Aus dem Biomedizinischen Centrum
der Ludwig-Maximilians-Universität München
Medizinische Fakultät
Lehrstuhl für Molekularbiologie
Vorstand: Prof. Dr. rer. nat. Peter B. Becker

The *Drosophila* speciation factor HMR localizes to genomic insulator sites

Dissertation
zum Erwerb des Doktorgrades der Naturwissenschaften
an der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

vorgelegt von
Thomas Andreas Gerland
aus München
Jahr 2017

Gedruckt mit Genehmigung der Medizinischen Fakultät
der Ludwig-Maximilians-Universität München

Betreuer:	Prof. Dr. rer. nat. Axel Imhof
Zweitgutachter:	Prof. Dr. André Brändli
Dekan:	Prof. Dr. med. dent. Reinhard Hickel
Tag der mündlichen Prüfung:	14.11.2017

Eidesstattliche Versicherung

Gerland, Thomas Andreas

Ich erkläre hiermit an Eides statt,
dass ich die vorliegende Dissertation mit dem Thema

“The *Drosophila* speciation factor HMR localizes to genomic insulator sites”

selbständig verfasst, mich außer der angegebenen keiner weiteren Hilfsmittel bedient und alle Erkenntnisse, die aus dem Schrifttum ganz oder annähernd übernommen sind, als solche kenntlich gemacht und nach ihrer Herkunft unter Bezeichnung der Fundstelle einzeln nachgewiesen habe.

Ich erkläre des Weiteren, dass die hier vorgelegte Dissertation nicht in gleicher oder in ähnlicher Form bei einer anderen Stelle zur Erlangung eines akademischen Grades eingereicht wurde.

Ort, Datum	Unterschrift Doktorandin/Doktorand

Wesentliche Teile dieser Arbeit sind veröffentlicht in:

PLoS ONE, 2017 February 16, doi:10.1371/journal.pone.0171798
The *Drosophila* speciation factor HMR localizes to genomic insulator sites
Gerland T. A., Sun B., Smialowski P., Lukacs A., Thomae A. W., Imhof A.

Mitwirkungen:

Bioinformatische und statistische Datenanalyse durchgeführt in Zusammenarbeit mit
Bo Sun, Dr. Pawel Smialowski und Dr. Tobias Straub

Next Generation Sequencing durchgeführt in Zusammenarbeit mit
Dr. Stefan Krebs und Dr. Helmut Blum, Genomics unit of LAFUGA

CRISPR/Cas9-basiertes genome editing durchgeführt in Zusammenarbeit mit
Prof. Dr. Klaus Förstemann

Immunohistologische Studien durchgeführt in Zusammenarbeit mit
Dr. Andreas W. Thomae und Natalia Kochanova

Table of contents

SUMMARY

Diversity of life on Earth is a result of the emergence and the extinction of species. A crucial step in species formation is the reproductive isolation of populations that diverged from a common ancestor. The incapability of diverging populations to produce viable and fertile offspring prevents the exchange of genetic material, and therefore, gives rise to the formation of new species. Reproductive isolation can be caused by genes. As it turned out, many of these genes encode chromatin components that cause severe problems in a hybrid background. These findings link chromatin biology to species formation, but how evolutionary changes at the level of chromatin mediate reproductive isolation is still unclear.

A particularly well-characterized model to study speciation is the gene pair *Hybrid male rescue* (*Hmr*) and *Lethal hybrid rescue* (*Lhr*) in the fruit fly. The presence of *Hmr* and *Lhr* causes lethality of male hybrids from *Drosophila melanogaster* mothers and *D. simulans* fathers. HMR and LHR form a chromatin-residing complex that localizes to centromeric heterochromatin and plays an important role in chromosome segregation in *D. melanogaster*. In hybrids, widespread mislocalization of this complex is suggested to cause lethality by a misregulation of *de novo* target loci. However, a detailed molecular description of HMR binding sites and their putative role in HMR's function in pure species and in hybrid background remains to be elucidated.

This study identifies and describes the genome-wide binding properties of HMR in *D. melanogaster* Schneider S2 cells. For this, we performed chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) on HMR and endogenously epitope-tagged HMR. We demonstrate an extensive localization of HMR to genomic insulators and propose that HMR is targeted to these genomic sites by residing insulator protein complexes. Insulator proteins serve as regulatory elements of chromatin and transcription and successively gained new members during the evolution of *Drosophila*. Using protein knockdown strategies that affect the insulator protein complex structure, we demonstrate a loss of HMR binding to the affected genomic insulator sites once the insulator complex is disrupted. We found that HMR is associated with two classes of genomic insulators: *gypsy* and BEAF-32 insulators. The *gypsy* insulator sites are associated with the DNA-binding protein Su(Hw) and are located at the *gypsy* and *gypsy-twin* repeats and at euchromatic sites across the genome. Coincidently, HMR is also found at the *gypsy* and *gtwin* repeat

regions as well as at non-repetitive regions at the chromosome arms such as the 1A-2 locus. Another subset of HMR binding sites associates with BEAF-32, an insulator protein with direct DNA-binding activity. In contrary to *gypsy* insulators, HMR and BEAF-32 localize at boundaries that separate heterochromatin domains from active gene bodies. These genes are less transcribed upon the loss of HMR or BEAF-32 indicating that HMR and BEAF-32 promote transcription at these genomic sites. As the loss of HMR or BEAF-32 was reported to cause similar phenotypes, we speculate that HMR and BEAF-32 act in a common pathway. Even though the underlying mechanism remains to be elucidated, it is tempting to speculate that HMR acquired this function during *Drosophila* evolution.

