REVIEW

How to study enhancers in non-traditional insect models

Yoshinori Tomoyasu[1],* and Marc S. Halfon[2],*

ABSTRACT

Transcriptional enhancers are central to the function and evolution of genes and gene regulation. At the organismal level, enhancers play a crucial role in coordinating tissue- and context-dependent gene expression. At the population level, changes in enhancers are thought to be a major driving force that facilitates evolution of diverse traits. An amazing array of diverse traits seen in insect morphology, physiology and behavior has been the subject of research for centuries. Although enhancer studies in insects outside of *Drosophila* have been limited, recent advances in functional genomic approaches have begun to make such studies possible in an increasing selection of insect species. Here, instead of comprehensively reviewing currently available technologies for enhancer studies in established model organisms such as *Drosophila*, we focus on a subset of computational and experimental approaches that are likely applicable to non-*Drosophila* insects, and discuss the pros and cons of each approach. We discuss the importance of validating enhancer function and evaluate several possible validation methods, such as reporter assays and genome editing. Key points and potential pitfalls when establishing a reporter assay system in non-traditional insect models are also discussed. We close with a discussion of how to advance enhancer studies in insects, both by improving computational approaches and by expanding the genetic toolbox in various insects. Through these discussions, this Review provides a conceptual framework for studying the function and evolution of enhancers in non-traditional insect models.

KEY WORDS: Enhancers, *Cis*-regulation, Non-traditional insect models, Reporter assay, *Tribolium*, SCRMshaw

Introduction

Understanding how genes are regulated is fundamental to various disciplines of biology. In the field of insect science, molecular mechanisms underlying gene regulation are best studied in the fruit fly *Drosophila melanogaster*. With a suite of sophisticated genetic tools available in this insect (Hales et al., 2015), scientists have been able to decipher complex interactions among genes and their protein products, revealing comprehensive networks of gene interactions and regulations (i.e. gene regulatory networks, GRNs) for various tissues and contexts.

Two innovations in the last two decades have drastically changed the way we study insects beyond *Drosophila*. The first is the advancement of next-generation sequencing technology, which allows researchers to gather genomic and transcriptomic information from the insect they study relatively easily and even sometimes prior to detailed 'wet' investigation. The second is the

[1]Department of Biology, Miami University, Oxford, OH 45056, USA. [2]Department of Biochemistry, University at Buffalo-State University of New York, Buffalo, NY 14203, USA.

*Authors for correspondence (tomoyay@miamioh.edu; mshalfon@buffalo.edu)

Y.T., 0000-0001-9824-3454; M.S.H., 0000-0002-4149-2705

application of RNA interference (RNAi), and more recently, CRISPR/Cas9 (clustered regularly interspaced short palindromic repeats/CRISPR associated protein 9) genome editing, in various insects (reviewed in Bellés, 2010; Gilles and Averof, 2014). These gene knockdown/knockout techniques now allow loss-of-function (LOF) analyses in many (albeit not all) insects without the need for creating mutants through traditional means, and often with the capability of controlling the timing of gene disruption. With these critical advances, we can now study various tissues and contexts of non-traditional model insects or even non-model insects at the detailed molecular level (Reardon, 2019).

Although researchers are gradually stepping out from *Drosophila* to explore molecular mechanisms underlying various intriguing processes found in other insects, the knowledge obtained from *Drosophila* studies continues to play a critical role in insect science. Researchers often use *Drosophila* GRNs as a starting point (i.e. *Drosophila* paradigm) and investigate the function of the genes that are homologous/orthologous to the genes in the *Drosophila* GRNs through RNAi-based LOF studies, often combined with expression analyses, in their insects. This approach has been very fruitful in gaining new insights into gene function and regulation, as well as into the evolution of GRNs among insects, that are difficult to obtain through studying *Drosophila* alone (Bellés, 2010).

When discussing GRNs, there are two types of components: *trans* and *cis* (Fig. 1A). *trans* components are transcription factors (TFs) and their upstream regulators that provide instructive cues to cells for patterning, differentiation and various other biological processes. In contrast, *cis* components are non-coding DNA elements that integrate the upstream *trans* information and determine the expression of the genes downstream in the GRNs. Enhancers (often also called *cis*-regulatory elements or *cis*-regulatory modules) are a class of *cis* components that play a central role in determining spatial and temporal gene expression (Blackwood and Kadonaga, 1998; Buffry et al., 2016; Cho, 2012; Long et al., 2016; Pennacchio et al., 2013; Rickels and Shilatifard, 2018). As mentioned, most current studies in insects outside of *Drosophila* utilize RNAi-based LOF analyses (or knocking out coding genes via CRISPR/Cas9) as a central approach. This allows for an investigation of GRNs from the *trans* point of view (Fig. 1A), by inhibiting the function of *trans* components and assessing their influence on GRNs. However, although it is at least as important to study GRNs from the *cis* perspective to gain a comprehensive view of gene regulatory mechanisms, the lack of a reliable method to identify enhancers in non-*Drosophila* insects has made it difficult to study the function and evolution of *cis* components beyond *Drosophila* species.

There are several reasons as to why studying enhancers is so challenging in non-*Drosophila* insects. First, compared with *trans* components, *cis* components, especially enhancers, are extremely labile (Li et al., 2007), which makes identification of enhancers based on sequence conservation challenging even among closely related species (Papatsenko et al., 2006) and nearly impossible among species with divergence time beyond ~60 million years

1

A

B
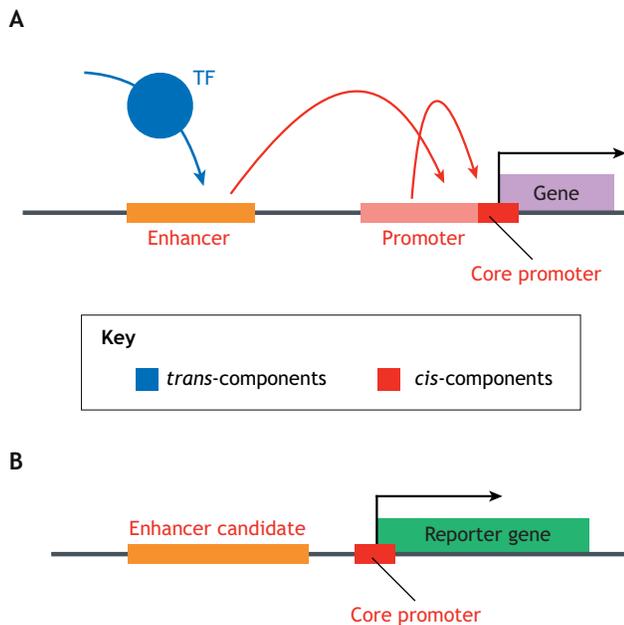
Fig. 1. Gene regulation and reporter assay. (A) *trans* and *cis* components in gene regulation. RNAi allows functional analyses from the *trans* point of view. (B) A typical reporter assay configuration. Note that the core promoter (red) is a short stretch (∼80 bp) of DNA sequence where the general transcription factors and RNA polymerase are assembled for gene expression. A core promoter itself is typically not sufficient to initiate transcription unless an active enhancer (either proximally or distally located) facilitates the assembly of transcription initiation factors at the core promoter. The term 'promoter' (pink and red) is often used to describe the region immediately upstream of the transcription start site when this region contains both the core promoter (red) and a proximally located enhancer (orange), and is sufficient to drive gene expression.

(Kazemian et al., 2014; Li et al., 2007). This is especially problematic for insects, which underwent an early radiation (winged insects had diversified into at least 10 orders by the early Permian, 300 million years ago; Kukalova-Peck, 1991) and have short generation times, leading to limited non-coding homology beyond the genus or family level. Second, functional validation of enhancers often requires the use of modern genetic and genomic tools, which are currently largely absent from most non-*Drosophila* insects. Because of these hurdles, investigations into the function and evolution of enhancers have been quite limited in insects outside of *Drosophila*, despite the clear awareness among researchers that changes in enhancers and other *cis*-regulatory elements play a crucial role in facilitating evolution and diversification of various traits among insects and other organisms (reviewed in Carroll, 2008).

Typically, an enhancer study consists of two steps: (1) the identification of possible enhancer regions (either focusing on a single gene of interest or genome-wide), and (2) the validation and further downstream functional evaluation of enhancer activity *in vivo*. Some approaches allow functional validation of enhancer activity in the first step, while others require separate *in vivo* validation experiments. In this Review, we will first summarize currently available approaches to identify possible enhancer regions in insect genomes. Although many of these approaches are technology and resource intensive (i.e. model system-centered), recent advances in genomics and computational biology have started making some of these approaches more accessible to researchers that use insects other than *Drosophila* as their model.

We will discuss the pros and cons of each of these approaches when applied to non-traditional model insects. We will then turn our attention to *in vivo* validation of enhancer activity in non-traditional insect models and discuss several possible approaches for enhancer validation, such as reporter assays and CRISPR/Cas9-based genome editing. We will use our recent attempt to establish a cross-species compatible reporter construct as a case study, and discuss some of the key points in establishing a reporter assay system in insects outside of *Drosophila*. Lastly, we will touch on our current effort to advance enhancer studies in insects, both by improving the computational approach to identify possible enhancer regions in insect genomes and by expanding the genetic toolbox for enhancer studies in various insects.

Experimental approaches to identifying possible enhancer regions in insect genomes
Classic reporter assay
By definition, enhancers are short DNA sequences that act in *cis* and increase the transcription of a nearby gene regardless of their orientation and the distance from the gene they regulate (Blackwood and Kadonaga, 1998; Pennacchio et al., 2013). The reporter assay takes advantage of this feature and places a candidate enhancer in front of a marker gene that can be easily visualized (i.e. a reporter gene), such as the *lacZ* gene of *Escherichia coli* or fluorescent protein genes, along with a core promoter (Fig. 1B). The enhancer activity of this 'reporter construct' can then be assayed *in vivo* by visualizing the expression of the reporter gene in various tissues and contexts. This approach was first used to investigate the regulation of several segmentation genes in *Drosophila*, such as *fushi tarazu* (*ftz*) (Hiromi et al., 1985) and *even skipped* (*eve*) (Goto et al., 1989; Harding et al., 1989), and now has become the 'gold standard' approach when evaluating the activity of enhancers in *Drosophila* and other traditional model organisms (Suryamohan and Halfon, 2015). However, this approach is often inefficient and incomplete as a method to identify enhancers, as it requires the generation of many transgenic lines to be able to survey a sufficient length of the genome, many of which will inevitably only provide negative results.

Genome-wide reporter assay
Despite the arduous and time-consuming nature of the reporter assay, this approach has been used in a genome-wide fashion in *Drosophila*. Flylight and Fly Enhancers are the two major projects attempting genome-wide enhancer identification through reporter assays, with a focus on brain development and embryogenesis, respectively (Jenett et al., 2012; Kvon et al., 2014; Pfeiffer et al., 2008). The Flylight collection was also used to describe genome-wide enhancer activities in several different developmental contexts (Jenett et al., 2012; Jory et al., 2012; Tokusumi et al., 2017). These projects have identified thousands of functionally validated enhancers and provided us with an overall outlook of the *cis*-regulatory landscape in the *Drosophila* genome. Furthermore, over 10,000 lines generated through these projects use the yeast Gal4 transcription factor as the reporter, which allows application of the Gal4-UAS bipartite expression system (Brand and Perrimon, 1993) to enable researchers to misexpress genes and trace the lineage of cells and tissues with unprecedented precision in *Drosophila*.

The reporter assay system has also been used in a high-throughput setting. An example of a high-throughput approach used in *Drosophila* is STARR-seq (self-transcribing active regulatory region sequencing) (Arnold et al., 2013; Muerdter et al., 2015). STARR-seq utilizes a library of reporter constructs, which covers

the entirety of the *Drosophila* genome >10 fold. The reporter constructs are designed with the candidate enhancer sequences placed downstream of the core promoter. As the result, active enhancers are directly transcribed, thus serving double-duty as their own reporter genes when transfected cultured cells are subjected to RNA sequencing (RNA-seq) analysis (Fig. 2C). Furthermore, STARR-seq enables identification of enhancers in a quantitative manner, because the number of RNA-seq reads corresponding to each candidate enhancer sequence is directly proportional to the strength of the enhancer activity of the genome fragment.

Phylogenetic footprinting

Phylogenetic footprinting is based on the concept that functionally important sequences within a genome, even outside of the coding regions (such as TF binding sites), should be evolutionarily conserved. The fast-evolving nature of insect genomes appears to make the alignment of genomic sequences among multiple insect species challenging. Nonetheless, this approach can be powerful at identifying enhancers when genome sequences from a set of closely related species are available. For example, multiple sequenced genomes within the genus *Drosophila*, along with the available genome alignments for these species, have made it possible to quickly identify blocks of conserved sequence outside of the coding regions (Frazer et al., 2004; Mayor et al., 2000; Papatsenko et al., 2006; Sosinsky et al., 2007; Stark et al., 2007). However, several studies functionally validating the conserved non-coding sequences point toward a consensus that conservation alone might not be sufficient to efficiently identify enhancers (i.e. not all demonstrated enhancers are well conserved and not all conserved non-coding sequences appear to function as enhancers) (Bergman et al., 2002; Kharchenko et al., 2011; Li et al., 2007; Richards et al., 2005; Roy et al., 2010).

