Alternative splicing and its regulation in the malaria vector *Anopheles gambiae*

Bárbara Díaz-Terenti[1], Elena Gómez-Díaz[1*]

[1]Instituto de Parasitología y Biomedicina López-Neyra, Consejo Superior de Investigaciones Científicas (IPBLN, CSIC), Granada, Spain

*** Correspondence:**
Elena Gómez-Díaz
elena.gomez@csic.es

Abstract

Alternative splicing (AS) is a highly conserved mechanism that allows to expand the coding capacity of the genome by modifying how multiple isoforms are expressed or used to generate different phenotypes. Despite its importance in physiology and disease, genome-wide studies of AS are lacking in most insects, including mosquitoes. And, even for model organisms, chromatin associated processes involved in the regulation AS are poorly known. In this study, we investigated AS in the malaria mosquito *Anopheles gambiae* in the context tissue-specific gene expression and mosquito responses to a *Plasmodium falciparum* infection, as well as the relationship between patterns of differential isoform expression and usage with chromatin accessibility changes. For this, we combined RNA-seq and ATAC-seq data from *A. gambiae* midguts and salivary glands, and from infected and non-infected midguts. We report differences between tissues in the expression of 456 isoforms and in the use of 211 isoforms, highlighting the role of AS in tissue-specific gene expression regulation. Secondly, we find a clear and significant association between chromatin accessibility states and tissue-specific patterns of AS. The analysis of differential accessible regions located at splicing sites permitted the identification of several motifs resembling the binding sites of *Drosophila* TFs potentially involved in mosquito AS regulation. Finally, the genome-wide analysis of tissue-dependent enhancer activity revealed that approximately 20% of *A. gambiae* transcriptional enhancers annotate to a differentially expressed or used isoform and that their activation status is linked to AS differences between tissues. This research illuminates the role of AS in gene expression in vector mosquitoes, and identifies regulatory regions potentially involved in AS regulation, which could reveal novel strategies for vector control.

Keywords: Mosquitoes, Plasmodium, Isoform expression, Isoform usage, Chromatin accessibility

Abbreviations [1]

[1] **Abbreviations**: DEMG, Differential isoform expression of multisoform genes; DUI, Differential isoform usage of multisoform genes; SS, splicing sites.

1 Introduction

Alternative splicing (AS from here) is an important source of phenotypic variation in eukaryotes that allows the generation multiple transcripts from a single gene, expanding the coding capacity and functionality of the genome. AS regulates gene expression by acting on mRNA transcripts at a post-transcriptional level regulating cell differentiation, temporal and spatial expression of the isoforms, the activation of the RNA Pol II or the on-off regulation by nonsense-mediated decay (Kelemen et al., 2013; Nilsen and Graveley, 2010; Shenasa and Hertel, 2019; Ule and Blencowe, 2019; Wang and Cooper, 2007; Wright et al., 2022). Another function of AS is regulating isoform usage in combination with other mechanisms as selective degradation of isoforms or alternative polyadenylation sites (Ye et al., 2018). Isoform usage is defined as, the participation of each isoform in the gene expression, so when these proportions switch between two conditions, some isoforms are differentially used (Chen et al., 2022; De La Fuente et al., 2020; Li et al., 2020a; Merino et al., 2019; Vitting-Seerup and Sandelin, 2017). As a result, the function of gene might change because different variant proteins are produced, while the total gene expression remain unaltered.

AS is mediated by four types of basic events: alternative 5′ splice-site, alternative 3′ splice-site, cassette-exon inclusion or skipping, and intron retention. These processes can be combined to obtain dozens of isoforms from a single gene. The type of events and mechanisms of AS vary significantly amongst species. Despite of this, AS is highly conserved in eukaryotes playing fundamental functions in health and disease, adaptation to the environment, and defence against pathogens (Bush et al., 2017; Kelemen et al., 2013; Li et al., 2020a; Martinez and Lynch, 2013; Nilsen and Graveley, 2010; Song et al., 2020; Ule and Blencowe, 2019; Verta and Jacobs, 2022; Vitting-Seerup and Sandelin, 2017; Wright et al., 2022; Zhao, 2019).

A simple way to study AS is to quantify and compare the expression of isoforms that are generated in different tissues or experimental conditions (Glaus et al., 2012; Katz et al., 2010). Transcriptomic data obtained by RNA-seq is mostly used for gene level analysis (Vitting-Seerup and Sandelin, 2017; Zhao, 2019), but offers the possibility to study AS to analyze and count known transcripts or to detect new splicing events (Zhao, 2019).

Regulation of AS is intrinsically related to chromatin dynamics. To obtain diverse transcripts, *cis* and *trans*-regulatory elements participate both in the generation of transcripts and in marking splice sites (Bush et al., 2017; Shenasa and Hertel, 2019; Verta and Jacobs, 2022; Zhao, 2019). In addition, AS regulation can be affected by posttranslational histone modifications, nucleosome positioning, and chromatin accessibility (Davuluri et al., 2008; Klemm et al., 2019; Luco et al., 2011; Montes et al., 2012; Nussinov et al., 2021). However, how these different epigenetic layers interact, and which are the mechanisms that underlie the great diversification of gene functions through AS, has been poorly investigated (Davuluri et al., 2008; Klemm et al., 2019; Luco et al., 2011). The integrated analysis of transcriptomic and epigenetic data could bring new insights into how AS is regulated and how it changes under different conditions.

AS has been widely studied in model organisms like *Drosophila melanogaster*, *Danio renio* or *Caenorhabditis elegans* (Lin et al., 2022; Venables et al., 2012; Zahler, 2005). In *Drosophila melanogaster* AS plays important roles in sex-determination, muscle, neural development among other biological processes (Venables et al., 2012). Only one study has examined the relationship between AS and chromatin structure in *Drosophila* gonads. That work revealed co-enrichment of sex-specific isoforms and RNA splicing factors together with chromatin regulators and histone modifying enzymes in the undifferentiated cells (Gan et al., 2010). Despite of this, there is a lack of knowledge on how chromatin operates at the level of AS to

regulate gene function under different physiological or cellular conditions (Kelemen et al., 2013; Luco et al., 2011; Nilsen and Graveley, 2010).

Mosquitoes are major disease vectors and a big threat for human health worldwide. Given that AS is evolutionarily conserved, one would expect that AS functions and mechanisms are maintained across insects. There is, however, marked inter-species variation in alternative splicing that probably resulted through selective pressures exerted by environment and lifestyle (Kurtz and Armitage, 2006; Malko et al., 2006). In *Aedes aegypti*, previous studies have focused on sex differentiation (Calkins et al., 2015; Salvemini et al., 2013, 2011). In the human malaria mosquito *Anopheles gambiae*, the study of AS is limited to a few genes, such as *doublesex* and *Dscam* (Djihinto et al., 2022; Krzywinska et al., 2016; Sinkins, 2016; Tsujimoto et al., 2013), tissues like salivary glands (Dixit et al., 2009) or signalling pathways (Meister et al., 2005). These studies suggest that AS could play a crucial role in vector competence and mosquito immunity. However, a more systematic analysis onto the role of AS and its regulation in mosquitoes, is lacking.

In an earlier study, we combined transcriptomic and chromatin accessibility data of infected and non-infected *A. gambiae* salivary glands and midguts, to identify *cis*-regulatory networks involved in tissue-specific gene expression regulation and mosquito responses to malaria infection (Ruiz et al., 2019). In this study, we used this data to interrogate which isoforms belonging to multisoform genes are differentially regulated in different mosquito tissues infected with the malaria parasite *Plasmodium falciparum*, and whether chromatin-associated mechanisms underlie these events (Love et al., 2014; Vitting-Seerup et al., 2019). We contend that a deeper understanding of the mechanisms AS regulation can reveal key *cis* and *trans* regulatory elements, like splicing transcription factors and enhancers, involved in mosquito responses to *P. falciparum* infection and/or vector competence.

2 Material and Methods

2.1 Genes susceptible of alternative splicing and experimental design

To study the role of Alternative Splicing in *A. gambiae* we considered genes that are susceptible to be regulated by this mechanism. Only 1,263 genes in the *A. gambiae* genome (13,094 genes) have more than one transcript and could be classified as multisoform genes.

We used 12 RNA-seq and 4 ATAC-seq datasets corresponding to 8 *A. gambiae* infected and 4 non-infected (control) samples from two mosquito tissues (midguts and salivary glands), obtained in a previous study, some of which have already been published (José L. Ruiz et al., 2021; Ruiz et al., 2019). The samples are distributed in three different comparisons: infection vs control in midguts and salivary glands, and infected midguts vs infected salivary glands, having each of them 2 replicates per condition.

2.2 RNA-seq and ATAC-seq processing analysis

Quality control of RNA-seq data was performed with FastQC v0.11.9 ("Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data," n.d.). A replicability test based on correlation analysis and principal component analysis (PCA), support the use of the different samples as replicates (Figure S1).

We used SortMeRNA v4.2.0 (Kopylova et al., 2012) to determine and eliminate possible ribosomal RNA contamination. Filtered reads were then mapped to the reference genome (Anopheles gambiae PEST genome, version P4.12) available in VectorBase (release 53) (Giraldo-Calderón et al., 2015), using Salmon v.0.11.2 (Patro et al., 2017). This software is specifically used for the alignment and quantification of isoforms (Zhang et al., 2016) and was run using the standard parameters *-l A --validateMappings --go 4 --mismatchSeedSkip 5*. For this analysis we applied a conservative filter, only isoforms with a count value greater than 10 were considered as being expressed. A total number of 1,561 (Inf MG vs. Inf SG comparison), 1,394 (Inf vs. Ctrl MG), 1,679 (Inf vs. Ctrl SG) isoforms passed this threshold and were included in all subsequent analysis.