Overall, we demonstrate a novel link between HMR and insulator proteins, a class of proteins that successively gained new factors during *Drosophila* evolution. Our findings provide new molecular insights for the speciation research field and further promote the chromatin research field by characterizing a biologically relevant example of chromatin factors.

ZUSAMMENFASSUNG

Die Vielfalt des Lebens auf unserem Planeten ist das Ergebnis eines fortlaufenden Entstehens und Verschwindens biologischer Arten. Populationen, die sich auseinander entwickelt haben, sind nicht mehr in der Lage überlebensfähige und fortpflanzungsfähige Nachkommen hervorzubringen. Diese sogenannte reproduktive Isolation spielt bei der Entstehung neuer Arten eine entscheidende Rolle und kann durch Gene verursacht werden. Viele dieser Gene codieren Chromatin-Komponenten, welche schwerwiegende Probleme in Hybriden verursachen. Auf diese Weise sind Chromatin-Biologie und Artbildung eng miteinander verknüpft. Wie aber Veränderungen des Chromatins während der Evolution letztendlich in einer Inkompatibilität der Arten enden können, ist unklar.

Ein besonders gut charakterisiertes Model bei der Untersuchung der Artbildung sind die beiden Gene *Hybrid male rescue* (*Hmr*) und *Lethal hybrid rescue* (*Lhr*) in der Fruchtfliege. Die Anwesenheit beider Gene führt dazu, dass männliche Hybride mit einer *Drosophila melanogaster* Mutter und einem *D. simulans* Vater früh in ihrer Entwicklung sterben. Die Genprodukte HMR und LHR bilden einen Proteinkomplex, der am Heterochromatin des Centromers zu finden ist und eine wichtige Rolle bei der Chromosomen-Segregation in *D. melanogaster* spielt. In Hybriden führt eine Bindung dieses Proteinkomplexes an falsche Stellen im Genom vermutlich zu einer Fehlregulierung dieser Bindestellen und letztlich zum Tod des Tieres. Derzeit ist jedoch nur sehr wenig über die Natur solcher Bindestellen bekannt. Wie die Bindestellen von HMR auf molekularer Ebene aussehen, und welche Auswirkungen diese auf die Funktion von HMR in der reinen Spezies und in Hybriden haben, muss noch erforscht werden.

Diese Studie identifiziert und beschreibt die genomweiten Bindeeigenschaften von HMR in *D. melanogaster* Schneider S2 Zellen. Dafür haben wir Chromatin-Immunpräzipitation und genomweite DNA Sequenzierung verwendet, speziell für HMR und endogen Epitop-getaggtes HMR. Wir zeigen eine umfassende Bindung von HMR an genomische Isolatoren und schlagen mit unseren Daten eine Rekrutierung von HMR an diese Bindestellen durch den anwesenden Komplex aus Isolator-Proteinen vor. Die Familie der Isolator-Proteine bildet ein wichtiges regulatorisches Element des Chromatins und der zellulären Transkription und hat während der Evolution in *Drosophila* ständig neue Mitglieder erhalten. Mit Hilfe von Protein-Knockdown-Experimenten, welche die Struktur von Isolator-Proteinenkomplexen

spezifisch beeinflussen, zeigen wir den Einfluss dieser Komplexe auf die HMR-Bindung. HMR bindet an zwei verschiedene Klassen genomischer Isolatoren: *gypsy*- und BEAF-32-Isolatoren. Die *gypsy*-Isolatoren sind mit dem DNA-bindenden Protein Su(Hw) assoziiert und befinden sind an den repetitiven Sequenzen *gypsy* und *gtwin* sowie an weiteren Stellen entlang des Genoms. HMR lokalisiert sowohl an *gypsy* und *gtwin* als auch an nicht-repetitive Sequenzen entlang der Chromosomenarme wie etwa die Region 1A-2. Zusätzlich bindet HMR an eine weitere Klasse von Isolatoren, nämlich solche, die mit dem DNA-bindenden Protein BEAF-32 assoziiert sind. Im Gegensatz zu den *gypsy*-Isolatoren lokalisieren HMR und BEAF-32 an die Grenze zwischen Heterochromatin und aktiv transkribierten Genen. Interessanterweise werden diese Gene weniger stark transkribiert sobald HMR oder BEAF-32 verloren gehen, was darauf hindeutet, dass HMR und BEAF-32 die Transkription dieser Gene unterstützen. Da der Verlust eines dieser Proteine außerdem in einem ähnlichen Phänotyp mündet, liegt die Vermutung nahe, dass HMR und BEAF-32 in ihrer Funktion zusammenarbeiten könnten. Obwohl der zugrundeliegende Mechanismus dieser Funktion noch weiter erforscht werden muss, ist es gut möglich, dass HMR eben diese Funktion erst während der Evolution in *Drosophila* erhalten hat.