Chromatin profiling

Changes in chromatin status through epigenetic modifications are critical to facilitate precise gene regulation (reviewed in Klemm et al., 2019). Various *cis*-regulatory elements, including enhancers, are 'open' (i.e. nucleosome free) when they are active, so TFs have access
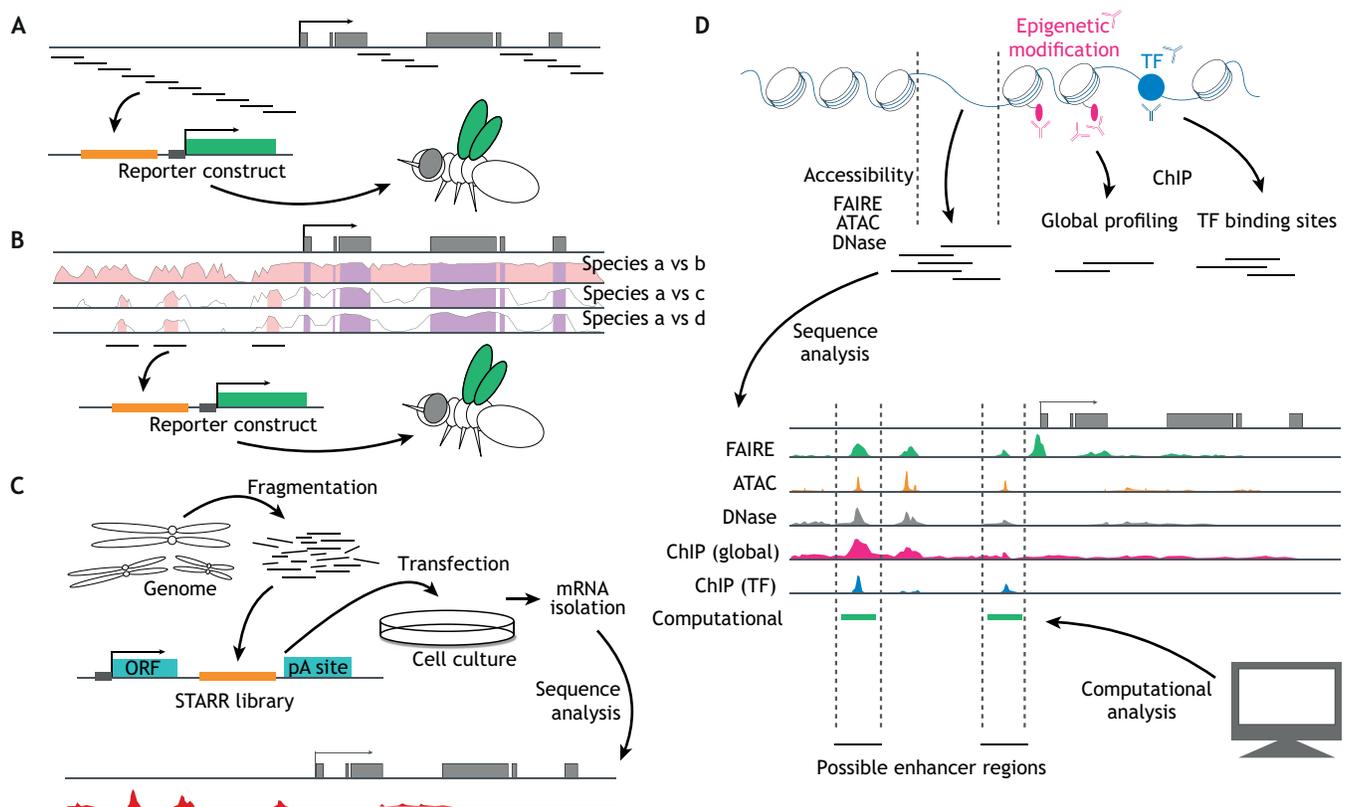

Fig. 2. Enhancer identification methods for non-traditional insect models. (A) Classic reporter assay and (B) reporter assay aided by phylogenetic footprinting. The diagram illustrates an example of searching for a wing enhancer, in which a genomic fragment containing a wing enhancer (orange in the reporter construct) drives reporter gene expression (green) in the wing of the transgenic insect. The classic reporter assay requires a survey of a large genomic region (upstream, downstream, introns and sometimes even exons of the gene of interest) to identify an enhancer (A). In phylogenetic footprinting, pair-wise comparisons of genomic sequences among several closely related species (e.g. species a–d in B) allow identification of evolutionarily conserved regions of the genome (VISTA plot in B; the height of the peak corresponds to the degree of conservation; blue highlights the coding region). Evolutionary conservation outside of the coding sequence (highlighted in pink in the VISTA plot) may imply the presence of functional *cis*-elements, and therefore could help narrow the search. (C) STARR-seq. The STARR reporter is designed in a way that the inserted genomic fragment is transcribed when the fragment acts as an enhancer (orange in the reporter construct) upon transfection into cultured cells. This allows identification of enhancers in a quantitative and genome-wide manner through RNA-seq, as the number of RNA-seq reads corresponding to each candidate enhancer (depicted as red peaks) is directly proportional to the strength of the enhancer activity of the genome fragment. ORF, open reading frame; pA site, polyadenylation site. (D) Enhancer identification through chromatin profiling and computational approaches. FAIRE-seq, ATAC-seq and DNase-seq allow identification of open chromatin regions, while ChIP-seq can be used either to profile genome-wide epigenetic modifications or to identify the binding sites of a transcription factor (TF) of interest. Outcomes of these analyses are often presented as peaks along the genome, where the peak height represents the number of sequence reads that were mapped to the corresponding genomic region. Note that enhancer regions identified through chromatin profiling or computational approaches are still predictions that require functional validation.

to these regions. Several methods exploit this feature of the genome and identify possible enhancer regions through chromatin profiling (reviewed in Klemm et al., 2019; Meyer and Liu, 2014; Suryamohan and Halfon, 2015). DNase-seq (DNase I hypersensitive sites sequencing) uses high sensitivity to DNase as the indicator of open chromatin regions (Boyle et al., 2008). FAIRE-seq (formaldehyde-assisted isolation of regulatory elements, combined with sequencing) and ATAC-seq (assay for transposase-accessible chromatin using sequencing) also identify open chromatin regions. FAIRE-seq uses organic phase separation chemistry to isolate nucleosome-free chromatin away from nucleosome-containing DNA (Giresi et al., 2007; McKay, 2019; McKay and Lieb, 2013), while ATAC-seq uses transposase accessibility as the indicator of open chromatin (Buenrostro et al., 2013).

Another type of chromatin-related method that might be useful to identify enhancers is chromosome conformation capture (3C) (reviewed in de Wit and de Laat, 2012). Hi-C (3C combined with high-throughput sequencing) allows genome-wide investigation of the spatial chromatin organization, including long-distance interactions among multiple loci (Lieberman-Aiden et al., 2009). This technique can be useful in identifying enhancers by analyzing promoter–enhancer interactions (Ron et al., 2017).

Antibody-based enhancer identification

A number of antibody-based methods have proven useful when identifying possible enhancer regions (reviewed in Suryamohan and Halfon, 2015). ChIP-seq (chromatin immunoprecipitation followed by sequencing) is a widely used technique to either (1) identify the binding sites of a specific TF or (2) gain a genome-wide chromatin profile (Ghavi-Helm and Furlong, 2012; Ghavi-Helm et al., 2016; Park, 2009). For the former application, antibodies that specifically recognize the TF of interest are used to identify the regions that are occupied by the TF throughout the genome. Those binding sites are often indicative of the enhancers that are regulated by the investigated TF. For the latter application of ChIP-seq, antibodies against global chromatin modification markers are used. For example, antibodies against histone H3 with its lysine at position 27 acetylated (H3 K27Ac) can be used to identify active chromatin regions, while antibodies against histone H3 with its K27 trimethylated (H3 K27me3) are often useful to identify inactive regions in the genome (Bannister and Kouzarides, 2011). Antibodies against histone acetyltransferase p300 are also often used to identify active chromatin regions (Kharchenko et al., 2011; Nègre et al., 2011; Visel et al., 2009).

More recently, a new antibody-based method, CUT&RUN (cleavage under targets and release using nuclease combined with sequencing), has been developed (Meers et al., 2019; Skene and Henikoff, 2017; Skene et al., 2018). Briefly, in this method, unfixed permeabilized tissues/cells are incubated with antibodies that target a protein of interest (such as TFs or histones with a specific modification). Then, protein A conjugated micrococcal nuclease (MNase) is added, which binds to the antibody and cuts the DNA in its vicinity. The released DNA fragments are isolated through size selection and used for sequencing. CUT&RUN allows researchers to obtain data equivalent to ChIP-seq, but with fewer procedures and much less input tissue.

CRISPR/Cas9-based screening

Unlike RNAi, CRISPR/Cas9-based genome disruption can interrogate not only the function of transcriptionally active regions of the genome but also the non-coding portions, such as enhancers. Several high-throughput strategies have been established to identify possible enhancer regions through CRISPR/Cas9-based genome disruption, many of which use a tiling approach in a cultured cell setting and comprehensively survey a locus of interest with a collection of short guide RNAs (sgRNAs) designed to cover the entirety of the locus (reviewed in Catarino and Stark, 2018; Klein et al., 2018; Lopes et al., 2016). More recently, the next generation of CRISPR/Cas9 technologies, such as CRISPR-based transcriptional activation (CRISPRa) or interference (CRISPRi), have allowed researchers to manipulate the transcription of endogenous loci by taking advantage of the sequence specificity of the CRISPR/Cas9 system (reviewed in Adli, 2018; Pickar-Oliver and Gersbach, 2019). In brief, these techniques utilize a nuclease-inactive version of the Cas9 protein (dCas9) fused with either a transcription activation domain, such as VP64 or p300, or a repressive chromatin modifier domain, such as Krüppel-associated box (KRAB). These dCas9–effector fusion proteins can facilitate transcriptional regulation at any desired genomic site guided by an sgRNA (Gilbert et al., 2013). The initial studies utilizing the dCas9–effector fusion proteins focused on the transcribed regions of the genome [coding regions as well as long non-coding RNA (lncRNA) loci] (e.g. Ewen-Campen et al., 2017; Jia et al., 2018; Lin et al., 2015); however, this technique was later successfully used to modulate enhancer functions (e.g. Thakore et al., 2015, reviewed in Klein et al., 2018; Lopes et al., 2016). Considering that dCas9–effector techniques have successfully been used in a genome-wide fashion (albeit currently limited to a cultured cell setting), these techniques should be adoptable to identify endogenous enhancers in insects, especially if the context of interest can be studied in cell culture.

Computational enhancer prediction through integration of multiple enhancer features

Evolution of computational approaches

Early approaches to computational enhancer prediction often relied on a limited degree of knowledge of enhancer features, such as evolutionary conservation (i.e. phylogenetic footprints) and/or the tendency of TF binding motifs to cluster within an enhancer (Berman et al., 2002; Halfon et al., 2002; Markstein et al., 2002; also reviewed in Halfon and Michelson, 2002; Markstein and Levine, 2002). These studies resulted in successful identification of enhancers in *Drosophila*, especially during embryogenesis, but success rates were low and false-positive prediction rates high. In recent years, the field has started to coalesce around supervised machine learning approaches that are trained using one or more features from a known set of enhancers. These features can include the DNA sequence itself, epigenetic information such as histone methylation and acetylation status, DNA methylation status and nucleosome positioning, transcription factor and co-factor binding, and evidence of transcription (e.g. of 'enhancer RNAs'), among others. Support vector machines (SVMs) and random forest classifiers remain common approaches (e.g. Arbel et al., 2019; Chen et al., 2018; He et al., 2017; Le et al., 2019; Liu et al., 2018), although 'deep learning' approaches using artificial neural networks (ANNs) have been increasing in popularity as these methods become more mature and more feasible with current advances in computing power (e.g. Chen et al., 2018; Li et al., 2018; Liu et al., 2016; Min et al., 2017; Yang et al., 2017). In-depth reviews of computational enhancer discovery approaches have been provided elsewhere (Kleftogiannis et al., 2016; Lim et al., 2018; Suryamohan and Halfon, 2015), and the interested reader is directed to these for detailed treatment.

Generic versus specific enhancer prediction

In general, the current computational approaches can be classified into two types: 'generic' and 'specific.' Generic approaches rely on

characteristics likely to be common among all enhancers regardless of particular spatio-temporal specificity, such as histone modifications and chromatin accessibility. This broad applicability means that a method trained on a single set of known enhancers in a particular cell line, or functioning under a given set of biological conditions, is still likely to be effective for enhancer discovery in a different tissue, cell line or physiological milieu. Indeed, these characteristics may be able to carry over across vast evolutionary distances, allowing models trained on insect enhancers to be used for mammalian enhancer discovery (Sethi et al., 2018 preprint), and presumably vice-versa. However, generic approaches primarily provide a large list of sequences with predicted enhancer function, but no information as to what spatial, temporal or physiological characteristics these putative enhancers may have. Specific approaches, in contrast, attempt to discover discrete subsets of enhancers with common activity. While these may include general features such as chromatin accessibility or histone modification status obtained through the use of chromatin profiling techniques (such as FAIRE-seq or ATAC-seq) performed on a specific tissue of interest, they also include specific features such as presence of particular bound TFs or their binding sites, or the DNA sequence itself. Although specific approaches in general find fewer enhancers overall than generic approaches, true-positive prediction rates tend to be similar for both types of methods.

Enhancer prediction independent of experimentally derived features

When considering application to non-traditional insect models, methods that require training based on multiple experimentally derived features are of considerably less utility, as these data sets are often not available, and rarely, if ever, exist for multiple cell types or conditions. Therefore, approaches that rely solely on genome sequence are likely to be the most appealing to researchers that use non-traditional insect models. A number of these are available, most of which fall into the 'specific' enhancer discovery class (Chen et al., 2018; Kazemian and Halfon, 2019; Le et al., 2019; Liu et al., 2018). In general, these approaches deconstruct the training sequences into a set of small (e.g. 4–8 nucleotides) subsequences, or 'k-mers', which are then evaluated against a similarly deconstructed set of non-enhancer background sequences. With the notable exception of SCRMshaw (Kantorovitz et al., 2009; Kazemian and Halfon, 2019; Kazemian et al., 2011, 2014), most such approaches have not been tested with respect to insect genomes, including that of *Drosophila* (a somewhat ironic situation given the unmatched availability of empirically confirmed *Drosophila* enhancers for use as training data; Rivera et al., 2019). Although methods demonstrated to work using vertebrate genomes are expected to function equally well in insects, comparing efficacies is difficult given the different training and validation regimens applied. An evaluation platform for assessing methods using a uniform set of *Drosophila* training and validation data has recently been described (Asma and Halfon, 2019), and a critical comparison of various approaches would be a valuable addition to the field.

