciences and Algarve Biomedical Center, Universidade do Algarve, 8005-139 Faro, Portugal
[4]Center for Integrative Genomics, Faculty of Biology and Medicine, University of Lausanne, CH-1015 Lausanne, Switzerland
[5]Institute of Pharmaceutical and Biomedical Sciences, Johannes Gutenberg-University Mainz, 55128 Mainz, Germany
[6]Department of Medical Sciences and Institute for Biomedicine (iBiMED), Universidade de Aveiro, 3810-193 Aveiro, Portugal

ABSTRACT

The NineTeen Complex (NTC), also known as pre-mRNA-processing factor 19 (Prp19) complex, regulates distinct spliceosome conformational changes necessary for splicing. During *Drosophila* midblastula transition, splicing is particularly sensitive to mutations in NTC-subunit Fandango, which suggests differential requirements of NTC during development. We show that NTC-subunit Salsa, the *Drosophila* ortholog of human RNA helicase Aquarius, is rate-limiting for splicing of a subset of small first introns during oogenesis, including the first intron of *gurken*. Germline depletion of Salsa and splice site mutations within *gurken* first intron impair both adult female fertility and oocyte dorsal–ventral patterning, due to an abnormal expression of Gurken. Supporting causality, the fertility and dorsal–ventral patterning defects observed after Salsa depletion could be suppressed by the expression of a *gurken* construct without its first intron. Altogether, our results suggest that one of the key rate-limiting functions of Salsa during oogenesis is to ensure the correct expression and efficient splicing of the first intron of *gurken* mRNA. Retention of *gurken* first intron compromises the function of this gene most likely because it undermines the correct structure and function of the transcript 5′UTR.

Keywords: splicing; *Drosophila*; female fertility; dorsal–ventral patterning; Gurken

INTRODUCTION

The spliceosome is a highly dynamic molecular machine, composed of five small nuclear ribonucleoproteins (snRNPs) that sequentially associate to the precursor mRNA (pre-mRNA) during the splicing reaction (Martinho et al. 2015; Papasaikas and Valcarcel 2016). Each snRNP (U1, U2, U4, U5, and U6) contains a U-rich snRNA and a unique group of proteins. Although spliceosome assembly is ordered (U1 > U2 > U4/U5/U6 > NineTeen Complex), the splicing reaction is without an apparent irreversible and/or rate-limiting step (Hoskins et al. 2011), with commitment to splicing progressively increased as snRNPs and NTC bind to the pre-mRNA (Hoskins et al. 2011).

The spliceosomal NineTeen Complex (NTC), also known as Pre-mRNA-processing factor 19 (Prp19) complex, regulates distinct spliceosome conformational changes necessary for efficient pre-mRNA splicing (Hogg et al. 2010; Chanarat and Strasser 2013). NTC composition is dynamic and comprises a subset of conserved core subunits and many transiently associated ones (Hogg et al. 2010). NTC associates with the spliceosome during its activation and just before the first transesterification (Hogg et al. 2010). Interestingly, NTC also has a significant role in the crosstalk between transcription, cotranscriptional processing of the nascent RNA, and DNA repair, as distinct NTC subunits have been reported to be important for

Corresponding authors: rgmartinho@ua.pt, nmorais@medicina.ulisboa.pt
Article is online at http://www.rnajournal.org/cgi/doi/10.1261/rna.077446.120.

© 2020 Rathore et al. This article is distributed exclusively by the RNA Society for the first 12 months after the full-issue publication date (see http://rnajournal.cshlp.org/site/misc/terms.xhtml). After 12 months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

transcriptional elongation and genomic stability (Chanarat and Strasser 2013; Martinho et al. 2015; Mahajan 2016). Human NTC-subunits PRP19, XAB2, and CDC5L are important for transcriptional elongation, transcription-coupled DNA repair, and activation of the ATM-related (ATR)-dependent DNA damage checkpoint (Kuraoka et al. 2008; Marechal et al. 2014; Wan and Huang 2014). RNA Polymerase II (RNA Pol II) also promotes cotranscriptional splicing activation through the recruitment of NTC (David et al. 2011).

Human Aquarius (AQR) (also known as intron-binding protein 160, IBP160) is an ATP-dependent RNA helicase that associates with NTC during spliceosome activation and formation of the activated B complex (B^{ACT}) (Hirose et al. 2006; De et al. 2015; Haselbach et al. 2018). AQR binds to introns independently of sequence, but usually upstream of the branch-site (BS) and close to the associated U2 snRNP SF3a and SF3b proteins, being essential for intron-binding complex formation and efficient splicing (Hirose et al. 2006; De et al. 2015). AQR has also been suggested to be important for deposition of the exon junction complex (EJC) during the splicing reaction (Ideue et al. 2007) and formation of intron-encoded snoRNAs (Hirose et al. 2006), suggesting it regulates the cross-talk between splicing and other RNA processing events.

Splicing during *Drosophila* early embryonic development is notably sensitive to mutations in NTC-subunit Fandango (Guilgur et al. 2014), suggesting differential requirements of NTC during development (Martinho et al. 2015). To test this possibility, we decided to investigate the role of other NTC-subunits during *Drosophila* oogenesis and early embryonic development. We focused our initial work on uncharacterized gene *CG31368*, which encodes the *Drosophila* ortholog of human Aquarius (Herold et al. 2009; De et al. 2015). Since there is already a nonrelated *Drosophila* protease named *aquarius* (*CG14061*), we decided to rename *CG31368* as *salsa*. Our working hypothesis is that *salsa*, similar to its *Caenorhabditis elegans* ortholog *emb-4* (Akay et al. 2017), is likely to have important developmental functions.

During *Drosophila* oogenesis, *gurken* mRNA localizes to the posterior cortex of the developing oocyte and Gurken signal is restricted to the underlying posterior follicle cells (Schupbach 1987; Neuman-Silberberg and Schupbach 1993; Gonzalez-Reyes et al. 1995; Roth et al. 1995). In response to a signal from the posterior follicle cells, there is a considerable reorganization of the cytoskeleton and a microtubule-dependent migration of the oocyte nucleus to the anterior cortex (Gonzalez-Reyes and St Johnston 1994; Gonzalez-Reyes et al. 1995; Roth et al. 1995; Zhao et al. 2012). The anteriorly localized nucleus defines the dorsal-anterior region and provides the first detectable dorsal–ventral (D/V) asymmetry of the oocyte, with the expression of both *gurken* mRNA and protein restricted to the cytoplasmic perinuclear region of the oocyte (Neuman-Silberberg and Schupbach 1993; Gonzalez-Reyes et al. 1995; Roth et al. 1995).

gurken mRNA is transcribed in the supporting nurse cells (Caceres and Nilson 2005) and actively transported to the dorsal-anterior region of the oocyte by a dynein-mediated transport (MacDougall et al. 2003; Delanoue et al. 2007). The oocyte dorsal-anterior localization of *gurken* mRNA relies on multiple elements localized to the transcript 5′UTR, 3′UTR and open-reading frame (Saunders and Cohen 1999; Thio et al. 2000; Van De Bor et al. 2005). Although this localization is crucial for its efficient translation (Neuman-Silberberg and Schupbach 1993; Roth et al. 1995; Davidson et al. 2016), the precise contribution of each element for RNA localization is still a matter of debate.

D/V patterning of the developing *Drosophila* egg is dependent on the dorsal-anterior localization of Gurken during mid-oogenesis (Neuman-Silberberg and Schupbach 1993; Roth et al. 1995; Huynh and St Johnston 2004). Gurken is the ligand for the Epidermal growth factor receptor (Egfr) that locates to the apical surface of follicle cells that surround the developing oocyte (Schupbach 1987; Neuman-Silberberg and Schupbach 1993; Gonzalez-Reyes et al. 1995; Roth et al. 1995). Activation of Egfr modifies the cell fate of the dorsal follicle cells and restricts the formation of Spätzle ligand to the ventral region of the oocyte (Morisato and Anderson 1994; Schneider et al. 1994), which is essential for normal morphogenesis of the eggshell dorsal appendages.

Here we found that Salsa, the *Drosophila* ortholog of AQR, is rate-limiting for efficient splicing of a subset of small first introns, including the first intron of *gurken*. Consistent with the functional relevance of *gurken* splicing defects, mutations within the splice sites of the first intron of *gurken* impair the function of this gene. Female germline depletion of Salsa and splice mutations within *gurken* first intron were both associated to a decrease in female fertility, significant D/V patterning defects of the eggshell and abnormal expression of Gurken during oogenesis. Supporting causality, expression of a *gurken* construct without its first intron suppressed the female fertility and D/V patterning defects observed after Salsa depletion. Altogether our results suggest that one of the key rate-limiting functions of Salsa during oogenesis is to ensure the correct expression and efficient splicing of the first intron of *gurken* mRNA.

RESULTS

Salsa is required for splicing of a subset of introns

Splicing during *Drosophila* early embryonic development is particularly sensitive to mutations in NTC-subunit Fandango (Guilgur et al. 2014), suggesting differential requirements of NTC during development (Martinho et al.

2015). To test this possibility, we decided to investigate the role of other NTC-subunits during *Drosophila* oogenesis and early embryonic development. We focused our initial work on the gene *CG31368*. Bidirectional blast analysis showed that *CG31368* is the ortholog of human Aquarius (AQR) (Herold et al. 2009) (Flybase). CG31368 physically interacts with NTC-subunits Fandango and Prp19 (Guilgur et al. 2014), and this interaction is independent of RNA (Supplemental Fig. S1). Since there is already a nonrelated *Drosophila* protease named Aquarius, and given its interaction with Fandango, we decided to rename *CG31368* as *salsa*.

In order to address the effect of Salsa on splicing we performed transcriptome analyses of control and Salsa-depleted ovaries. To deplete Salsa specifically within the female germline, the *nanos*-Gal4 (*nos*-Gal4) germline driver (Van Doren et al. 1998) and the upstream activating sequence (UAS)/Gal4 system (Brand and Perrimon 1993; Rorth 1998; Ni et al. 2011) were used to express short hairpin RNAs (shRNAs) against *salsa*. For control we used a short hairpin against mCherry. We focused our analysis on introns with good sequencing coverage and performed differential intron retention analysis between control and Salsa-depleted samples (for more experimental detail see Materials and Methods). A *salsa* RNAi hairpin (*salsa* RNAi-1) with a mild dorsal–ventral (D/V) patterning defect but without detectable egg chamber necrosis was used for this study in order to minimize the identification of introns whose splicing was nonspecifically affected by Salsa depletion.

Genome-wide analysis of the splicing defects observed after germline-specific depletion of Salsa showed that this RNA helicase is required for splicing of a small subset of introns, being the large majority of alternative splicing defects related to increased levels of intron retention (IR) upon Salsa depletion (Supplemental Fig. S2). Considering as biologically relevant a difference in percent of intron retention (PIR) of at least 10% (see Materials and Methods), we identified a significant enrichment for increased intron retention, with 47 introns increasingly retained and only 6 introns less retained in Salsa-depleted samples (red/purple and yellow points in Fig. 1A; Supplemental Fig. S3A, respectively) (Supplemental Table S1). This proportion (47/53; 0.89) is significantly different from 0.5 (proportion's test *P*-value = 1.78×10^{-8}). The fact that most identified cases were related with increased levels of intron retention is consistent with the recently proposed function of human Aquarius before splicing catalysis for increased pre-mRNA splicing efficiency (De et al. 2015).

Analysis of the introns that were more retained after Salsa depletion showed that this RNA helicase is required for efficient splicing, among others, of the first introns of the dorsal–ventral patterning gene *gurken* (Fig. 1B), the sex-determining gene *transformer 2* (*tra2*) (Supplemental Fig. S3B), *suppressor of variegation 205* (*Su[var205]*), *phosphoribosylamidotransferase* (*Prat*), *CG18428* and *multiprotein bridging factor 1* (*mbf1*), the second 5′UTR-localized intron of *CG7168*, the fourth *CG31211* intron, the third *CG5746* intron and the fourth *baboon* (*babo*) intron (red points in Fig. 1A; Supplemental Fig. S3A; Supplemental Table S1).

Salsa is required for splicing of small first introns

Since gene architecture influences splicing kinetics (Sterner et al. 1996; Fox-Walsh et al. 2005; Pai et al. 2017) and given the fact that contrary to the first intron, splicing of the second and third introns of *gurken* was not affected by Salsa depletion (Fig. 1B), we decided to investigate if there was any particular feature significantly associated to introns whose splicing was sensitive to Salsa depletion. A comparison between differentially retained introns after Salsa depletion detected a significant bias for small proximal introns, as the affected introns had a higher probability of being first introns (28/47; 0.6) compared to control (62/169; 0.37, proportion's test *P*-value = 0.0011) (Fig. 2A; Supplemental Fig. S4) or to commonly skipped introns (636/4078; 0.16, proportion's test *P*-value $<2.2 \times 10^{-16}$) (Fig. 2A; Supplemental Fig. S4), and were typically smaller than control introns (Wilcoxon's rank sum test with continuity correction *P*-value = 0.0046) (Fig. 2B,C; Supplemental Fig. S5A).

Recruitment of U1 to the proximal region of nascent pre-mRNAs is facilitated by interaction with the cap-binding complex (CBC) (Izaurralde et al. 1994; Lewis et al. 1996; Gornemann et al. 2005; Pabis et al. 2013), which protects against premature polyadenylation (Ashe et al. 1997; Gunderson et al. 1998; Guo et al. 2011) and facilitates splicing of 5′end localized introns (Sakurai et al. 2002; Lin and Zhang 2005). The effect of CBC on splicing enhancement is distance dependent (Sakurai et al. 2002), with a significant positive correlation between transcription start site (TSS) distance and 5′splice site (5′ss) strength (Lepennetier and Catania 2016). Consistent with the hypothesis that Salsa is rate-limiting for CBC-dependent splicing, alternatively spliced introns with increased retention after Salsa depletion showed a bias for shorter distances between the transcription start site (TSS) and their 5′ss (average of 1060 base pairs) when compared to control introns (average of 3632 base pairs, Wilcoxon's *P*-value = 0.0014) (Fig. 2D). Nevertheless, since this distance bias was not significant among first introns themselves (Fig. 2E; Supplemental Fig. S5B), suggesting it is mostly related with a first intron position bias and not the distance per se, and since there was no significant bias for weaker 5′splice sites among the affected introns (Supplemental Fig. S6A), it is unlikely that the observed biases are related with CBC-dependent splicing. Interestingly, whereas affected introns showed a bias to weaker 3′splice sites when compared to control introns (Wilcoxon's *P*-value = 0.018) (Fig. 2F,G;