Zusammenfassend zeigen wir auf diese Weise zum ersten Mal einen Zusammenhang zwischen HMR und Isolator-Proteinen, einer Familie von Proteinen, die während der Evolution in *Drosophila* neue Mitglieder erhalten hat. Unsere Ergebnisse stellen wichtige Informationen für das Forschungsfeld der Artenentwicklung zur Verfügung und bringen darüber hinaus durch die Charakterisierung eines biologisch hochrelevanten Beispiels von Chromatin-Proteinen, das Feld der Chromatin-Forschung weiter voran.

1 INTRODUCTION

1.1 Speciation and hybrid incompatibility can be caused by genes

Evolutionary biology seeks to explain the great diversity of life around us. In the first paragraph of his book *On the Origin of Species*, Darwin expresses his hope to "throw some light on […] that mystery of mysteries". This was in 1859 (Darwin, 1859). In fact, biodiversity on Earth is the result of repeated formation and extinction of species. A crucial step in species formation is the reproductive isolation of populations that evolved from a common ancestor. Reproductive isolation is caused by infertility or lethality of the offspring from two sibling species. Mayr defined species as "groups of interbreeding natural populations that are reproductively isolated from other such groups" (Mayr, 1942). Reproductive barriers abolish or restrict the gene flow, the exchange of individual alleles, between populations. The genetic and phenotypic integrity of such populations is maintained by reproductive isolation even in geographical proximity (Coyne and Orr, 2004; Price, 2008; Seehausen et al., 2014). Reproductive isolation can occur through external forces, such as geographical barriers, or can be caused by intrinsic postzygotic barriers. In the latter case, incompatibility is caused by genes, or more generic, by genomic regions, that evolved independently in separate populations. But how can such hybrid incompatibility evolve without negatively affecting the pure species? The theoretical work on that started with the concept of Dobzhansky (Dobzhansky, 1936) and Muller (Muller and Pontecorvo, 1942). The Dobzhansky-Muller model (**Figure 1**) proposes that hybrid incompatibilities are caused by epistatic interactions of genes or genomic regions (*a/A* and *b/B*) that evolved independently in each of the hybridizing species: once the ancestral population is split into two, *a* can evolve into *A* in population 1 and *b* can evolve into *B* in population 2. *A* and *B* are not compatible but as *A* and *B* evolved independently, *A* and *B* are individually not negatively selected (Coyne and Orr, 2004). In the last few years, several of these hybrid incompatibility genes were mapped and subjected for investigation. So far, the speciation research field still lacks a detailed understanding of their molecular properties and function in pure species as well as of their deleterious role in hybrids.

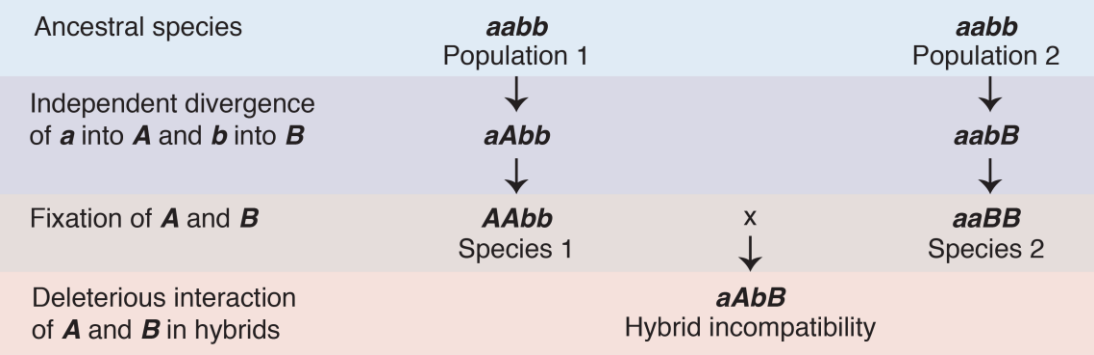

Figure 1. Genetic incompatibility can cause reproductive isolation. The Dobzhansky-Muller model explains how hybrid incompatibilities can be caused by the incompatibility of two genomic regions, *A* and *B*. The two genomic regions evolved independently (*aa* to *AA* and *bb* to *BB*) in two separate populations. The *A*-*B* interaction is not present in the pure species and therefore not subject to negative selection but causes severe problems in hybrids. Adapted from (Wu and Ting, 2004).

1.2 *Hmr* and *Lhr* are hybrid incompatibility genes that encode a chromatin residing complex

Hybrid incompatibility genes function normally within their own genetic background, but their divergence causes incompatible interactions in hybrids. The sibling fly species *D. melanogaster* and *D. simulans* (**Figure 2 A**) served as the main model system to identify hybrid incompatibility genes. Hybrid males from *D. melanogaster* mothers and *D. simulans* fathers dye at late larval stage and do not develop into adults, while hybrid females are sterile (Sturtevant, 1920). A mutation either in the *D. simulans* gene *Lethal hybrid rescue* (*Lhr*) (Watanabe and Kawanishi, 1979) or in the X-linked *D. melanogaster* gene *Hybrid male rescue* (*Hmr*) (Barbash et al., 2003) rescue the hybrid phenotypes (**Figure 2 B**). Hybrid male lethality results from a genetic interaction of the *Hmr* allele that diverged in *D. melanogaster* (Hmr_{mel}) and the *Lhr* allele that diverged in *D. simulans* (Lhr_{sim}) (Brideau et al., 2006). In fact, the gene pair *Hmr* and *Lhr* is the first described Dobzhansky-Muller gene pair (Brideau et al., 2006) and served as a popular model to study hybrid incompatibility on a molecular level (Satyaki et al., 2014; Thomae et al., 2013; Wei et al., 2014). *Hmr* and *Lhr* encode proteins that interact and form a protein complex that associates to chromatin (Satyaki et al., 2014; Thomae et al., 2013).