Considerations when choosing enhancer identification methods for non-traditional insect models

As showcased above, there are a variety of approaches that allow identification of possible enhancer regions from the genomes of
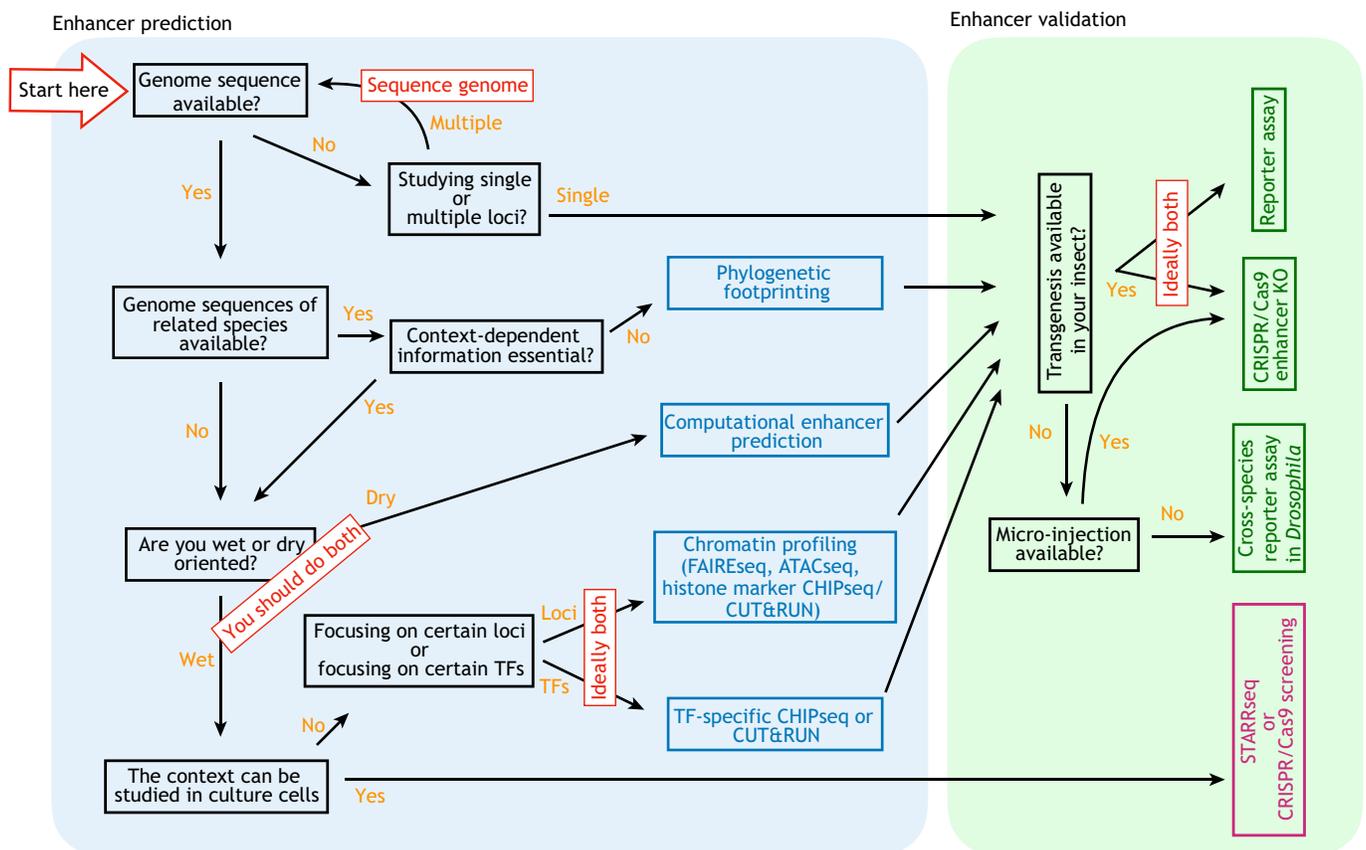

Fig. 3. A flow chart on enhancer prediction and validation. Blue boxes indicate the approaches that allow identification of candidate enhancer regions. The function of these enhancer candidates can be validated through the approaches indicated by green boxes. Some approaches allow both identification and validation simultaneously (pink box), especially when the context of interest can be studied in cultured cells.

insects (and many more that we could not cover here; see Suryamohan and Halfon, 2015 for a more comprehensive review on currently available techniques). However, the options are limited when using non-traditional insect models owing to the early stage of genetic and genomic resource development in these insects. Below, we focus on several options that are more likely applicable to non-traditional insect models, and discuss key points to consider when choosing a method depending on the insect used or the context studied (Figs 2 and 3).

Choice of experimental approaches

Reporter assays

Perhaps the first factor that influences the decision as to which approach to take would be whether one is interested in analyzing multiple loci or focusing on just one locus. A brute-force classic reporter assay-based survey for enhancers is a feasible option when focusing on analyzing the regulation of just one gene (Fig. 2A), assuming that either a valid reporter assay system is available in the study insect or the assay can be performed in *Drosophila* (see 'Validating and investigating enhancer function' below for more detailed discussion). However, the position of enhancers in relation to the gene of interest is quite unpredictable, from tens of thousands of base pairs upstream or downstream of the gene they regulate to inside of an intron or even sometimes within an exon (for example, see Arnold et al., 2013; Kvon et al., 2014), making the reporter assay-based enhancer search time-consuming, tedious and quite risky. Therefore, considering that genome sequences of many insects are now available and long-read sequencing technologies continue to advance (reviewed in Levy and Myers, 2016), it is beneficial to utilize some additional experimental approaches to identify possible enhancer regions even when focusing on only a single locus.

When genome sequences of a set of closely related species (such as within the same genus) are available or can be obtained, phylogenetic footprinting can mitigate the risk of a reporter assay-based enhancer search by providing candidate regions based on evolutionary conservation (Fig. 2B). However, as mentioned, conservation is not always reliable for predicting enhancers. Also, phylogenetic footprinting does not provide any context-dependent information (such as which enhancers are active in which tissues, at which time points or under which physiological conditions).

Nonetheless, evolutionary conservation may be informative in a certain genus/family of insects, making phylogenetic footprinting a possible option to consider, particularly as a means of refining the boundaries of a putative enhancer sequence predicted by other methods.

Unfortunately, a genome-wide reporter assay is currently not an option for most insects, as it requires a large workforce, a well-annotated genome, and a highly established and efficient transgenic technique (the Flylight project in *Drosophila*, for example). However, some unique circumstances make a high-throughput reporter assay a feasible option for genome-wide enhancer identification. For instance, STARR-seq is an option when the context of interest can be studied using a cultured cell line or perhaps even with a tissue that is culturable and transfectable *in vitro* (Fig. 2C). Therefore, although contexts are limited, when applicable, a high-throughput reporter assay can be a very powerful approach that will allow a comprehensive identification of functionally validated enhancers.

When performing a reporter assay-based enhancer search, whether a brute-force survey, aided by phylogenetic footprinting, or a high-throughput approach, there is a significant caveat in regard to the backbone structure of the reporter construct used in the assay, such as the choice of core promoter. There is no guarantee that reporter constructs previously established in *Drosophila* are transferable to a different insect of interest. We discuss this point, along with other potential challenges of establishing a reporter assay system in non-traditional insect models, in a later section (see 'Validating and investigating enhancer function').

Chromatin profiling

When a well-annotated genome sequence is available, chromatin profiling can provide rich information about the *cis*-regulatory landscape of the genome of the study insect. Currently, FAIRE-seq and ATAC-seq appear to be the primary approaches when investigating non-traditional insect models (Fig. 2D) [e.g. *Aedes* (Behura et al., 2016), *Anopheles* (Pérez-Zamorano et al., 2017), *Tribolium* (Lai et al., 2018), *Heliconius* (Lewis and Reed, 2019), *Junonia* (van der Burg et al., 2019), *Bombyx* (Zhang et al., 2017c, 2019)] as they do not require any special reagents and can be performed with a relatively small amount of input tissue. For example, we previously performed FAIRE-seq with tissues of the
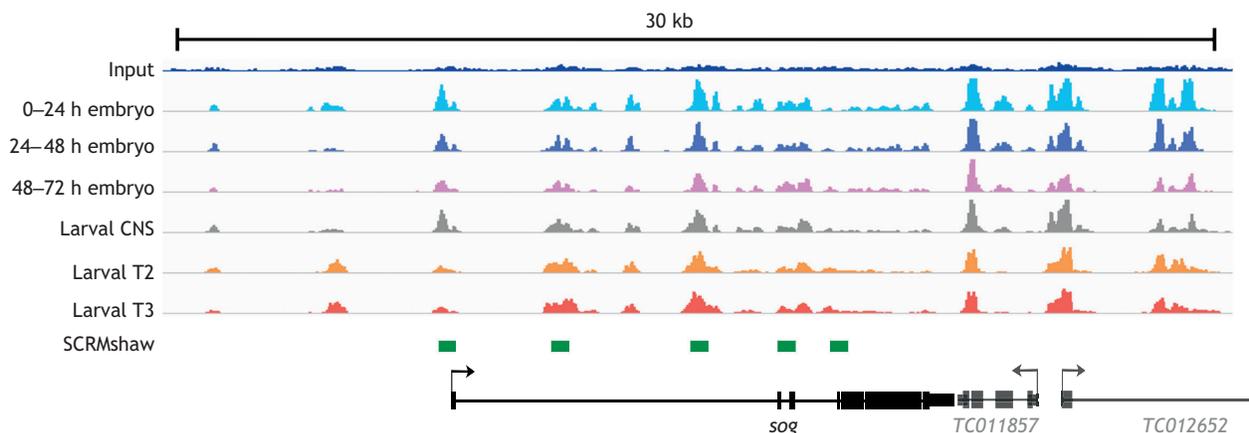

Fig. 4. A significant overlap between FAIRE peaks and SCRMshaw predictions. FAIRE profiles and SCRMshaw predictions at the *Tribolium sog* locus in six different tissues/stages. More examples of the overlap between FAIRE peaks and SCRMshaw predictions can be found in fig. S3 of Lai et al. (2018).

red flour beetle (*Tribolium castaneum*) and successfully obtained genome-wide chromatin profiles from various tissues and stages of this insect (Fig. 4) (Lai et al., 2018). More than 40,000 open chromatin regions in the *Tribolium* genome are detected, and comparison of the profiles across the samples revealed a distinct set of open chromatin regions in each tissue and at each stage. Many of these context-dependent openings likely correspond to tissue- and timing-specific enhancers, thus demonstrating the usefulness of chromatin profiling when studying non-traditional model insects.

With the use of antibodies against global chromatin modification markers, antibody-based methods, such as ChIP-seq, are also powerful at obtaining a genome-wide chromatin landscape (Fig. 2D). These techniques can reveal context-dependent and/or tissue-specific chromatin profiles, which is very useful when identifying possible enhancer regions that are active uniquely in a certain context. The requirement of a large amount of input tissues has been a significant limiting factor when using ChIP-seq (for instance, thousands of discs are likely required for one biological replicate if ChIP-seq is performed with *Drosophila* imaginal discs); however, CUT&RUN might now allow researchers to perform an equivalent analysis with a much smaller amount of input tissue. Through a combination of these chromatin profiling techniques, the Reed lab has revealed the genome-wide chromatin landscape of *Heliconius* butterfly wings, a beautiful example of the use of chromatin profiling in a non-traditional insect model (Lewis and Reed, 2019; Lewis et al., 2016).

ChIP-seq can also be used for identifying the binding sites of a particular TF throughout the genome (Fig. 2D). TF binding sites detected by ChIP-seq are often instructive when identifying context-dependent enhancers that are under the regulation of the investigated TF; thus this approach is quite advantageous when you know which TF to study, or which TF possibly regulates the gene of interest. The requirement for a high-quality 'ChIP-compatible' antibody against the TF of interest and the need for a large amount of input tissue have been significant drawbacks of this technique; however, the latter can now be bypassed by using CUT&RUN. One important

---

Box 1. Influence of the status of genome assemblies on computational enhancer prediction

An important point to consider when applying computational enhancer prediction to non-traditional insect models is the status of their genome assemblies and gene annotations. Although the count of sequenced insect species is currently ~470 (i5k: Sequencing Five Thousand Arthropod Genomes; http://i5k.github.io/arthropod_genomes_at_ncbi), assemblies are of varying quality, ranging from the extremely well-assembled *Drosophila melanogaster* (contig N50=21 Mb) to the poorly assembled meadow spittlebug *Philaenus spumarius* (contig N50=319 bp), and fewer than 40% have accompanying gene annotation (Li et al., 2019). How effective is enhancer discovery when genome assemblies are highly incomplete? Testing SCRMshaw with simulated dis-assembly of the *Drosophila* genome has revealed that contig N50s of at least 23,000 bp (which encompasses the upper 50% of current insect assemblies) are sufficient for effective SCRMshaw prediction, with minor loss of sensitivity and negligible increase in false-positive rates (Asma and Halfon, 2019). Therefore, highly complete genome assembly does not appear to be a prerequisite for successful enhancer prediction by SCRMshaw. Requirements for gene annotation are more difficult to assess. Annotation is not strictly necessary for enhancer prediction, but can certainly facilitate it. For example, SCRMshaw disregards coding sequences to focus on the regions that more likely contain enhancers, i.e. non-coding regions. The effect of gene annotation quality on computational enhancer prediction has not been explored.