For the analysis of ATAC-seq data in relation to AS, we took the nucleosome free reads, the ATAC peaks resulting from the analysis with MACS2 (Zhang et al., 2008) and the differentially accessible peaks resulting from the analysis with Diffbind (Stark and Brown, n.d.) from an article previously published by our group (José L Ruiz et al., 2021). We then quantified the ATAC-seq nucleosome-free signal enrichment at genomic regions of interest. The count of the ATAC reads was performed using BEDTools v2.30.0 ("bedtools: a powerful toolset for genome arithmetic — bedtools 2.30.0 documentation," n.d.) with the parameter *intersect* and the option ***-c*** and read counts were normalized (RPKM).

2.3 AS mechanisms and differential analyses

The counts generated by Salmon were used to perform various analyses. First, to study the most prevailing alternative splicing mechanisms in each condition we used the R package IsoformSwitchAnalyceR v1.12.0 (Vitting-Seerup et al., 2019). We applied the function *analyzeAlternativeSplicing* with the argument *onlySwitchingGenes=F*. We then compared the frequencies of the events for the DEMG and DUI groups between conditions (tissues and infection status) to determine how are they altered.

The differential expression analysis was performed using the R package (v4.0.2) DESeq2 v1.30.0, (Love et al., 2014). Taking multisoform genes, we considered differentially expressed isoforms (DEMG), those with a significancy *P-value* below 0.05 and a LogFoldChange greater than 1 or lower than -1 (corresponding to a change of the 10%). In the comparison between Inf MG vs. Inf SG a LogFC < 0 means overexpression in MG, whereas a LogFC > 0 means overexpression in SG. In the comparison between Inf vs. Control a LogFC > 0 means more expression in the infected tissue.

The study of the differential usage of the isoforms was performed using IsoformSwitchAnalyceR v1.12.0 (Vitting-Seerup et al., 2019). Isoforms significantly differentially used (DUI) were those with a significancy P-value below 0.05. In this case we used the parameter, difference in isoform fraction (dIF) to measure the magnitude of change. The isoform fraction (IF) is a measurement that determines the proportion of the total expression of a parent gene that comes from a particular isoform (isoform_exp / gene_exp). In the comparison between Inf MG vs. Inf SG a dIF < 0 means more used in MG, and dIF > 0 means more used in SG. In the comparison between Inf vs. Control a dIF < 0 means more usage in the infected tissue.

We performed Gene Ontology (GO) terms overrepresentation tests for the sets of differentially expressed or used multisoform genes between Infected midguts *vs.* Infected salivary glands using VectorBase Overrepresentation Test and ImmunoDB database to know which isoforms are immunologically relevant ("EZlab | Computational Evolutionary Genomics group," n.d.; Giraldo-Calderón et al., 2015). We chose the p-value as a measure of the likelihood that a

certain GO term appear among the genes in your results more often than it appears in the set of all genes for that organism (background) and we applied a threshold of p-value < 0.05.

2.4 Study of the relationship between expression and accessibility

For DEMG group of isoforms we discarded genes that overlap more than a 25% with other genes. Then we use the Spearman's correlation test from R v4.2.0 to analyze the association between isoform expression levels and chromatin accessibility enrichment in various genomic regions where the AS mechanisms would operate: The isoform promoter, which comprises 1Kbp from the TSS (end 5' UTR) based in the location of the majority of MACS2 peaks (Figure S2) and the isoform body including all exons and introns. For those isoforms that lack a TSS annotated, we considered the ATG as the reference point.

We used a R script to classify the multisoform genes into three groups - high, medium, and low - depending on the expression level of the isoforms. The threshold values for this classification were determined by dividing the signal into three quantile groups based on their means, using the Hmisc::cut2 R function. We then correlated isoform expression and chromatin accessibility for each group separately. The software *ngsplot* was used to visualize as Average profile plots the chromatin accessibility (nucleosome free region signal) at the TSS in the different groups of isoforms classified by their expression level (high, medium, low) (Shen et al., 2014).

In the case of the DUI group, we analyse the association between isoform usage and chromatin accessibility focusing on sites of alternative isoform splicing. To do this, we first searched for significantly differential peaks between conditions (previously obtained by DiffBind) that coincide with splicing sites (SS) (+/- 100bp) and counted how many of those peaks annotate to isoforms classified as DUI.

Taking the peaks that are located in a splicing site and annotate to DUI isoforms, we performed a motif analysis using HOMER software v4.11 ("Homer Software and Data Download," n.d.) We used *findMotifGenome.pl* to determine known and *de novo* motifs overrepresented in our sequences, and then used *annotatePeaks.pl* to link the motifs back to the DUI isoforms and the splicing sites. Out of the results obtained by HOMER, only motifs which significance p-values less than 1e-25 were considered.

2.5 Enhancers associated to AS in *A. gambiae*

We used the set of *A. gambiae* enhancers characterized in a previous study (José L. Ruiz et al., 2021), to interrogate whether these regulatory sequences could be involved in AS events based on the integrative analysis of the ATAC-seq and RNA-seq data. The list included 1,402 *A. gambiae* enhancers predicted by homology with *Drosophila* enhancers and, 2,866 *Anopheles* enhancers identified either computationally or experimentally by different studies (José L. Ruiz et al., 2021). Of this list, we selected only the enhancers that appeared to be active in at least one of the conditions, that is, enhancers that have a MACS2 chromatin accessibility peak and appear enriched in H3K27ac. This resulted in a list of 811 active enhancers which we overlapped with our DEMG and DUI groups using BEDtools *intersect*. We also classified these enhancers as intragenic or intergenic using BEDtools *intersect*.

Finally, we quantified which portion of the active enhancers annotated to DEMG or DUI isoforms display a change in accessibility, that is, that are differentially accessible between tissues, coinciding with the change in the expression of the isoform/s.

3 Results

3.1 Alternative splicing (AS) regulates tissue-specific gene expression in infected A. gambiae

We first describe the types of events that lead to differential isoform expression (DEMG) and/or isoform usage (DUI) in our set of *A. gambiae* multisoform genes. Our results show Exon Skipping (ES) and Alternative Start Site (ATSS) as the two most frequent types of events independently of the group of isoforms and comparison. We report differences in the frequency of these events between conditions in the DEMG group for both ES and ATSS: 68 ES events in *Inf MG vs*. 102 ES events in *Inf SG* and in the case of ATSS, 89 ATSS events in *Inf MG vs*. 105 ATSS events in *Inf SG*. In the DUI group, only the frequency of ATSS when comparing Inf MG vs. Inf SG is moderately altered (*43 ATSS events in SG compared to 51 events in MG*).

Considering the 1,263 multisoform genes detected in *A. gambiae*, we aimed to identify alternative splicing events and their role in gene expression regulation in different tissues and infection conditions. We perform two types of analysis: one to identify differentially expressed isoforms belonging to multisoform genes (DEMG). The other analysis detects isoforms differentially used (DUI); isoforms that are altered in their relative proportion to the total expression of the gene.

For these two analyses we first compared infected and non-infected (control) mosquito midguts. Out of the 848 multisoform genes present in this dataset (with counts > 10 as the detectability threshold), we found 32 isoforms that are altered in the context of a *P. falciparum* midgut infection. Of these, 19 (2%) correspond to DEMG and 13 (1,5%) are DUI. In salivary glands, we detected 977 expressed multisoform genes and 91 differentially regulated isoforms; 56 (6%) of which are DEMG and 35 (3,5%) DUI.

Next, we examined differences in AS regulation between tissues: midguts and salivary glands infected with *P. falciparum*. We detected 920 multisoform genes: 456 (50%) DEMG and 211 (23%) DUI. In addition to these categories, it is possible to find multisoform genes in which one or more isoforms are differentially expressed and one or more isoforms are differentially used (**Error! Reference source not found.**a). In our results, this overlaps between DEMG and DUI involves 22% (105) of the multisoform genes that display AS when comparing midguts and salivary glands (**Error! Reference source not found.**b), and less than 1% when comparing infected and non-infected mosquito tissues (Supplementary table S1 and S2).

To test for differential functionality between tissues, we performed GO enrichment analysis in each group of tissue-specific isoforms and type of multisoform genes (DEMG and DUI). In midguts, the DEMG group is enriched GO terms categorized as catabolic process, cytoskeleton organization, as well as terms related to immune responses. Similarly, in salivary glands, the DEMG genes display functions involved in ion transport, cytoskeleton organization, and mosquito immunity (Supplementary table S3).

DUI isoforms overused in midguts display functions related to catabolic process of carbohydrates and aminoacids, melanization responses, muscle contraction and regulation. In salivary glands we find similar functions for DUI isoforms in relation muscle function, cellular homeostasis and immunological responses related to the pathways JAK-STAT (Supplementary table S3).

We compared our differentially expressed and used isoforms with genes previously reported by other authors which play important functions in mosquito immunity (Table S4) (Sreenivasamurthy et al., 2013). We found 8 genes in our DEMG/DUI datasets that encode for

proteins involved in anti-*Plasmodium* mosquito responses (STAT, PGRPLC, OXR1). One example are the AGAP005203-RA and AGAP005203-RE isoforms of the PGRPLC gene identified in the present work as differentially expressed (DEMG). The RA isoform is overexpressed during midgut infection, while RE isoform, is overrepresented in infected salivary glands compared to control. RB and RC appear similarly expressed and the rest of isoforms have values below the expression threshold, so we consider them as non-expressed (**Error! Reference source not found.**c).

3.2 Alternative Splicing is positively correlated with chromatin accessibility in malaria-infected *A. gambiae* tissues

3.2.1 Model and predictions

The relationship between the gene expression and chromatin structure has been examined in a few model organisms (Luco et al., 2011). These previous studies showed that various epigenetic mechanisms such as nucleosome positioning, post-transcriptional modifications (Luco et al., 2011) or the modification of the chromatin accessibility (Li et al., 2020b; Nussinov et al., 2021), alter or correlate with AS dynamics. In *A. gambiae*, an earlier work investigated the relationship between chromatin accessibility and gene expression regulation at gene level (José L. Ruiz et al., 2021), while here we focus on gene regulation occurring at the isoform level via alternative splicing.

We have made some predictions on how chromatin structure may be altered when comparing different mosquito tissues in which AS would play a role in gene expression regulation (see previous section), and the pattern expected considering DEMG and DUI genes. Our expectation is to find a positive relationship between chromatin accessibility differences and the differential expression and/or usage of the isoforms.