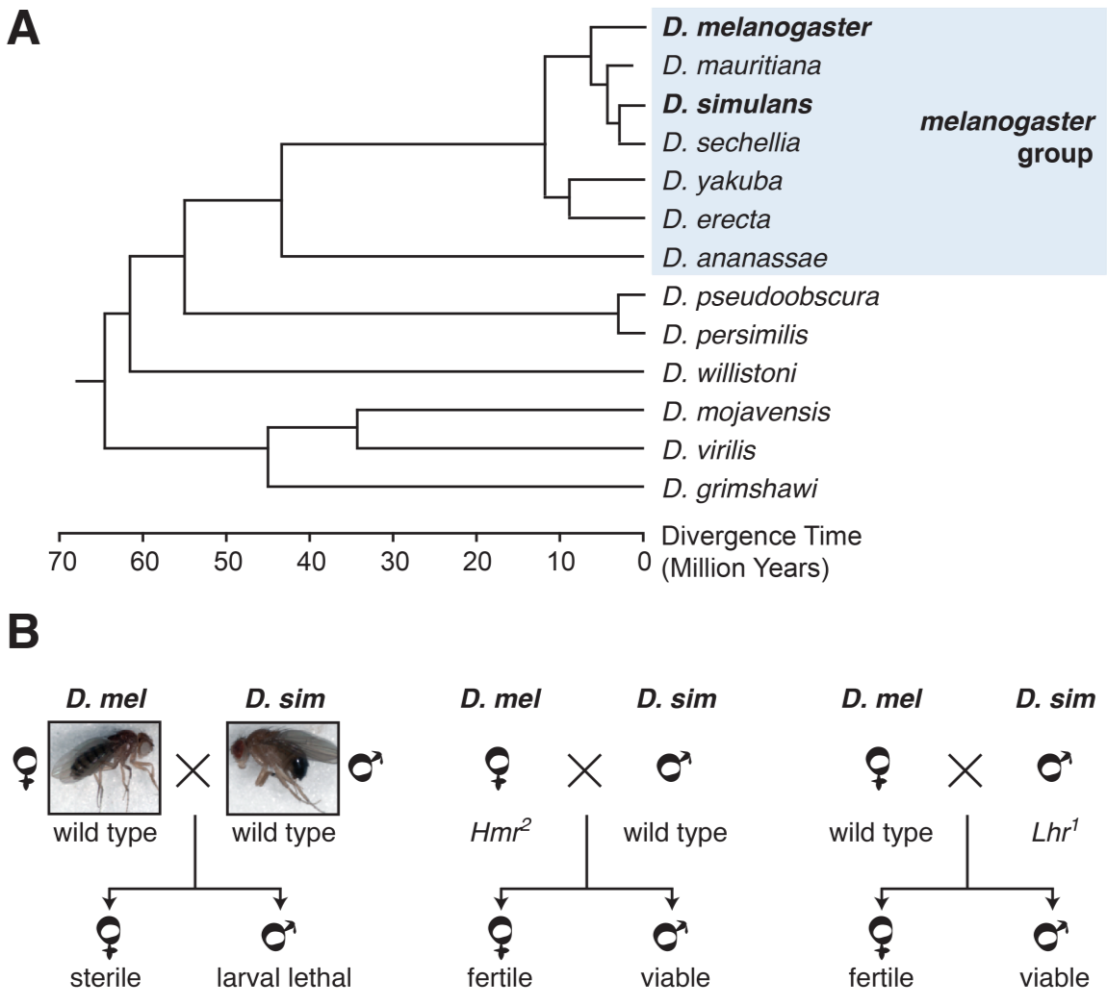

Figure 2. *Hmr* and *Lhr* are hybrid incompatibility genes and reproductively isolate *D. melanogaster* and *D. simulans*. (A) Phylogenetic tree of *Drosophila* species. *D. melanogaster* and *D. simulans* (shaded in grey) are very closely related and reproductively isolated. Adapted from (Granzotto et al., 2009) and (Seetharam and Stuart, 2013) **(B)** Speciation involves incompatibility of genes. *Hmr* and *Lhr* cause *D. melanogaster* and *D. simulans* hybrid incompatibility. *Hmr* has diverged in *D. melanogaster*, *Lhr* has diverged in *D. simulans*. The two genes cause reproductive isolation in hybrids. Hmr^2 and Lhr^1 alleles suppress the lethality of hybrid males and partially restore fertility of hybrid females. Pictures of flies kindly provided by Andrea Lukacs. Adapted from (Sawamura et al., 1993).