---

point that requires attention when using TF binding sites to identify enhancers relates to the affinity of TFs to DNA. Recent studies have revealed that low-affinity binding of TFs to DNA can also be critical for gene regulation (Crocker et al., 2015, 2016). These low-affinity TF binding sites might not be readily detected by ChIP-seq and other antibody-based methods (as these techniques rely on strong binding of TFs to DNA), presenting a risk of missing biologically relevant TF binding sites. Moreover, low-affinity TF binding sites are often not evolutionary conserved (Crocker et al., 2015), further compounding the difficulty of finding enhancers.

It is worth emphasizing that the 'enhancers' identified through chromatin profiling described in this section, as well as the computationally identified enhancers (Fig. 2D, next section), are all still predictions. Therefore, it is imperative to functionally validate these candidate enhancer regions.

Computational approaches for non-traditional insect models

Computational approaches are an attractive option for use with non-traditional insect models in that they are quick, inexpensive, and in many cases do not rely on extensive empirically derived genomic data. As mentioned, approaches that rely solely on genome sequence (see Box 1 for discussion about how the status of genome assemblies and gene annotations influence enhancer prediction), such as SCRMshaw, are the most appealing. However, there is a significant caveat when applying supervised sequence-based approaches to non-traditional insect models: the acute dearth of training data, as few insect enhancers are known outside of *Drosophila*. Interestingly, SCRMshaw trained with known *Drosophila* enhancers was demonstrated to effectively discover enhancers throughout the 345 Mya range of holometabolous insects (Kazemian et al., 2014), indicating that *Drosophila* enhancers can be useful as training data at least for the genomes of the Holometabola. When SCRMshaw-predicted enhancers from other insects, including bees, wasps, beetles and mosquitoes, are tested in reporter gene assays in transgenic *Drosophila*, they validate at rates similar to those seen from within-species prediction of *Drosophila* enhancers (Kazemian et al., 2014; Suryamohan et al., 2016). Direct testing of *Tribolium* enhancers in transgenic *Tribolium* confirms that SCRMshaw can find bona fide enhancers cross-species (Lai et al., 2018). Although not tested in insects, Chen et al. (2018) similarly demonstrate that a *k*-mer-based prediction method trained using data from a single species can be used for enhancer discovery across a range of mammalian genomes. When trained on a tissue-specific enhancer set, their method performed better at discovering enhancers in the same tissue in other species than in different tissues of the same species. Together, these studies indicate that specific enhancer characteristics could be learned and applied in a cross-species setting, and therefore *k*-mer-based enhancer predictions will be useful when studying non-traditional insect models.

Integrating experimental and computational approaches

As discussed above, each approach has its strengths and weaknesses. Therefore, the use of multiple strategies (ideally both experimental and computational), and comparison across the outcomes of several different approaches, is likely to be the most fruitful in narrowing down candidate regions to be functionally validated. In *Tribolium*, we compared the FAIRE profiles with SCRMshaw predictions and found surprisingly high overlaps between these two datasets (Fig. 4) (Lai et al., 2018). However, in the case of *Tribolium* (but we think this can be generalizable), chromatin profiling provided too many candidate regions (>40,000

peaks across samples), while *k*-mer-based computational prediction was too stringent and identified a relatively small number of candidate enhancers (~1200 regions). Nonetheless, having two independent enhancer prediction approaches greatly helped us narrow down the enhancers for functional validation. Adding more tissues and/or using more homogeneous tissues/cell types for chromatin profiling, along with enhancing *k*-mer-based computational prediction through the use of improved training data and other refinements, will help increase resolution when identifying candidate regions for context-specific enhancers.

Validating and investigating enhancer function

Unless a reporter assay system is used to screen for enhancers (Fig. 2A–C), the enhancer regions identified through the methods described above (Fig. 2D), either chromatin profiling or computational approaches, are still predictions that require functional validation. In this section, we discuss several possible validation approaches when studying enhancers of non-traditional insect models, such as the reporter assay and CRISPR/Cas9-based genome editing. We also highlight some key issues and potential pitfalls when establishing a reporter assay system in insects outside of *Drosophila*.

Testing activity of non-*Drosophila* enhancers in *Drosophila*

As mentioned, confirmation of enhancer activity *in vivo* via a reporter assay is widely considered to be the gold standard when validating enhancer function (Fig. 1B). Since the first application of a reporter assay in *Drosophila* in the 1980s (Goto et al., 1989; Harding et al., 1989; Hiromi et al., 1985), reporter assays have been used to investigate the regulation and evolution of numerous genes in *Drosophila*, identifying over 20,000 enhancers (Rivera et al., 2019) and generating a large variety of useful reporter constructs. Although now feasible in a growing number of species (Fraser, 2012), making transgenic lines is a laborious task when using non-traditional insect models. Therefore, considering the ease of making transgenic lines in *Drosophila* (in part thanks to low-cost commercial injection services) and the availability of established reporter assay systems, the logical first step is to test the activity of possible enhancer regions identified from non-*D. melanogaster* insects (including various species in the genus *Drosophila*) in *D. melanogaster*. This approach has been quite successful when studying enhancer evolution among multiple *Drosophila* species (e.g. Frankel et al., 2011; Gompel et al., 2005; also see Rebeiz and Williams, 2017; Stern and Frankel, 2013 for review). Some studies have even demonstrated that enhancers from insect orders outside of Diptera work in *Drosophila*. For example, enhancers of some developmental genes in beetles, honeybees and even spiders were demonstrated to be active in their expected contexts in *Drosophila* (e.g. Ayyar et al., 2010; Cande et al., 2009a,b; Kazemian et al., 2014; Lai et al., 2018; Prasad et al., 2016; Wolff et al., 1998; Zinzen et al., 2006). These studies show the power of the cross-species reporter assay using *Drosophila* as an *in vivo* test tube, but with one unavoidable concern: are these non-*Drosophila* enhancers really showing biologically relevant activities in *Drosophila*? Owing to this obvious caveat of using a cross-species reporter assay, it is ideal if the enhancer activity is also tested in the native species.

Establishing a reporter assay in non-traditional insect models

We often assume that the reporter constructs and other genetic tools established in *Drosophila* are readily transferable to other insects. However, considering the deep divergence and the vast diversity among insect orders, there is no guarantee that these reporter

---

Box 2. A quest to identify a proper core promoter in *Tribolium*

Since the early time of reporter assays, the core promoter of the *Heat Shock Protein 70* (*Hsp70*) gene has been widely used in *Drosophila* when an exogenous core promoter is required (Goto et al., 1989; Harding et al., 1989; Hiromi and Gehring, 1987). This core promoter functioned properly in other insects when 3xP3, a synthetic enhancer composed of three repeats of the P3 Pax6 homeodomain binding site (Mishra et al., 2010; Sheng et al., 1997), was used to drive expression in eye- and nervous-related tissues as a marker of transgenesis [e.g. *Drosophila virilis* (Horn and Wimmer, 2000), *Tribolium* (Berghammer et al., 1999; Lorenzen et al., 2003), *Bombyx mori* (Thomas et al., 2002)]. The *Dm-hsp70* core promoter was also used for a reporter assay in *Tribolium* to identify embryonic enhancers that regulate the *Tribolium hairy* gene (Eckert et al., 2004). However, when establishing the Gal4/UAS system in *Tribolium*, Schinko et al. (2010) found that the *Dm-hsp70* core promoter does not reliably work in their constructs. SCP1 (Super Core Promoter 1, a composition of *Drosophila* and viral core promoter motifs that was designed to drive a high level of transcription in Hela cells; Juven-Gershon et al., 2006) also did not work well when used in a UAS construct either in *Tribolium* or in *Drosophila* (Schinko et al., 2010). These outcomes led Schinko et al. (2010) to try *Tribolium*-native promoters, the core promoters of *Tc-hsp68* (*Tc-bhsp68*) and *Tc-hairy*. Although the *Tc-hairy* core promoter failed to work properly (it drove an unexpected nervous-system-specific expression), the *Tc-bhsp68* core promoter worked well with their UAS constructs in *Tribolium*. Based on the extensive characterization of core promoter compatibility in *Tribolium* by Schinko et al. (2010), we thought that the *Tc-bhsp68* core promoter would be a safe choice when we attempted to establish a reporter assay system in *Tribolium*. However, surprisingly, the *Tc-bhsp68* core promoter failed to work reliably in *Tribolium* in a reporter construct with a wing enhancer of the *nubbin* gene (*Tc-nub*), even though the same construct worked well in *Drosophila* (Lai et al., 2018). We decided to try DSCP (*Drosophila* Synthetic Core Promoter), the core promoter that was established for a genome-wide reporter assay in *Drosophila* (the FlyLight project) (Pfeiffer et al., 2008). This core promoter is a chimera of the SCP and the *Drosophila eve* gene core promoter, which was shown to work more efficiently, and with a more diverse array of developmental enhancers, as compared with gene-specific core promoters in *Drosophila* (Pfeiffer et al., 2008). The DSCP also was found to work preferentially with developmental enhancers (versus housekeeping gene enhancers) when tested with STARR-seq (Zabidi et al., 2015). The DSCP we used was a variation of this promoter, containing a fragment of *Dm-hsp70* promoter downstream of the transcription initiation site in addition to the motifs taken from SCP and the *eve* core promoter (cloned from the construct used in McKay and Lieb, 2013; see fig. 6 of Lai et al., 2018 for the sequence and annotation of this promoter). This DSCP worked well in our reporter construct both in *Tribolium* and *Drosophila* and in two very different developmental contexts (wing development and embryogenesis) in *Tribolium*, thus allowing us to establish a cross-species compatible reporter assay system (Lai et al., 2018).

It is currently unknown why the *Dm-hsp70* and *Tc-bhsp68* core promoters did not work properly when used in some transgenic constructs in *Tribolium*. Interestingly (and confusingly), these core promoters drove various patterns of enhancer–trap expression when inserted in the *Tribolium* genome (Lai et al., 2018; Lorenzen et al., 2003; Trauner et al., 2009). This indicates that these core promoters are capable of working with a diverse array of enhancers in *Tribolium*, even though they did not work properly in the tested artificially configured transgenic constructs. Also worth mentioning is the use of gene-specific promoters (not to be confused with 'core' promoters, see Fig. 1A). Promoters of several housekeeping genes have been successfully used to drive gene expression in *Tribolium* (Gilles et al., 2019; Lai et al., 2018; Lorenzen et al., 2002; Rylee et al., 2018; Sarrazin et al., 2012; Schinko et al., 2012; Siebert et al., 2008; Strobl et al., 2018). However, when we tested the *Tc-nub* promoter with the *Tc-nub* wing enhancer in a reporter construct, this construct failed to drive any expression either in *Tribolium* or in *Drosophila* (Lai et al., 2018). These confusing outcomes regarding promoters might be related to enhancer–core promoter compatibility and/or an optimal distance between the enhancer and the core promoter, which will require detailed investigation in the future.

constructs will function properly in other insects, especially those outside of the order Diptera. In fact, several groups including ourselves have encountered various interesting issues when transferring *Drosophila* constructs to *Tribolium* (Lai et al., 2018; Schinko et al., 2010). One of the major issues was related to the choice of core promoter (also known as 'minimal' or 'basal' promoter; reviewed in Kadonaga, 2012; Vo Ngoc et al., 2019). The most widely used core promoter in insects, the core promoter of the *Drosophila Heat Shock Protein 70* (*Hsp70*) gene, did not work reliably in *Tribolium*, forcing researchers to look for an alternative core promoter. Schinko et al. (2010) identified that a *Tribolium*-native promoter, the core promoter of *Tc-hsp68* (*Tc-bhsp68*), works well for their Gal4/UAS system, while Lai et al. (2018) determined that a variation of the DSCP (*Drosophila* Synthetic Core Promoter) properly drives gene expression when placed in a reporter construct (see Box 2 for details). Other components of transgenic constructs, such the choice of untranslated regions (UTRs) or the inclusion of exogenous genes that are often used in modern genetics, could also present problems. For instance, the yeast Gal4 gene has been used routinely in the gene misexpression system (Gal4/UAS system) in *Drosophila* (Brand and Perrimon, 1993) and has been successfully transferred to mosquitoes (Kokoza and Raikhel, 2011; Lynd and Lycett, 2012; O'Brochta et al., 2012) and silk moths (Imamura et al., 2003) without any major modification. However, when Schinko et al. (2010) worked on transferring the Gal4/UAS system to *Tribolium*, they noticed that the full-length *Gal4* gene does not work in *Tribolium*. We also confirmed this, even though the *Gal4* gene is transcribed in *Tribolium* (K. D. Deem and Y. Tomoyasu, unpublished data). Schinko et al. (2010) also tried two shorter and more active versions of Gal4, Gal4-VP16 and Gal4Δ (Ma and Ptashne, 1987; Viktorinová and Wimmer, 2007), both of which worked in *Tribolium*. These outcomes regarding the core promoter and other components highlight the potential difficulty of establishing a reporter assay in non-traditional insect models; however, we hope that the various issues we and others have encountered (such as the deep rabbit hole of core promoters; Box 2) will serve as guidance when working on other insects.

One aspect that often makes this type of technology transfer so challenging is the absence of reliable positive controls in non-traditional insect models. For instance, we did not have any known *Tribolium* enhancers that work in the context we study, forcing us to assume that the potential enhancer we identified through a cross-species assay was in fact a true functional *Tribolium* enhancer and blindly use it as a positive control when we were troubleshooting our reporter constructs (Lai et al., 2018). The enhancers we validated through our cross-species reporter assay (such as Tc-Nub1L) are functional in both *Drosophila* (Diptera) and *Tribolium* (Coleoptera); thus these enhancers might serve as positive controls in a wide range of insects (at least in Holometabola). Hopefully the number of positive control enhancers will quickly increase as more enhancer studies are performed in new insect models.