For the DEMG group, we expect to find an increased accessibility in the promoter of one or more isoforms in the tissue where the expression of the isoform/s is higher (Figure 2a). In the DUI group, we expect to find changes in chromatin accessibility at regulatory regions coinciding or near alternative splicing sites in the isoform more used in each tissue, as is shown in **Error! Reference source not found.**b. For example, for an exon skipping event we expect more accessibility in the surroundings of the exon to be added to the transcript. In our data the event more used is the ES and the majority of the MACS2 peaks are located in the isoform body. Alternatively, this association can also be evident at the promoter of differentially used isoforms for AS events like ES, ATTS and IR (Li et al., 2020b).

3.3 Tissue-specific isoform expression is positively associated to differential chromatin accessibility

First, we tested if higher expression of the isoforms is associated with higher levels of chromatin accessibility in the isoform promoter (1kb upstream of the TSS/ATG of the isoform) or the isoform body. The results for the DEMG group show a positive and significant correlation between isoform expression and chromatin accessibility for the isoform body, both in salivary glands (*rho* = 0,328, *p-value* = 8,095e-07) and midguts (*rho* = 0,369, *p-value* = 3.009e-07), but not for the promoter region (Table S5A).

Next, we examined whether the relationship is quantitative, that is, if the levels of expression of the isoform differentially expressed (categorized as high, medium, or low) correlate with the

degree of chromatin accessibility in the isoform promoter. This relationship appears not significant either in midguts or salivary glands (*rho* < 0,02, *p-value* > 0.1 for all levels and tissues, Table S5B) (Figure S3).

For the analysis of chromatin accessibility in relation to differential usage, we used data of differential accessibility peaks between conditions published previously by our group (José L. Ruiz et al., 2021). We then selected DiffBind peaks that are located at the splicing site or in the region nearby (+/-100 Kb), and that annotate to DUI isoforms. We report 98 DUI isoforms that display a DiffBind peak located at or nearby splicing sites, representing the 54,4% of the total number of isoforms differentially used.

3.3.1.1 Motif analysis

On the set of DiffBind peaks that are located at or nearby splicing sites for DUI isoforms we perform a motif analysis using HOMER. The purpose of this analysis is to detect overrepresented motifs that point to transcription factors involved in mosquito AS regulation.

Using the hierarchical clustering of MEME *de novo motif discovery* option, several motif hits with distinct similarities with *Drosophila* sequences appear overrepresented (summary in Table 1) (Table S6B). For the subset of DUI isoforms, we find *de novo* motifs related with brachyentron (byn), implicated in midgut development, snail (sna), which participates in epithelial transition, grainy head (grh) involved in epithelial cell fate or chromatin remodellers like the Chromatin-linked adaptor for MSL proteins (CLAMP).

Now analyzing the set of "known" motifs, we found only two motifs overrepresented. The motif CATCMCTA, which is a *Drosophila* promoter element of unknown function, and the motif GGYCATAAAW corresponding to the caudal homeobox of *Drosophila* (Table S6A).

3.3.2 Mosquito enhancers involved in AS

We aimed to study if *A. gambiae* previously identified transcriptional enhancers also participates in the regulation of AS. Such a relationship has been shown in human or mouse where a 5% of the enhancers are associated to AS changes (Shiau et al., 2021). For this analysis, we used the *A. gambiae* enhancers described by our lab in a previous study (José L. Ruiz et al., 2021). This dataset included 4,268 enhancers identified computationally by orthology with *Drosophila* enhancers, or experimentally discovered by others in mosquitoes, and that were validated by ours using the combined analysis of differential chromatin accessibility and histone modifications marking, as well as gene expression changes between tissues (José L Ruiz et al., 2021). Of these mosquito enhancers, 811 appear active in our dataset.

In the DEMG group we found a total of 125 accessible enhancers that annotate to 76 genes (21,1% of the 360 DEMG multisoform genes) (Supplementary Table S7A). Of these 125 accessible (active) enhancers, 26 of them coincide with a Diffbind peak, that is, they are differentially accessible between infected midguts and salivary glands. The GO terms associated with the genes targeted by these active enhancers, are related with muscle development, homeostasis, neuron development and different classes of immune responses (Table S8A).

For the DUI group we identified 39 accessible enhancer regions that annotate to 23 genes (19,8% of the 116 DUI multisoform genes) (Supplementary Table S7B). Out of these enhancers, 11 of them exhibit Diffbind peaks, indicating differential accessibility between infected midguts

and salivary glands. The GO terms appear associated with neural development, homeostasis, lipidic metabolic processes and responses and metabolism of superoxide (Table S8B).

Lastly, we examine the location of the enhancers annotated to DEMG and DUI isoforms. For both, the DEMG and DUI groups, the number of intragenic enhancers is almost double of the intergenic (DEMG: 44 intergenic and 81 intragenic enhancers; DUI: 14 intergenic and 25 intragenic enhancers).

One example is the gene AGAP005234 that encodes for a copper-zinc superoxide dismutase. This protein has functions involved in mosquito immunity, including anti-Plasmodium responses, as well as response to oxygen radicals. (Marikovsky et al., 2003; Molina-Cruz et al., 2008). The gene contains two different isoforms (AGAP005234-RA and AGAP005234-RB), RB isoform is differentially expressed and used in salivary glands and RA is differentially used in midguts. The enhancer associated to this gene is intergenic, located 2 Kbp upstream of the start of the gene, and its accessibility in higher in SG coinciding with the expression and use that is also higher in SG (**Error! Reference source not found.**).

4 Discussion

Alternative splicing has been relatively little studied in mosquitoes, and the relationship of this process with chromatin dynamics remains mostly unknown. In this study we investigated alternative splicing in *A. gambiae* and the role of chromatin structure in its regulation. We compared patterns of AS in different conditions, first between two mosquito tissues: midguts and salivary glands, and second between *P. falciparum* infected and non-infected (control) mosquito midguts and salivary glands.

The first part of this work aimed at identifying how prevalent is AS in mosquitoes and how AS patterns change between tissues and infection status. From a total of 1,263 *A. gambiae* genes susceptible of AS, we detect around a 67-77% being differentially expressed or used at the level of isoforms.

These differences were largely due AS events at the tissue-level. Indeed, when comparing the two infected tissues, the 49% of isoforms (456) appeared differentially expressed (DEMG) and 22,9% (211) differentially used (DUI), suggesting that AS plays an important role in regulating tissue-specific expression in *A. gambiae*. The role of AS in tissue specific processes has been reported in different organisms like in humans, *D. melanogaster*, *C. elegans* (Ragle et al., 2015; Telonis-Scott et al., 2009; Venables et al., 2012; Wang et al., 2008) or even in *Anopheles stephensi* (Sreenivasamurthy et al., 2017).

However, while in this study the AS patterns vary significantly between tissues, differences were minimal when comparing infected and uninfected tissues for both DEMG and DUI genes (only 32 DEMG and DUI isoforms detected in case of midgut infection and 91 DEMG and DUI isoforms detected in case of salivary glands infection). These results can be explained by the inter-sample variability in these comparisons as revealed by the PCA analysis. These differences between samples may be reducing the statistical power of the expression and differential use tests, resulting in low numbers of DUI and DEMG isoforms.

Among the tissue-specific multisoform genes, we report several genes with immune functions that have a specific isoform differentially expressed in midguts and another isoform differentially expressed in salivary glands, meaning that each has a different function depending on the tissue in which it is expressed. An example of this is the AGAP005203 gene, that encodes for a Peptidoglycan recognition protein (PGRPLC) and is involved in the Imd pathway (Meister

et al., 2009). This protein has been reported to be implicated in anti-*Plasmodium* responses at the level of the mosquito midgut, but without little information about the functionality of their isoforms (Sreenivasamurthy et al., 2013). This gene consists of six isoforms corresponding to three different variants. PGRPLC gene produces three primary protein isoforms each with a distinct PGRP domain and an optional 75-nucleotide cassette located at the 3' end of the shared exon 3. Consequently, each isoform is present in two forms, the longer version is 25 amino acids longer than the shorter version. AGAP005203 has 2 isoforms differentially expressed in our data all of them corresponding to the shorter version of the variants. RA form has been reported to be overexpressed during midgut infection and has been proposed to be involved in the mosquito immune response against *Plasmodium* (Meister et al., 2009). On the other side, the RE isoform, was overexpressed in salivary glands, and although its function in this tissue is still unknown. previous work reported that the RD and RE isoforms are positively regulated in response to stress and antimicrobial responses (Chen et al., 2009; Lin et al., 2007; Meister et al., 2009).

Chromatin dynamics and transcriptional regulation are intimately related, but its role in the regulation of alternative splicing remains little known. There are very few studies that investigated how epigenetic mechanisms such as histone modifications or nucleosome positioning, would contribute to changes in the AS pattern. This is despite the functional consequences that AS events can have during development and disease (Petrova et al., 2022b; Simon et al., 2014). Previous evidence in model organisms reported local changes in nucleosome positioning histone modifications and chromatin accessibility near the exon that undergoes splicing (Chodavarapu et al., 2010; Li et al., 2020b; Naftelberg et al., n.d.).

Here, we propose a model for *A. gambiae* to describe the association between isoform expression and/or usage and chromatin structure dynamics. For DEMG isoforms we expect increased promoter accessibility in the isoform that is more expressed (Li et al., 2020b; José L. Ruiz et al., 2021). For the DUI isoforms, we expect accessibility changes associates to the differential use, but localized at or nearby AS regulatory regions (splicing sites). In agreement with this model, our results indicate that there is a relationship between the differential expression and the use of isoforms with changes in chromatin accessibility enrichment suggesting that chromatin structure would play a role in AS-mediated gene expression regulation in *A. gambiae*. However, this relationship was only significant in the region of the isoform body. This is supported by previous studies in mammals and plants, reporting higher chromatin accessibility coinciding with the region of the intron retention (Petrova et al., 2022a; Ullah et al., 2018).