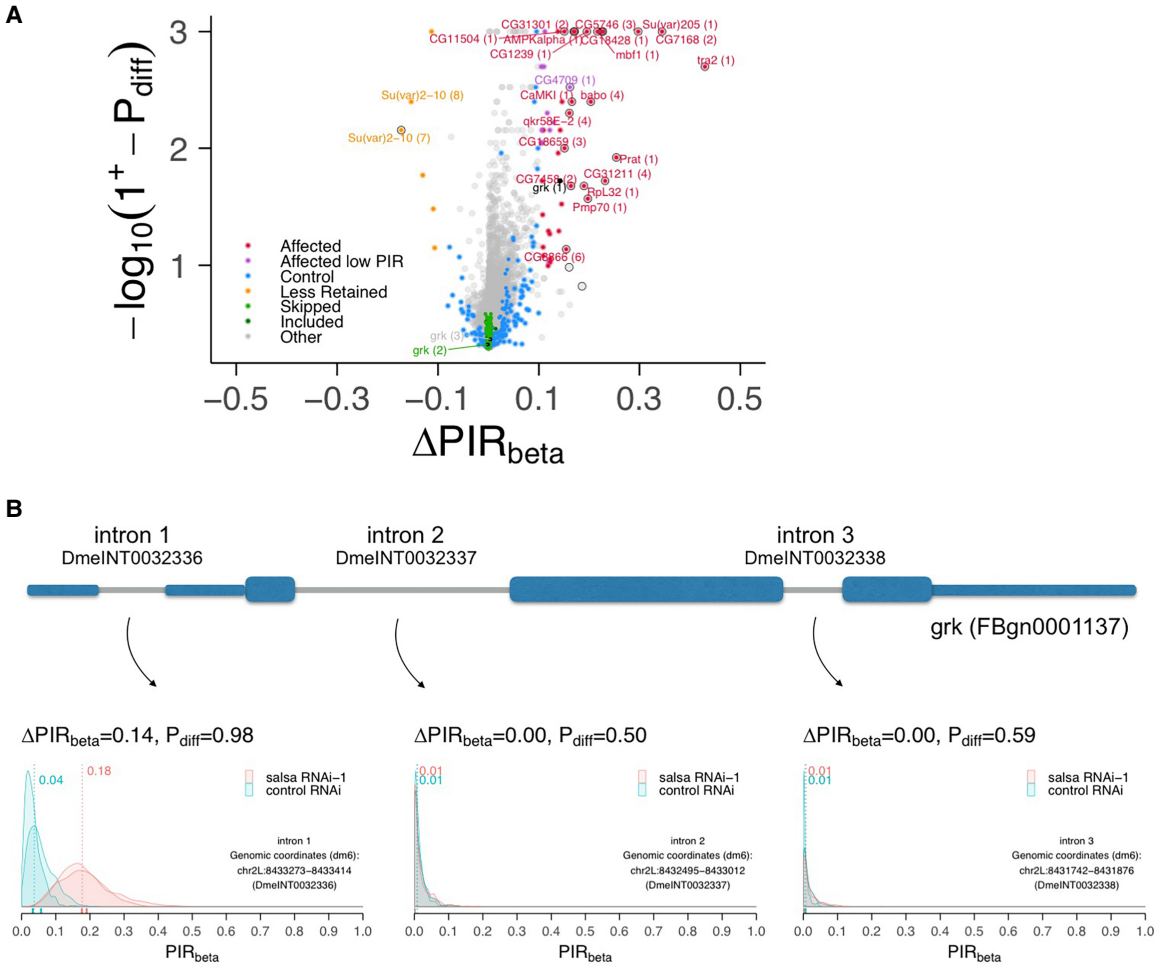

FIGURE 1. A subset of introns shows increased levels of retention (IR) upon Salsa depletion. (*A*) Volcano plot for beta distribution-based differential intron retention analysis, with differences in inclusion levels between Salsa-depleted (*salsa* RNAi-1) and control samples (mCherry RNAi) ($\Delta PIR_{beta} = PIR_{salsa\ RNAi} - PIR_{control}$) in the *x*-axis and probability of differential retention [$-\log_{10}(1^+ - P_{diff})$] in the *y*-axis, with $1^+ = 1.001$ used to avoid infinite values when $P_{diff} = 1$. Differentially retained introns (*Affected*, *Affected low PIR*, and *Less Retained*), identified by $P_{diff} \geq 0.9$, are highlighted in red, purple and yellow, respectively. Events with an absolute ΔPIR_{beta} greater than 0.15 are labeled with gene name and the transcript's intron position. *Control*, *Included*, and *Skipped* introns are also highlighted for comparison, as well as *gurken* intron retention events. Differential retention of the first intron of *gurken* is shown in black as this intron was filtered out of the global differential intron retention analysis due to insufficient coverage in one of the samples (see Materials and Methods). (*B*) Diagram of simplified *gurken* (*grk*) gene structure with exons (blue) and introns (gray) and density plots (smoothed histograms) reflecting points emitted from beta distributions used to model *gurken* intron 1 (*left* panel), 2 (*middle*), and 3 (*right*) retention for each sample and PIR_{beta} estimates for Salsa-depleted and control conditions (dashed vertical lines). Rug plots *below* the density curves reflect PIR values per sample. Event identifiers and genomic coordinates of the respective alternative sequences are part of the VAST-DB dm6 annotation (Tapial et al. 2017).

Supplemental Fig. S5C), there was, however, no bias for higher GC content (Supplemental Fig. S6B,C). Altogether our data suggests that Salsa is required for efficient splicing of first introns, typically smaller and with weaker 3′splice sites, but independent of TSS distance.

Salsa regulates the expression levels of a very small number of genes

Human NTC-subunit Xab2, the ortholog of *Drosophila* Fandango (Guilgur et al. 2014), has been reported to be rate-limiting, besides splicing, for transcription and transcription-coupled repair (Kuraoka et al. 2008). Since Salsa physically interacts with Fandango (Guilgur et al. 2014), it was possible that transcription would be similarly affected in Salsa-depleted ovaries. Differential gene expression analysis revealed that Salsa is significantly rate-limiting for the expression levels of a very small number of genes (Supplemental Fig. S7A). Salsa positively regulates the expression levels of *CG4570* and *Activity-regulated cytoskeleton associated protein 1* (*Arc1*) (significantly down-regulated in Salsa-depleted ovaries) (Supplemental Fig. S7A;

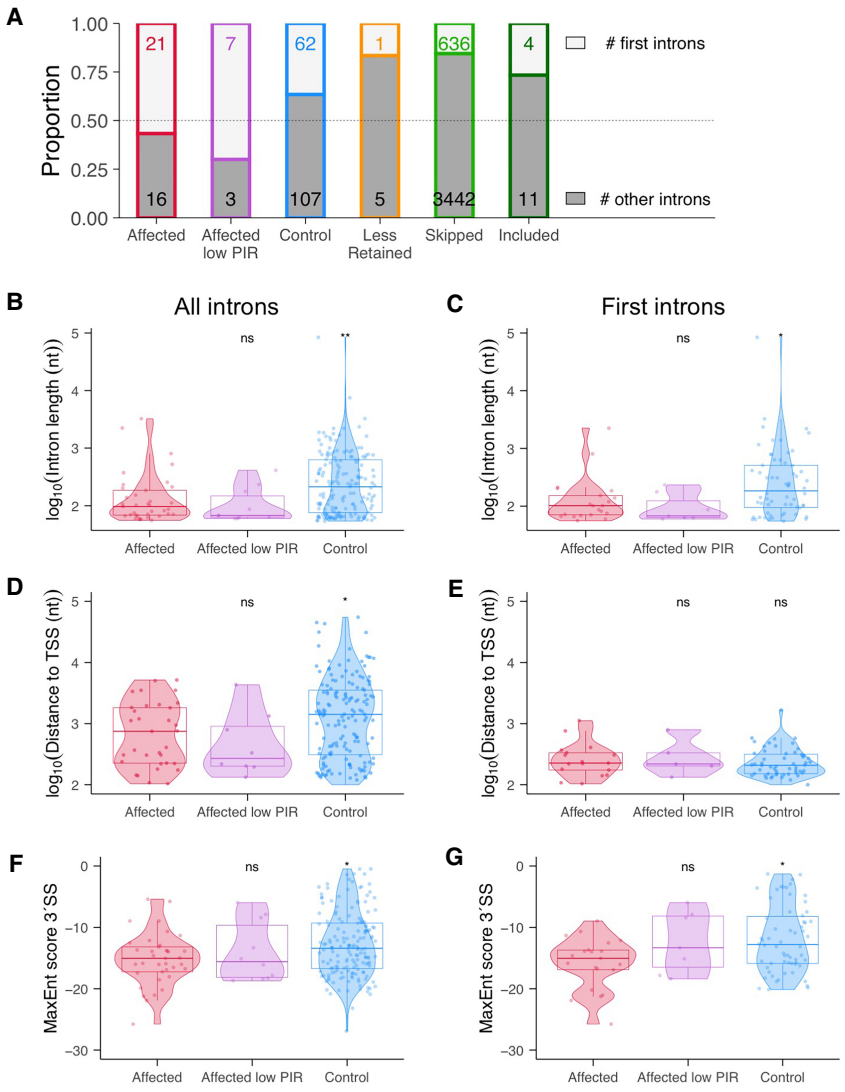

FIGURE 2. Small first introns show a bias for increased intron retention after Salsa depletion. (*A*) Proportion of first (white) and all other (gray) introns within each intron class, with the number of first introns/number of other introns indicated on *top*/*bottom* of each bar in color/black, respectively. (*B–G*) Violin plots and boxplots (highlighting the median, 25th and 75th percentiles per intron class) summarizing distributions of observations (each point refers to one intron) considering all introns (*left* panels) and first introns only (*right* panels) of: (*B*,*C*) $\log_{10}$ of intron length (nt); (*D*,*E*) $\log_{10}$ of distance (nt) between intron start and the transcript transcription start site (TSS); (*F*,*G*) Maximum entropy (MaxEnt) scores (Yeo and Burge 2004) for estimation of 3′ splice site efficiency, obtained with MATT (Gohr and Irimia 2019). Comparison of metrics in *B–G* was performed with *Affected* as the reference: (*) $P < 0.05$, (**) $P < 0.01$ and ns, nonsignificant (Wilcoxon rank-sum test with continuity correction).

Supplemental Table S2), whereas negatively regulates the expression of *death resistor Adh domain containing target* (*Drat*), *Cyp4p2*, *sevenless* (*sev*) and *Insulin-like peptide 6* (*Ilp6*) (significantly up-regulated upon Salsa depletion) (Supplemental Fig. S7A; Supplemental Table S2). Significantly differentially expressed genes (B-statistic > 0) are highlighted in red, while *salsa* (*CG31368*) is highlighted in black as a positive control (Supplemental Fig. S7A). Since there were no significant alterations in alternative splicing of these genes (Supplemental Fig. S8), then these gene expression changes are more likely the result of a minor role of Salsa in transcription and/or an indirect consequence of the splicing defects observed after Salsa depletion. Nonsense-mediated decay pathway genes showed no differences in expression levels between conditions (blue points in Supplemental Fig. S7A–G), suggesting the decay pathway is similarly efficient in control and Salsa-depleted samples. Altogether these results suggest that Salsa is not rate-limiting for overall transcription and the observed splicing defects did not elicit a genome-wide change of the transcriptome steady-state stability. This is consistent with previous results on *fandango* mutants where transcription was not obviously affected (Guilgur et al. 2014), and raises the possibility that contrary to human NTC, *Drosophila* NTC does not have any major role on transcription.

Salsa is required for female fertility and eggshell dorsal–ventral patterning

Since Salsa is required for efficient splicing of a subset of small first introns, including the first intron of *gurken* (Fig. 1B), we decided to investigate the function of this RNA helicase during oogenesis. To confirm that the observed RNAi phenotypes were due to Salsa depletion, and not a consequence of an off-target effect of the used *salsa* RNAi construct, three nonoverlapping shRNAs were used against *salsa* (*salsa* RNAi-1, *salsa* RNAi-2, and *salsa* RNAi-3). Since germline expression of salsa RNAi-3 was associated to significant levels of egg chamber necrosis, salsa RNAi-1 and salsa RNAi-2 constructs were mostly used in this manuscript to avoid nonspecific phenotypes. Depletion of Salsa within the female germline resulted in a strong reduction of *salsa* mRNA levels in the ovaries and unfertilized eggs/embryos (Fig. 3A; Supplemental Fig. S9A). Depletion of *salsa* mRNA was stronger in unfertilized eggs/embryos (maternal contribution) than in the ovaries (Supplemental Fig. S9A). This is most likely because Salsa is expressed in

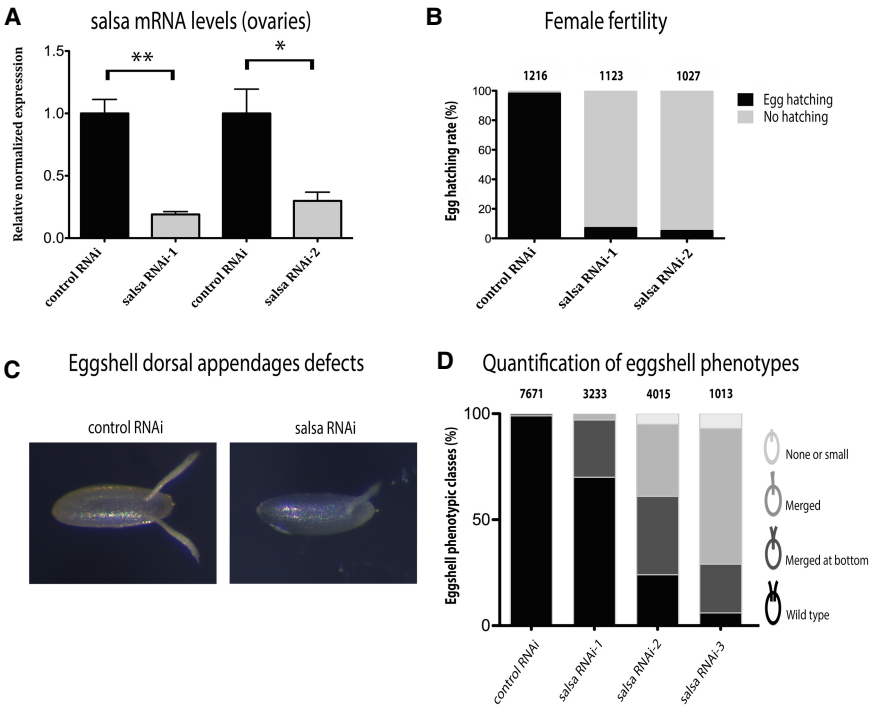

FIGURE 3. Salsa is required for female fertility and eggshell dorsal–ventral patterning. (*A*) Efficient depletion of *salsa* mRNA after germline-specific RNAi (*nanos*-Gal4 and UAS-*salsa* RNAi). RT-qPCR analysis of *salsa* mRNA in control (mCherry RNAi) and Salsa-depleted ovaries (two nonoverlapping shRNAs: *salsa* RNAi-1 and *salsa* RNAi-2). Relative normalized expression corresponds to values normalized with two distinct reference genes (*β-actin* and *GAPDH*) and relative to negative control (mCherry RNAi). At least three biological replicates were used for all shown results. Error bars indicate standard deviation. (*B*) Female fertility (egg hatching) was significantly reduced after germline-specific depletion of Salsa. Virgin females were crossed with wild-type males, and female fertility was calculated by the frequency of egg hatching 48 h post oviposition. Number of eggs scored for each experiment are indicated *above* the bar plot. (*C*, *D*) Germline-specific depletion of Salsa (three nonoverlapping shRNAs: *salsa* RNAi-1, *salsa* RNAi-2, and *salsa* RNAi-3) impaired eggshell dorsal–ventral patterning. (*C*) A significant proportion of the eggs laid by females whose germline was depleted for Salsa showed ventralized eggshells, with highly abnormal (fused or partially fused) dorsal appendages (*D*) Quantification of eggshell dorsal appendages defects (eggshell ventralization). Observed phenotypes were categorized in four different phenotypic classes based on the eggshell dorsal appendages: class 0: wild-type dorsal appendages—two individualized dorsal appendages); class 1: dorsal appendages only fused at bottom; class 2: fused dorsal appendages—spindle phenotype; and class 3: short eggs (dumpless-like phenotype) without or with extremely short dorsal appendages. Number of eggs scored for each experiment are indicated *above* the bar plot. Examples of scored phenotypes are shown in Supplemental Figure S10.

the soma (modEncode project; Graveley et al. 2011) and the dissected ovaries contain supporting somatic cells not affected by the germline-specific RNAi depletion. We next tested the effect of depleting Salsa in adult female fertility and observed that, after being mated with wild-type males, there was a strong reduction in egg hatching (Fig. 3B).

The reduced fertility of Salsa-depleted females was most likely related to the observed eggshell dorsal appendages defects (Fig. 3C,D), as this is suggestive of dorsal–ventral (D/V) patterning defects (Neuman-Silberberg and Schupbach 1993, 1994). For phenotypic quantification, we grouped the observed eggshell dorsal appendages phenotypes in four different classes: class 0: wild-type dorsal appendages, class 1: dorsal appendages only fused at bottom, class 2: fused dorsal appendages (spindle phenotype), and class 3: short eggs (dumpless-like phenotype) without or with extremely short dorsal appendages. Examples of scored phenotypes are shown in Supplemental Figure S10. Eggs laid by control females (mCherry RNAi) showed wild-type dorsal appendages, suggestive of normal D/V patterning of the oocyte, whereas a significant proportion of the eggs laid by females whose germline was depleted for Salsa showed ventralized eggshells, with highly abnormal (fused or partially fused) dorsal appendages (Fig. 3C,D; for more experimental detail see Materials and Methods). The ventralized eggshell phenotypes were observed for all three *salsa* RNAi constructs. Suggesting that the dorsal–ventral patterning defects observed with *salsa* RNAi-1 and *salsa* RNAi-2 hairpins are hypomorphic phenotypes, the ventralized phenotype of *salsa* RNA-3 hairpin was more severe. In addition, there were significant levels of egg chamber necrosis with *salsa* RNAi-3 and after germline mutant clones of *salsa* (using the publicly available transposon allele PBac{RB}CG31368^{e00215}). Altogether these results show that Salsa is required within the germline for female fertility, possibly because it is important for D/V patterning of the developing oocyte. The dumpless-like phenotype also suggests residual levels of nurse cells dumping defects (Ferreira et al. 2014).