1.3 Heterochromatin, a species barrier and driver for hybrid incompatibility

D. melanogaster and its sibling species served as a popular model to study hybrid incompatibility. However, hybrid incompatibility genes were also identified in other organisms, ranging from yeast (Lee et al., 2008) to mammals (Mihola et al., 2009). Interestingly, many of these genes encode chromatin components, chromatin modifying enzymes or repetitive DNA elements that strongly interfere with chromatin

structure and organization (reviewed in (Sawamura, 2012)). Chromatin is the combination of proteins, noncoding RNA and chromosomal DNA of eukaryotes. On the one hand, chromatin allows high compaction of DNA and, on the other hand, chromatin permits regulated transcription of DNA throughout the cell cycle (Bohmdorfer and Wierzbicki, 2015; Holoch and Moazed, 2015; Meller et al., 2015; Pollard et al., 2007). In the first half of the 20th century, chromosomal DNA was classified into euchromatin and heterochromatin. This classification was based on the observation that one fraction of chromatin – euchromatin – changes the degree of condensation throughout the cell cycle, whereas the other fraction – heterochromatin – remains condensed (Heitz, 1930) (reviewed in (Eissenberg and Reuter, 2009)). Heterochromatin is mainly present at the centromere and telomere regions containing long stretches of repetitive DNA, low-complex highly repetitive satellite DNA or transposable elements. These DNA elements evolve rapidly and are embedded in heterochromatin to prevent their transcription. Even closely related species strongly differ in the sequence and copy number of such repetitive elements (Clark et al., 2007; Le et al., 1995; Lerat et al., 2011; Vieira and Biemont, 2004). Several hybrid incompatibility proteins are connected to heterochromatin (reviewed in (Sawamura, 2012)). These proteins show signs of adaptive evolution, which means that their alleles underwent positive selection during evolution (Bayes and Malik, 2009; Maheshwari and Barbash, 2012; Phadnis and Orr, 2009). This finding propels a model by which hybrid incompatibility results from the adaptive coevolution of chromatin-associated factors and heterochromatic DNA (Brown and O'Neill, 2010; Crespi and Nosil, 2013; Johnson, 2010; Presgraves, 2010; Sawamura, 2012). Changes in sequence and copy number of heterochromatic DNA trigger the adaptive evolution of the corresponding regulatory factors and ensure chromatin integrity in the pure species (**Figure 3 A**). In the hybrid background, such molecular arms races may result in genetic incompatibility: Referring to the model proposed by Dobzhansky and Muller (**Figure 1**), the hybrid incompatibility (HI) protein and its genomic target sites coevolved in species 1. The species 2 genome evolved independently from species 1. Therefore, heterochromatic DNA and their regulatory factors are not compatible any more among the sibling species and result in chromatin misregulation once they are brought together in hybrids (**Figure 3 B**).

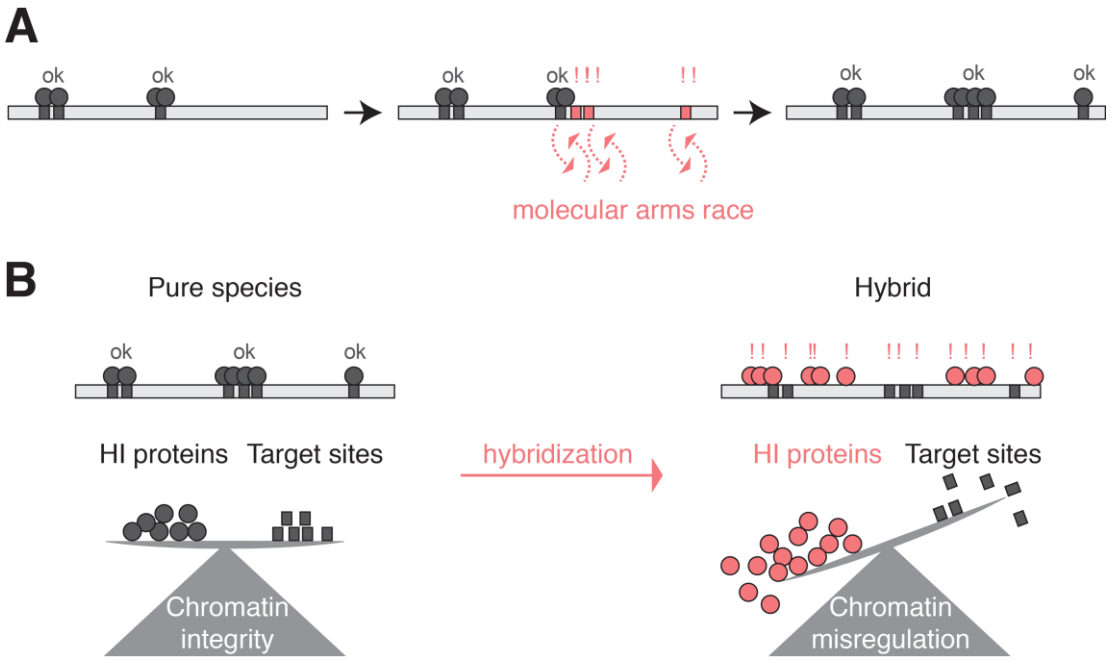

Figure 3. Chromatin-related molecular arms race in pure species can drive the evolution of hybrid incompatibility. (A) Changes in heterochromatic DNA, such as changes in sequence and copy number of selfish DNA elements, trigger the adaptive evolution of the corresponding regulatory factors. **(B)** The genes involved in this molecular arms race rapidly evolve and ensure chromatin integrity in pure species but become incompatible with sibling species genome and cause chromatin misregulation in hybrid background.