In addition, changes in accessibility along the isoform body have also been reported to alter the speed of RNA polymerase II (RNA pol II) (Saldi et al., 2016) or, alternatively, impact the recruitment of splicing factors through direct or indirect recognition of the different chromatin marks (Schor et al., 2013). Indeed, an increase in chromatin accessibility in histone-depleted genes was shown to alter the elongation rate of Pol II, resulting in several transcripts with increased intron retention and changes in alternative splicing (Jimeno-González et al., 2015).

For DUI isoforms, we already predict that changes at the level of splicing sites would be more important. Our results confirmed the existence of differential accessibility peaks at splicing sites for around 54% of the isoforms. What we cannot predict from our results, however, are the precise mechanism and the functional consequences of this change, that is, whether a local increase in chromatin accessibility at the splicing site would result in one or another alternatively spliced isoform due the selective binding of splicing repressor versus an inducer. Indeed, according to a recent study, the type of event and the choice of a particular isoform would be multifactorial and depend on a trade-off between chromatin accessibility at regulatory

sequences (such as splicing sites), and variable RNA pol II elongation rates (Pai and Luca, 2019).

Motif analysis can give some insights on the underlying mechanisms of AS regulation. Looking at the *de novo* motifs identified in the DEMG and DUI isoforms, we found overrepresented motifs that resemble several transcription and slicing factors of *Drosophila* such as snail (sna), grainy head (grh) and chromatin-linked adaptor for MSL proteins (CLAMP). In the case of CLAMP, this is a chromatin remodeling protein with many different functions in regulating RNA Pol II dynamics, chromatin architecture, and alternative splicing (Schuettengruber et al., 2009) (Urban et al., 2017). It has been described for example, that the interaction of CLAMP with the proteins Squid (Sqd) and Syncrip (Syp), members of the heterogeneous nuclear ribonucleoproteins (hnRNPs), has a role in the regulation of the AS in *Drosophila* (Blanchette et al., 2009).

Another aspect that we approached in this study is the relationship between the enhancer function, based on histone PTMs and accessibility patterns, and AS events in *A. gambiae*. Previous evidence suggests that enhancers can mediate isoform-specific expression by regulating the production and binding of antisense transcripts, as well as by regulating messenger RNA 3′ end processing (Kwon et al., 2022; Onodera et al., 2012; Shiau et al., 2021). In such cases, enhancers are expected to be located intragenically, near the exons or introns that undergo AS (Lam and Hertel, 2002; Matlin et al., 2005).

When we look at the presence of enhancers in our AS susceptible isoforms (DEMG and DUI), we discovered that approximately 19-23% of differentially expressed and used isoforms have an enhancer annotated, and approximately 18-20% of these are active in our study, i.e. accessible. This observation suggests a relationship of the enhancers with the expression and use of the isoforms. Indeed, we also report that 20% of the enhancers active in DEMG and 28% in DUI correspond to differentially accessible enhancers, i.e. they have differential activation status, between tissues. Furthermore, most of our accessible enhancers (64,6% of the total of both groups) appear to be intragenic. Although their function in alternative splicing still needs to be experimentally confirmed, the fact that their activation status changes according to changes in the expression and use of the isoforms to which they are annotated, suggests a possible role in AS-mediated isoform expression regulation.

One example of a AS gene with an associated enhancer that is differentially accessible between conditions is the immune gene AGAP005234. In *A. gambiae* the function of this gene is associated the production of reactive oxygen species (ROS), that have been implicated in the immune response to pathogens (Marikovsky et al., 2003), including a *Plasmodium* infection (Molina-Cruz et al., 2008). Our data shows that the RB isoform is more used and expressed in salivary glands, whereas the RA isoform is only used in midguts. There is an enhancer annotated to the AGAP005234 gene that is more accessible in salivary glands, suggesting a possible regulatory function of this enhancer in determining the choice of which of the isoforms of this gene is used (RA or RB) or differentially expressed (RB).

To conclude, this is, to our knowledge, the first study that analyze comprehensively alternative splicing and its regulation in *A. gambiae*. We report genome-wide alterations in isoform expression and usage between different tissues and, less so, linked to the infection status. Also, the study of the relationship between AS with chromatin, points to the importance of chromatin structure dynamics at splicing sites in tissue-specific gene expression regulation. Finally, the identification of active enhancers associated to differentially expressed and used isoforms suggests a possible role of these regulatory elements on the splicing process. Our study lays the foundation for further research on the mechanisms of alternative splicing in *A. gambiae*,

including the functional characterization of regulatory elements like enhancers and splicing factors as potential candidates for vector control.

5 Acknowledgements

We are grateful to Dr Lisa Ranford-Cartwright and Dr José L. Ruiz for sharing data. We thank the bioinformatics unit of the IPBLN-CSIC for computing facilities and advice on the analysis.

6 Funding sources

This work was supported by the Spanish Ministry of Science and Innovation [grant no. PID2019-111109RB-I00]. BDT is funded by a doctoral training fellowship [grant no. PRE2020-095076].

7 References

Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data [WWW Document], n.d. URL https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 11.11.22).

bedtools: a powerful toolset for genome arithmetic — bedtools 2.30.0 documentation [WWW Document], n.d. URL https://bedtools.readthedocs.io/en/latest/ (accessed 12.1.22).

Blanchette, M., Green, R.E., MacArthur, S., Brooks, A.N., Brenner, S.E., Eisen, M.B., Rio, D.C., 2009. Genome-wide analysis of alternative pre-mRNA splicing and RNA binding specificities of the Drosophila hnRNP A/B family members. Mol Cell 33, 438. https://doi.org/10.1016/J.MOLCEL.2009.01.022

Bush, S.J., Chen, L., Tovar-Corona, J.M., Urrutia, A.O., 2017. Alternative splicing and the evolution of phenotypic novelty. Philosophical Transactions of the Royal Society B: Biological Sciences 372. https://doi.org/10.1098/RSTB.2015.0474

Calkins, T.L., Woods-Acevedo, M.A., Hildebrandt, O., Piermarini, P.M., 2015. The molecular and immunochemical expression of innexins in the yellow fever mosquito, Aedes aegypti: insights into putative life stage- and tissue-specific functions of gap junctions. Comp Biochem Physiol B Biochem Mol Biol 183, 11. https://doi.org/10.1016/J.CBPB.2014.11.013

Chen, L., Chen, K., Hong, Y., Xing, L., Zhang, J., Zhang, K., Zhang, Z., 2022. The landscape of isoform switches in sepsis: a multicenter cohort study. Scientific Reports 2022 12:1 12, 1–16. https://doi.org/10.1038/s41598-022-14231-9

Chen, Y., Ling, E., Weng, Z., 2009. Functional characterization of PGRP-LC1 of Anopheles gambiae through deletion and RNA interference. Insect Sci 16, 443–453. https://doi.org/10.1111/J.1744-7917.2009.01296.X

Chodavarapu, R.K., Feng, S., Bernatavichute, Y. V., Chen, P.Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., Casero, D., Bernal, M., Huijser, P., Clark, A.T., Krämer, U., Merchant, S.S., Zhang, X., Jacobsen, S.E., Pellegrini, M., 2010. Relationship between nucleosome positioning and DNA methylation. Nature 2010 466:7304 466, 388–392. https://doi.org/10.1038/nature09147

Davuluri, R. V., Suzuki, Y., Sugano, S., Plass, C., Huang, T.H.M., 2008. The functional consequences of alternative promoter use in mammalian genomes. Trends in Genetics 24, 167–177. https://doi.org/10.1016/J.TIG.2008.01.008

De La Fuente, L., Arzalluz-Luque, Á., Tardáguila, M., Del Risco, H., Martí, C., Tarazona, S., Salguero, P., Scott, R., Lerma, A., Alastrue-Agudo, A., Bonilla, P., Newman, J.R.B., Kosugi, S., McIntyre, L.M., Moreno-Manzano, V., Conesa, A., 2020. TappAS: A comprehensive computational framework for the analysis of the functional impact of differential splicing. Genome Biol 21, 1–32. https://doi.org/10.1186/S13059-020-02028-W/FIGURES/7

Dixit, R., Sharma, A., Mourya, D.T., Kamaraju, R., Patole, M.S., Shouche, Y.S., 2009. Salivary gland transcriptome analysis during Plasmodium infection in malaria vector Anopheles stephensi. International Journal of Infectious Diseases 13, 636–646. https://doi.org/10.1016/J.IJID.2008.07.027

Djihinto, O., Saizonou, H.D.M., Djogbenou, L.S., Nolan, T., Lee, Y., 2022. Single nucleotide polymorphism (SNP) in the doublesex (dsx) gene splice sites and relevance for its alternative splicing in the malaria vector Anopheles gambiae [version 1; peer review: 1 approved with reservations, 1 not approved]. https://doi.org/10.12688/wellcomeopenres.17572.1

EZlab | Computational Evolutionary Genomics group [WWW Document], n.d. URL https://www.ezlab.org/ (accessed 11.14.22).

Gan, Q., Chepelev, I., Wei, G., Tarayrah, L., Cui, K., Zhao, K., Chen, X., 2010. Dynamic regulation of alternative splicing and chromatin structure in Drosophila gonads revealed by RNA-seq. Cell Res 20, 763. https://doi.org/10.1038/CR.2010.64

Giraldo-Calderón, G.I., Emrich, S.J., MacCallum, R.M., Maslen, G., Emrich, S., Collins, F., Dialynas, E., Topalis, P., Ho, N., Gesing, S., Madey, G., Collins, F.H., Lawson, D., Kersey, P., Allen, J., Christensen, M., Hughes, D., Koscielny, G., Langridge, N., Gallego, E.L., Megy, K., Wilson, D., Gelbart, B., Emmert, D., Russo, S., Zhou, P., Christophides, G., Brockman, A., Kirmitzoglou, I., MacCallum, B., Tiirikka, T., Louis, K., Dritsou, V., Mitraka, E., Werner-Washburn, M., Baker, P., Platero, H., Aguilar, A., Bogol, S., Campbell, D., Carmichael, R., Cieslak, D., Davis, G., Konopinski, N., Nabrzyski, J., Reinking, C., Sheehan, A., Szakonyi, S., Wieck, R., 2015. VectorBase: an updated bioinformatics resource for invertebrate vectors and other organisms related with human diseases. Nucleic Acids Res 43, D707–D713. https://doi.org/10.1093/NAR/GKU1117

Glaus, P., Honkela, A., Rattray, M., 2012. Identifying differentially expressed transcripts from RNA-seq data with biological variation. Bioinformatics 28, 1721–1728. https://doi.org/10.1093/BIOINFORMATICS/BTS260

Homer Software and Data Download [WWW Document], n.d. URL http://homer.ucsd.edu/homer/motif/ (accessed 11.14.22).