Functional analysis of enhancers through genome editing
Although validation by a reporter assay continues to be the gold standard when studying enhancers, reporter assays do come with several caveats, even beyond the species-specific issues discussed above. For instance, the distance between the enhancer and the core promoter within a reporter construct has been observed in some cases to affect proper enhancer activity (e.g. Small et al., 1993; Swanson et al., 2010). Compatibility between enhancers and core promoters is another potential issue, which can drastically influence the outcome of the assay (e.g. Pfeiffer et al., 2008; also see Zabidi

et al., 2015 for an extreme case of compatibility issues between two core promoters, reviewed in Atkinson and Halfon, 2014). Considering these caveats, in addition to the potential hassles described above when establishing a reporter assay system in non-*Drosophila* insects, an LOF approach through CRISPR/Cas9-based disruption of enhancers is an attractive alternative when validating enhancer function (also discussed in Duester, 2019).

Several CRISPR/Cas9-based methods have been used to investigate the function of enhancers in *Drosophila*, some of which are described in the previous section. For instance, Xu et al. (2017) have used a split-drive configuration of a CRISPR/Cas9-based gene drive system (dubbed as CopyCat) to replace the wing vein enhancer of the *knirps* (*kni*) gene in *Drosophila* with that of a mutant allele or the homologous enhancers of other dipteran species. Also, some next-generation CRISPR/Cas technologies, such as CRISPRa, have been successfully used *in vivo* using *Drosophila* (Ewen-Campen et al., 2017; Jia et al., 2018; Lin et al., 2015). Although not CRISPR/Cas9-based, Crocker and Stern (2013) used a method that is conceptually similar to dCas9-effector technologies (transcription activator-like effectors, TALEs) to interrogate enhancer function in *Drosophila*, demonstrating that the dCas9-effector technologies can be quite useful when studying insect enhancers.

Although these and other elaborated CRISPR/Cas technologies (reviewed in Bier et al., 2018) are attractive, a simple CRISPR/Cas9-based knockout via NHEJ (non-homologous end joining) may be most appealing to researchers that use non-traditional insect models owing to the challenging nature of implementing these new technologies in insects outside of *Drosophila*. Since the first breakthrough application of the CRISPR/Cas9 system in genome editing in 2012 and 2013 (Cong et al., 2013; Jinek et al., 2012), CRISPR/Cas9-based knockout techniques have already been applied to various orders of non-traditional model insects and other arthropod species [e.g. Coleoptera (Gilles et al., 2015), Lepidoptera (Connahs et al., 2019; Mazo-Vargas et al., 2017; Prakash and Monteiro, 2018; Wei et al., 2014; Zhang and Reed, 2016), Hymenoptera (Trible et al., 2017; Yan et al., 2017), Orthoptera (Watanabe et al., 2017), Zygentoma (Ohde et al., 2018) and crustacean species (Martin et al., 2016; Nakanishi et al., 2014), reviewed in Gantz and Akbari, 2018; Gilles and Averof, 2014], showing the relative ease of adopting this technique in insects.

The idea of using CRISPR/Cas9 knockout to validate enhancer function is straightforward: remove/disrupt the candidate enhancer from the genome and evaluate the resulting phenotype. There are several points that require attention when designing an enhancer knockout experiment using non-traditional insect models. From a technical point of view, researchers need to decide whether to analyze the mutant phenotype in the somatic cells of G0 insects or in established mutant lines. Analyzing in G0 is beneficial for insects with difficulties in husbandry and/or a longer generation time, as it allows bypassing multiple generations of crosses to establish a mutant strain and analyzing the outcome immediately in the individuals that received CRISPR/Cas9 injection. This approach has been very successful in butterflies (e.g. Connahs et al., 2019; Mazo-Vargas et al., 2017; Prakash and Monteiro, 2018; Zhang and Reed, 2016; Zhang et al., 2017a,b) as well as in other species such as in the crustacean *Parhyale* (Bruce and Patel, 2018 preprint; Clark-Hachtel and Tomoyasu, 2017 preprint; Martin et al., 2016). However, to be able to detect mutant phenotypes in G0, (1) genome editing events need to happen in a large enough number of somatic cells and (2) mutant phenotypes need to be clearly visible (e.g. pigmentation

and patterning defect, loss of tissues) in the context of study. In other words, the outcome is interpretable likely only when genome editing works properly and results in a clear mutant phenotype, thus presenting a potential caveat of analyzing mutant phenotypes in G0. Establishing a mutant line is a safer option if genetics of your insects allows, and if a valid screening scheme is available to identify and track mutations. A stable line allows for more quantitative measures to be brought to bear, such as qRT-PCR to measure changes in gene expression. Replacing and/or disrupting the targeted enhancer with a marker construct (such as a fluorescent gene driven by 3xP3) might allow an easier tracking of the mutation compared with the mutation created through a simple NHEJ knockout, which likely requires genomic PCR-based methods to identify the mutation (unless the mutation results in a haplo-insufficient, visible and viable phenotype). Although the HDR (homology directed repair) knock-in approach appears to suffer from low efficiency when used in non-traditional insect models (C. M. Clark-Hachtel, K. D. Deem and Y. Tomoyasu, unpublished data), various strategies have been developed to improve success rates (e.g. Aird et al., 2018; Savic et al., 2018, also see Bier et al., 2018; Liu et al., 2019 for review of additional methods to increase the HDR efficiency), some of which might be worth pursuing when performing HDR knock-in in insects. In addition, NHEJ knock-in (Auer et al., 2014; Watanabe et al., 2017), including the CRISPaint (CRISPR-assisted insertion tagging) system (Bosch et al., 2020; Schmid-Burgk et al., 2016), or MMEJ (microhomology mediated end joining) knock-in (e.g. Nakade et al., 2014) might facilitate a more efficient disruption of enhancers while also tagging the genome-editing event with a visible marker.

A further aspect that requires attention when knocking out enhancers is related to the nature of enhancers. It has been shown that multiple enhancers sometimes act redundantly (dubbed as 'shadow enhancers'), which fosters a robust and stable gene expression (Perry et al., 2010) and might also facilitate evolution of novel traits (Hong et al., 2008). Enhancer redundancy appears to be pervasive among developmental genes (Cannavò et al., 2016). This may cause an issue when performing an enhancer knockout, as targeting one enhancer might not be sufficient to cause any visible abnormalities, which is especially problematic when analyzing in G0. Whether these features of enhancers found in *Drosophila* are conserved in other insects is not known, which is one of the very reasons why it is essential to study enhancers in various organisms. Also, considering that knocking out enhancers also presents some caveats, it would be ideal to validate enhancer function via multiple methods, such as a reporter assay and knockout, through which both necessity and sufficiency of an enhancer can be evaluated (discussed in detail in Catarino and Stark, 2018; Halfon, 2019).

Pushing the frontiers of enhancer studies in non-traditional insect models

With the recent confluence of effective enhancer-discovery approaches established in *Drosophila*, and the sequencing of numerous insect genomes (Thomas et al., 2018 preprint), the time is ripe to start broadening the investigation of enhancers and other *cis*-regulatory mechanisms to a wider range of insects. There are several critical areas that should receive high priority for active development to make enhancer studies readily possible in non-traditional insect models.

On the computational side, these areas include (1) the use of improved training data by incorporating the latest available data (such as the data from the REDfly database; Rivera et al., 2019), and

(2) better integration of computational and experimental enhancer-discovery methods (such as ATAC-seq and FAIRE-seq). Another interesting area to pursue would be to leverage comparative genomics in improving the accuracy of computational predictions. As mentioned, enhancer sequences are frequently alignable between closely related species, but conservation diminishes with increasing evolutionary distance. Nevertheless, enhancer locations are often maintained in equivalent locations despite the inability to directly align the enhancer sequences themselves (Cande et al., 2009a; Kazemian et al., 2014). By prioritizing those enhancer predictions among moderately related species that fall into the same approximate genomic location, we should be able to not only reduce false-positive predictions, but also build up sets of evolutionarily related but non-alignable enhancers. The latter will provide an unprecedented resource for probing the evolution of regulatory sequences. With its average 75% success rate (Kazemian et al., 2014), SCRMshaw has emerged as a promising first-line tool for identifying enhancer sequences in non-traditional insect models, whose prediction capacity will be further improved through these areas of focus.

Although advances in computational and genomics approaches have begun to enable enhancer studies in insects outside of *Drosophila*, the underdeveloped nature of functional genetics and genomics tools in non-traditional insect models still hinders researchers from functionally dissecting diverse mechanisms underlying *cis*-regulation and investigating how changes in *cis*-regulation have contributed (and are contributing) to the evolution of various traits at the detailed molecular level. The cross-species compatible reporter assay construct we previously established (Lai et al., 2018) is a step forward towards performing enhancer studies in various insects, and there are several areas that we can explore to continue our progress for a better implementation of a reporter assay system and other functional genetics tools in non-traditional insect models. The first area is related to core promoters. As we discussed extensively in Box 2, it is crucial to choose the right core promoter that fits the gene, context and species of the study. The modified DSCP we used in our study worked well in two species (a coleopteran and a dipteran, representing a large span of the Holometabola) and in two developmental contexts (appendage development and embryogenesis), suggesting that this core promoter can be used in a wide taxonomy of insects. Nonetheless, the DSCP is mainly constructed with *Drosophila* core promoter motifs (Juven-Gershon et al., 2006; Lai et al., 2018; Pfeiffer et al., 2008); it might therefore be possible to tailor a more efficient core promoter for each species by a similar strategy. It may also be possible to design a universal core promoter that works across multiple orders of insects. The second area is to expand the genetic toolkit for non-*Drosophila* insects. Tissue- and context-specific enhancers identified through enhancer studies will be essential when developing additional genetic tools and resources. With the use of these enhancers combined with some modifications to reporter constructs, we will be able to build various modern genetic tools useful for lineage tracing, gene misexpression, tissue-specific RNAi and CRISPR, and beyond. Developing these functional genetic tools will further accelerate investigation of enhancers as well as of many other aspects of biology in non-traditional insect models.

Armed with ample genetic, genomic and computational tools and resources, studies in *Drosophila* have revealed a wealth of intriguing aspects of enhancer function and evolution. In this Review, we mainly focused on how to 'find' enhancers in non-traditional insect models. This is just the first step in investigating the amazing array

of diverse traits found in insects from the *cis* point of view, a widely unexplored area of biology.

References

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* **9**, 1911. doi:10.1038/s41467-018-04252-2

Aird, E. J., Lovendahl, K. N., St Martin, A., Harris, R. S. and Gordon, W. R. (2018). Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template. *Commun. Biol.* **1**, 54. doi:10.1038/s42003-018-0054-2

Arbel, H., Basu, S., Fisher, W. W., Hammonds, A. S., Wan, K. H., Park, S., Weiszmann, R., Booth, B. W., Keranen, S. V., Henriquez, C. et al. (2019). Exploiting regulatory heterogeneity to systematically identify enhancers with high accuracy. *Proc. Natl Acad. Sci. USA* **116**, 900-908. doi:10.1073/pnas.1808833115

Arnold, C. D., Gerlach, D., Stelzer, C., Boryn, L. M., Rath, M. and Stark, A. (2013). Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074-1077. doi:10.1126/science.1232542

Asma, H. and Halfon, M. S. (2019). Computational enhancer prediction: evaluation and improvements. *BMC Bioinformatics* **20**, 174. doi:10.1186/s12859-019-2781-x

Atkinson, T. J. and Halfon, M. S. (2014). Regulation of gene expression in the genomic context. *Comput. Struct. Biotechnol. J.* **9**, e201401001. doi:10.5936/csbj.201401001

Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. and Del Bene, F. (2014). Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Res.* **24**, 142-153. doi:10.1101/gr.161638.113

Ayyar, S., Negre, B., Simpson, P. and Stollewerk, A. (2010). An arthropod cis-regulatory element functioning in sensory organ precursor development dates back to the Cambrian. *BMC Biol.* **8**, 127. doi:10.1186/1741-7007-8-127

Bannister, A. J. and Kouzarides, T. (2011). Regulation of chromatin by histone modifications. *Cell Res.* **21**, 381-395. doi:10.1038/cr.2011.22

Behura, S. K., Sarro, J., Li, P., Mysore, K., Severson, D. W., Emrich, S. J. and Duman-Scheel, M. (2016). High-throughput *cis*-regulatory element discovery in the vector mosquito *Aedes aegypti*. *BMC Genomics* **17**, 341. doi:10.1186/s12864-016-2468-x

Bellés, X. (2010). Beyond *Drosophila*: RNAi *in vivo* and functional genomics in insects. *Annu. Rev. Entomol.* **55**, 111-128. doi:10.1146/annurev-ento-112408-085301

Berghammer, A. J., Klingler, M. and Wimmer, E. A. (1999). A universal marker for transgenic insects. *Nature* **402**, 370-371. doi:10.1038/46463

Bergman, C. M., Pfeiffer, B. D., Rincón-Limas, D. E., Hoskins, R. A., Gnirke, A., Mungall, C. J., Wang, A. M., Kronmiller, B., Pacleb, J., Park, S. et al. (2002). Assessing the impact of comparative genomic sequence data on the functional annotation of the *Drosophila* genome. *Genome Biol.* **3**, RESEARCH0086. doi:10.1186/gb-2002-3-12-research0086

Berman, B. P., Nibu, Y., Pfeiffer, B. D., Tomancak, P., Celniker, S. E., Levine, M., Rubin, G. M. and Eisen, M. B. (2002). Exploiting transcription factor binding site clustering to identify cis-regulatory modules involved in pattern formation in the *Drosophila* genome. *Proc. Natl Acad. Sci. USA* **99**, 757-762. doi:10.1073/pnas.231608898

Bier, E., Harrison, M. M., O'Connor-Giles, K. M. and Wildonger, J. (2018). Advances in engineering the fly genome with the CRISPR-Cas system. *Genetics* **208**, 1-18. doi:10.1534/genetics.117.1113

Blackwood, E. M. and Kadonaga, J. T. (1998). Going the distance: a current view of enhancer action. *Science* **281**, 60-63. doi:10.1126/science.281.5373.60

Bosch, J. A., Colbeth, R., Zirin, J. and Perrimon, N. (2020). Gene knock-ins in *Drosophila* using homology-independent insertion of universal donor plasmids. *Genetics* **214**, 75-89. doi:10.1534/genetics.119.302819

Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z., Furey, T. S. and Crawford, G. E. (2008). High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311-322. doi:10.1016/j.cell.2007.12.014

Brand, A. H. and Perrimon, N. (1993). Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401-415.