The heterochromatin functions are diverse. Genes placed in heterochromatin proximity can get transcriptionally silenced (also known as position-effect-variegation (Spofford, 1976) (reviewed in (Eissenberg and Reuter, 2009)). This observation, together with the fact that some genes must be embedded in heterochromatin to be actively transcribed, suggests that heterochromatin plays a crucial role in gene regulation (Biggs et al., 1994; Devlin et al., 1990) (reviewed in (Yasuhara and Wakimoto, 2006)). In addition to its regulative function, heterochromatin is crucial in mediating chromosome segregation during meiosis. Chromosomes that do not recombine, namely the X and the 4th chromosome in *D. melanogaster*, are paired by heterochromatic threads which eventually allow proper meiotic chromosome segregation (Dernburg et al., 1996; Hughes et al., 2009; Karpen et al., 1996; Theurkauf and Hawley, 1992). Such threads have also been identified in mammalian centromere regions during mitosis (Baumann et al., 2007; Chan et al., 2007).

1.4 Several hybrid incompatibility proteins are heterochromatin components

Given the functional importance of heterochromatin in mitosis and meiosis and the rapid evolution of its underlying DNA sequences, heterochromatin potentially serves as a key player in setting up species barriers. Species-specific heterochromatin regions cause hybrid incompatibility, which is demonstrated by the *Zygotic hybrid rescue* (*Zhr*) locus. *Zhr* causes lethality in female hybrids of *D. simulans* females and *D. mel*anogaster males (Sawamura et al., 1993) (inverse to the cross rescued by *Hmr*/*Lhr*, **Figure 2 B**). *Zhr* maps to a *359-bp* repeat-containing region on the *D. melanogaster* male X chromosome (Ferree and Barbash, 2009; Sawamura and Yamamoto, 1997). This region causes mitotic segregation defects and death of female hybrid embryos (Ferree and Barbash, 2009). Besides heterochromatic DNA sequence, proteins that bind and regulate such sequences can cause reproductive isolation. *Odysseus-site homeobox* (*OdsH*) encodes a protein that binds to repeat-rich regions of the *D. simulans* X and 4th chromosome and opens heterochromatic regions (Bayes and Malik, 2009; Ting et al., 1998). A gain of binding sites for the *D. mauritiana* OdsH to the *D. simulans* Y chromosome is suggested to affect heterochromatin packaging, which causes male sterility in hybrids (Bayes and Malik, 2009). Other examples for hybrid incompatibility genes identified in *Drosophila* are *Nup160* and *Nup96*, which both encode components of the nuclear pore complex (NPC). NPCs are giant tunnel-like structures in the nuclear membrane that mediate mutual exchange of molecules from cytosol to nucleus and *vice versa*. NPCs interact with chromatin, contribute to chromatin structure and gene regulation (Capelson et al., 2010; Grossman et al., 2012; Kalverda and Fornerod, 2010; Liang and Hetzer, 2011). The NPC's architecture and function is remarkably conserved (Ryan and Wente, 2000; Vasu and Forbes, 2001; Yang et al., 1998). In contrast to other NPC components, Nup160 and Nup96 evolve rapidly and cause *D. melanogaster* and *D. simulans* hybrid lethality (Presgraves et al., 2003; Tang and Presgraves, 2009). As NPCs are involved in chromosome segregation, it was speculated that Nup160 and Nup96 adaptively evolved to recognize repetitive DNA elements in centromeric heterochromatin (Sawamura, 2012).

Even though most HI genes were identified in *Drosophila*, reproductive isolation and chromatin are connected in mammals too. The first HI gene identified in mammals encodes PR domain-containing protein 9 (Prdm9), a DNA-binding protein with

histone H3 lysine 4 (H3K4) trimethyltransferase activity. In mouse, Prdm9 causes sterility in hybrid males by inducing meiotic arrest (Mihola et al., 2009).

These examples show that *Hmr* and *Lhr* are not the only *Drosophila* genes that cause reproductive isolation. It is important to mention that hybrid incompatibility is the result of a complex interaction between the sibling species genomes. Hybrid incompatibility cannot be simply reduced to the divergence of two individual genomic regions as the Dobzhansky-Muller model proposes. In fact, male flies of *D. melanogaster* that express endogenous *Hmr* and transgenic Lhr_{sim} are viable, whereas hybrid males that carry Hmr_{mel} and Lhr_{sim} are not (Brideau et al., 2006). This indicates that additional regions on the *D. simulans* genome cause hybrid male lethality. The identification of such genes is ongoing. Recently, Phadnis *et al.* identified another gene involved in hybrid incompatibility, the *glutathione-S-transferase-containing FLYWCH zinc finger protein* (*gfzf*) (Phadnis et al., 2015). The cell cycle regulator *gfzf* causes hybrid male inviability by inducing meiotic arrest (Phadnis et al., 2015). A summary of hybrid incompatibility genes identified so far and mentioned in this work is given in **Table 1**.