Jimeno-González, S., Payán-Bravo, L., Muñoz-Cabello, A.M., Guijo, M., Gutierrez, G., Prado, F., Reyes, J.C., 2015. Defective histone supply causes changes in RNA polymerase II elongation rate and cotranscriptional pre-mRNA splicing. Proc Natl Acad Sci U S A 112, 14840–14845. https://doi.org/10.1073/PNAS.1506760112/-/DCSUPPLEMENTAL

Katz, Y., Wang, E.T., Airoldi, E.M., Burge, C.B., 2010. Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nature Methods 2010 7:12 7, 1009–1015. https://doi.org/10.1038/nmeth.1528

Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., Stamm, S., 2013. Function of alternative splicing. Gene 514, 1. https://doi.org/10.1016/J.GENE.2012.07.083

Klemm, S.L., Shipony, Z., Greenleaf, W.J., 2019. Chromatin accessibility and the regulatory epigenome. Nature Reviews Genetics 2018 20:4 20, 207–220. https://doi.org/10.1038/s41576-018-0089-8

Kopylova, E., Noé, L., Touzet, H., 2012. SortMeRNA: fast and accurate filtering of ribosomal RNAs in metatranscriptomic data. Bioinformatics 28, 3211–3217. https://doi.org/10.1093/BIOINFORMATICS/BTS611

Krzywinska, E., Dennison, N.J., Lycett, G.J., Krzywinski, J., 2016. A maleness gene in the malaria mosquito Anopheles gambiae. Science (1979) 353, 67–69. https://doi.org/10.1126/SCIENCE.AAF5605/SUPPL_FILE/KRZYWINSKA_SM.PDF

Kurtz, J., Armitage, S.A.O., 2006. Alternative adaptive immunity in invertebrates. Trends Immunol 27, 493–496. https://doi.org/10.1016/J.IT.2006.09.001

Kwon, B., Fansler, M.M., Patel, N.D., Lee, J., Ma, W., Mayr, C., 2022. Enhancers regulate 3′ end processing activity to control expression of alternative 3′UTR isoforms. Nat Commun 13. https://doi.org/10.1038/S41467-022-30525-Y

Lam, B.J., Hertel, K.J., 2002. A general role for splicing enhancers in exon definition. RNA 8, 1233. https://doi.org/10.1017/S1355838202028030

Li, Y., Liu, Y., Yang, H., Zhang, T., Naruse, K., Tu, Q., 2020a. Dynamic transcriptional and chromatin accessibility landscape of medaka embryogenesis. Genome Res 30, 924–937. https://doi.org/10.1101/GR.258871.119/-/DC1

Li, Y., Liu, Y., Yang, H., Zhang, T., Naruse, K., Tu, Q., 2020b. Dynamic transcriptional and chromatin accessibility landscape of medaka embryogenesis. Genome Res 30, 924–937. https://doi.org/10.1101/GR.258871.119/-/DC1

Lin, H., Zhang, L., Luna, C., Hoa, N.T., Zheng, L., 2007. A splice variant of PGRP-LC required for expression of antimicrobial peptides in Anopheles gambiae. Insect Sci 14, 185–192. https://doi.org/10.1111/J.1744-7917.2007.00142.X

Lin, X., Liu, F., Meng, K., Liu, H., Zhao, Y., Chen, Y., Hu, W., Luo, D., 2022. Comprehensive Transcriptome Analysis Reveals Sex-Specific Alternative Splicing Events in Zebrafish Gonads. Life 12, 1441. https://doi.org/10.3390/LIFE12091441/S1

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15, 1–21. https://doi.org/10.1186/S13059-014-0550-8/FIGURES/9

Luco, R.F., Allo, M., Schor, I.E., Kornblihtt, A.R., Misteli, T., 2011. Epigenetics in Alternative Pre-mRNA Splicing. Cell 144, 16–26. https://doi.org/10.1016/J.CELL.2010.11.056

Malko, D.B., Makeev, V.J., Mironov, A.A., Gelfand, M.S., 2006. Evolution of exon–intron structure and alternative splicing in fruit flies and malarial mosquito genomes. Genome Res 16, 505–509. https://doi.org/10.1101/GR.4236606

Marikovsky, M., Ziv, V., Nevo, N., Harris-Cerruti, C., Mahler, O., 2003. Cu/Zn Superoxide Dismutase Plays Important Role in Immune Response. The Journal of Immunology 170, 2993–3001. https://doi.org/10.4049/JIMMUNOL.170.6.2993

Martinez, N.M., Lynch, K.W., 2013. Control of alternative splicing in immune responses: many regulators, many predictions, much still to learn. Immunol Rev 253, 216–236. https://doi.org/10.1111/IMR.12047

Matlin, A.J., Clark, F., Smith, C.W.J., 2005. Understanding alternative splicing: towards a cellular code. Nature Reviews Molecular Cell Biology 2005 6:5 6, 386–398. https://doi.org/10.1038/nrm1645

Meister, S., Agianian, B., Turlure, F., Relógio, A., Morlais, I., Kafatos, F.C., Christophides, G.K., 2009. Anopheles gambiae PGRPLC-Mediated Defense against Bacteria Modulates Infections with Malaria Parasites. PLoS Pathog 5. https://doi.org/10.1371/JOURNAL.PPAT.1000542

Meister, S., Kanzok, S.M., Zheng, X.L., Luna, C., Li, T.R., Hoa, N.T., Clayton, J.R., White, K.P., Kafatos, F.C., Christophides, G.K., Zheng, L., 2005. Immune signaling pathways regulating bacterial and malaria parasite infection of the mosquito Anopheles gambiae. Proc Natl Acad Sci U S A 102, 11420–11425. https://doi.org/10.1073/PNAS.0504950102/SUPPL_FILE/04950FIG8.PDF

Merino, G.A., Conesa, A., Fernández, E.A., 2019. A benchmarking of workflows for detecting differential splicing and differential expression at isoform level in human RNA-seq studies. Brief Bioinform 20, 471–481. https://doi.org/10.1093/BIB/BBX122

Molina-Cruz, A., DeJong, R.J., Charles, B., Gupta, L., Kumar, S., Jaramillo-Gutierrez, G., Barillas-Mury, C., 2008. Reactive oxygen species modulate Anopheles gambiae immunity against bacteria and Plasmodium. Journal of Biological Chemistry 283, 3217–3223. https://doi.org/10.1074/jbc.M705873200

Montes, M., Becerra, S., Sánchez-Álvarez, M., Suñé, C., 2012. Functional coupling of transcription and splicing. Gene 501, 104–117. https://doi.org/10.1016/J.GENE.2012.04.006

Naftelberg, S., Schor, I.E., Ast, G., Kornblihtt, A.R., n.d. Regulation of Alternative Splicing Through Coupling with Transcription and Chromatin Structure. https://doi.org/10.1146/annurev-biochem-060614-034242

Nilsen, T.W., Graveley, B.R., 2010. Expansion of the eukaryotic proteome by alternative splicing. Nature 2010 463:7280 463, 457–463. https://doi.org/10.1038/nature08909

Nussinov, R., Zhang, M., Maloney, R., Jang, H., 2021. Ras isoform-specific expression, chromatin accessibility, and signaling. Biophysical Reviews 2021 13:4 13, 489–505. https://doi.org/10.1007/S12551-021-00817-6

Onodera, C.S., Underwood, J.G., Katzman, S., Jacobs, F., Greenberg, D., Salama, S.R., Haussler, D., 2012. Gene Isoform Specificity through Enhancer-Associated Antisense Transcription. PLoS One 7, 43511. https://doi.org/10.1371/JOURNAL.PONE.0043511

Pai, A.A., Luca, F., 2019. Environmental influences on RNA processing: biochemical, molecular and genetic regulators of cellular response. Wiley Interdiscip Rev RNA 10, e1503. https://doi.org/10.1002/WRNA.1503

Patro, R., Duggal, G., Love, M.I., Irizarry, R.A., Kingsford, C., 2017. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods 2017 14:4 14, 417–419. https://doi.org/10.1038/nmeth.4197

Petrova, V., Song, R., Consortium, D., Nordström, K.J. V, Walter, J., Wong, J.J.L., Armstrong, N.J., Rasko, J.E.J., Schmitz, U., 2022a. Increased chromatin accessibility facilitates intron retention in specific cell differentiation states. Nucleic Acids Res 50, 11563. https://doi.org/10.1093/NAR/GKAC994

Petrova, V., Song, R., Consortium, D., NordströmNordstr, K.J., Walter, orn, J-L Wong, J., Armstrong, N.J., Rasko, J.E., Schmitz, U., 2022b. Increased chromatin accessibility facilitates intron retention in specific cell differentiation states. Nucleic Acids Res. https://doi.org/10.1093/NAR/GKAC994

Ragle, J.M., Katzman, S., Akers, T.F., Barberan-Soler, S., Zahler, A.M., 2015. Coordinated tissue-specific regulation of adjacent alternative 3′ splice sites in C. elegans. Genome Res 25, 982. https://doi.org/10.1101/GR.186783.114

Ruiz, José L., Ranford-Cartwright, L.C., Gómez-Díaz, E., 2021. The regulatory genome of the malaria vector Anopheles gambiae: integrating chromatin accessibility and gene expression. NAR Genom Bioinform 3. https://doi.org/10.1093/NARGAB/LQAA113