Bruce, H. S. and Patel, N. H. (2018). Insect wings and body wall evolved from ancient leg segments. *bioRxiv*. doi:10.1101/244541

Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. and Greenleaf, W. J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213-1218. doi:10.1038/nmeth.2688

Buffry, A. D., Mendes, C. C. and McGregor, A. P. (2016). The functionality and evolution of eukaryotic transcriptional enhancers. *Adv. Genet.* **96**, 143-206. doi:10.1016/bs.adgen.2016.08.004

Cande, J., Goltsev, Y. and Levine, M. S. (2009a). Conservation of enhancer location in divergent insects. *Proc. Natl. Acad. Sci. USA* **106**, 14414-14419. doi:10.1073/pnas.0905754106

Cande, J. D., Chopra, V. S. and Levine, M. (2009b). Evolving enhancer-promoter interactions within the tinman complex of the flour beetle, *Tribolium castaneum*. *Development* **136**, 3153-3160. doi:10.1242/dev.038034

Cannavò, E., Khoueiry, P., Garfield, D. A., Geeleher, P., Zichner, T., Gustafson, E. H., Ciglar, L., Korbel, J. O. and Furlong, E. E. M. (2016). Shadow enhancers are pervasive features of developmental regulatory networks. *Curr. Biol.* **26**, 38-51. doi:10.1016/j.cub.2015.11.034

Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* **134**, 25-36. doi:10.1016/j.cell.2008.06.030

Catarino, R. R. and Stark, A. (2018). Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* **32**, 202-223. doi:10.1101/gad.310367.117

Chen, L., Fish, A. E. and Capra, J. A. (2018). Prediction of gene regulatory enhancers across species reveals evolutionarily conserved sequence properties. *PLoS Comput. Biol.* **14**, e1006484. doi:10.1371/journal.pcbi.1006484

Cho, K. W. Y. (2012). Enhancers. *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 469-478. doi:10.1002/wdev.53

Clark-Hachtel, C. M. and Tomoyasu, Y. (2017). Two sets of wing homologs in the crustacean, Parhyale hawaiensis. *bioRxiv*, 1-9. doi:10.1101/236281

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A. et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819-823. doi:10.1126/science.1231143

Connahs, H., Tlili, S., van Creij, J., Loo, T. Y. J., Banerjee, T. D., Saunders, T. E. and Monteiro, A. (2019). Activation of butterfly eyespots by Distal-less is consistent with a reaction-diffusion process. *Development* **146**, dev169367. doi:10.1242/dev.169367

Crocker, J. and Stern, D. L. (2013). TALE-mediated modulation of transcriptional enhancers in vivo. *Nat. Methods* **10**, 762-767. doi:10.1038/nmeth.2543

Crocker, J., Abe, N., Rinaldi, L., McGregor, A. P., Frankel, N., Wang, S., Alsawadi, A., Valenti, P., Plaza, S., Payre, F. et al. (2015). Low affinity binding site clusters confer hox specificity and regulatory robustness. *Cell* **160**, 191-203. doi:10.1016/j.cell.2014.11.041

Crocker, J., Noon, E. P.-B. and Stern, D. L. (2016). The soft touch: low-affinity transcription factor binding sites in development and evolution. *Curr. Top. Dev. Biol.* **117**, 455-469. doi:10.1016/bs.ctdb.2015.11.018

de Wit, E. and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11-24. doi:10.1101/gad.179804.111

Duester, G. (2019). Knocking out enhancers to enhance epigenetic research. *Trends Genet.* **35**, 89. doi:10.1016/j.tig.2018.10.001

Eckert, C., Aranda, M., Wolff, C. and Tautz, D. (2004). Separable stripe enhancer elements for the pair-rule gene hairy in the beetle *Tribolium*. *EMBO Rep.* **5**, 638-642. doi:10.1038/sj.embor.7400148

Ewen-Campen, B., Yang-Zhou, D., Fernandes, V. R., González, D. P., Liu, L.-P., Tao, R., Ren, X., Sun, J., Hu, Y., Zirin, J. et al. (2017). Optimized strategy for in vivo Cas9-activation in *Drosophila*. *Proc. Natl. Acad. Sci. USA* **114**, 9409-9414. doi:10.1073/pnas.1707635114

Frankel, N., Erezyilmaz, D. F., McGregor, A. P., Wang, S., Payre, F. and Stern, D. L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* **474**, 598-603. doi:10.1038/nature10200

Fraser, M. J. (2012). Insect transgenesis: current applications and future prospects. *Annu. Rev. Entomol.* **57**, 267-289. doi:10.1146/annurev.ento.54.110807.090545

Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. and Dubchak, I. (2004). VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, W273-W279. doi:10.1093/nar/gkh458

Gantz, V. M. and Akbari, O. S. (2018). Gene editing technologies and applications for insects. *Curr. Opin. Insect Sci.* **28**, 66-72. doi:10.1016/j.cois.2018.05.006

Ghavi-Helm, Y. and Furlong, E. E. M. (2012). Analyzing transcription factor occupancy during embryo development using ChIP-seq. *Methods Mol. Biol.* **786**, 229-245. doi:10.1007/978-1-61779-292-2_14

Ghavi-Helm, Y., Zhao, B. and Furlong, E. E. M. (2016). Chromatin immunoprecipitation for analyzing transcription factor binding and histone modifications in *Drosophila*. *Methods Mol. Biol.* **1478**, 263-277. doi:10.1007/978-1-4939-6371-3_16

Gilbert, L. A., Larson, M. H., Morsut, L., Liu, Z., Brar, G. A., Torres, S. E., Stern-Ginossar, N., Brandman, O., Whitehead, E. H., Doudna, J. A. et al. (2013).

CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* **154**, 442-451. doi:10.1016/j.cell.2013.06.044

Gilles, A. F. and Averof, M. (2014). Functional genetics for all: engineered nucleases, CRISPR and the gene editing revolution. *Evodevo* **5**, 43. doi:10.1186/2041-9139-5-43

Gilles, A. F., Schinko, J. B. and Averof, M. (2015). Efficient CRISPR-mediated gene targeting and transgene replacement in the beetle *Tribolium castaneum*. *Development* **142**, 2832-2839. doi:10.1242/dev.125054

Gilles, A. F., Schinko, J. B., Schacht, M. I., Enjolras, C. and Averof, M. (2019). Clonal analysis by tunable CRISPR-mediated excision. *Development* **146**, dev170969. doi:10.1242/dev.170969

Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. and Lieb, J. D. (2007). FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877-885. doi:10.1101/gr.5533506

Gompel, N., Prud'homme, B., Wittkopp, P. J., Kassner, V. A. and Carroll, S. B. (2005). Chance caught on the wing: cis-regulatory evolution and the origin of pigment patterns in *Drosophila*. *Nature* **433**, 481-487. doi:10.1038/nature03235

Goto, T., Macdonald, P. and Maniatis, T. (1989). Early and late periodic patterns of even skipped expression are controlled by distinct regulatory elements that respond to different spatial cues. *Cell* **57**, 413-422. doi:10.1016/0092-8674(89)90916-1

Hales, K. G., Korey, C. A., Larracuente, A. M. and Roberts, D. M. (2015). Genetics on the fly: a primer on the Drosophila model system. *Genetics* **201**, 815-842. doi:10.1534/genetics.115.183392

Halfon, M. S. (2019). Studying transcriptional enhancers: the founder fallacy, validation creep, and other biases. *Trends Genet.* **35**, 93-103. doi:10.1016/j.tig.2018.11.004

Halfon, M. S. and Michelson, A. M. (2002). Exploring genetic regulatory networks in metazoan development: methods and models. *Physiol. Genomics* **10**, 131-143. doi:10.1152/physiolgenomics.00072.2002

Halfon, M. S., Grad, Y., Church, G. M. and Michelson, A. M. (2002). Computation-based discovery of related transcriptional regulatory modules and motifs using an experimentally validated combinatorial model. *Genome Res.* **12**, 1019-1028.

Harding, K., Hoey, T., Warrior, R. and Levine, M. (1989). Autoregulatory and gap gene response elements of the even-skipped promoter of *Drosophila*. *EMBO J.* **8**, 1205-1212. doi:10.1002/j.1460-2075.1989.tb03493.x

He, Y., Gorkin, D. U., Dickel, D. E., Nery, J. R., Castanon, R. G., Lee, A. Y., Shen, Y., Visel, A., Pennacchio, L. A., Ren, B. et al. (2017). Improved regulatory element prediction based on tissue-specific local epigenomic signatures. *Proc. Natl. Acad. Sci. USA* **114**, E1633-E1640. doi:10.1073/pnas.1618353114

Hiromi, Y. and Gehring, W. J. (1987). Regulation and function of the *Drosophila* segmentation gene *fushi tarazu*. *Cell* **50**, 963-974. doi:10.1016/0092-8674(87)90523-X

Hiromi, Y., Kuroiwa, A. and Gehring, W. J. (1985). Control elements of the *Drosophila* segmentation gene *fushi tarazu*. *Cell* **43**, 603-613. doi:10.1016/0092-8674(85)90232-6

Hong, J.-W., Hendrix, D. A. and Levine, M. S. (2008). Shadow enhancers as a source of evolutionary novelty. *Science* **321**, 1314-1314. doi:10.1126/science.1160631

Horn, C. and Wimmer, E. A. (2000). A versatile vector set for animal transgenesis. *Dev. Genes Evol.* **210**, 630-637. doi:10.1007/s004270000110

Imamura, M., Nakai, J., Inoue, S., Quan, G. X., Kanda, T. and Tamura, T. (2003). Targeted gene expression using the GAL4/UAS system in the silkworm *Bombyx mori*. *Genetics* **165**, 1329-1340.

Jenett, A., Rubin, G. M., Ngo, T.-T. B., Shepherd, D., Murphy, C., Dionne, H., Pfeiffer, B. D., Cavallaro, A., Hall, D., Jeter, J. et al. (2012). A GAL4-driver line resource for *Drosophila* neurobiology. *Cell Rep.* **2**, 991-1001. doi:10.1016/j.celrep.2012.09.011

Jia, Y., Xu, R.-G., Ren, X., Ewen-Campen, B., Rajakumar, R., Zirin, J., Yang-Zhou, D., Zhu, R., Wang, F., Mao, D. et al. (2018). Next-generation CRISPR/Cas9 transcriptional activation in *Drosophila* using flySAM. *Proc. Natl. Acad. Sci. USA* **115**, 4719-4724. doi:10.1073/pnas.1800677115

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816-821. doi:10.1126/science.1225829

Jory, A., Estella, C., Giorgianni, M. W., Slattery, M., Laverty, T. R., Rubin, G. M. and Mann, R. S. (2012). A survey of 6,300 genomic fragments for cis-regulatory activity in the imaginal discs of *Drosophila melanogaster*. *Cell Rep.* **2**, 1014-1024. doi:10.1016/j.celrep.2012.09.010

Juven-Gershon, T., Cheng, S. and Kadonaga, J. T. (2006). Rational design of a super core promoter that enhances gene expression. *Nat. Methods* **3**, 917-922. doi:10.1038/nmeth937

Kadonaga, J. T. (2012). Perspectives on the RNA polymerase II core promoter. *Wiley Interdiscip. Rev. Dev. Biol.* **1**, 40-51. doi:10.1002/wdev.21

Kantorovitz, M. R., Kazemian, M., Kinston, S., Miranda-Saavedra, D., Zhu, Q., Robinson, G. E., Göttgens, B., Halfon, M. S. and Sinha, S. (2009). Motif-blind, genome-wide discovery of cis-regulatory modules in *Drosophila* and mouse. *Dev. Cell* **17**, 568-579. doi:10.1016/j.devcel.2009.09.002

Kazemian, M. and Halfon, M. S. (2019). CRM discovery beyond model insects. *Methods Mol. Biol.* **1858**, 117-139. doi:10.1007/978-1-4939-8775-7_10

Kazemian, M., Zhu, Q., Halfon, M. S. and Sinha, S. (2011). Improved accuracy of supervised CRM discovery with interpolated Markov models and cross-species comparison. *Nucleic Acids Res.* **39**, 9463-9472. doi:10.1093/nar/gkr621

Kazemian, M., Suryamohan, K., Chen, J.-Y., Zhang, Y., Samee, M. A. H., Halfon, M. S. and Sinha, S. (2014). Evidence for deep regulatory similarities in early developmental programs across highly diverged insects. *Genome Biol. Evol.* **6**, 2301-2320. doi:10.1093/gbe/evu184

Kharchenko, P. V., Alekseyenko, A. A., Schwartz, Y. B., Minoda, A., Riddle, N. C., Ernst, J., Sabo, P. J., Larschan, E., Gorchakov, A. A., Gu, T. et al. (2011). Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**, 480-485. doi:10.1038/nature09725

Kleftogiannis, D., Kalnis, P. and Bajic, V. B. (2016). Progress and challenges in bioinformatics approaches for enhancer identification. *Brief. Bioinform.* **17**, 967-979. doi:10.1093/bib/bbv101

Klein, J. C., Chen, W., Gasperini, M. and Shendure, J. (2018). Identifying novel enhancer elements with CRISPR-based screens. *ACS Chem. Biol.* **13**, 326-332. doi:10.1021/acschembio.7b00778

Klemm, S. L., Shipony, Z. and Greenleaf, W. J. (2019). Chromatin accessibility and the regulatory epigenome. *Nat. Rev. Genet.* **20**, 207-220. doi:10.1038/s41576-018-0089-8

Kokoza, V. A. and Raikhel, A. S. (2011). Targeted gene expression in the transgenic *Aedes aegypti* using the binary Gal4-UAS system. *Insect Biochem. Mol. Biol.* **41**, 637-644. doi:10.1016/j.ibmb.2011.04.004

Kukalova-Peck, J. (1991). Fossil history and the evolution of hexapod structures. In *The Insects of Australia: A Textbook for Students and Research Workers* (ed. I. D. Naumann), pp. 141-179. Melbourne University Press.