Salsa is required for splicing of the first intron of *gurken* mRNA

The *gurken* gene encodes a single transcript with three introns (Neuman-Silberberg and Schupbach 1993), with the first intron localized to the gene 5′UTR, and the second and third introns localized to the gene open-reading frame. Since Salsa-depleted eggshell phenotypes were reminiscent of the ones observed in hypomorphic alleles

of *gurken* (Schupbach 1987), and since our RNA-seq results suggested that Salsa is required for efficient splicing of the first intron of *gurken* (Fig. 1B), we hypothesized that *salsa* D/V patterning phenotypes were due to defects of *gurken* expression.

To confirm the *gurken* splicing defects, we performed reverse transcription-quantitative PCR (RT-qPCR) to quantify the relative levels of intron retention after Salsa depletion. Specific primers flanking the 3′ splice site of all three introns of *gurken* were designed so that signal amplification would only occur in the presence of the intron (for used primer sequence see Supplemental Table S3). Consistent with our previous results, after depletion of Salsa there was a significant increase in the retention of the first intron, but not of the second and third introns, of *gurken* mRNA (Fig. 4A). To accurately determine gurken first intron unspliced/spliced ratio, we performed reverse transcriptase droplet digital PCR (RT-ddPCR). Specific primers and probe sets for unspliced or spliced first intron *gurken* transcripts were designed (for used primers and probes sequence see Supplemental Table S3). Interestingly, whereas in control ovaries (mCherry RNAi) the first intron of *gurken* unspliced/spliced ratio was only 0.06 ± 0.0072 (Fig. 4B,C), after Salsa depletion (*salsa* RNAi-2), this ratio increased to 0.33 ± 0.015 (Fig. 4B,C). Altogether, these results show that Salsa is required for efficient splicing of the first intron of *gurken*.

Salsa is required for dorsal-anterior localization of *gurken* mRNA

The correct dorsal-anterior localization of *gurken* mRNA has been reported to rely on multiple elements localized to the transcript 5′UTR, 3′UTR and open-reading frame (Saunders and Cohen 1999; Thio et al. 2000; Van De Bor et al. 2005). Since the 5′UTR of *gurken* mRNA was suggested to be necessary for its correct dorsal-anterior localization (Saunders and Cohen 1999; Thio et al. 2000), and Salsa depletion was associated with retention of the 5′UTR-localized first intron, we hypothesized that Salsa was required for dorsal-anterior localization of *gurken* mRNA.

To test this hypothesis, we performed fluorescence in situ hybridization with antisense RNA probes against *gurken* mRNA. For phenotypic quantification, we grouped the observed *gurken* mRNA localization phenotypes in three different classes: (i) normal localization, (ii) partial localization, and (iii) no localization. Examples of scored phenotypes are shown in Supplemental Figure S11 (for more experimental detail see Materials and

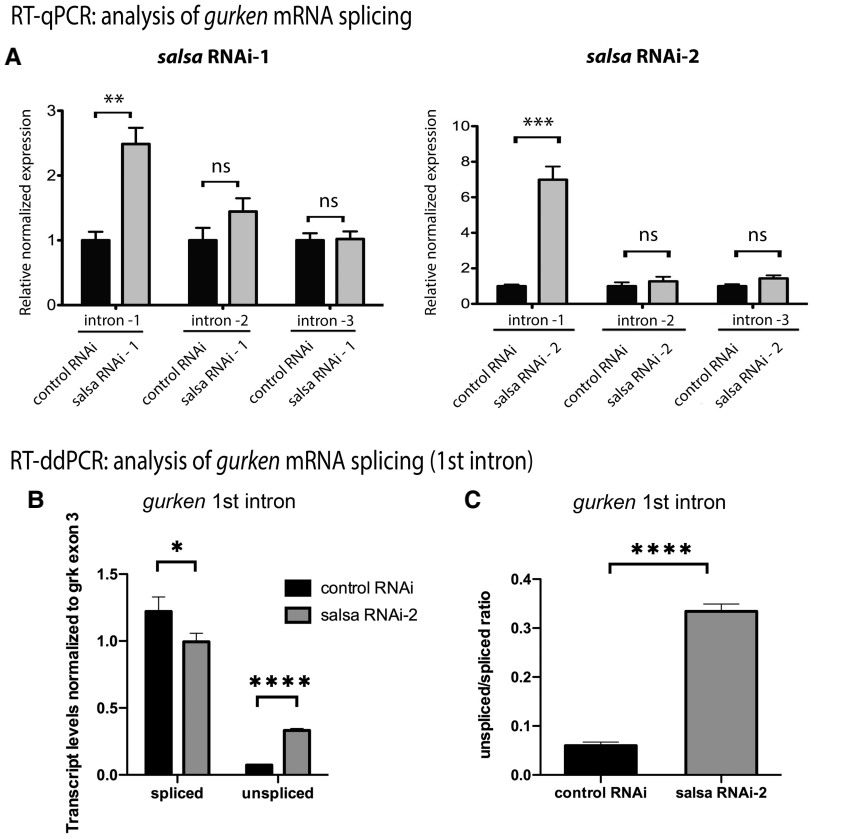

FIGURE 4. Salsa is required for splicing of the first intron of *gurken* mRNA. Germline-specific depletion of Salsa is associated with a significant increase in the retention of the first intron, but not of the second and third introns, of *gurken* mRNA. (*A*) RT-qPCR analysis of *gurken* transcript using intron–exon primers and an iScript cDNA library (random hexamer and oligo [dT] reaction mix). Germline-specific depletion of Salsa induced a significant retention of the first intron of *gurken* transcript, whereas the second and third introns were correctly spliced. Shown are the fold changes of intron retention. Relative normalized expression corresponds to values normalized with two distinct reference genes (*β-actin* and *GAPDH*) and relative to control conditions (mCherry RNAi). At least three biological replicates were used for each data set. Error bars indicate standard deviation. (*B*,*C*) Reverse transcription droplet digital PCR (RT-ddPCR) of *gurken* to accurately determine first intron unspliced/spliced ratio. Whereas in control ovaries (mCherry RNAi) the unspliced/spliced ratio of the first intron of *gurken* was only 0.06 ± 0.0072, after Salsa depletion (salsa RNAi-2) it raised to 0.33 ± 0.015. (*B*) Accurate quantification, by ddPCR, of spliced and unspliced isoforms of *gurken* first intron, relative to reference sequence: After Salsa depletion, there was an enrichment in unspliced intron 1 (unspliced *t*-test, $P < 0.00001$) and the corresponding decrease in intron 1 removal by splicing (spliced *t*-test, $P = 0.011$). (*C*) Graphic representation of the increase in *gurken* first intron ratio unspliced/spliced after Salsa depletion (salsa RNAi-2) (*t*-test, $P < 0.00001$). Two biological replicates were used for each RT-ddPCR data set. Error bars indicate standard deviation.

Methods). Consistent with the observed eggshell dorsal appendages defects, depletion of Salsa was associated with a reduction of the dorsal-anterior localization of *gurken* mRNA. Whereas egg chambers from control females (mCherry RNAi) showed a normal dorsal-anterior localization of *gurken* mRNA during mid-oogenesis (Fig. 5A,B), a significant proportion of egg chambers from females whose germline were depleted for Salsa showed a detectable reduction of dorsal-anterior localized *gurken* mRNA (Fig. 5A,B). Interestingly, the reduction of localized *gurken* transcript was proportional to the severity of eggshell ventralization and *gurken* mRNA splicing defects (compare *salsa* RNAi-1 and RNAi-2 in Figs. 3D, 4A, 5A,B).

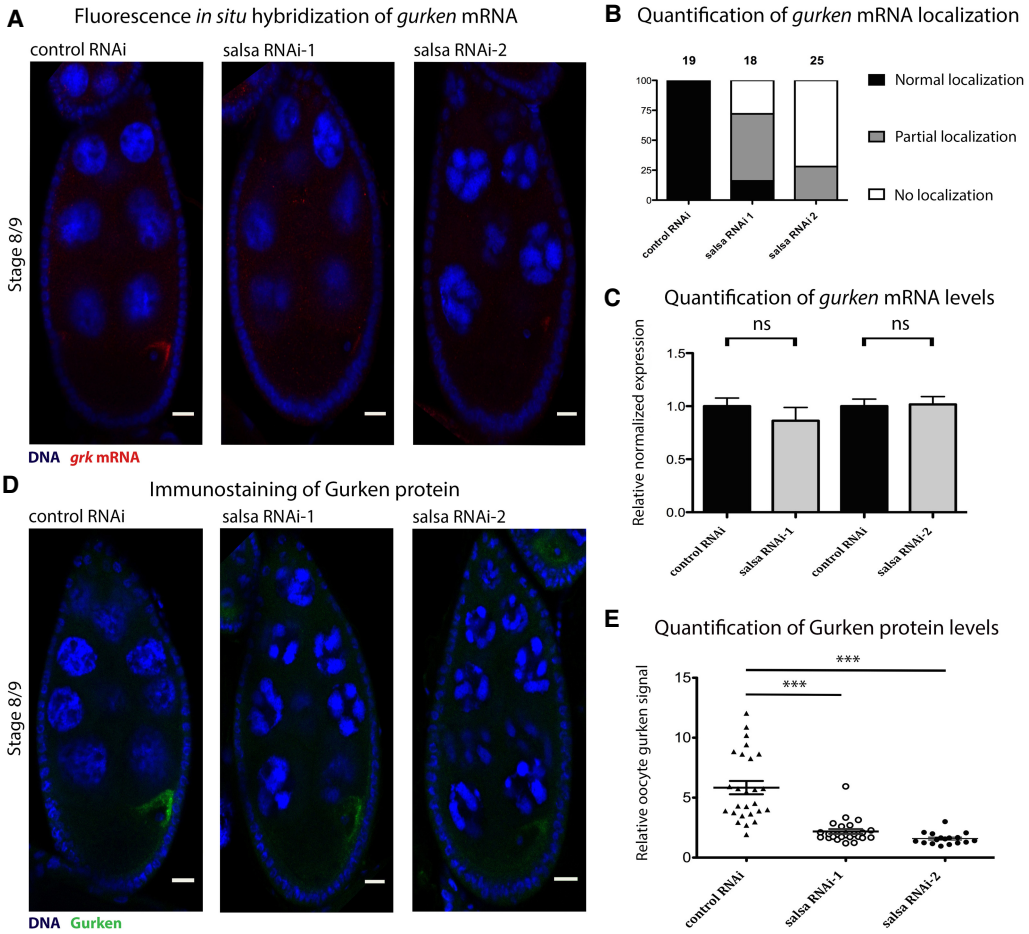

FIGURE 5. Salsa is required for dorsal-anterior expression of Gurken. (*A*,*B*) Salsa is required for oocyte dorsal-anterior localization of *gurken* mRNA. (*A*) Germline-specific depletion of Salsa impaired dorsal-anterior localization of *gurken* mRNA during mid-oogenesis. Fragmented digoxigenin-labeled antisense RNA probes were used to detect *gurken* mRNA in situ. Visualization of probes was done using an anti-Digoxigenin Cy3 secondary antibody. DNA (blue) and *gurken* mRNA (red). Scale bar, 10 µm. (*B*) Quantification of anterior dorsal localization defects of *gurken* mRNA in stage 8/9 egg chambers using three phenotypic classes: "normal localization," "partial localization," and "no localization." Additional examples of scored phenotypes are shown in Supplemental Figure S11. Negative control (mCherry RNAi): 100% "normal DV localization" ($n = 19$); Salsa depletion (*salsa* RNAi-1): 16% "normal DV localization," 56% "partial DV localization," and 28% "no DV localization" ($n = 18$); Salsa depletion (*salsa* RNAi-2): 0% "normal DV localization," 28% "partial DV localization," and 72% shows "no DV localization" ($n = 25$). (*C*) Salsa is not required for stability of *gurken* mRNA. Real-time qPCR analysis detected no significant reduction of total levels *gurken* mRNA after depletion of Salsa. Relative normalized expression corresponds to values normalized with two distinct reference genes (*β-actin* and *GAPDH*) and relative to negative control (mCherry RNAi). At least three biological replicates were used for all data sets. Error bars indicate standard deviation. (*D*,*E*) Salsa is required for oocyte dorsal-anterior localization of Gurken. (*D*) Immunostaining of Gurken during mid-oogenesis (stage 8/9 egg chambers). Gurken was detected with an anti-Gurken monoclonal antibody and DNA was visualized with DAPI staining. DNA (Blue) and Gurken (green). Scale bar 10 µm. (*E*) Relative oocyte Gurken dorsal anterior signal (arbitrary units [a.u.]) corresponds to Gurken signal pixel intensity (average of three different measurements taken from the oocyte dorsal-anterior region with the highest perinuclear signal) divided by the respective background signal (average of three different measurements from the respective nurse cells cytoplasm). To minimize sample variation, all measurements were obtained from maximum intensity projections obtained from confocal *Z* stacks of stage 8/9 egg chambers. Each dot represents an individual stage 8/9-egg chamber. *Horizontal* lines specify mean values and asterisks indicate significant difference (two-tailed unpaired *t*-test; $P < 0.001$).

Since abnormally processed mRNAs are usually targeted for degradation (Popp and Maquat 2013), we next tested if the steady-state stability of *gurken* mRNA was reduced after salsa depletion, as this would easily explain the reduced levels of localized *gurken* transcript. Using RT-qPCR, we failed to detect any significant reduction in the total levels of *gurken* mRNA in the ovaries after depletion of Salsa (Fig. 5C; Supplemental Fig. S9B). Consistently, *gurken* was also not found to be differentially expressed in Salsa-depleted samples when compared to control (logFC = −0.64, and B-statistics = −4.36) in our RNA-seq experiment. Suggesting that Salsa depletion does not prevent nuclear export of *gurken* mRNA from the nurse cells, significant levels of intron retention could be detected within the maternally loaded *gurken* mRNAs from mature oocytes/early embryos (Supplemental Fig. S9C).

Salsa is required for dorsal-anterior localization of Gurken protein

Since dorsal-anterior localization of *gurken* mRNA was reduced after depletion of Salsa (Fig. 5A,B), we hypothesized that a similar reduction was likely to occur for Gurken protein. To test this hypothesis, we performed an immunostaining of Gurken using an anti-Gurken antibody (Queenan et al. 1999) in *Drosophila* egg chambers. Gurken accumulated at high levels in the cytoplasmic perinuclear dorsal-anterior region of control oocytes (mCherry RNAi) (Fig. 5D), whereas such expression was significantly reduced after depletion of Salsa (Fig. 5D,E; see Materials and Methods for Gurken levels quantification). Since we failed to detect the total levels of Gurken protein in the ovaries by western blot, using the currently available antibodies, it is unclear if this reduction results from reduced levels of protein translation and/or mislocalization. Altogether, our results show that Salsa is required for normal dorsal-anterior expression of Gurken, most likely because it is rate limiting for efficient splicing of the first intron of *gurken* transcript.

Retention of the first intron of *gurken* impairs Gurken expression and function

If Salsa is rate limiting for dorsal-anterior expression of Gurken because it is required for efficient splicing of the first intron of *gurken*, then mutations within its 5′ and 3′ splice sites should similarly impair the function and expression of this gene. To examine such a possibility, we investigated if a genomic construct of *gurken* with its own endogenous promoter but with or without splice site mutations within the transcript´s first intron (for more experimental detail see Materials and Methods) could complement two amorphic/loss-of-function alleles of *gurken*. $grk^{deltaFRT}$ is an amorphic deletion allele of *gurken* (Lan et al. 2010), whereas grk^{HF} contains a premature stop codon that severely truncates Gurken protein (Thio et al. 2000). Females homozygous for $grk^{deltaFRT}$ or transheterozygous for $grk^{deltaFRT}/grk^{HF}$ are viable but sterile (Thio et al. 2000; Lan et al. 2010), as mutant egg chambers and eggshells show severe antero-posterior and dorsoventral patterning defects (Thio et al. 2000; Lan et al. 2010) and egg laying is severely compromised (Lan et al. 2010).

Supporting the hypothesis that splicing of the first intron of *gurken* is important for the expression and function of this gene, a wild-type genomic construct of gurken (grk^{WT}), but not an identical construct carrying first intron splice site mutations ($grk^{splice\ site\ mutant}$) (splice site mutation: GT/AG to TT), could efficiently rescue the female sterility, eggshell dorsal appendages defects, and abnormal Gurken expression of *gurken* mutant females (Fig. 6A,B, D,E). It should be noticed that a proportion of the L1 larvae rescued by the $grk^{splice\ site\ mutant}$ transgene (Fig. 6A) died after hatching. Interestingly, and although these mutations completely abrogated splicing of the first intron of gurken (Fig. 6G), there was no detectable reduction of the expression levels of gurken transcript (Fig. 6F).