Ruiz, José L, Ranford-Cartwright, L.C., Gómez-Díaz, E., 2021. The regulatory genome of the malaria vector Anopheles gambiae: integrating chromatin accessibility and gene expression. NAR Genom Bioinform 3. https://doi.org/10.1093/NARGAB/LQAA113

Ruiz, J.L., Yerbanga, R.S., Lefèvre, T., Ouedraogo, J.B., Corces, V.G., Gómez-Díaz, E., 2019. Chromatin changes in Anopheles gambiae induced by Plasmodium falciparum infection 06 Biological Sciences 0604 Genetics 11 Medical and Health Sciences 1108 Medical Microbiology. Epigenetics Chromatin 12, 1–18. https://doi.org/10.1186/S13072-018-0250-9/TABLES/1

Saldi, T., Cortazar, M.A., Sheridan, R.M., Bentley, D.L., 2016. Coupling of RNA polymerase II transcription elongation with pre-mRNA splicing. J Mol Biol 428, 2623. https://doi.org/10.1016/J.JMB.2016.04.017

Salvemini, M., D'Amato, R., Petrella, V., Aceto, S., Nimmo, D., Neira, M., Alphey, L., Polito, L.C., Saccone, G., 2013. The Orthologue of the Fruitfly Sex Behaviour Gene Fruitless in the Mosquito Aedes aegypti: Evolution of Genomic Organisation and Alternative Splicing. PLoS One 8, 48554. https://doi.org/10.1371/JOURNAL.PONE.0048554

Salvemini, M., Mauro, U., Lombardo, F., Milano, A., Zazzaro, V., Arcà, B., Polito, L.C., Saccone, G., 2011. Genomic organization and splicing evolution of the doublesex gene, a Drosophila regulator of sexual differentiation, in the dengue and yellow fever mosquito Aedes aegypti. BMC Evol Biol 11, 41. https://doi.org/10.1186/1471-2148-11-41

Schor, I.E., Gómez Acuña, L.I., Kornblihtt, A.R., 2013. Coupling between transcription and alternative splicing. Cancer Treat Res 1–24. https://doi.org/10.1007/978-3-642-31659-3_1/COVER

Schuettengruber, B., Ganapathi, M., Leblanc, B., Portoso, M., Jaschek, R., Tolhuis, B., Van Lohuizen, M., Tanay, A., Cavalli, G., 2009. Functional Anatomy of Polycomb and Trithorax Chromatin Landscapes in Drosophila Embryos. PLoS Biol 7, 1000013. https://doi.org/10.1371/JOURNAL.PBIO.1000013

Shen, L., Shao, N., Liu, X., Nestler, E., 2014. Ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 15, 1–14. https://doi.org/10.1186/1471-2164-15-284/FIGURES/6

Shenasa, H., Hertel, K.J., 2019. Combinatorial regulation of alternative splicing. Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1862, 194392. https://doi.org/10.1016/J.BBAGRM.2019.06.003

Shiau, C.K., Huang, J.H., Liu, Y.T., Tsai, H.K., 2021. Genome-wide identification of associations between enhancer and alternative splicing in human and mouse. BMC Genomics 22. https://doi.org/10.1186/S12864-022-08537-1

Simon, J.M., Hacker, K.E., Singh, D., Brannon, A.R., Parker, J.S., Weiser, M., Ho, T.H., Kuan, P.F., Jonasch, E., Furey, T.S., Prins, J.F., Lieb, J.D., Rathmell, W.K., Davis, I.J., 2014. Variation in chromatin accessibility in human kidney cancer links H3K36 methyltransferase loss with widespread RNA processing defects. Genome Res 24, 241–250. https://doi.org/10.1101/GR.158253.113

Sinkins, S.P., 2016. Yob makes mosquitoes male. Science (1979) 353, 33–34. https://doi.org/10.1126/SCIENCE.AAG1012/ASSET/61061E9A-55CE-4212-8E99-003D05315234/ASSETS/GRAPHIC/353_33_F2.JPEG

Song, L., Pan, Z., Chen, L., Dai, Y., Wan, J., Ye, H., Nguyen, H.T., Zhang, G., Chen, H., 2020. Analysis of Whole Transcriptome RNA-seq Data Reveals Many Alternative Splicing Events in Soybean Roots under Drought Stress Conditions. Genes 2020, Vol. 11, Page 1520 11, 1520. https://doi.org/10.3390/GENES11121520

Sreenivasamurthy, S.K., Dey, G., Ramu, M., Kumar, M., Gupta, M.K., Mohanty, A.K., Harsha, H., Sharma, P., Kumar, N., Pandey, A., Kumar, A., Prasad, T.K., 2013. A compendium of molecules involved in vector-pathogen interactions pertaining to malaria. Malar J 12, 216. https://doi.org/10.1186/1475-2875-12-216

Sreenivasamurthy, S.K., Madugundu, A.K., Patil, A.H., Dey, G., Mohanty, A.K., Kumar, M., Patel, K., Wang, C., Kumar, A., Pandey, A., Prasad, T.S.K., 2017. Mosquito-Borne Diseases and Omics: Tissue-Restricted Expression and Alternative Splicing Revealed by Transcriptome Profiling of Anopheles stephensi. OMICS 21, 488–497. https://doi.org/10.1089/OMI.2017.0073

Stark, R., Brown, G., n.d. DiffBind: Differential binding analysis of ChIP-Seq peak data.

Telonis-Scott, M., Kopp, A., Wayne, M.L., Nuzhdin, S. V., McIntyre, L.M., 2009. Sex-Specific Splicing in Drosophila: Widespread Occurrence, Tissue Specificity and Evolutionary Conservation. Genetics 181, 421. https://doi.org/10.1534/GENETICS.108.096743

Tsujimoto, H., Liu, K., Linser, P.J., Agre, P., Rasgon, J.L., 2013. Organ-Specific Splice Variants of Aquaporin Water Channel AgAQP1 in the Malaria Vector Anopheles gambiae. PLoS One 8. https://doi.org/10.1371/JOURNAL.PONE.0075888

Ule, J., Blencowe, B.J., 2019. Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution. Mol Cell 76, 329–345. https://doi.org/10.1016/J.MOLCEL.2019.09.017

Ullah, F., Hamilton, M., Reddy, A.S.N., Ben-Hur, A., 2018. Exploring the relationship between intron retention and chromatin accessibility in plants. BMC Genomics 19, 21. https://doi.org/10.1186/S12864-017-4393-Z

Urban, J.A., Urban, J.M., Kuzu, G., Larschan, E.N., 2017. The Drosophila CLAMP protein associates with diverse proteins on chromatin. PLoS One 12. https://doi.org/10.1371/JOURNAL.PONE.0189772

Venables, J.P., Tazi, J., Juge, F., 2012. Regulated functional alternative splicing in Drosophila. Nucleic Acids Res 40, 1–10. https://doi.org/10.1093/NAR/GKR648

Verta, J.P., Jacobs, A., 2022. The role of alternative splicing in adaptation and evolution. Trends Ecol Evol 37, 299–308. https://doi.org/10.1016/J.TREE.2021.11.010

Vitting-Seerup, K., Sandelin, A., 2017. The landscape of isoform switches in human cancers. Molecular Cancer Research 15, 1206–1220. https://doi.org/10.1158/1541-7786.MCR-16-0459/14585/AM/THE-LANDSCAPE-OF-ISOFORM-SWITCHES-IN-HUMAN

Vitting-Seerup, K., Sandelin, A., Berger, B., 2019. IsoformSwitchAnalyzeR: analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. Bioinformatics 35, 4469–4471. https://doi.org/10.1093/BIOINFORMATICS/BTZ247

Wang, E.T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S.F., Schroth, G.P., Burge, C.B., 2008. Alternative isoform regulation in human tissue transcriptomes. Nature 2008 456:7221 456, 470–476. https://doi.org/10.1038/nature07509

Wang, G.S., Cooper, T.A., 2007. Splicing in disease: disruption of the splicing code and the decoding machinery. Nature Reviews Genetics 2007 8:10 8, 749–761. https://doi.org/10.1038/nrg2164

Wright, C.J., Smith, C.W.J., Jiggins, C.D., 2022. Alternative splicing as a source of phenotypic diversity. Nature Reviews Genetics 2022 1–14. https://doi.org/10.1038/s41576-022-00514-4

Ye, C.J., Chen, J., Villani, A.C., Gate, R.E., Subramaniam, M., Bhangale, T., Lee, M.N., Raj, T., Raychowdhury, R., Li, W., Rogel, N., Simmons, S., Imboywa, S.H., Chipendo, P.I., McCabe, C., Lee, M.H., Frohlich, I.Y., Stranger, B.E., de Jager, P.L., Regev, A., Behrens, T., Hacohen, N., 2018. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of ERAP2 transcripts under balancing selection. Genome Res 28, 1812–1825. https://doi.org/10.1101/GR.240390.118

Zahler, A.M., 2005. Alternative splicing in C. elegans. WormBook 1–13. https://doi.org/10.1895/WORMBOOK.1.31.1

Zhang, C., Zhang, B., Vincent, M.S., Zhao, S., 2016. Bioinformatics Tools for RNA-seq Gene and Isoform Quantification. https://doi.org/10.4172/2469-9853.1000140

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., Shirley, X.S., 2008. Model-based Analysis of ChIP-Seq (MACS). Genome Biol 9, R137. https://doi.org/10.1186/GB-2008-9-9-R137

Zhao, S., 2019. Alternative splicing, RNA-seq and drug discovery. Drug Discov Today 24, 1258–1267. https://doi.org/10.1016/J.DRUDIS.2019.03.030

Tables

Motif	Name	P-value	Flybase ID	Flybase name	Nº of DUI isoforms with the motif
CACCWATT	sna/MA0086.2	1E-34	FBgn0003448	snail	22
CTTGCAAGGT	byn/dmmpmm	1E-31	FBgn0011723	brachyenteron	21
ACARAGAAAT	sd/dmmpmm	1E-28	FBgn0003345	scalloped	26
ACACAAAG	cad/dmmpmm	1E-27	FBgn0000251	caudal	26
GAGTCTTTAT	br-Z4/dmmpmm	1E-27	FBgn0283451	broad	20
GAAAACAGTT	grh/MA1457.1	1E-27	FBgn0259211	grainy head	21
AAACAGTAAAAG	br-Z2/dmmpmm	1E-27	FBgn0283451	broad	20
CAAGAAGA	Aef1/dmmpmm	1E-27	FBgn0005694	Adult enhancer factor 1	25
GAMCGAGH	Clamp/MA1700.1	1E-27	FBgn0032979	Chromatin-linked adaptor for MSL proteins	24
TTATAAYT	br-Z1/dmmpmm	1E-27	FBgn0283451	broad	22

Table 1. Ten first “de novo” motifs detected with Homer that appear overrepresented in the splicing sites of DEMG or DUI that has a differential accessibility between tissues. Additional information in Supplementary table S6.