Kvon, E. Z., Kazmar, T., Stampfel, G., Yáñez-Cuna, J. O., Pagani, M., Schernhuber, K., Dickson, B. J. and Stark, A. (2014). Genome-scale functional characterization of Drosophila developmental enhancers in vivo. *Nature* **512**, 91-95. doi:10.1038/nature13395

Lai, Y.-T., Deem, K. D., Borràs-Castells, F., Sambrani, N., Rudolf, H., Suryamohan, K., El-Sherif, E., Halfon, M. S., McKay, D. J. and Tomoyasu, Y. (2018). Enhancer identification and activity evaluation in the red flour beetle, *Tribolium castaneum*. *Development* **145**, dev160663. doi:10.1242/dev.160663

Le, N. Q. K., Yapp, E. K. Y., Ho, Q.-T., Nagasundaram, N., Ou, Y.-Y. and Yeh, H.-Y. (2019). iEnhancer-5Step: Identifying enhancers using hidden information of DNA sequences via Chou's 5-step rule and word embedding. *Anal. Biochem.* **571**, 53-61. doi:10.1016/j.ab.2019.02.017

Levy, S. E. and Myers, R. M. (2016). Advancements in next-generation sequencing. *Annu. Rev. Genomics Hum. Genet.* **17**, 95-115. doi:10.1146/annurev-genom-083115-022413

Lewis, J. J. and Reed, R. D. (2019). Genome-wide regulatory adaptation shapes population-level genomic landscapes in *Heliconius*. *Mol. Biol. Evol.* **36**, 159-173. doi:10.1093/molbev/msy094

Lewis, J. J., van der Burg, K. R. L., Mazo-Vargas, A. and Reed, R. D. (2016). ChIP-Seq-annotated *Heliconius erato* genome highlights patterns of cis-regulatory evolution in Lepidoptera. *Cell Rep.* **16**, 2855-2863. doi:10.1016/j.celrep.2016.08.042

Li, L., Zhu, Q., He, X., Sinha, S. and Halfon, M. S. (2007). Large-scale analysis of transcriptional cis-regulatory modules reveals both common features and distinct subclasses. *Genome Biol.* **8**, R101. doi:10.1186/gb-2007-8-6-r101

Li, Y., Shi, W. and Wasserman, W. W. (2018). Genome-wide prediction of cis-regulatory regions using supervised deep learning methods. *BMC Bioinformatics* **19**, 202. doi:10.1186/s12859-018-2187-1

Li, F., Zhao, X., Li, M., He, K., Huang, C., Zhou, Y., Li, Z. and Walters, J. R. (2019). Insect genomes: progress and challenges. *Insect Mol. Biol.* **28**, 739-758. doi:10.1111/imb.12599

Lieberman-Aiden, E., van Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O. et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-293. doi:10.1126/science.1181369

Lim, L. W. K., Chung, H. H., Chong, Y. L. and Lee, N. K. (2018). A survey of recently emerged genome-wide computational enhancer predictor tools. *Comput. Biol. Chem.* **74**, 132-141. doi:10.1016/j.compbiolchem.2018.03.019

Lin, S., Ewen-Campen, B., Ni, X., Housden, B. E. and Perrimon, N. (2015). *In vivo* transcriptional activation using CRISPR/Cas9 in *Drosophila*. *Genetics* **201**, 433-442. doi:10.1534/genetics.115.181065

Liu, F., Li, H., Ren, C., Bo, X. and Shu, W. (2016). PEDLA: predicting enhancers with a deep learning-based algorithmic framework. *Sci. Rep.* **6**, 28517. doi:10.1038/srep28517

Liu, B., Li, K., Huang, D.-S. and Chou, K.-C. (2018). iEnhancer-EL: identifying enhancers and their strength with ensemble learning approach. *Bioinformatics* **34**, 3835-3842. doi:10.1093/bioinformatics/bty458

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D. and Ma, W. (2019). Methodologies for improving HDR efficiency. *Front. Genet.* **9**, 691. doi:10.3389/fgene.2018.00691

Long, H. K., Prescott, S. L. and Wysocka, J. (2016). Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* **167**, 1170-1187. doi:10.1016/j.cell.2016.09.018

Lopes, R., Korkmaz, G. and Agami, R. (2016). Applying CRISPR-Cas9 tools to identify and characterize transcriptional enhancers. *Nat. Rev. Mol. Cell Biol.* **17**, 597-604. doi:10.1038/nrm.2016.79

Lorenzen, M. D., Brown, S. J., Denell, R. E. and Beeman, R. W. (2002). Transgene expression from the *Tribolium castaneum* Polyubiquitin promoter. *Insect Mol. Biol.* **11**, 399-407. doi:10.1046/j.1365-2583.2002.00349.x

Lorenzen, M. D., Berghammer, A. J., Brown, S. J., Denell, R. E., Klingler, M. and Beeman, R. W. (2003). piggyBac-mediated germline transformation in the beetle *Tribolium castaneum*. *Insect Mol. Biol.* **12**, 433-440. doi:10.1046/j.1365-2583. 2003.00427.x

Lynd, A. and Lycett, G. J. (2012). Development of the bi-partite Gal4-UAS system in the African malaria mosquito, *Anopheles gambiae*. *PLoS ONE* **7**, e31552. doi:10.1371/journal.pone.0031552

Ma, J. and Ptashne, M. (1987). Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* **48**, 847-853. doi:10.1016/0092-8674(87)90081-X

Markstein, M. and Levine, M. (2002). Decoding cis-regulatory DNAs in the *Drosophila* genome. *Curr. Opin. Genet. Dev.* **12**, 601-606. doi:10.1016/S0959-437X(02)00345-3

Markstein, M., Markstein, P., Markstein, V. and Levine, M. S. (2002). Genome-wide analysis of clustered Dorsal binding sites identifies putative target genes in the *Drosophila* embryo. *Proc. Natl Acad. Sci. USA* **99**, 763-768. doi:10.1073/pnas. 012591199

Martin, A., Serano, J. M., Jarvis, E., Bruce, H. S., Wang, J., Ray, S., Barker, C. A., O'Connell, L. C. and Patel, N. H. (2016). CRISPR/Cas9 mutagenesis reveals versatile roles of Hox genes in crustacean limb specification and evolution. *Curr. Biol.* **26**, 14-26. doi:10.1016/j.cub.2015.11.021

Mayor, C., Brudno, M., Schwartz, J. R., Poliakov, A., Rubin, E. M., Frazer, K. A., Pachter, L. S. and Dubchak, I. (2000). VISTA: visualizing global DNA sequence alignments of arbitrary length. *Bioinformatics* **16**, 1046-1047. doi:10.1093/bioinformatics/16.11.1046

Mazo-Vargas, A., Concha, C., Livraghi, L., Massardo, D., Wallbank, R. W. R., Zhang, L., Papador, J. D., Martinez-Najera, D., Jiggins, C. D., Kronforst, M. R. et al. (2017). Macroevolutionary shifts of *WntA* function potentiate butterfly wing-pattern diversity. *Proc. Natl. Acad. Sci.* **114**, 10701-10706. doi:10.1073/pnas. 1708149114

McKay, D. J. (2019). Using Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) to identify functional regulatory DNA in insect genomes. *Methods Mol. Biol.* **1858**, 89-97. doi:10.1007/978-1-4939-8775-7_8

McKay, D. J. and Lieb, J. D. (2013). A common set of DNA regulatory elements shapes *Drosophila* appendages. *Dev. Cell* **27**, 306-318. doi:10.1016/j.devcel. 2013.10.009

Meers, M. P., Bryson, T. D., Henikoff, J. G. and Henikoff, S. (2019). Improved CUT&RUN chromatin profiling tools. *eLife* **8**, e46314. doi:10.7554/eLife.46314

Meyer, C. A. and Liu, X. S. (2014). Identifying and mitigating bias in next-generation sequencing methods for chromatin biology. *Nat Rev. Genet.* **15**, 709-721. doi:10. 1038/nrg3788

Min, X., Zeng, W., Chen, S., Chen, N., Chen, T. and Jiang, R. (2017). Predicting enhancers with deep convolutional neural networks. *BMC Bioinformatics* **18**, 478. doi:10.1186/s12859-017-1878-3

Mishra, M., Oke, A., Lebel, C., McDonald, E. C., Plummer, Z., Cook, T. A. and Zelhof, A. C. (2010). Pph13 and Orthodenticle define a dual regulatory pathway for photoreceptor cell morphogenesis and function. *Development* **137**, 2895-2904. doi:10.1242/dev.051722

Muerdter, F., Boryń, Ł. M. and Arnold, C. D. (2015). STARR-seq - principles and applications. *Genomics* **106**, 145-150. doi:10.1016/j.ygeno.2015.06.001

Nakade, S., Tsubota, T., Sakane, Y., Kume, S., Sakamoto, N., Obara, M., Daimon, T., Sezutsu, H., Yamamoto, T., Sakuma, T. et al. (2014). Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nat. Commun.* **5**, 5560. doi:10.1038/ncomms6560

Nakanishi, T., Kato, Y., Matsuura, T. and Watanabe, H. (2014). CRISPR/Cas-mediated targeted mutagenesis in *Daphnia magna*. *PLoS ONE* **9**, e98363. doi:10. 1371/journal.pone.0098363

Nègre, N., Brown, C. D., Ma, L., Bristow, C. A., Miller, S. W., Wagner, U., Kheradpour, P., Eaton, M. L., Loriaux, P., Sealfon, R. et al. (2011). A cis-regulatory map of the *Drosophila* genome. *Nature* **471**, 527-531. doi:10.1038/nature09990

O'Brochta, D. A., Pilitt, K. L., Harrell, R. A., Aluvihare, C. and Alford, R. T. (2012). Gal4-based enhancer-trapping in the malaria mosquito *Anopheles stephensi*. *G3 (Bethesda).* **2**, 1305-1315. doi:10.1534/g3.112.003582

Ohde, T., Takehana, Y., Shiotsuki, T. and Niimi, T. (2018). CRISPR/Cas9-based heritable targeted mutagenesis in *Thermobia domestica*: a genetic tool in an apterygote development model of wing evolution. *Arthropod Struct. Dev.* **47**, 362-369. doi:10.1016/j.asd.2018.06.003

Papatsenko, D., Kislyuk, A., Levine, M. and Dubchak, I. (2006). Conservation patterns in different functional sequence categories of divergent *Drosophila* species. *Genomics* **88**, 431-442. doi:10.1016/j.ygeno.2006.03.012

Park, P. J. (2009). ChIP–seq: advantages and challenges of a maturing technology. *Nat. Rev. Genet.* **10**, 669-680. doi:10.1038/nrg2641

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. and Bejerano, G. (2013). Enhancers: five essential questions. *Nat. Rev. Genet.* **14**, 288-295. doi:10. 1038/nrg3458

Pérez-Zamorano, B., Rosas-Madrigal, S., Lozano, O. A. M., Castillo Méndez, M. and Valverde-Garduño, V. (2017). Identification of cis-regulatory sequences reveals potential participation of lola and Deaf1 transcription factors in *Anopheles gambiae* innate immune response. *PLoS ONE* **12**, e0186435. doi:10.1371/ journal.pone.0186435

Perry, M. W., Boettiger, A. N., Bothma, J. P. and Levine, M. (2010). Shadow enhancers foster robustness of *Drosophila* gastrulation. *Curr. Biol.* **20**, 1562-1567. doi:10.1016/j.cub.2010.07.043

Pfeiffer, B. D., Jenett, A., Hammonds, A. S., Ngo, T.-T. B., Misra, S., Murphy, C., Scully, A., Carlson, J. W., Wan, K. H., Laverty, T. R. et al. (2008). Tools for neuroanatomy and neurogenetics in *Drosophila*. *Proc. Natl Acad. Sci. USA* **105**, 9715-9720. doi:10.1073/pnas.0803697105

Pickar-Oliver, A. and Gersbach, C. A. (2019). The next generation of CRISPR–Cas technologies and applications. *Nat. Rev. Mol. Cell Biol.* **20**, 490-507. doi:10. 1038/s41580-019-0131-5

Prakash, A. and Monteiro, A. (2018). apterous A specifies dorsal wing patterns and sexual traits in butterflies. *Proc. R. Soc. B* **285**, 20172685. doi:10.1098/rspb.2017. 2685

Prasad, N., Tarikere, S., Khanale, D., Habib, F. and Shashidhara, L. S. (2016). A comparative genomic analysis of targets of Hox protein Ultrabithorax amongst distant insect species. *Sci. Rep.* **6**, 27885. doi:10.1038/srep27885

Reardon, S. (2019). CRISPR gene-editing creates wave of exotic model organisms. *Nature* **568**, 441-442. doi:10.1038/d41586-019-01300-9