Deletion of the first intron of *gurken* does not impair Gurken expression and function

Splicing has been previously reported to be important for transcript localization. For example, the first intron of *oskar* is required for posterior localization of *oskar* mRNA, as its splicing recruits the exon junction complex (EJC) and generates a short RNA stem–loop structure both important for kinesin 1-dependent transport to the posterior (Hachet and Ephrussi 2004; Ghosh et al. 2012). In order to investigate if splicing of the first intron of *gurken* is *per se* important for *gurken* mRNA localization and function, we decided to test if a genomic construct of *gurken* with its own endogenous promoter but without the first intron could complement two amorphic/ loss-of-function alleles of *gurken*. Deletion of the first intron of *gurken* did not impair *gurken* expression levels (Supplemental Fig. S12). Since both wild-type and first intron-deleted genomic constructs (respectively, grk^{WT} and $grk^{no\ 1st\ intron}$) could similarly rescue fertility (Fig. 6A) and eggshell dorsal appendages (Fig. 6B,C) of *gurken* mutant females, we concluded that, contrary to *oskar*, deletion of the first intron of *gurken* does not impair the expression and function of this gene. Altogether these results suggest that splicing of the first intron of *gurken* is important for the expression and function of this gene, not because splicing is per se rate-limiting for transcript localization, but because retention of the first intron is likely to disturb the structure of the 5′UTR of gurken mRNA.

Deletion of the first intron of *gurken* rescues the D/V patterning defects after Salsa depletion

Our results show that Salsa is rate-limiting for efficient splicing of the first intron of *gurken* and its depletion is

associated with significant defects in Gurken dorsal-anterior expression. Since the 5′UTR of *gurken* mRNA was reported to be required for mRNA localization (Saunders and Cohen 1999; Thio et al. 2000), *gurken* first intron localizes to the transcript 5′UTR and its retention impairs the function of this gene (Fig. 6A,B), we reasoned that expression of a genomic construct of *gurken* lacking its first intron (*gurken*$^{\text{no 1st intron}}$) would suppress the D/V patterning defects after Salsa depletion. Consistently, expression of a genomic construct of *gurken* lacking its first intron (*grk*$^{\text{no 1st intron}}$) efficiently suppressed the ventralized eggshell phenotypes observed after depletion of Salsa (compare *salsa* RNAi with *salsa* RNAi + 1× (*grk*$^{\text{no 1st intron}}$) in Fig. 7A). Since expression of a wild-type genomic construct of *gurken* (*grk*$^{\text{WT}}$) also suppressed, albeit less efficiently, the eggshell defects after Salsa depletion (compare *salsa* RNAi with *salsa* RNAi + 1× (*grk*$^{\text{WT}}$) in Fig. 7A), it is still possible that Salsa regulates Gurken dorsal-anterior expression by additional mechanisms beyond splicing of *gurken* first intron. Significantly, the egg hatching defects after salsa depletion could be partially suppressed by the expression of a genomic construct of *gurken* lacking its first intron (*gurken* no first intron), but not by the expression of a wild-type *gurken* construct (Fig. 7B). This suggests that although depletion of Salsa impairs efficient splicing of multiple introns, *gurken* mRNA is one of the key rate-limiting transcripts regulated by this RNA helicase during oogenesis. Altogether our results strongly suggest that the

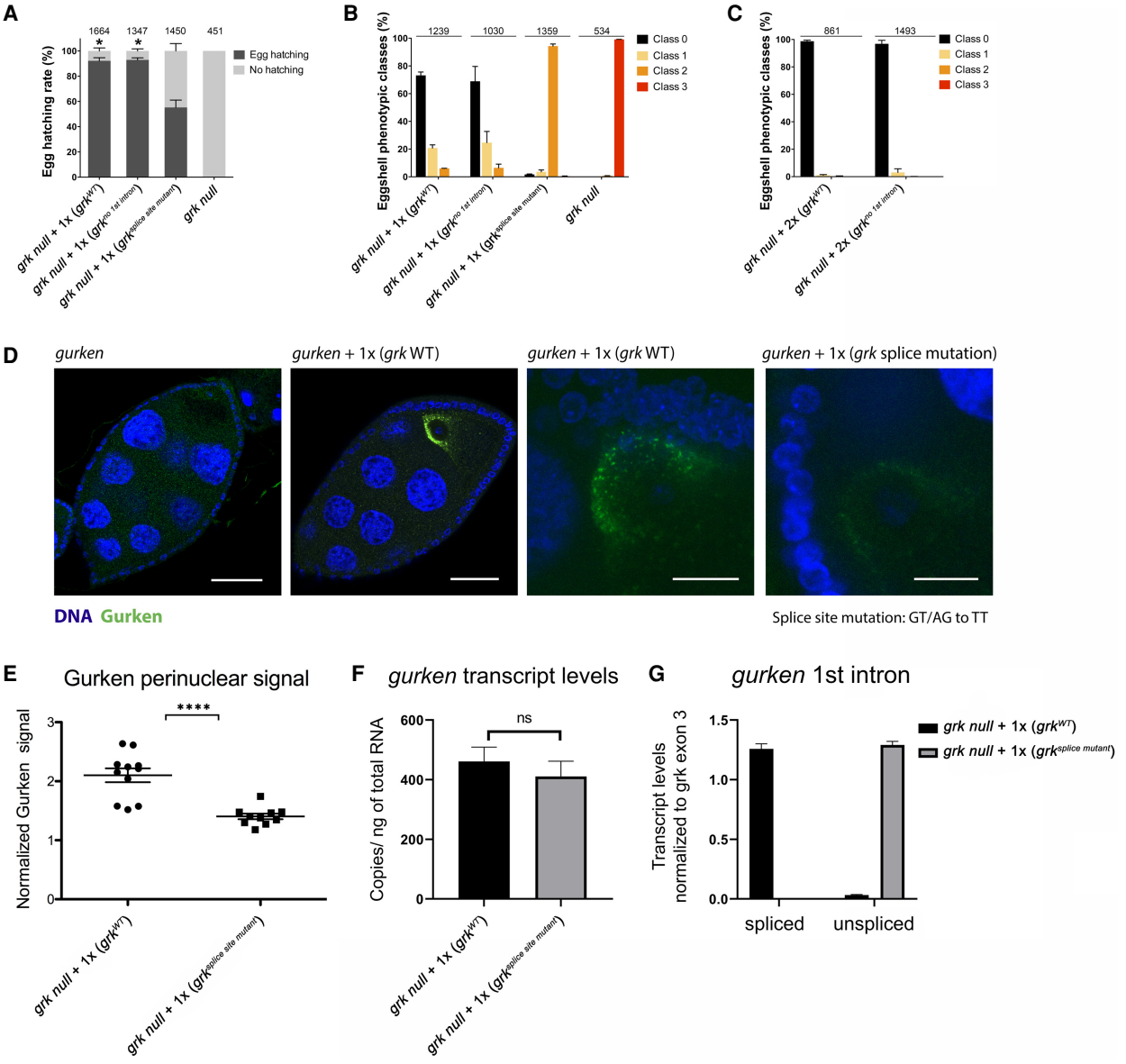

FIGURE 6. (*See legend on next page*)

female fertility and D/V eggshell patterning defects observed after Salsa-depletion are mostly due to defects of Gurken expression and splicing.

Salsa is required for eggshell dorsalization after increased *gurken* dosage

D/V patterning is sensitive to *gurken* dosage (Neuman-Silberberg and Schupbach 1994). Consistently, expression of a genomic construct containing wild-type *gurken* (grk^{WT}), in an otherwise wild-type genetic background (three copies of *gurken*: two endogenous and one corresponding to the genomic transgene), induced significant eggshell dorsalization defects (Fig. 7C; dorsalization corresponds to class −2 and −1), whereas 2 copies of grk^{WT} fully rescued the D/V patterning defects of a loss-of-function allele of gurken (compare Figs. 6C, 7C). Depletion of Salsa partially suppressed the dorsalization phenotype after three copies of *gurken* (compare wild-type genomic construct of gurken [grk^{WT}] in Fig. 7A,C), further confirming the requirement of Salsa for dorsal-anterior expression of Gurken during oogenesis.

D/V patterning defects after Salsa depletion are not due to a persistent activation of the meiotic checkpoint

Accumulation of persistent DNA damage during oogenesis leads to meiotic checkpoint activation, which severely disrupts oocyte nucleus morphology (Abdu et al. 2002; Klattenhoff et al. 2007) and embryonic anterior–posterior and dorsal–ventral axis specification by inhibiting the function of Vasa (Ghabrial and Schupbach 1999; Klattenhoff et al. 2007). Vasa is a DEAD box RNA helicase required for efficient translation of *gurken* (Styhler et al. 1998).

Although depletion of Salsa impairs dorsal-anterior expression of *gurken* and D/V patterning of the eggshell is severely compromised, the microtubule-dependent migration of the posterior oocyte nucleus to the anterior cortex was mostly unaffected (Fig. 5; Supplemental Fig. S11). Furthermore, and contrary to the highly abnormal oocyte nucleus observed after the accumulation of persistent DNA damage (Abdu et al. 2002), we failed to detect any oocyte nuclear abnormalities after Salsa depletion (Supplemental Fig. S13), suggesting that the observed

FIGURE 6. Splicing of the first intron of *gurken* is required for Gurken expression and function. Females homozygous for $grk^{deltaFRT}$ are viable but sterile, with egg laying severely compromised (Lan et al. 2010). $grk^{deltaFRT}$ corresponds to an amorphic allele of *gurken* (deletion mutant) (Lan et al. 2010), whereas grk^{HF} contains a single base pair change that introduces a precocious stop codon that severely truncates Gurken protein (Thio et al. 2000). (*A*) Expression of one copy of a genomic construct of *gurken* with its own endogenous promoter, with (grk^{WT}) or without ($grk^{no\ 1st\ intron}$) the transcript first intron, rescued fertility (egg hatching) of *grk null* ($grk^{deltaFRT}/grk^{HF}$) females. Expression of one copy of a genomic construct of *gurken* with mutated first intron splice sites ($grk^{splice\ site\ mutant}$) only partially rescued egg hatching and a proportion of the rescued L1 larvae died after eclosion. Number of eggs scored for each experiment are indicated *above* the bar plot and are the sum of two independent experiments. *t*-test analysis shows that expression of grk^{WT} ($P = 0.0144$) or $grk^{no\ 1st\ intron}$ ($P = 0.0128$) significantly increased egg hatching rate when compared to the expression of $grk^{splice\ site\ mutant}$. (*B*) Expression of one copy of a genomic construct of *gurken* with its own endogenous promoter, with (grk^{WT}) or without ($grk^{no\ 1st\ intron}$) the transcript first intron, rescued eggshell dorsal appendages defects (eggshell ventralization) of eggs laid by *grk null* ($grk^{deltaFRT}/grk^{HF}$) females. Expression of one copy of a genomic construct of *gurken* with mutated first intron splice sites ($grk^{splice\ site\ mutant}$) only weakly rescued eggshell ventralization. Number of eggs scored for each experiment are indicated *above* the bar plot and are the sum of two independent experiments. (*C*) Expression of two copies of a genomic construct of *gurken* with its own endogenous promoter, with (grk^{WT}) or without ($grk^{no\ 1st\ intron}$) the transcript first intron, fully rescued eggshell dorsal appendages defects (eggshell ventralization) of eggs laid by *grk null* ($grk^{deltaFRT}/grk^{deltaFRT}$) females. Number of eggs scored for each experiment are indicated *above* the bar plot and are the sum of two independent experiments. (*B*,*C*) For quantification of eggshell dorsal appendages defects (eggshell ventralization), observed phenotypes were categorized in four different phenotypic classes based on the eggshell dorsal appendages: class 0: wild-type dorsal appendages—two individualized dorsal appendages); class 1: dorsal appendages only fused at bottom; class 2: fused dorsal appendages—spindle phenotype; and class 3: short eggs (dumpless-like phenotype) without or with extremely short dorsal appendages. Examples of scored phenotypes are shown in Supplemental Figure S10. (*D*,*E*) Eggs from *grk null* ($grk^{deltaFRT}/grk^{HF}$) females lack detectable Gurken protein. Expression of one copy of a wild-type genomic construct of *gurken* (grk^{WT}) rescued dorsal anterior expression of Gurken. Retention of the first intron due to splice site mutations ($grk^{splice\ site\ mutant}$) significantly decreased Gurken dorsal anterior expression. First intron 5′ and 3′ splice sites (respectively, GT and AG) were both mutated to TT. Gurken was detected with an anti-Gurken monoclonal antibody and DNA was visualized with DAPI staining. DNA (Blue) and Gurken (green). Scale bars, 30 and 15 µm (respectively, first two and last two images). (*E*) Relative oocyte Gurken dorsal anterior signal (arbitrary units [a.u.]) corresponds to Gurken signal pixel intensity (average of three different measurements taken from the oocyte dorsal-anterior region with the highest perinuclear signal) divided by the respective background signal (average of three different measurements from the respective oocyte cytoplasm). To minimize sample variation all measurements were obtained from maximum intensity projections obtained from confocal Z stacks of stage 8/9/10 egg chambers. Each dot represents an individual egg chamber. *Horizontal* lines specify mean values and asterisks indicate significant difference (two-tailed unpaired *t*-test; $P < 0.0001$). (*F*) Expression of a genomic construct of *gurken*, without or with first intron splice site mutations (respectively, grk^{WT} and $grk^{splice\ site\ mutant}$), in a *grk null* ($grk^{deltaFRT}/grk^{deltaFRT}$) background, generated equivalent levels of *gurken* transcripts. Respectively, RT-ddPCR (see Materials and Methods) measured 460.9 ± 48.6 and 410.6 ± 51 copies of *gurken* mRNA (exon3) per ng of total RNA. (*G*) Splicing of the first intron of *gurken* was completely abrogated by splice site mutations ($grk^{splice\ site\ mutant}$). Measurement of the *gurken* first intron spliced and unspliced isoforms relative to exon 3. Splice site mutant shows total retention, while in wild-type it was only a 0.03 ± 0.006 for unspliced ratio. Wild-type constructs show 1.26 ± 0.044 for first intron spliced transcripts, whereas the splice site mutant shows 1.29 ± 0.031 for unspliced transcripts. The percentage of unspliced first intron was calculated as the ratio unspliced/(spliced + unspliced). For grk^{WT}, the unspliced ratio is 2.5 ± 0.41%, whereas $grk^{splice\ site\ mutant}$ completely abrogates first intron splicing ($grk^{splice\ site\ mutant} = 100 \pm 0.002\%$). At least two biological replicates were done for each experiment. Error bars indicate standard deviation.