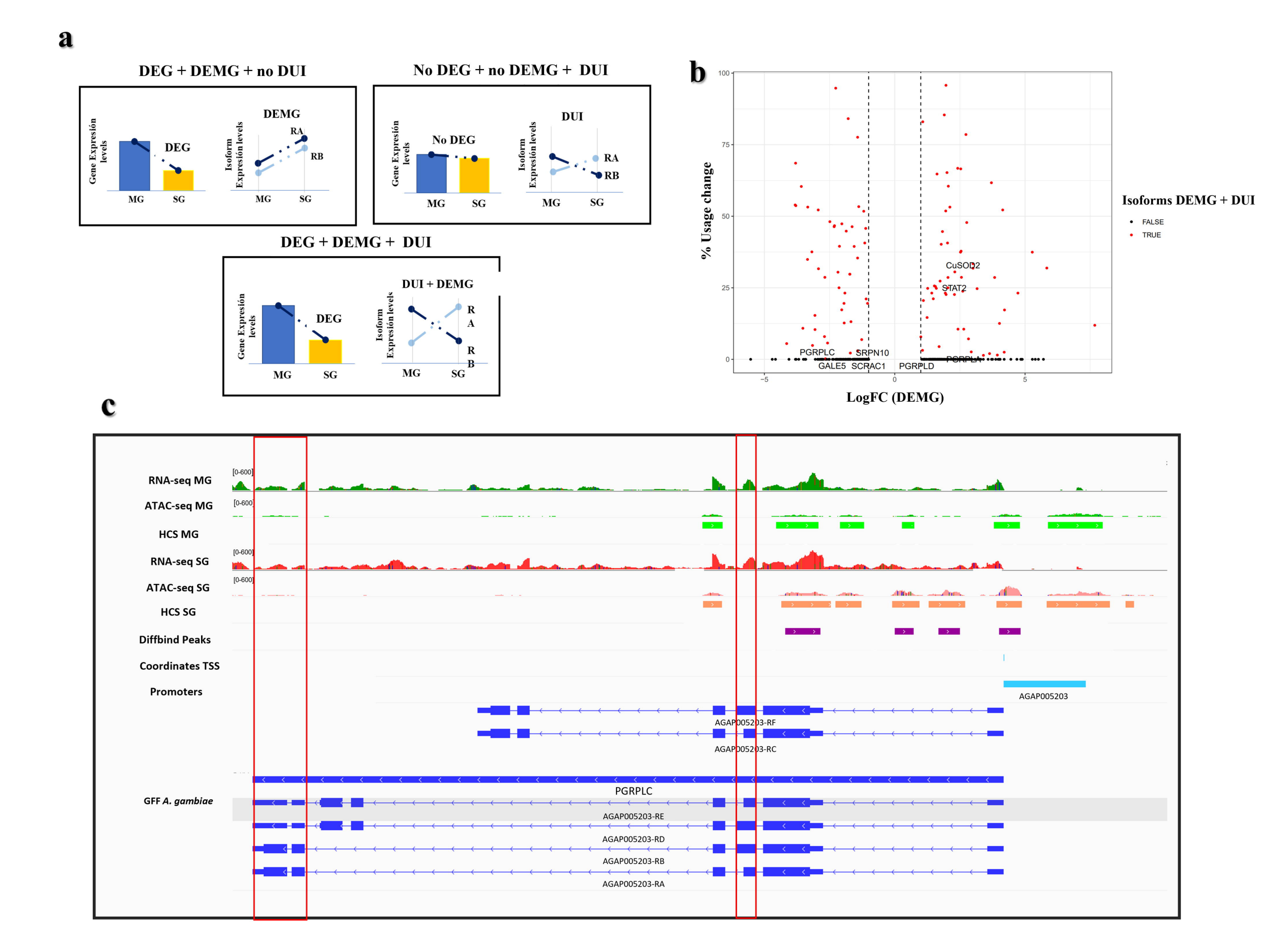
Figure legends

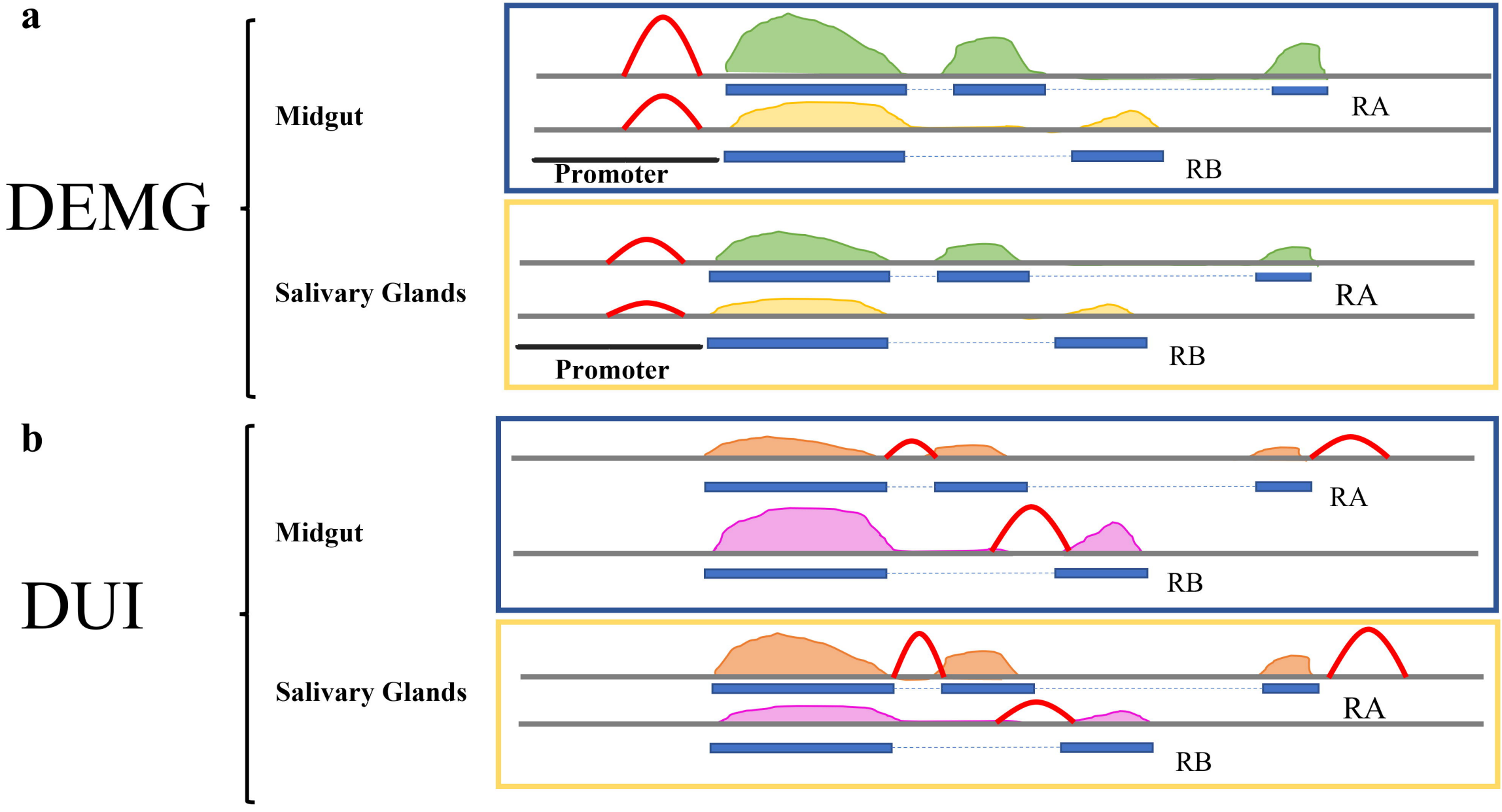
Figure 1. Patterns of differentially expressed multisoform genes (DEMG) and differentially used isoforms between tissues. (A) Diagram illustrating possible DEMG and DUI patterns expected in our data (B) Volcano plot showing total usage change (i.e., expression redistribution between isoforms) versus log-transformed values of isoform expression fold change between tissues. Isoforms that are DUI and DEMG are shown in red whereas black dots correspond to differentially expressed isoforms. Labels identify isoforms corresponding to *A. gambiae* genes with immune functions. (C) IGV representation of the AGAP005203 (PGRPLC) gene. Red lines highlight differential expression and accessibility at RA and RE differential exons.