Table 1. Summary of hybrid incompatibility genes mentioned in this work.

Gene	Species	Rapidly evolving	Comment	Reference
gfzf	*D. melanogaster/ D. simulans*	yes	Cell cycle regulator	(Phadnis et al., 2015)
Hmr	*D. melanogaster/ D. simulans*	yes	Interaction with LHR and heterochromatin proteins; centromeric heterochromatin-binding; role in centromere function	(Barbash et al., 2003; Thomae et al., 2013)
Lhr	*D. melanogaster/ D. simulans*	yes	Interaction with HMR and heterochromatin proteins; centromeric heterochromatin-binding; role in centromere function	(Brideau et al., 2006; Thomae et al., 2013)
Nup160	*D. melanogaster/ D. simulans*	yes	Centromeric heterochromatin-binding?	(Sawamura et al., 2010; Tang and Presgraves, 2009)
Nup96	*D. melanogaster/ D. simulans*	yes	Centromeric heterochromatin-binding?	(Presgraves et al., 2003)
OdsH	*D. melanogaster/ D. simulans*	yes	Heterochromatin-binding	(Bayes and Malik, 2009; Sun et al., 2004; Ting et al., 1998)

Ovd	*D. pseudoobscura/ D. bogotana*	yes	Heterochromatin-binding; role in segregation distortion	(Phadnis and Orr, 2009)
Prdm9	*M. m. musculus/ M. m. domesticus*	yes	Histone H3 methyltransferase	(Mihola et al., 2009; Oliver et al., 2009; Thomas et al., 2009)
Zhr	*D. melanogaster/ D. simulans*	yes	Centromeric *359*-bp repeat region	(Ferree and Barbash, 2009; Sawamura et al., 1993)

1.5 HMR, LHR and heterochromatic proteins interact and function in a dosage-dependent manner

There are several examples for heterochromatin being involved in hybrid incompatibility (**Table 1**). This suggests chromatin as a driver in evolutionary processes and provides hints for a molecular arms race at the level of chromatin and its regulation. However, the molecular details on these processes are still poorly understood. Some mechanistic insights have been gained from studies on *Hmr* and *Lhr*. *Hmr* was among the first genes identified to cause hybrid inviability. Population genetic analysis demonstrated that *Hmr* and its genetically incompatible locus *Lhr* evolved both under positive selection (Barbash et al., 2004; Maheshwari et al., 2008). Notably, despite their deleterious function in hybrids, neither *Hmr* nor *Lhr* are essential for pure species viability. Loss-of-function mutations of *Hmr* and *Lhr* in *D. melanogaster* reduce female fertility (Aruna et al., 2009; Satyaki et al., 2014). *Hmr* orthologs from *D. simulans* and *D. mauritiana* can partially rescue this fertility phenotype (Aruna et al., 2009; Satyaki et al., 2014). On the contrary, only the *D. melanogaster Hmr* is lethal to hybrids, as hybrid males rescued by Hmr_{mel} mutation are not killed by *Hmr* orthologs from sibling species (Barbash et al., 2004). This and the fact that hybrid female fertility in the case of *Hmr* mutant parents is partially rescued indicate that the role of *Hmr* in hybrids is not directly related to the pure species function. Instead, *Hmr* seems to gain new deleterious roles in hybrids.

Hmr encodes a member of the Myb/SANT-like domain in ADF1 (MADF) protein family and contains four MADF domains that diverged between *Hmr* orthologs (Maheshwari et al., 2008). Originally, the MADF domain was identified in ADF1 where it mediates ADF1 DNA binding activity (Cutler et al., 1998; England et al., 1992). MADF domains are generally associated with DNA or chromatin binding suggesting that HMR potentially diverged in its chromatin binding specificity and

function (Aasland et al., 1996; Maheshwari et al., 2008). In fact, a single mutation in *D. melanogaster Hmr*'s third MADF domain, the Hmr^2 allele, abrogates HMR's centromere binding and rescues the hybrid phenotype (Barbash et al., 2003; Thomae et al., 2013) suggesting that HMR's association to chromatin plays a crucial role in the hybrid gain-of-function. Apart from the N-terminal MADF domains, *Hmr* encodes a C-terminal BEAF, Su(var)3-7 and Stonewall-like (BESS) domain (Brideau et al., 2006). The family of MADF-BESS domain containing proteins, with MADF at the N-terminal part and BESS at the C-terminal part, consists of at least 16 members in *D. melanogaster* (Shukla et al., 2014). Interestingly, phylogenetic analysis revealed that the MADF-BESS proteins evolve rapidly and increased substantially in number during *Drosophila* evolution, presumably through gene duplication mechanisms (Shukla et al., 2014). Numerous of these genes, including *Hmr* and *Overdrive*, are involved in hybrid incompatibility.

Hmr and *Lhr* are *Drosophila*-specific and rapidly evolving (Barbash et al., 2003; Brideau et al., 2006). Strikingly, these two genes not only genetically interact to cause hybrid phenotype, but also encode two proteins that physically interact to form a protein complex *in vivo*. Compared to *Hmr*, the *Lhr* gene encodes a small protein that does not contain a MADF domain but harbors a BESS domain too (Brideau et al., 2006). The BESS domain mediates protein-protein interaction, in the case of HMR's and LHR's BESS domain their mutual interaction (Bhaskar and Courey, 2002; Brideau et al., 2006; Thomae et al., 2013).