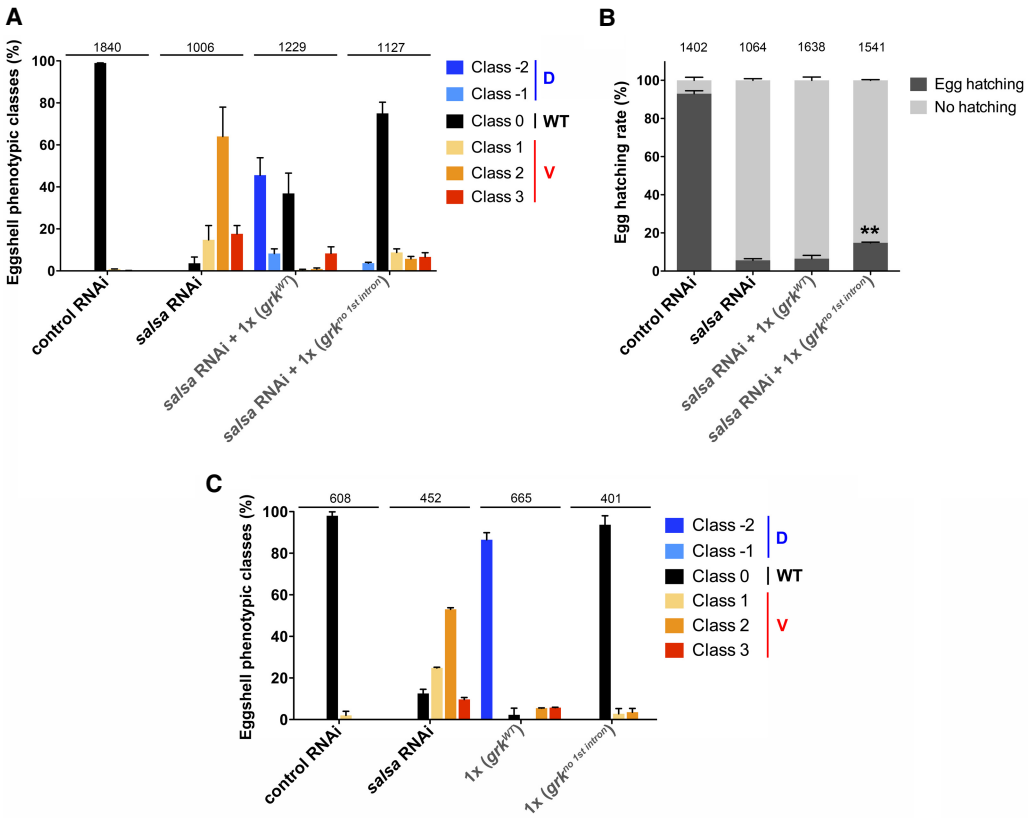

FIGURE 7. Deletion of the first intron of *gurken* rescues the D/V patterning defects of Salsa depletion. (*A*) Eggs laid by control females (mCherry RNAi) showed wild-type dorsal appendages (WT), whereas a significant proportion of the eggs laid by females whose germline was depleted for Salsa (*nanos*-Gal4 and *salsa* RNAi-2) showed eggshell ventralization (V), with highly abnormal (fused or partially fused) dorsal appendages (classes 1 and 2). Expression of a genomic construct of *gurken* with its own endogenous promoter, with (grk^{WT}) or without the transcript first intron ($grk^{no\ 1st\ intron}$), suppressed the eggshell ventralization defects of eggs laid by females whose germline was depleted for Salsa (*salsa* RNAi-2). Suppression of $grk^{no\ 1st\ intron}$ was significantly better than grk^{WT}, with a higher frequency of wild-type eggshells (class 0). Number of eggs scored for each experiment is indicated *above* the bar plot and is the sum of two independent experiments. (*B*) Control females (mCherry RNAi) are fertile (egg hatching), whereas a significant proportion of the eggs laid by females whose germline was depleted for Salsa (*nanos*-Gal4 and *salsa* RNAi-2) did not hatch. While expression of one copy of a genomic construct of wild-type *gurken* (grk^{WT}) did not rescue fertility, expression of one copy of a similar genomic construct of *gurken* without the transcript first intron ($grk^{no\ 1st\ intron}$) partially rescued the fertility defects after Salsa depletion (*salsa* RNAi-2), (**) $P = 0.0053$; *t*-test analysis. Number of eggs scored for each experiment are indicated *above* the bar plot and are the sum of two independent experiments. (*C*) Expression of a genomic construct of *gurken* (grk^{WT}) in an otherwise wild-type genetic background (1× grk^{WT}; three copies of *gurken*: two endogenous and one corresponding to the genomic transgene) induced a significant eggshell dorsalization (D) (classes −2 and −1). Eggshell dorsalization was not observed after expression of a genomic construct of *gurken* without the first intron of *gurken* mRNA (1× $grk^{no\ 1st\ intron}$). Number of eggs scored for each experiment is indicated *above* the bar plot and is the sum of two independent experiments. For quantification of eggshell dorsal appendages defects, the observed phenotypes were categorized in six different phenotypic classes based on the eggshell dorsal appendages. Class 0 corresponds to wild-type eggshell dorsal appendages (WT). Eggshell ventralization (V) of the eggshells corresponds to classes 1 to 3. Eggshell dorsalization (D) corresponds to classes −1 and −2. Class 0 (wild-type dorsal appendages): two individualized dorsal appendages); class 1: dorsal appendages only fused at bottom; class 2: fused dorsal appendages—spindle phenotype; and class 3: short eggs (dumpless-like phenotype) without or with extremely short dorsal appendages. Class −1: corresponds to short eggs with dorsalized appendages; class −2: corresponds to a broad and thick crown of appendage material or a broad/fused appendage that covers the entire width of the eggshell. Examples of scored phenotypes are shown in Supplemental Figure S10.

D/V patterning phenotypes are not mediated by persistent activation of the meiotic checkpoint.

Salsa is required for nurse cells chromosome dispersal

During early oogenesis, the nurse cell chromosomes undergo several rounds of endoreplication that results in a characteristic polytene blob-like appearance (Dej and Spradling 1999). This chromatin architecture disappears by stage 5/6 with the dispersal of the nurse cells polytene chromosomes. Besides the previously described D/V patterning defects (eggshell dorsalization), mutants for *squid* (*sqd*), *hrb27C*, and *half pint* (*hfp*) also show defects in nurse cells chromosome dispersal (Van Buskirk and Schupbach 2002; Goodrich et al. 2004), clearly suggesting that these

proteins regulate multiple processes during oogenesis. Interestingly, whereas alternative splicing of *ovarian tumor* (*otu*) was abnormal in *sqd* and *hfp* mutants, ectopic expression of the 104 kDa Otu isoform was sufficient to rescue the nurse cells chromosome dispersal phenotype of these mutants (Van Buskirk and Schupbach 2002; Goodrich et al. 2004).

Ovaries depleted for Salsa showed nurse cells chromosome dispersal defects (Fig. 5D) identical to the ones previously observed *in sqd* and *hfp* mutants (Van Buskirk and Schupbach 2002; Goodrich et al. 2004). Nonetheless, Salsa is most likely not functionally related with Sqd and Hfp, as Salsa depletion is associated with eggshell ventralization, and not dorsalization, and it did not significantly impair alternative splicing of *otu* (Supplemental Fig. S14).

Our results suggest that although the female fertility defects after Salsa depletion are clearly related to an abnormal splicing and expression of *gurken*, other phenotypes, like the nurse cells chromosome dispersal defects may also give a contribution for such reduction of female fertility.

DISCUSSION

Salsa regulates expression and efficient splicing of *gurken*

Oocyte dorsal-anterior localization of *gurken* mRNA relies on multiple elements localized to the transcript 5′UTR, 3′UTR and open-reading frame (Saunders and Cohen 1999; Thio et al. 2000; Van De Bor et al. 2005), yet the relative importance of each element for mRNA localization is still unclear. The 5′ and 3′-UTRs of *gurken* were reported to be required for dorsal-anterior localization of *gurken* transcript (Saunders and Cohen 1999). Furthermore, and using a genomic *gurken* construct with a lacZ reporter inserted within the gene open-reading frame, it was shown that whereas *gurken* 5′UTR is required for transcript oocyte accumulation, its coding region and 3′UTR are necessary for its posterior and dorsal-anterior localization (Thio et al. 2000). Nevertheless, it was recently reported, using an oocyte injection assay, that a small stem–loop located within the open-reading frame was necessary and sufficient for *gurken* transcript localization (Van De Bor et al. 2005).

Our results show that efficient splicing of the first intron of *gurken* is required for mRNA dorsal-anterior localization and dorsal–ventral patterning. This is most likely because retention of the first intron impairs the secondary RNA structure of *gurken* 5′UTR, and the function of a closely located RNA element important for its localization (Saunders and Cohen 1999). Splicing of the first intron of *gurken* is also likely to facilitate Gurken protein expression, as deletion of the first intron of gurken suppresses the dorsalization phenotype associated with increased copy number of *gurken* gene without affecting the levels of gurken mRNA. Our results therefore fully support the role of *gurken* 5′UTR in mRNA localization within the oocyte, and strongly suggest that Salsa-dependent splicing of the first intron of *gurken* mRNA is important for the correct expression and function of this gene.

Salsa facilitates efficient splicing of small first introns

The precise function of human Aquarius in splicing is still poorly understood. This RNA helicase is recruited to the spliceosome as a pentameric complex known as intron-binding complex (IBC), which also contains hSyf1 (also known as Xab2), hIsy1, CypE, and CCDC16 (De et al. 2015). Coimmunoprecipitation experiments suggest a large interaction interface between IBC and U2 snRNP, within the activated spliceosome (B^{act} stage) and just before the first splicing reaction. Although Aquarius ability to bind and hydrolyze ATP is important for spliceosome activation and splicing efficiency, the role of its RNA unwinding activity is less clear.

Our work has identified a small subset of introns whose splicing is particularly sensitive to depletion of Salsa (the *Drosophila* ortholog of human Aquarius). The fact that splicing was only affected in a small number of introns is consistent with the observation that immunodepletion of human Aquarius from nuclear extracts only weakly impaired splicing in vitro (De et al. 2015). This suggests that although this RNA helicase is apparently not critical for overall splicing, during female gametogenesis there is a subset of introns whose efficient removal relies on the function of this enzyme.

Analysis of the introns whose splicing was sensitive to Salsa depletion showed a clear bias for small first introns with weak 3′splice sites, independently of their distance to the transcription start site (TSS), 5′splice site strength and GC content. The bias for small introns suggests that Salsa is mostly rate-limiting when introns are recognized by intron definition (Pai et al. 2017), where the initial pairing between U1 and U2 snRNPs occurs across the intron. Furthermore, the bias for introns with weak 3′splice sites is in accordance with the extensive interaction between IBC and U2 snRNP in the activated spliceosome, and implies that depletion of Salsa is likely to impair, at least in a subset of introns, U2 snRNP function during splicing. The absence of any detectable bias for short distances between the TSS and 5′splice site, when evaluating affected and control first introns, or any bias for weak 5′splice site strength, suggests that Salsa is not likely rate-limiting for Cap-Binding Complex-mediated splicing (Qiu et al. 2007).

Drosophila first introns are more likely to be cotranscriptionally retained than internal and terminal introns (Khodor et al. 2011). This is not consistent with the kinetic competition model (Bentley 2014), where the fastest processes are the ones most likely to occur, suggesting additional

constraints to first intron splicing. Although the precise nature of such constraints is still poorly understood, binding of transcriptional initiation factors to the 5′splice site-associated U1snRNP (Damgaard et al. 2008; Bentley 2014) potentially restricts splicing efficiency, as it might impair the initial pairing between U1 and U2 snRNPs. Our working hypothesis is that Salsa is required for splicing of small first introns with weak 3′splice site because this enzyme facilitates U2 snRNP function, minimizing the interference effect of transcriptional initiation factors on splicing. Future work will help define the function of this RNA helicase and its contribution for differential gene expression during development.

MATERIALS AND METHODS

Fly husbandry

All flies were raised at 25°C, unless otherwise indicated, using standard techniques.

Drosophila RNAi stocks

Unless indicated, all stocks are available at the Bloomington Drosophila Stock Center. Depletion of Salsa was obtained using three different nonoverlapping dsRNA hairpins. *salsa* RNAi-2 stock is available in BDSC (Bloomington number 55172; hairpin reference; P {TRiP.HMC03852}attP40; Map: Chr 2,25C6, 2L: 5108448..5108448.), whereas *salsa* RNAi-1 and *salsa* RNAi-3 were custom made.

Generation of nonoverlapping *salsa* RNAi (*salsa* RNAi-1 and RNAi-3)

We designed two nonoverlapping RNAi hairpins against *salsa* (CG31368) using an algorithm that minimizes off target effects (Vert et al. 2006). Based on a miR1 scaffold, for the top strand oligo (TS) of *salsa*, ctagcagt nucleotides were added to the 5′ end of the passenger strand DNA and tagttatattcaagcata nucleotides were added between passenger strand DNA and the guide strand DNA. In the end gcg nucleotides were added to the 3′ end of guide strand DNA. The top strand oligo (TS) of *salsa* 1 and *salsa* 3 are the following:

salsa RNAi-1 TS: ctagcagtCGCTTGGATATGGACGATCTAtagtta tattcaagcataTAGATCGTCCATATCCAAGCGgcg

salsa RNAi-3 TS: ctagcagtCCACGATTATCTCCTACGCAAtagtt atattcaagcataTTGCGTAGGAGATAATCGTGGgcg

For the bottom strand oligo (BS) of *salsa*, aattcgc nucleotides were added to the 5′ end of the passenger strand DNA, and tatgcttgaatataacta nucleotides were added between passenger strand DNA and the guide strand DNA. In the end actg nucleotides were added to the 3′ end of guide strand DNA. The bottom strand oligo (BS) of *salsa* 1 and *salsa* 3 are the following:

salsa RNAi-1 BS: aattcgcCGCTTGGATATGGACGATCTAtatgctt gaatataactaTAGATCGTCCATATCCAAGCGactg

salsa RNAi-3 BS: aattcgcCCACGATTATCTCCTACGCAAtatgctt gaatataactaTTGCGTAGGAGATAATCGTGGactg

Annealing the top and bottom strand oligos: 10 µL top strand oligo (10 µM) and 10 µL bottom strand oligo (10 µM) were added into 80 µL annealing buffer (10 mM Tris-HCl, pH 7.5, 0.1 M NaCl, 1 mM EDTA). The reaction mix was incubated at 95°C for 5 min, and then, slowly cooled down to room temperature. The resulting DNA fragment has overhangs for *NheI* and *EcoRI*. The resulting DNA fragments were directly cloned into a VALIUM22 vector, which has been linearized by *NheI* and *EcoRI* enzymes. For the DNA ligation reaction, 6 µL of the annealing product mixed with 2 µL of 10× ligation buffer and 1 µL of T4 DNA ligase (1 U/µL). Finally, 1 µL of 40 ng/µL backbone (gel purified VALIUM22 digested with *NheI* and *EcoRI*) was added to the final reaction. The final volume of 20 µL was made up with ddH20, mixed carefully and incubated overnight at 16°C. The reaction was stored at −20°C, until further use.

For transformation, 10 µL of ligation reaction was transformed into 50 µL TOP10 competent cells using the electroporation method. Clones were selected through PCR using the standard primers for pVALIUM22 (Supplemental Table S3). The correct shRNA constructs were further confirmed by sequencing. The primer used for sequencing is shown in Supplemental Table S3. Plasmid amplification and isolation of the correct shRNA constructs was done using the standard protocol from the NZY Miniprep isolation kit. (NZY Tech, Genes and Enzymes). For production of transgenic *Drosophila* stocks, isolated plasmids were sent for injection (BestGene, Chino Hills). All transgenic stocks were confirmed by DNA sequencing.

Germline-specific depletion of Salsa

Salsa was specifically depleted within the female germline by using the germline-specific driver *nanos*-Gal4 (*nos*-Gal4) (Van Doren et al. 1998) and the upstream activating sequence (UAS)/Gal4 system (Brand and Perrimon 1993; Rorth 1998; Ni et al. 2011).

Ventralized eggshell phenotypes

Flies were cultured at 25°C and transferred to a new tube every 3 d. Zero to three days old F1 females ($n = 20$) were crossed with wild-type males ($n = 5$) and added to an egg collection cage supplemented with fresh yeast. Ventralized eggshell phenotype was scored at least 2 d after assembly of the collection cage (for optimal egg laying) and using fresh egg collections (4–5 h) to facilitate unequivocal scoring.

Egg hatching

One to three days old F1 females ($n = 20$) were crossed with wild-type males ($n = 5$) and added to an egg collection cage supplemented with fresh yeast. Eggs were collected using freshly prepared apple juice plates. Plates were incubated for 48 h at 25°C and egg hatching was scored.

Genomic construct of *gurken*

Drosophila melanogaster Flybase genome was used as reference for all *gurken* genomic constructs. Used sequences are identical to the most up-do-date release (FB2020_03, released Jun 16, 2020). All constructs include 600 nt upstream of the annotated transcriptional start site (endogenous promoter) and 200 nt downstream from the transcriptional end site. All constructs were cloned into a pWallium22 plasmid and integrated in the genome using phiC31 mediated integration. All constructs were integrated in the attP2 docking site. Genomic construct of *gurken* lacking its first intron ($grk^{no\ 1st\ intron}$) is identical to wild-type construct (grk^{WT}), except that the first intron was fully deleted and the first and second exons fused in-frame. Genomic construct of *gurken* with first intron splice site mutations ($grk^{splice\ site\ mutant}$) is identical to wild-type construct, except that the first intron 5′ and 3′splice sites (respectively, GT and AG) were both mutated to TT. DNA synthesis was done by GenScript Biotech (Piscataway). For production of transgenic *Drosophila* stocks, isolated plasmids were sent for injection (BestGene). All transgenic stocks were confirmed by DNA sequencing.