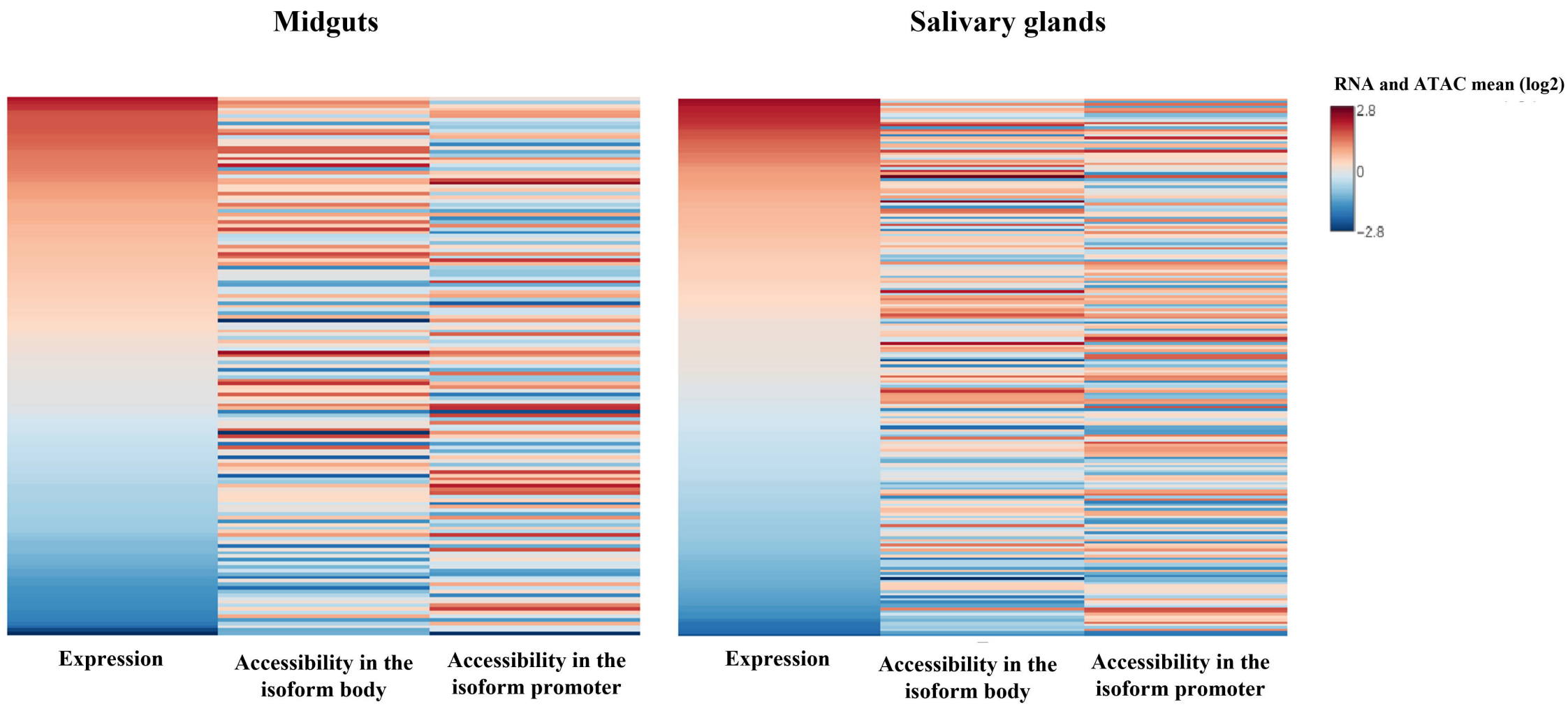
Figure 2. Model for the relationship between isoform expression, isoform usage and chromatin accessibility. The ATAC-seq peaks are shown in red. A) For the DEMG group the expectation is that the accessibility of the promoter will increase in the tissue where the expression of the isoform is higher b) For the DUI group, we expect to find chromatin accessibility changes near the splicing sites coinciding with the different usage of the isoform/s.

Figure 3. Differential chromatin accessibility between tissues in the isoform body correlates with changes in isoform expression. Heatmap showing ATAC-seq nucleosome-free enrichment at promoters or in the body of DEMG isoforms and their expression levels. There is a positive and significant correlation between chromatin accessibility at the isoform body and isoform expression. Genes are ordered by normalized RNA-seq isoform expression (RPKM). Changes tend to occur in the same direction, and there is a positive and significant correlation between the magnitude of changes in accessibility and expression at the isoform body in both tissues.

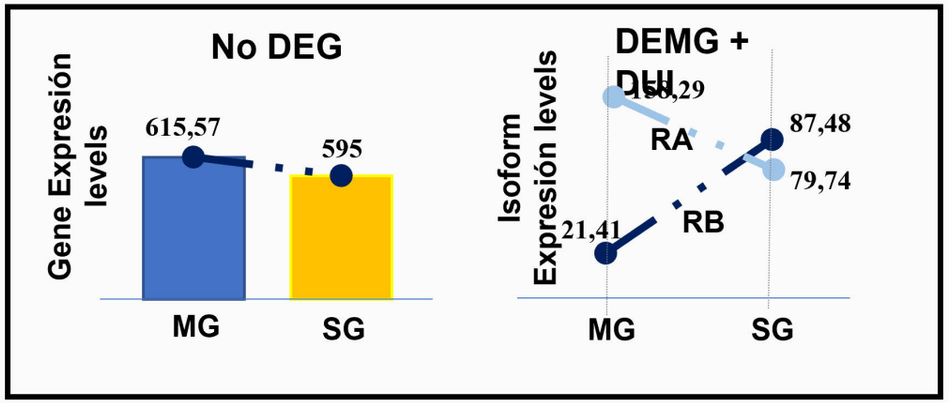
Figure 4. Enhancer function and AS in the region containing the encoding gene AGAP005234 (A) Scheme summarizing gene and isoform expression between tissues for the AGAP005234 gene. (b) IGV screenshot showing the AGAP005234 gene annotated enhancer which is differentially accessible between tissues coinciding with the differential expression and usage of their isoforms. All tracks are shown at equal scale.


a
DEG + DEMG + no DUI
Gene Expresión levels
DEG
MG
SG
Isoform Expresión levels
DEMG
RA
RB
No DEG + no DEMG + DUI
No DEG
DUI
DEG + DEMG + DUI
DUI + DEMG
b
% Usage change
LogFC (DEMG)
Isoforms DEMG + DUI
FALSE
TRUE
CuSOD2
STAT2
PGRPLC
SRPN10
GALE5
SCRAC1
PGRPLD
PGRPLA
c
RNA-seq MG
ATAC-seq MG
HCS MG
RNA-seq SG
ATAC-seq SG
HCS SG
Diffbind Peaks
Coordinates TSS
Promoters
AGAP005203
AGAP005203-RF
AGAP005203-RC
GFF A. gambiae
PGRPLC
AGAP005203-RE
AGAP005203-RD
AGAP005203-RB
AGAP005203-RA

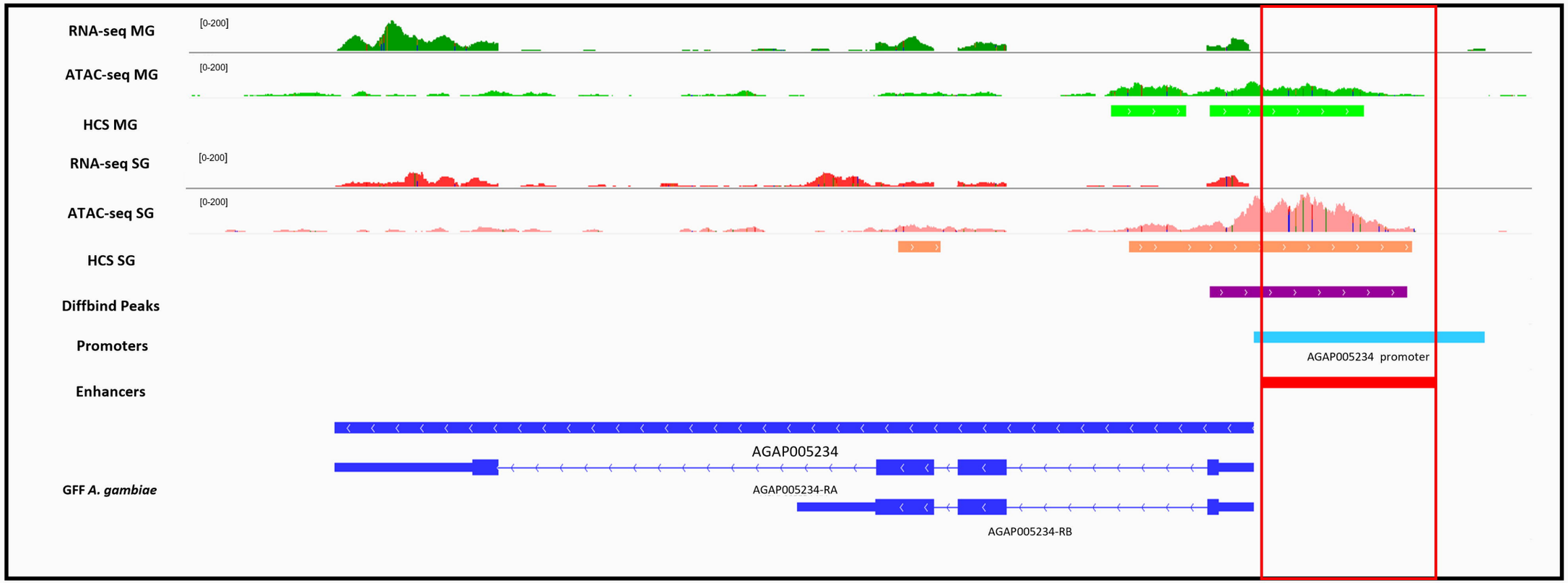

a
DEMG
Midgut
Salivary Glands
RA
RB
Promoter
RA
RB
Promoter
b
DUI
Midgut
Salivary Glands
RA
RB
RA
RB


Midguts
Salivary glands
RNA and ATAC mean (log2)
2.8
0
-2.8
Expression
Accessibility in the isoform body
Accessibility in the isoform promoter
Expression
Accessibility in the isoform body
Accessibility in the isoform promoter

a


No DEG
Gene Expresión levels
615,57
595
MG
SG
DEMG + DIU
Isoform Expresión levels
158,29
RA
87,48
79,74
21,41
RB
MG
SG

b


RNA-seq MG
[0-200]
ATAC-seq MG
[0-200]
HCS MG
RNA-seq SG
[0-200]
ATAC-seq SG
[0-200]
HCS SG
Diffbind Peaks
Promoters
AGAP005234 promoter
Enhancers
AGAP005234
GFF A. gambiae
AGAP005234-RA
AGAP005234-RB