Rebeiz, M. and Williams, T. M. (2017). Using *Drosophila* pigmentation traits to study the mechanisms of cis-regulatory evolution. *Curr. Opin. insect Sci.* **19**, 1-7. doi:10.1016/j.cois.2016.10.002

Richards, S., Liu, Y., Bettencourt, B. R., Hradecky, P., Letovsky, S., Nielsen, R., Thornton, K., Hubisz, M. J., Chen, R., Meisel, R. P. et al. (2005). Comparative genome sequencing of *Drosophila pseudoobscura*: chromosomal, gene, and cis-element evolution. *Genome Res.* **15**, 1-18. doi:10.1101/gr.3059305

Rickels, R. and Shilatifard, A. (2018). Enhancer logic and mechanics in development and disease. *Trends Cell Biol.* **28**, 608-630. doi:10.1016/j.tcb. 2018.04.003

Rivera, J., Keränen, S. V. E., Gallo, S. M. and Halfon, M. S. (2019). REDfly: the transcriptional regulatory element database for *Drosophila*. *Nucleic Acids Res.* **47**, D828-D834. doi:10.1093/nar/gky957

Ron, G., Globerson, Y., Moran, D. and Kaplan, T. (2017). Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nat. Commun.* **8**, 2237. doi:10.1038/s41467-017-02386-3

Roy, S., Ernst, J., Kharchenko, P. V., Kheradpour, P., Negre, N., Eaton, M. L., Landolin, J. M., Bristow, C. A., Ma, L., Lin, M. F. et al. (2010). Identification of functional elements and regulatory circuits by Drosophila modENCODE. *Science* **330**, 1787-1797. doi:10.1126/science.1198374

Rylee, J. C., Siniard, D. J., Doucette, K., Zentner, G. E. and Zelhof, A. C. (2018). Expanding the genetic toolkit of *Tribolium castaneum*. *PLoS ONE* **13**, e0195977. doi:10.1371/journal.pone.0195977

Sarrazin, A. F., Peel, A. D. and Averof, M. (2012). A segmentation clock with two-segment periodicity in insects. *Science* **336**, 338-341. doi:10.1126/science. 1218256

Savic, N., Ringnalda, F. C., Lindsay, H., Berk, C., Bargsten, K., Li, Y., Neri, D., Robinson, M. D., Ciaudo, C., Hall, J. et al. (2018). Covalent linkage of the DNA repair template to the CRISPR-Cas9 nuclease enhances homology-directed repair. *eLife* **7**, e33761. doi:10.7554/eLife.33761

Schinko, J. B., Weber, M., Viktorinova, I., Kiupakis, A., Averof, M., Klingler, M., Wimmer, E. A. and Bucher, G. (2010). Functionality of the GAL4/UAS system in *Tribolium* requires the use of endogenous core promoters. *BMC Dev. Biol.* **10**, 53. doi:10.1186/1471-213X-10-53

Schinko, J. B., Hillebrand, K. and Bucher, G. (2012). Heat shock-mediated misexpression of genes in the beetle *Tribolium castaneum*. *Dev. Genes Evol.* **222**, 287-298. doi:10.1007/s00427-012-0412-x

Schmid-Burgk, J. L., Höning, K., Ebert, T. S. and Hornung, V. (2016). CRISPaint allows modular base-specific gene tagging using a ligase-4-dependent mechanism. *Nat. Commun.* **7**, 12338. doi:10.1038/ncomms12338

Sethi, A., Gu, M., Gumusgoz, E., Chan, L., Yan, K.-K., Rozowsky, J., Barozzi, I., Afzal, V., Akiyama, J., Plajzer-Frick, I. et al. (2018). A cross-organism framework for supervised enhancer prediction with epigenetic pattern recognition and targeted validation. *bioRxiv*, 385237. doi:10.1101/385237

Sheng, G., Thouvenot, E., Schmucker, D., Wilson, D. S. and Desplan, C. (1997). Direct regulation of rhodopsin 1 by Pax-6/eyeless in *Drosophila*: evidence for a conserved function in photoreceptors. *Genes Dev.* **11**, 1122-1131. doi:10.1101/ gad.11.9.1122

Siebert, K. S., Lorenzen, M. D., Brown, S. J., Park, Y. and Beeman, R. W. (2008). Tubulin superfamily genes in *Tribolium castaneum* and the use of a Tubulin promoter to drive transgene expression. *Insect Biochem. Mol. Biol.* **38**, 749-755. doi:10.1016/j.ibmb.2008.04.007

Skene, P. J. and Henikoff, S. (2017). An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *eLife* **6**, e21856. doi:10.7554/eLife.21856

Skene, P. J., Henikoff, J. G. and Henikoff, S. (2018). Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006-1019. doi:10.1038/nprot.2018.015

Small, S., Arnosti, D. N. and Levine, M. (1993). Spacing ensures autonomous expression of different stripe enhancers in the even-skipped promoter. *Development* **119**, 762-772.

Sosinsky, A., Honig, B., Mann, R. S. and Califano, A. (2007). Discovering transcriptional regulatory regions in Drosophila by a nonalignment method for phylogenetic footprinting. *Proc. Natl Acad. Sci. USA* **104**, 6305-6310. doi:10.1073/pnas.0701614104

Stark, A., Lin, M. F., Kheradpour, P., Pedersen, J. S., Parts, L., Carlson, J. W., Crosby, M. A., Rasmussen, M. D., Roy, S., Deoras, A. N. et al. (2007). Discovery of functional elements in 12 *Drosophila* genomes using evolutionary signatures. *Nature* **450**, 219-232. doi:10.1038/nature06340

Stern, D. L. and Frankel, N. (2013). The structure and evolution of cis-regulatory regions: the shavenbaby story. *Philos. Trans. R. Soc. B* **368**, 20130028. doi:10.1098/rstb.2013.0028

Strobl, F., Anderl, A. and Stelzer, E. H. K. (2018). A universal vector concept for a direct genotyping of transgenic organisms and a systematic creation of homozygous lines. *eLife* **7**, e31677. doi:10.7554/eLife.31677

Suryamohan, K. and Halfon, M. S. (2015). Identifying transcriptional *cis*-regulatory modules in animal genomes. *Wiley Interdiscip. Rev. Dev. Biol.* **4**, 59-84. doi:10.1002/wdev.168

Suryamohan, K., Hanson, C., Andrews, E., Sinha, S., Scheel, M. D. and Halfon, M. S. (2016). Redeployment of a conserved gene regulatory network during *Aedes aegypti* development. *Dev. Biol.* **416**, 402-413. doi:10.1016/j.ydbio.2016.06.031

Swanson, C. I., Evans, N. C. and Barolo, S. (2010). Structural rules and complex regulatory circuitry constrain expression of a Notch- and EGFR-regulated eye enhancer. *Dev. Cell* **18**, 359-370. doi:10.1016/j.devcel.2009.12.026

Thakore, P. I., D'Ippolito, A. M., Song, L., Safi, A., Shivakumar, N. K., Kabadi, A. M., Reddy, T. E., Crawford, G. E. and Gersbach, C. A. (2015). Highly specific epigenome editing by CRISPR-Cas9 repressors for silencing of distal regulatory elements. *Nat. Methods* **12**, 1143-1149. doi:10.1038/nmeth.3630

Thomas, J.-L., Da Rocha, M., Besse, A., Mauchamp, B. and Chavancy, G. (2002). 3xP3-EGFP marker facilitates screening for transgenic silkworm *Bombyx mori* L. from the embryonic stage onwards. *Insect Biochem. Mol. Biol.* **32**, 247-253. doi:10.1016/S0965-1748(01)00150-3

Thomas, G. W. C., Dohmen, E., Hughes, D. S. T., Murali, S. C., Poelchau, M., Glastad, K., Anstead, C. A., Ayoub, N. A., Batterham, P., Bellair, M. et al. (2018). The genomic basis of arthropod diversity. *bioRxiv*, 382945.

Tokusumi, T., Tokusumi, Y., Brahier, M. S., Lam, V., Stoller-Conrad, J. R., Kroeger, P. T. and Schulz, R. A. (2017). Screening and analysis of Janelia FlyLight Project enhancer-Gal4 strains identifies multiple gene enhancers active during hematopoiesis in normal and Wasp-challenged *Drosophila* larvae. *G3 (Bethesda)* **7**, 437-448. doi:10.1534/g3.116.034439

Trauner, J., Schinko, J., Lorenzen, M. D., Shippy, T. D., Wimmer, E. A., Beeman, R. W., Klingler, M., Bucher, G. and Brown, S. J. (2009). Large-scale insertional mutagenesis of a coleopteran stored grain pest, the red flour beetle *Tribolium castaneum*, identifies embryonic lethal mutations and enhancer traps. *BMC Biol.* **7**, 73. doi:10.1186/1741-7007-7-73

Trible, W., Olivos-Cisneros, L., McKenzie, S. K., Saragosti, J., Chang, N.-C., Matthews, B. J., Oxley, P. R. and Kronauer, D. J. C. (2017). orco mutagenesis causes loss of antennal lobe glomeruli and impaired social behavior in ants. *Cell* **170**, 727-735.e10. doi:10.1016/j.cell.2017.07.001

van der Burg, K. R. L., Lewis, J. J., Martin, A., Nijhout, H. F., Danko, C. G. and Reed, R. D. (2019). Contrasting roles of transcription factors spineless and EcR in the highly dynamic chromatin landscape of butterfly wing metamorphosis. *Cell Rep.* **27**, 1027-1038.e3. doi:10.1016/j.celrep.2019.03.092

Viktorinová, I. and Wimmer, E. A. (2007). Comparative analysis of binary expression systems for directed gene expression in transgenic insects. *Insect Biochem. Mol. Biol.* **37**, 246-254. doi:10.1016/j.ibmb.2006.11.010

Visel, A., Blow, M. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature* **457**, 854-858. doi:10.1038/nature07730

Vo Ngoc, L., Kassavetis, G. A. and Kadonaga, J. T. (2019). The RNA polymerase II core promoter in *Drosophila*. *Genetics* **212**, 13-24. doi:10.1534/genetics.119.302021

Watanabe, T., Noji, S. and Mito, T. (2017). Genome editing in the cricket, *Gryllus bimaculatus*. *Methods Mol. Biol.* **1630**, 219-233. doi:10.1007/978-1-4939-7128-2_18

Wei, W., Xin, H., Roy, B., Dai, J., Miao, Y. and Gao, G. (2014). Heritable genome editing with CRISPR/Cas9 in the silkworm, *Bombyx mori*. *PLoS ONE* **9**, e101210. doi:10.1371/journal.pone.0101210

Wolff, C., Schröder, R., Schulz, C., Tautz, D. and Klingler, M. (1998). Regulation of the *Tribolium* homologues of caudal and hunchback in *Drosophila*: evidence for maternal gradient systems in a short germ embryo. *Development* **125**, 3645-3654.

Xu, X.-R. S., Gantz, V. M., Siomava, N. and Bier, E. (2017). CRISPR/Cas9 and active genetics-based trans-species replacement of the endogenous *Drosophila* kni-L2 CRM reveals unexpected complexity. *eLife* **6**, e30281. doi:10.7554/eLife.30281

Yan, H., Opachaloemphan, C., Mancini, G., Yang, H., Gallitto, M., Mlejnek, J., Leibholz, A., Haight, K., Ghaninia, M., Huo, L. et al. (2017). An engineered orco mutation produces aberrant social behavior and defective neural development in ants. *Cell* **170**, 736-747.e9. doi:10.1016/j.cell.2017.06.051

Yang, B., Liu, F., Ren, C., Ouyang, Z., Xie, Z., Bo, X. and Shu, W. (2017). BiRen: predicting enhancers with a deep-learning-based model using the DNA sequence alone. *Bioinformatics* **33**, 1930-1936. doi:10.1093/bioinformatics/btx105

Zabidi, M. A., Arnold, C. D., Schernhuber, K., Pagani, M., Rath, M., Frank, O. and Stark, A. (2015). Enhancer–core-promoter specificity separates developmental and housekeeping gene regulation. *Nature* **518**, 556-559. doi:10.1038/nature13994

Zhang, L. and Reed, R. D. (2016). Genome editing in butterflies reveals that spalt promotes and Distal-less represses eyespot colour patterns. *Nat. Commun.* **7**, 11769. doi:10.1038/ncomms11769

Zhang, L., Martin, A., Perry, M. W., van der Burg, K. R. L., Matsuoka, Y., Monteiro, A. and Reed, R. D. (2017a). Genetic basis of melanin pigmentation in butterfly wings. *Genetics* **205**, 1537-1550. doi:10.1534/genetics.116.196451

Zhang, L., Mazo-Vargas, A. and Reed, R. D. (2017b). Single master regulatory gene coordinates the evolution and development of butterfly color and iridescence. *Proc. Natl Acad. Sci. USA* **114**, 10707-10712. doi:10.1073/pnas.1709058114

Zhang, Q., Cheng, T., Jin, S., Guo, Y., Wu, Y., Liu, D., Xu, X., Sun, Y., Li, Z., He, H. et al. (2017c). Genome-wide open chromatin regions and their effects on the regulation of silk protein genes in *Bombyx mori*. *Sci. Rep.* **7**, 12919. doi:10.1038/s41598-017-13186-6

Zhang, Q., Cheng, T., Sun, Y., Wang, Y., Feng, T., Li, X., Liu, L., Li, Z., Liu, C., Xia, Q. et al. (2019). Synergism of open chromatin regions involved in regulating genes in *Bombyx mori*. *Insect Biochem. Mol. Biol.* **110**, 10-18. doi:10.1016/j.ibmb.2019.04.014

Zinzen, R. P., Cande, J., Ronshaugen, M., Papatsenko, D. and Levine, M. (2006). Evolution of the ventral midline in insect embryos. *Dev. Cell* **11**, 895-902. doi:10.1016/j.devcel.2006.10.012