Ovaries immunostaining

Adult ovaries (20 ovary pairs per sample per experiment) were processed according to standard procedures (Ferreira et al. 2014; Navarro-Costa et al. 2016; Prudencio et al. 2018). All incubation steps involve a gentle mix with an orbital or nutator mixer. Briefly, ovaries were dissected from 2–4-d-old females in ice cold Phosphate-buffered saline (PBS) and fixed for 20 min with a 3:1 mix of heptane (Fluka) and fixative aqueous mix. The fixative mix consisted of 4% formaldehyde (Polysciences) in PBS + 0.5% NP-40 (Sigma) solution. Following three 10 min washes in PBST (PBS + 0.2% Tween 20), ovarioles were gently detached with a tungsten needle. Ovaries were subsequently blocked for 1 h in PBS supplemented with 1% Triton X-100 (Sigma), 1% (w/v) bovine serum albumin (BSA; Sigma), and 1% (w/v) donkey serum (Sigma). Primary antibody was performed overnight at 4°C in PBST supplemented with 1% BSA and 1% donkey serum (BBT solution). After three 10 min washes in PBST, ovaries were incubated for 1 h at the room temperature with the appropriate secondary antibodies diluted in BBT solution. DNA was subsequently stained for 30 min at the room temperature using 1:10.000 DAPI in PBST. Prior to mounting, ovaries were washed three times, 10 min each, with PBST. Ovaries were mounted in Dako Faramount Aqueous Mounting Medium (Dako) and were visualized using a Zeiss LSM710 Confocal microscope.

Quantification of ovaries Gurken immunostaining

For quantification of Gurken signal at the dorsal anterior region of the oocyte, maximum intensity projections of serial confocal optical sections of individual egg chambers stained with an anti-Gurken mouse antibody (Queenan et al. 1999) (clone 1D12; dilution 1/30) were obtained using the Image J program (Grouped Zprojector, maximum pixel intensity). The average Gurken signal intensity within three oocyte dorsal anterior regions with highest levels of localized Gurken was obtained using ImageJ software. For each egg chamber, obtained values were divided by the respective average background signal intensity value obtained within three cytoplasmic regions of the nurse cells or the oocyte.

Preparation of a DIG-labelled *gurken* probe

A plasmid carrying a *gurken* cDNA sequence (Neuman-Silberberg and Schupbach 1993) was linearized with Sal I restriction enzyme and analyzed on agarose gel electrophoresis. The linearized plasmid was isolated and purified with a gel purification kit (Promega), using standard procedures. The purified linearized plasmid was subsequently quantified with a spectrophotometer (Thermo Scientific NanoDrop 2000). Synthesis of DIG-labeled antisense *gurken* RNA probes, was made according to the following reaction: 2 µL of linearized plasmid (1 µg) was mixed with the following reagents, in this order and at room temperature: 1 µL of 10× transcription buffer (Roche), 1 µL DIG-NTP mix (Roche), 0.3 µL of RNase inhibitor (Roche) and 1 µL T7 RNA polymerase (10 U/µL; Roche) (de Las Heras et al. 2009). The final reaction volume was set up to 10µl of DEPC water. The final reaction mixture was subsequently incubated at 37°C for 2 h. Subsequently, 40 µL DEPC water and 50 µL 2× Carbonate Buffer (120 mM-Na_2CO_3; 80 mM $NaHcO_3$ at pH 10.2) were added to the probes and the mix was incubated at 65°C for 40 min. Hydrolyzed probes were precipitated by adding 100 µL 0.2 M NaAc (pH 6), 10 µL of 4 M LiCl, 10 µL of tRNA (Roche; stock solution 20 mg/µL), and 600 µL of absolute ethanol, on a −20°C overnight incubation. Mix was centrifugated at 4°C for 15 min. Pellet was washed twice with 70% EtOH, dried in air (10 min), and dissolved in Hybridization MIX buffer (50% formamide, 5× SSC solution, 0.1% Tween-20, 100 µg/mL heparin, 100 µg/mL ssDNA). Probes were stored at −20°C before use.

Fluorescent in situ hybridization of *Drosophila* ovaries

Adult ovaries from 2–4-d-old females were dissected and fixed according to standard procedures (Ferreira et al. 2014; Navarro-Costa et al. 2016; Prudencio et al. 2018). All incubation steps involve a gentle mix with a nutator mixer. Ovaries were dissected from 2–4-d-old females in ice cold Phosphate-buffered saline (PBS) and fixed for 20 min with a 3:1 mix of heptane (Fluka) and fixative aqueous mix. The fixative mix consisted of 4% formaldehyde (Polysciences) in PBS + 0.5% NP-40 (Sigma) solution. After fix, ovarioles were gently teased apart with forceps and dehydrated in methanol by washes of 5 min with the following solutions: 30% methanol/PBT (PBS + 0.2% Tween-20), 50% methanol/PBT and 70% of methanol/PBT. Ovaries were stored in 100% methanol at −20°C before usage.

Prior to usage, ovaries were rehydrated by washes with 70% Methanol/PBT; 50% Methanol/PBT and 30% Methanol/PBT. After rehydration, ovaries were washed with PBS + 0.1% Tween-20, for 5 min, three times and transferred to a prehybridization buffer (40% formamide, 4× SSC (20× SSC [RNase free], 0.1% Tween 20) and incubated for 1 h at the room temperature. Ovaries were then transferred into 200 µL hybridization buffer (50% formamide, 5× SSC [20× SSC (RNase free)], 0.1% Tween 20, 100 µg/mL heparin, 100 µg/mL ssDNA) and incubated for 1 h at 55°C. RNA probes were diluted at 5:100 in a preheated hybridization buffer. After removal of the hybridization solution,

50 µL of preheated diluted probes were added to the ovaries for overnight incubation at 55°C. The following day, used probes were collected, and ovaries were transferred to 1 mL prehybridization solution for a 30 min incubation at 55°C. They were subsequently washed five times with 1ml of PBS + 0.1% Tween-20 for 30 min each time, at the room temperature. After washes, ovaries were incubated in Roche Blocking Buffer (1:10 in PBS-0.1% Tween-20) for 1 h at the room temperature. The mouse anti-Dig (1:400) was diluted in Roche Blocking Buffer (1:10 in PBS-0.1% Tween-20) and incubated overnight at 4°C in a rocking shaker. The ovaries were washed three times with 1 mL PBS-0.1% Tween-20 for 10 min each at the room temperature and incubated with Roche Blocking Buffer as above for 30 min at the room temperature, following incubation (1 h at RT) with secondary antibody anti-mouse-Cy3 (1:1000) diluted in Roche Blocking Buffer as above. Ovaries were washed three times with 1 mL PBS-0,1% Tween-20, for 5 min each time, at room temperature. DNA was subsequently stained for 10 min at the room temperature using 1:10.000 DAPI in PBS-0.1% Tween-20. Prior to mounting, ovaries were washed in PBS-0.1% Tween-20 three times, 5 min each, then rinsed with PBS. Ovaries were mounted in Vectashield mounting medium and they were visualized using a Zeiss LSM710 Confocal microscope.

Quantification of *gurken* mRNA localization

For semi-quantitative quantification of dorsal-anterior localization of *gurken* mRNA, stage 8/9 oocytes were divided it into three different classes based on the *gurken* mRNA dorsal-anterior signal near to the oocyte nucleus: (a) normal dorsal-anterior localization of *gurken* mRNA, with a strong signal distributed in the periphery of the oocyte nucleus; (b) partial dorsal-anterior localization, with weaker but clearly detectable *gurken* mRNA nearby the periphery of the oocyte nucleus; (c) absence/almost absence of dorsal-anterior localization, with undetectable or hardly detectable *gurken* mRNA signal nearby the periphery of the oocyte nucleus.

Real-time quantitative PCR (real-time qPCR)

Optimization of primers efficiency for target genes and reference genes

Primers were designed using PrimerBLAST (NCBI) and optimized to an equal annealing temperature 55°C. Primer efficiencies were initially tested with *Drosophila* genomic DNA (gDNA). Fivefold serial dilutions of gDNA were made (40, 8, 1.6, 0.32, and 0.064 ng), and primer efficiencies for both target genes and reference genes were tested. Standard curves were constructed by the Cycle of threshold (C_t) (*y*-axis) versus log gDNA dilution (*x*-axis). The primer efficiency (E) of one cycle in the exponential phase was calculated according to the equation: $E = 10^{(-1/\text{slope})}-1$ (Nolan et al. 2006). Efficiency and regression curves for each primer set are shown in Supplemental Table S4. Specificity of primer pair amplification was confirmed by melting curve analysis and the Ct values were obtained from the dynamic range of the standard curves. Only primer sets with specificity and good efficiency were used for subsequent RT-qPCR analysis.

mRNA extraction and cDNA synthesis

Total RNA was extracted from 0–1 h embryos (after egg laying) or from 2–4-d-old adult female ovaries (after pupae eclosion), whose female germline was specifically depleted for salsa (salsa RNAi) or negative control (mCherry RNAi). Extraction was performed following standard procedures (PureLink RNA Mini Kit, Ambion). Genomic DNA was removed from RNA samples using PureLink DNase (Invitrogen) on-column method. Experion Electrophoresis (Bio-Rad) was used to assess the integrity and quantity of isolated RNA. Isolated RNA with less than 8 RQI was not considered for reverse transcription with iScript (as described above) and further Real-Time qPCR analysis. Three technical replicates were performed for each sample. A technical replicate corresponds to three different cDNAs synthesized from the same isolated RNA. Biological replicates correspond to RNAs obtained from distinct biological samples.

Quantitative PCR reaction (qPCR)

The equivalent to 4 ng cDNA was used as template in each qPCR reaction. qPCR reactions were performed on a CFX96 Real-Time system (Bio-Rad) using SsoFast EvaGreen Supermix (Bio-Rad). Two reference genes (Actin and GAPDH) were used for normalization. "no template," "no RT," and "NEG" were used as negative controls for testing the qPCR reaction mixture contamination. Genomic DNA was used as positive control. For each reaction, the RT-qPCR master mix included 10 µL of 2× EvaGreen mix, 1 µL of forward primer (10 µM), 1 µL of reverse primer (10 µM), and 4 µL of nuclease-free water. The master mix (16 µL) was added to a well of 96-multiwell plate (Bio-Rad) followed by the addition (when appropriate) of 4 µL cDNA (4 ng). PCR protocol: denaturation program (95°C for 30 sec), amplification program repeated for 60 cycles (denaturation: 95°C for 5 sec, annealing: 55°C for 5 sec + plate read for fluorescence measurement), melting curve program (65°C to 95°C with 0.5°C increments, for 5 sec + plate read for fluorescence measurement), and cooling program to 4°C.

Expression analysis

The threshold cycle (Ct) of each gene transcript was determined by setting the threshold line above background levels in the linear region of the exponential curve before the gene expression levels were measured. The baseline and fluorescence signals are adjusted automatically by the CFX96 Real-Time machine software (default parameters) to calculate the ΔCt value (Schmittgen and Livak 2008). The relative expression ratio (RE) was calculated using the difference between ΔCt value of negative control (mCherry RNAi) and salsa RNAi, also known as the $2^{-\Delta\Delta Ct}$ method.

$$RE = 2^{(-(\Delta Ct_{sample}(\text{target}-\text{reference})-\Delta Ct_{control}(\text{target}-\text{reference}))}.$$

ΔCt_{sample} is the Ct difference of target-reference genes in salsa RNAi; $\Delta Ct_{control}$ is the Ct difference of target-reference genes in control RNAi.

Total expression levels of the genes of interest were calculated by the mean of Ct values normalized to the expression levels of two reference genes (Actin and GAPDH) using the geNorm method (Vandesompele et al. 2002). Mean and standard deviation (SD) of RE values were calculated from three biological replicates.

Reverse transcription droplet digital PCR (RT-ddPCR)

The accurate quantification of the primary transcript and spliced isoform was achieved by one-step RT-ddPCR (Bio-Rad), with the Bio-Rad QX200 Droplet Digital PCR System and One-Step RT-ddPCR Advanced Kit for Probes, according to manufacturer's instructions. In short, 100 ng of purified total RNA, was mixed with 900 mM primers, 250 nM probes and the enzymes Reverse transcriptase and DNA polymerase and then partitioned into thousands of individual water-oil emulsion droplets. Primer-specific reverse transcription was carried out inside each droplet, followed by PCR amplification and Taqman probe detection. In a C1000 thermal cycler (Bio-Rad), reverse transcription was carried out for 60 min at 50°C, then the RT enzyme was denatured and simultaneously the DNA pol enzyme was activated by 10 min incubation at 95°C, followed by 46 cycles of denaturation for 35 sec at 95°C and annealing/elongation for 1 min at 60°C, and a final inactivation step for 10 min at 98°C. Primers and probes (IDT, Integrated DNA Technologies) were designed with IDT software tools available online. Unspliced GRK, exon1–intron 1: forward—TATAGCAGCTCCAGTACGTC, reverse—CTACACACTTGCATCTCCTTG, probe—TTGTTTCGTGTGTGTGCGTTCGTG; spliced GRK, exon1–exon2: forward—CCATATAGCAGCTCCATAAAATCC, reverse—TAAACGATCGAGGGATCGAG, probe—ATGTTTCAGCTTCTGTTGCGACGC; reference GRK, exon3: forward—CGACAAGGAGACAGAGATTCAG, reverse—TAACGCAGAGGCAGGAATG, probe—TGCCCTGCAGCGAAGCTTACAATA. Specific target sequences were detected with 6-FAM/ZEN/IBFQ labeled probes and the reference target was detected with a HEX/ZEN/IBFQ labeled probe. Positive fluorescent droplets were carrying the target RNA sequence. Individual droplet fluorescence was read in a QX200 Droplet reader and data was analyzed using Quantasoft Version 1.7.4.0917. Manual thresholds were applied to define positive fluorescent signal. A nontemplate control was included for each primer set. Each purified RNA sample was tested by a non-RT control with the pre-mRNA primer set, with no RT enzyme in the mix. Graphs and statistical *t*-tests were made using GraphPad Prism 8.2.

One advantage of One-Step ddPCR is that reverse transcription and amplification processes are carried out inside each individual droplet, avoiding bias. The primer-specific reverse transcription abrogates any reverse transcriptase bias and increases sensitivity. The separation of RNA in compartments avoids competition between splicing isoforms and therefore favors unbiased amplification of under-represented forms. The results of ddPCR are absolute and independent of primer amplification efficiency, with no need for efficiency determination or calibration curves.

Protein extraction

Zero to one hour embryos (after egg laying) or 2–4-d-old adult female ovaries (after pupae eclosion), whose female germline was specifically depleted for salsa (salsa RNAi) or negative control (mCherry RNAi), were homogenized with a pestle in ice-cold NB-lysis buffer (50 mM Tris [pH 7.5], 150 mM NaCl, 2 mM EDTA, 0.01% NP-40 [Igepal], 2 mM DTT, 10 mM NaF, and protease inhibitors [complete protease inhibitor cocktail, Roche]). The sample was centrifuged three times at 14,000 rpm, at 4°C, and for 5 min. Between each centrifugation, the supernatant was carefully collected into a new tube, avoiding the pellet and the upper lipid-rich layer. Bradford protein microassay (Bio-Rad) was performed to quantify the total amount of protein, and whenever needed, the total protein extract was frozen with liquid nitrogen and kept at −80°C until further usage.

Western blot analysis

For western blot analysis, total protein extracts were boiled for 5 min at 95°C in 2×Laemmli buffer (Sigma). Ten micrograms of protein were loaded per lane, and run in a SDS polyacrylamide gel electrophoresis gel (6% or 12%). Proteins were subsequently transferred onto Hybond-ECL nitrocellulose membrane (Amersham) using a Bio-Rad wet transfer apparatus (100 V for 60 min). Western blotting was performed using standard procedures. Briefly, the Hybond-ECL membrane was blocked overnight with 5% nonfat milk in PBS-T (PBS + 0.1% Tween-20) at 4°C with shaking. Primary antibodies were incubated overnight in 1% nonfat milk in PBS-T with gentle shaking at 4°C. Membranes were washed three times with PBS-T, for 15 min at room temperature and with strong shaking. Secondary antibodies were incubated for 2 h in 1% nonfat milk in PBS-T at the room temperature with gentle shaking. After three washes in PBS-T, for 5 min at room temperature, and one quick wash with PBS, proteins of interest were detected with an ECL Plus Western blotting detection system (GE Healthcare) and an ECL Hyperfilm (Amersham).

Expression of carboxy-terminal Myc-tagged Salsa

Cloning of Myc-tagged Salsa

For cloning of a carboxy-terminal Myc-tagged Salsa, the open reading frame of *CG31368* was amplified by Expand High Fidelity PCR System (Roche) from the cDNA clone RE 35509 (BDGP). Please note that a point mutation present in the original cDNA clone RE35509 was corrected by site-directed mutagenesis using the publically available Flybase sequence of *CG31368* as reference. The QuikChange Site-Directed Mutagenesis Kit (Agilent Technologies) was used for site-directed mutagenesis, according to manufacturer's standard protocol. For primers sequence see Supplemental Table S3. For PCR amplification we used the High Fidelity PCR system (Roche), with the following conditions: initial denaturation for the first cycle (95°C for 2 min), primary amplification program repeated for 10 cycles (denaturation: 94°C for 15 sec, annealing: 55°C for 30 sec and elongation: 68°C for 4 min), secondary amplification program repeated for 10 cycles (denaturation: 94°C for 15 sec, annealing: 55°C for 30 sec and elongation: 72°C for 4 min), final elongation for one cycle (72°C for 10 min), and cooling at 4°C for an unlimited time. The PCR product was cloned into pAMW plasmid (Gateway) using the Gateway system (Invitrogen).

Culture of Drosophila *S2 cells*

Drosophila Schneider 2 (S2) cells were cultured in Schneider's *Drosophila* complete medium: Schneider's insect medium (Sigma), supplemented with 1× L-glutamine, 1x PenStrep, and 10% Fetal bovine serum (Invitrogen) at 25°C.

Transfection of Myc-tagged Salsa

Transient transfection in *Drosophila* S2 cells with a carboxy-terminal Myc-tagged Salsa construct was done using FuGENE HD Transfection Reagent and standard procedures (Promega). Briefly, a mix of 100 µL of Serum Free Medium (SFM), 400 ng of DNA, and 4 µL of Fugene HD were incubated at the room temperature for 15 min. Meanwhile, cells were plated into six well plates at the concentration of 2.5×10^6 cells/well in serum free medium. The mix (DNA and Fugene) was added drop by drop on the top of cells. To stop the transfection reaction, 2 mL complete medium (with 10% Fetal bovine Serum) was added to the cells after 4 h of incubation at 25°C and cells were incubated for 48 h at 25°C.

Coimmunoprecipitation

Coimmunoprecipitation was done using protein extracts from *Drosophila* S2 cells expressing carboxy-terminal Myc-tagged Salsa or a negative control (empty Myc-plasmid). Briefly, 1 mg of total protein extract was diluted in 1 mL of ice-cold NB-lysis buffer, and it was incubated with anti-c-Myc Magnetic beads (Invitrogen) for 1 h at 4°C. Anti-c-Myc magnetic beads were then washed three times with NB-lysis buffer and resuspended in 100 µL NB-lysis buffer. Half of each sample was boiled for 5 min at 95°C with 2×Laemmli buffer (Sigma) for further western blot analysis (no RNase A-treated samples). The remaining half was incubated with RNase A (Thermo Fisher) for 30 min at 37°C. Beads were further washed three times with NB-lysis buffer and boiled for 5 min at 95°C with 2×Laemmli buffer (Sigma).

RNA-seq analysis

RNA isolation

Total RNA was extracted from 2–4-d-old adult female ovaries (after pupae eclosion), whose female germline was specifically depleted for salsa (salsa RNAi) or negative control (mCherry RNAi). Extraction was performed following standard procedures (PureLink RNA Mini Kit, Ambion). Genomic DNA was removed from RNA samples using PureLink DNase (Invitrogen) on-column method. Experion Electrophoresis (Bio-Rad) was used to assess the integrity and quantity of isolated RNA.

Preparation of cDNA library and RNA-seq

After the QC procedures, mRNA from eukaryotic organisms is enriched using oligo(dT) beads. First, the mRNA is fragmented randomly to short fragments of 250–300 bp in fragmentation buffer, then cDNA is synthesized by using mRNA template and random hexamers primer. After the first strand was synthesized, a custom second-strand synthesis buffer (Illumina), dNTPs (dUTP, dATP, dGTP, and dCTP) and DNA polymerase I were added to initiate the second-strand synthesis. Then followed by purification by AMPure XP beads, terminal repair, polyadenylation, sequencing adapter ligation, size selection, and degradation of second-strand U-contained cDNA by the USER enzyme. The strand-specific cDNA library was generated after the final PCR enrichment. cDNA library preparation and HiSeq paired-end sequencing (150 nt) was performed by Novogene

Differential intron retention analysis

Overall quality of raw RNA-seq data was analyzed using *fastqc* version 0.11.5 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) on the associated FASTQ files. *vast-tools* (Braunschweig et al. 2014; Irimia et al. 2014; Tapial et al. 2017) version 2.0.0 was used for quantification of alternative sequence inclusion levels from FASTQ files using VAST-DB annotation for *Drosophila melanogaster* genome assembly dm6 (Tapial et al. 2017; Torres-Mendez et al. 2019). To ensure intron retention quantification with sufficient and reliable read coverage, we considered only *vast-tools*' events with a minimum mappability-corrected read coverage score of "OK" across all samples, corresponding to at least 20 mappability-corrected reads mapping either to the sum of skipping splice junctions or to one of the two inclusion exon–intron junctions and at least 15 of such reads mapping to the other inclusion splice junctions, as described in *vast-tools*'s GitHub description page (https://github.com/vastgroup/vast-tools). Also, only intron retention events without statistically significant evidence for a high imbalance in read counts between both upstream and downstream exon–intron junctions and the intron body (Braunschweig et al. 2014) (binomial test's *P*-value ≥ 0.05) for all samples were considered. Percent intron retention (PIR) (Braunschweig et al. 2014) estimates are quantified as the ratio between the average of exon–intron junction reads that support intron inclusion (#inc) and the sum of these with exon–exon junction reads, supporting intron skipping (#exc).

Precision of PIR estimates from their supporting read coverage (#inc and #exc) was modeled using the beta distribution (conjugate prior probability distribution for the binomial), constrained to [0,1] and defined by shape parameters α and β and a mean of $\alpha/(\alpha+\beta)$. Given that, for α, $\beta \neq 0$ for the same mean, beta distributions get narrower with increasing shape parameters, emitting points from a beta distribution with α = #inc and β = #exc allows the scattering of emitted values to serve as a surrogate for that PIR's dispersion (given the original read coverage supporting it), while the distribution's mean value, #inc/(#inc + #exc), is an approximation of the empirical PIR. For each considered intron retention event with nonzero #inc and #exc, the R function *rbeta* was used to emit, for each sample, 500 values from a beta distribution with shape parameters α = #inc and β = #exc. However, while #inc or #exc can be equal to zero, beta distributions with α or$\beta = 0$ are single valued as 0 or 1, respectively, and are not appropriate to model inclusion levels. Therefore, an increment factor (incr) proportional to global coverage was added to zero #inc (PIR = 0) or #exc (PIR = 1) read counts, followed by emission, using *rbeta*, of 500 points from a beta distribution with α = #inc + incr and β = #exc + incr. Increment factors per coverage were set as the maximum value between 0.01 and 1 such that not more than 5% of simulated 500-point beta distributions with α = incr and β = coverage have a median that shows a deviation from theoretical PIR larger than 0.01.

For each considered event, *rbeta*-emitted values were grouped per condition (control and salsa RNAi), with their median values set as the respective summary PIRs (PIR_{beta}). The difference between these, $\Delta PIR_{beta} = PIR_{salsa\ RNAi} - PIR_{control}$, was used as the magnitude of differential intron inclusion. The significance of each $\Delta PIR_{beta}$was set as the ratio between the number of differences between salsa-depleted and control emitted values that

are greater than zero and the total number of differences, reflecting the probability of $PIR_{salsa\ RNAi}$ being greater than $PIR_{control}$ (P_{diff}).

Differentially retained introns were considered as those with $P_{diff} \geq 0.9$. From these, introns that were more retained upon salsa depletion than in controls ($\Delta PIR_{beta} > 0.1$) were classified as *Affected,* while those showing decreased retention upon salsa depletion ($\Delta PIR_{beta} < -0.1$) were classified as *Less Retained*. Irrespectively of the significance of PIR differences, *Affected* introns showing very low inclusion levels in Control samples (mean PIR < 0.05) were classified into a separate group, *Affected low PIR*. *Included* introns were set as those showing consistently high inclusion, i.e. with a PIR ≥ 0.75 for all samples and an absolute $\Delta PIR_{beta} < 0.02$, while *Skipped* introns were classified as those always found spliced out, having PIR = 0 for all samples. Finally, *Control* introns were classified as those showing $PIR_{salsa\ RNAi}$ and $PIR_{control}$ between 0.05 and 0.9 (alternatively retained introns) but their retention does not seem to be affected by salsa depletion (absolute $\Delta PIR_{beta} < 0.02$).

MATT (Gohr and Irimia 2019) was used to retrieve intronic features, such as intron, upstream or downstream exon length, splice site strength (Yeo and Burge 2004) and GC content across groups using function *get_ifeatures*. Ensembl Biomart (Kinsella et al. 2011) was used to extract transcription start sites per transcript to calculate distance between these and the intron start (5′ss). Differential alternative splicing analysis considering all event types annotated in *vast-tools* (Supplemental Fig. S2) was performed using the same filtering criteria as described above for intron retention.

Differential gene expression analysis

Gene expression was quantified with *vast-tools*, that aligns RNA-seq reads against a reference transcriptome. Differential gene expression analysis was performed with the limma R package (Ritchie et al. 2015). The magnitude of difference in expression was measured in $\log_2$ fold-change between control and salsa RNAi. Six genes were considered to be significantly differentially expressed (those with an associated positive empirical Bayes statistic giving the log-odds ratio of the gene being differentially expressed): CG4570 and Arc1 down-regulated and Drat, Cyp4p2, sev and Ilp6 up-regulated in salsa-depleted samples.

DATA DEPOSITION

All data generated or analyzed during this study are included in this published article (and its Supplemental Information Files). The RNA-seq data have been deposited in the Sequence Read Archive under accession number PRJNA655826.

SUPPLEMENTAL MATERIAL

Supplemental material is available for this article.

ACKNOWLEDGMENTS

We thank Julian König (IMB Mainz) for scientific discussions. We also thank Cláudia Florindo for assistance in fluorescence microscopy, the TRiP collection at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing several of the transgenic RNAi fly stocks used in this study, and Manuel Irimia (CRG Barcelona) for advance access to the *Drosophila* VAST-DB annotation. The Microscopy Unit was partially supported by Portuguese national funding (FCT: PPBI-POCI-01-0145-FEDER-022122). This work was developed with the support of the research infrastructure Congento (project LISBOA-01-0145-FEDER-022170). R.G.M. is supported by Portuguese national funding through Fundação para a Ciência e a Tecnologia (FCT grant refs. PTDC/BEX-BID/0395/2014, PTDC/BIA-BID/28441/2017, and UID/BIM/04773/2013 CBMR 1334). Nuno L. Barbosa-Morais is supported by European Molecular Biology Organization (EMBO) Installation Grant (3057), Investigador FCT Starting Grant (IF/00595/2014), UID/BIM/50005/2019, the project funded by Fundação para a Ciência e a Tecnologia (FCT)/Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) through Fundos do Orçamento de Estado and iMM Lisboa start-up funds. O.S.R. and B.M. were supported by Portuguese national funding through Fundação para a Ciência e a Tecnologia, respectively, PD/BD/52428/2013 and PD/BD/128342/2017, within the scope of the ProRegeM PhD program (ref. PD/00117/2012, CRM:0027030). O.S.R. was also supported by a Federation of European Biochemical societies (FEBS) short-term fellowship. M.A.-F. is supported by Fundação para a Ciência e Tecnologia (FCT) PhD fellowship (PD/BD/128283/2017) and Fundação AstraZeneca. R.D.S. is supported by Portuguese national funding through Fundação para a Ciência e a Tecnologia, ref. DL 57/2016/CP1361/CT0019. The funding bodies had no role in the design of this study, collection, analysis, and interpretation of data, and manuscript writing.

Author contributions: O.S.R.: investigation, writing (review and editing); R.D.S.: investigation; M.A-F: investigation, writing (review and editing); R.M.: investigation; B.M.: investigation; M.N.T.: investigation; P.P.: investigation, writing (review and editing); R.P.A.: conceptualization, writing (review and editing); J-Y.R.: conceptualization, writing (review and editing); N.L.B-M.: conceptualization, writing (review and editing); R.G.M.: conceptualization, writing (original draft, review, and editing).

Received August 12, 2020; accepted September 7, 2020.

REFERENCES

Abdu U, Brodsky M, Schupbach T. 2002. Activation of a meiotic checkpoint during *Drosophila* oogenesis regulates the translation of Gurken through Chk2/Mnk. *Curr Biol* **12:** 1645–1651. doi:10.1016/S0960-9822(02)01165-X

Akay A, Di Domenico T, Suen KM, Nabih A, Parada GE, Larance M, Medhi R, Berkyurek AC, Zhang X, Wedeles CJ, et al. 2017. The helicase aquarius/EMB-4 is required to overcome intronic barriers to allow nuclear RNAi pathways to heritably silence transcription. *Dev Cell* **42:** 241–255.e246. doi:10.1016/j.devcel.2017.07.002

Ashe MP, Pearson LH, Proudfoot NJ. 1997. The HIV-1 5′ LTR poly(A) site is inactivated by U1 snRNP interaction with the downstream major splice donor site. *EMBO J* **16:** 5752–5763. doi:10.1093/emboj/16.18.5752

Bentley DL. 2014. Coupling mRNA processing with transcription in time and space. *Nat Rev Genet* **15:** 163–175. doi:10.1038/nrg3662

Brand AH, Perrimon N. 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118:** 401–415.

Braunschweig U, Barbosa-Morais NL, Pan Q, Nachman EN, Alipanahi B, Gonatopoulos-Pournatzis T, Frey B, Irimia M, Blencowe BJ. 2014. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res* **24:** 1774–1786. doi:10.1101/gr.177790.114

Caceres L, Nilson LA. 2005. Production of *gurken* in the nurse cells is sufficient for axis determination in the *Drosophila* oocyte. *Development* **132:** 2345–2353. doi:10.1242/dev.01820

Chanarat S, Strasser K. 2013. Splicing and beyond: the many faces of the Prp19 complex. *Biochim Biophys Acta* **1833:** 2126–2134. doi:10.1016/j.bbamcr.2013.05.023

Damgaard CK, Kahns S, Lykke-Andersen S, Nielsen AL, Jensen TH, Kjems J. 2008. A 5′ splice site enhances the recruitment of basal transcription initiation factors in vivo. *Mol Cell* **29:** 271–278. doi:10.1016/j.molcel.2007.11.035

David CJ, Boyne AR, Millhouse SR, Manley JL. 2011. The RNA polymerase II C-terminal domain promotes splicing activation through recruitment of a U2AF65-Prp19 complex. *Genes Dev* **25:** 972–983. doi:10.1101/gad.2038011

Davidson A, Parton RM, Rabouille C, Weil TT, Davis I. 2016. Localized translation of *gurken*/TGF-α mRNA during axis specification is controlled by access to Orb/CPEB on processing bodies. *Cell Rep* **14:** 2451–2462. doi:10.1016/j.celrep.2016.02.038

De I, Bessonov S, Hofele R, dos Santos K, Will CL, Urlaub H, Luhrmann R, Pena V. 2015. The RNA helicase Aquarius exhibits structural adaptations mediating its recruitment to spliceosomes. *Nat Struct Mol Biol* **22:** 138–144. doi:10.1038/nsmb.2951

Dej KJ, Spradling AC. 1999. The endocycle controls nurse cell polytene chromosome structure during *Drosophila* oogenesis. *Development* **126:** 293–303.

Delanoue R, Herpers B, Soetaert J, Davis I, Rabouille C. 2007. *Drosophila* Squid/hnRNP helps Dynein switch from a *gurken* mRNA transport motor to an ultrastructural static anchor in sponge bodies. *Dev Cell* **13:** 523–538. doi:10.1016/j.devcel.2007.08.022

de Las Heras JM, Martinho RG, Lehmann R, Casanova J. 2009. A functional antagonism between the *pgc* germline repressor and torso in the development of somatic cells. *EMBO Rep* **10:** 1059–1065. doi:10.1038/embor.2009.128

Ferreira T, Prudencio P, Martinho RG. 2014. *Drosophila* protein kinase N (Pkn) is a negative regulator of actin-myosin activity during oogenesis. *Dev Biol* **394:** 277–291. doi:10.1016/j.ydbio.2014.08.008

Fox-Walsh KL, Dou Y, Lam BJ, Hung SP, Baldi PF, Hertel KJ. 2005. The architecture of pre-mRNAs affects mechanisms of splice-site pairing. *Proc Natl Acad Sci* **102:** 16176–16181. doi:10.1073/pnas.0508489102

Ghabrial A, Schupbach T. 1999. Activation of a meiotic checkpoint regulates translation of Gurken during *Drosophila* oogenesis. *Nat Cell Biol* **1:** 354–357. doi:10.1038/14046

Ghosh S, Marchand V, Gaspar I, Ephrussi A. 2012. Control of RNP motility and localization by a splicing-dependent structure in *oskar* mRNA. *Nat Struct Mol Biol* **19:** 441–449. doi:10.1038/nsmb.2257

Gohr A, Irimia M. 2019. Matt: Unix tools for alternative splicing analysis. *Bioinformatics* **35:** 130–132. doi:10.1093/bioinformatics/bty606

Gonzalez-Reyes A, St Johnston D. 1994. Role of oocyte position in establishment of anterior-posterior polarity in *Drosophila*. *Science* **266:** 639–642. doi:10.1126/science.7939717

Gonzalez-Reyes A, Elliott H, St Johnston D. 1995. Polarization of both major body axes in *Drosophila* by *gurken*-torpedo signalling. *Nature* **375:** 654–658. doi:10.1038/375654a0

Goodrich JS, Clouse KN, Schupbach T. 2004. Hrb27C, Sqd and Otu cooperatively regulate *gurken* RNA localization and mediate nurse cell chromosome dispersion in *Drosophila* oogenesis. *Development* **131:** 1949–1958. doi:10.1242/dev.01078

Gornemann J, Kotovic KM, Hujer K, Neugebauer KM. 2005. Cotranscriptional spliceosome assembly occurs in a stepwise fashion and requires the cap binding complex. *Mol Cell* **19:** 53–63. doi:10.1016/j.molcel.2005.05.007

Graveley BR, Brooks AN, Carlson JW, Duff MO, Landolin JM, Yang L, Artieri CG, van Baren MJ, Boley N, Booth BW, et al. 2011. The developmental transcriptome of *Drosophila melanogaster*. *Nature* **471:** 473–479. doi:10.1038/nature09715

Guilgur LG, Prudencio P, Sobral D, Liszekova D, Rosa A, Martinho RG. 2014. Requirement for highly efficient pre-mRNA splicing during *Drosophila* early embryonic development. *Elife* **3:** e02181. doi:10.7554/eLife.02181

Gunderson SI, Polycarpou-Schwarz M, Mattaj IW. 1998. U1 snRNP inhibits pre-mRNA polyadenylation through a direct interaction between U1 70K and poly(A) polymerase. *Mol Cell* **1:** 255–264. doi:10.1016/S1097-2765(00)80026-X

Guo J, Garrett M, Micklem G, Brogna S. 2011. Poly(A) signals located near the 5′ end of genes are silenced by a general mechanism that prevents premature 3′-end processing. *Mol Cell Biol* **31:** 639–651. doi:10.1128/MCB.00919-10

Hachet O, Ephrussi A. 2004. Splicing of *oskar* RNA in the nucleus is coupled to its cytoplasmic localization. *Nature* **428:** 959–963. doi:10.1038/nature02521

Haselbach D, Komarov I, Agafonov DE, Hartmuth K, Graf B, Dybkov O, Urlaub H, Kastner B, Luhrmann R, Stark H. 2018. Structure and conformational dynamics of the human spliceosomal B^{act} complex. *Cell* **172:** 454–464. doi:10.1016/j.cell.2018.01.010

Herold N, Will CL, Wolf E, Kastner B, Urlaub H, Luhrmann R. 2009. Conservation of the protein composition and electron microscopy structure of *Drosophila melanogaster* and human spliceosomal complexes. *Mol Cell Biol* **29:** 281–301. doi:10.1128/MCB.01415-08

Hirose T, Ideue T, Nagai M, Hagiwara M, Shu MD, Steitz JA. 2006. A spliceosomal intron binding protein, IBP160, links position-dependent assembly of intron-encoded box C/D snoRNP to pre-mRNA splicing. *Mol Cell* **23:** 673–684. doi:10.1016/j.molcel.2006.07.011

Hogg R, McGrail JC, O'Keefe RT. 2010. The function of the NineTeen Complex (NTC) in regulating spliceosome conformations and fidelity during pre-mRNA splicing. *Biochem Soc Trans* **38:** 1110–1115. doi:10.1042/BST0381110

Hoskins AA, Friedman LJ, Gallagher SS, Crawford DJ, Anderson EG, Wombacher R, Ramirez N, Cornish VW, Gelles J, Moore MJ. 2011. Ordered and dynamic assembly of single spliceosomes. *Science* **331:** 1289–1295. doi:10.1126/science.1198830

Huynh JR, St Johnston D. 2004. The origin of asymmetry: early polarisation of the *Drosophila* germline cyst and oocyte. *Curr Biol* **14:** R438–R449. doi:10.1016/j.cub.2004.05.040

Ideue T, Sasaki YT, Hagiwara M, Hirose T. 2007. Introns play an essential role in splicing-dependent formation of the exon junction complex. *Genes Dev* **21:** 1993–1998. doi:10.1101/gad.1557907

Irimia M, Weatheritt RJ, Ellis JD, Parikshak NN, Gonatopoulos-Pournatzis T, Babor M, Quesnel-Vallieres M, Tapial J, Raj B, O'Hanlon D, et al. 2014. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell* **159:** 1511–1523. doi:10.1016/j.cell.2014.11.035

Izaurralde E, Lewis J, McGuigan C, Jankowska M, Darzynkiewicz E, Mattaj IW. 1994. A nuclear cap binding protein complex involved in pre-mRNA splicing. *Cell* **78:** 657–668. doi:10.1016/0092-8674(94)90530-4

Khodor YL, Rodriguez J, Abruzzi KC, Tang CH, Marr MT II, Rosbash M. 2011. Nascent-seq indicates widespread cotranscriptional pre-mRNA splicing in *Drosophila*. *Genes Dev* **25:** 2502–2512. doi:10.1101/gad.178962.111

Kinsella RJ, Kahari A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A, et al. 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. *Database (Oxford)* **2011:** bar030. doi:10.1093/database/bar030

Klattenhoff C, Bratu DP, McGinnis-Schultz N, Koppetsch BS, Cook HA, Theurkauf WE. 2007. *Drosophila* rasiRNA pathway mutations disrupt embryonic axis specification through activation of an ATR/Chk2 DNA damage response. *Dev Cell* **12:** 45–55. doi:10.1016/j.devcel.2006.12.001

Kuraoka I, Ito S, Wada T, Hayashida M, Lee L, Saijo M, Nakatsu Y, Matsumoto M, Matsunaga T, Handa H, et al. 2008. Isolation of XAB2 complex involved in pre-mRNA splicing, transcription, and transcription-coupled repair. *J Biol Chem* **283:** 940–950. doi:10.1074/jbc.M706647200

Lan L, Lin S, Zhang S, Cohen RS. 2010. Evidence for a transport-trap mode of *Drosophila melanogaster gurken* mRNA localization. *PLoS ONE* **5:** e15448. doi:10.1371/journal.pone.0015448

Lepennetier G, Catania F. 2016. mRNA-associated processes and their influence on exon-intron structure in *Drosophila melanogaster*. *G3 (Bethesda)* **6:** 1617–1626. doi:10.1534/g3.116.029231

Lewis JD, Izaurralde E, Jarmolowski A, McGuigan C, Mattaj IW. 1996. A nuclear cap-binding complex facilitates association of U1 snRNP with the cap-proximal 5′ splice site. *Genes Dev* **10:** 1683–1698. doi:10.1101/gad.10.13.1683

Lin K, Zhang DY. 2005. The excess of 5′ introns in eukaryotic genomes. *Nucleic Acids Res* **33:** 6522–6527. doi:10.1093/nar/gki970

MacDougall N, Clark A, MacDougall E, Davis I. 2003. *Drosophila gurken* (TGFα) mRNA localizes as particles that move within the oocyte in two dynein-dependent steps. *Dev Cell* **4:** 307–319. doi:10.1016/S1534-5807(03)00058-3

Mahajan K. 2016. hPso4/hPrp19: a critical component of DNA repair and DNA damage checkpoint complexes. *Oncogene* **35:** 2279–2286. doi:10.1038/onc.2015.321

Marechal A, Li JM, Ji XY, Wu CS, Yazinski SA, Nguyen HD, Liu S, Jimenez AE, Jin J, Zou L. 2014. PRP19 transforms into a sensor of RPA-ssDNA after DNA damage and drives ATR activation via a ubiquitin-mediated circuitry. *Mol Cell* **53:** 235–246. doi:10.1016/j.molcel.2013.11.002

Martinho RG, Guilgur LG, Prudencio P. 2015. How gene expression in fast-proliferating cells keeps pace. *Bioessays* **37:** 514–524. doi:10.1002/bies.201400195

Morisato D, Anderson KV. 1994. The spatzle gene encodes a component of the extracellular signaling pathway establishing the dorsal-ventral pattern of the *Drosophila* embryo. *Cell* **76:** 677–688. doi:10.1016/0092-8674(94)90507-X

Navarro-Costa P, McCarthy A, Prudencio P, Greer C, Guilgur LG, Becker JD, Secombe J, Rangan P, Martinho RG. 2016. Early programming of the oocyte epigenome temporally controls late prophase I transcription and chromatin remodelling. *Nat Commun* **7:** 12331. doi:10.1038/ncomms12331

Neuman-Silberberg FS, Schupbach T. 1993. The *Drosophila* dorsoventral patterning gene *gurken* produces a dorsally localized RNA and encodes a TGF α-like protein. *Cell* **75:** 165–174. doi:10.1016/S0092-8674(05)80093-5

Neuman-Silberberg FS, Schupbach T. 1994. Dorsoventral axis formation in *Drosophila* depends on the correct dosage of the gene *gurken*. *Development* **120:** 2457–2463.

Ni JQ, Zhou R, Czech B, Liu LP, Holderbaum L, Yang-Zhou D, Shim HS, Tao R, Handler D, Karpowicz P, et al. 2011. A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nat Methods* **8:** 405–407. doi:10.1038/nmeth.1592

Nolan T, Hands RE, Bustin SA. 2006. Quantification of mRNA using real-time RT-PCR. *Nat Protoc* **1:** 1559–1582. doi:10.1038/nprot.2006.236

Pabis M, Neufeld N, Steiner MC, Bojic T, Shav-Tal Y, Neugebauer KM. 2013. The nuclear cap-binding complex interacts with the U4/U6.U5 tri-snRNP and promotes spliceosome assembly in mammalian cells. *RNA* **19:** 1054–1063. doi:10.1261/rna.037069.112

Pai AA, Henriques T, McCue K, Burkholder A, Adelman K, Burge CB. 2017. The kinetics of pre-mRNA splicing in the *Drosophila* genome and the influence of gene architecture. *Elife* **6:** e32537. doi:10.7554/eLife.32537

Papasaikas P, Valcarcel J. 2016. The spliceosome: the ultimate RNA chaperone and sculptor. *Trends Biochem Sci* **41:** 33–45. doi:10.1016/j.tibs.2015.11.003

Popp MW, Maquat LE. 2013. Organizing principles of mammalian nonsense-mediated mRNA decay. *Annu Rev Genet* **47:** 139–165. doi:10.1146/annurev-genet-111212-133424

Prudencio P, Guilgur LG, Sobral J, Becker JD, Martinho RG, Navarro-Costa P. 2018. The Trithorax group protein dMLL3/4 instructs the assembly of the zygotic genome at fertilization. *EMBO Rep* **19:** e45728. doi:10.15252/embr.201845728

Qiu J, Cheng F, Pintel D. 2007. Distance-dependent processing of adeno-associated virus type 5 RNA is controlled by 5′ exon definition. *J Virol* **81:** 7974–7984. doi:10.1128/JVI.00714-07

Queenan AM, Barcelo G, Van Buskirk C, Schupbach T. 1999. The transmembrane region of Gurken is not required for biological activity, but is necessary for transport to the oocyte membrane in *Drosophila*. *Mech Dev* **89:** 35–42. doi:10.1016/S0925-4773(99)00196-3

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43:** e47. doi:10.1093/nar/gkv007

Rorth P. 1998. Gal4 in the *Drosophila* female germline. *Mech Dev* **78:** 113–118. doi:10.1016/S0925-4773(98)00157-9

Roth S, Neuman-Silberberg FS, Barcelo G, Schupbach T. 1995. *cornichon* and the EGF receptor signaling process are necessary for both anterior-posterior and dorsal-ventral pattern formation in Drosophila. *Cell* **81:** 967–978. doi:10.1016/0092-8674(95)90016-0

Sakurai A, Fujimori S, Kochiwa H, Kitamura-Abe S, Washio T, Saito R, Carninci P, Hayashizaki Y, Tomita M. 2002. On biased distribution of introns in various eukaryotes. *Gene* **300:** 89–95. doi:10.1016/S0378-1119(02)01035-1

Saunders C, Cohen RS. 1999. The role of oocyte transcription, the 5′UTR, and translation repression and derepression in *Drosophila gurken* mRNA and protein localization. *Mol Cell* **3:** 43–54. doi:10.1016/S1097-2765(00)80173-2

Schmittgen TD, Livak KJ. 2008. Analyzing real-time PCR data by the comparative C_T method. *Nat Protoc* **3:** 1101–1108. doi:10.1038/nprot.2008.73

Schneider DS, Jin Y, Morisato D, Anderson KV. 1994. A processed form of the Spatzle protein defines dorsal-ventral polarity in the *Drosophila* embryo. *Development* **120:** 1243–1250.

Schupbach T. 1987. Germ line and soma cooperate during oogenesis to establish the dorsoventral pattern of egg shell and embryo in *Drosophila melanogaster*. *Cell* **49:** 699–707. doi:10.1016/0092-8674(87)90546-0

Sterner DA, Carlo T, Berget SM. 1996. Architectural limits on split genes. *Proc Natl Acad Sci* **93:** 15081–15085. doi:10.1073/pnas.93.26.15081

Styhler S, Nakamura A, Swan A, Suter B, Lasko P. 1998. vasa is required for GURKEN accumulation in the oocyte, and is involved in oocyte differentiation and germline cyst development. *Development* **125:** 1569–1578.

Tapial J, Ha KCH, Sterne-Weiler T, Gohr A, Braunschweig U, Hermoso-Pulido A, Quesnel-Vallieres M, Permanyer J, Sodaei R, Marquez Y, et al. 2017. An atlas of alternative splicing profiles and functional associations reveals new regulatory programs and genes that simultaneously express multiple major isoforms. *Genome Res* **27:** 1759–1768. doi:10.1101/gr.220962.117

Thio GL, Ray RP, Barcelo G, Schupbach T. 2000. Localization of *gurken* RNA in *Drosophila* oogenesis requires elements in the 5′ and 3′ regions of the transcript. *Dev Biol* **221:** 435–446. doi:10.1006/dbio.2000.9690

Torres-Mendez A, Bonnal S, Marquez Y, Roth J, Iglesias M, Permanyer J, Almudi I, O'Hanlon D, Guitart T, Soller M, et al. 2019. A novel protein domain in an ancestral splicing factor drove the evolution of neural microexons. *Nat Ecol Evol* **3:** 691–701. doi:10.1038/s41559-019-0813-6

Van Buskirk C, Schupbach T. 2002. Half pint regulates alternative splice site selection in *Drosophila*. *Dev Cell* **2:** 343–353. doi:10.1016/S1534-5807(02)00128-4

Van De Bor V, Hartswood E, Jones C, Finnegan D, Davis I. 2005. *gurken* and the I factor retrotransposon RNAs share common localization signals and machinery. *Dev Cell* **9:** 51–62. doi:10.1016/j.devcel.2005.04.012

Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. 2002. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol* **3:** RESEARCH0034. doi:10.1186/gb-2002-3-7-research0034

Van Doren M, Williamson AL, Lehmann R. 1998. Regulation of zygotic gene expression in *Drosophila* primordial germ cells. *Curr Biol* **8:** 243–246. doi:10.1016/S0960-9822(98)70091-0

Vert JP, Foveau N, Lajaunie C, Vandenbrouck Y. 2006. An accurate and interpretable model for siRNA efficacy prediction. *BMC Bioinformatics* **7:** 520. doi:10.1186/1471-2105-7-520

Wan L, Huang J. 2014. The PSO4 protein complex associates with replication protein A (RPA) and modulates the activation of ataxia telangiectasia-mutated and Rad3-related (ATR). *J Biol Chem* **289:** 6619–6626. doi:10.1074/jbc.M113.543439

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11:** 377–394. doi:10.1089/1066527041410418

Zhao T, Graham OS, Raposo A, St Johnston D. 2012. Growing microtubules push the oocyte nucleus to polarize the *Drosophila* dorsal-ventral axis. *Science* **336:** 999–1003. doi:10.1126/science.1219147