Biochemical analysis of HMR and LHR revealed numerous interaction partners. Among the most prominent ones is Heterochromatin Protein 1 a (HP1a) (Alekseyenko et al., 2014; Thomae et al., 2013), a protein that is heterochromatic and also termed Suppressor of variegation 205 [Su(var)205] due to its impact on position-effect-variegation (Ebert et al., 2004; Eissenberg et al., 1990). The loss of a repressor such as HP1a results in a loss of silencing of a heterochromatic reporter gene. Numerous chromatin modifications and the proteins involved in writing and reading these modifications are specifically associated with one type of chromatin. For heterochromatin, these are HP1a and di- or tri-methylated lysine 9 on the histone H3 tail (H3K9me2/3) (Dillon and Festenstein, 2002; Grewal and Rice, 2004; Richards and Elgin, 2002). HP1a is essential, highly conserved from yeast to human (Singh and Georgatos, 2002) and interacts with methylated H3K9 through its N-terminal chromo domain (CD) (Bannister et al., 2001; Jacobs and Khorasanizadeh, 2002; Jacobs et al., 2001; Lachner et al., 2001; Nielsen et al., 2002). HP1a's C-terminal chromo-shadow

domain (CSD) mediates HP1a multimerization and interactions with other proteins (Alekseyenko et al., 2014; Brasher et al., 2000; Li et al., 2003; Murzina et al., 1999; Stewart et al., 2005). HP1a interacts and cooperates with the proteins Su(var)3-7 and Su(var)3-9 to mediate transcriptional silencing at heterochromatic regions (Danzer and Wallrath, 2004; Greil et al., 2003; Schotta et al., 2002). In contrast to HP1a and Su(var)3-9, the protein Su(var)3-7 is *Drosophila*-specific, evolved rapidly and harbors a BESS domain (Jaquet et al., 2006) that mediates its interaction with Su(var)3-9 (Schotta et al., 2002). Additionally, Su(var)3-7 harbors DNA-binding zinc finger domains that recruit Su(var)3-7 and its interaction partners to satellite repeat containing pericentromeric and telomeric regions (Cleard et al., 1997; Delattre et al., 2000). The protein Su(var)3-9 localizes with both, Su(var)3-7 and HP1a to these heterochromatic regions (Schotta and Reuter, 2000). Su(var)3-9 is a histone methyl transferase that sets the histone H3K9 methylation mark (Schotta et al., 2002), which, in turn is recognized by HP1a, suggesting a mechanism for heterochromatin spreading.

The cellular protein dosage of HMR, LHR and heterochromatin-associated proteins is crucial for their cellular function. Su(var) proteins act as modifiers of position-effect-variegation in a dosage-dependent manner (Eissenberg et al., 1990; Eissenberg et al., 1992; Locke et al., 1988; Schotta et al., 2002). Reduced or increased amounts of Su(var)3-7 further affect the genomic localization of Dosage Compensation Complex (DCC) proteins and the expression of dosage-compensated genes (Spierer et al., 2008). Prior studies from Thomae et al. aimed to identify species-specific differences between HMR and LHR orthologues and highlighted different expression levels for HMR and LHR in the two species (Thomae et al., 2013). The HMR protein level is increased in *D. melanogaster*, whereas the LHR protein level is higher in *D. simulans* (Maheshwari and Barbash, 2012; Thomae et al., 2013). HMR and LHR do not only interact but also mutually stabilize each other (Satyaki et al., 2014; Thomae et al., 2013). In hybrids, high levels of HMR_{mel} and high levels of LHR_{sim} face each other. HMR_{mel} and LHR_{sim} interact, which in turn results in a highly increased level of the HMR/LHR complex and a lethal gain of function in hybrids (Thomae et al., 2013). The molecular basis for this lethal gain-of-function remains to be elucidated. Higher levels of HMR and LHR lead to a remarkably increased number of their binding sites, both in the hybrids' and in the pure species' background (Thomae et al., 2013), indicating that the gain-of-function relates to chromatin binding activity.

1.6 HMR and LHR associate with fast evolving centromeric and telomeric heterochromatin

So far, HMR's and LHR's genomic localization were studied using cytological staining where HMR and LHR locate in embryos at subdomains of pericentromeric heterochromatin (Satyaki et al., 2014). In these studies, HMR colocalizes with HP1a and H3K9me2 at *Dodeca* and *GC*-rich satellite repeats but not with the *359-bp* satellite repeat that causes hybrid incompatibility (*Zhr*, see above) (Satyaki et al., 2014). Further, LHR was described as part of classical heterochromatin present at pericentromeric regions (Filion et al., 2010; Greil et al., 2007). These studies localized ectopically epitope-tagged HMR and LHR, making the results prone to protein overexpression bias. The first, and to date only, studies localizing endogenous HMR and LHR come from Thomae et al., who found both proteins associated to the centromere in *D. melanogaster* cell lines and mitotic cells of the larvae wing imaginal discs, which points to a role of the HMR/LHR complex in mitosis (Thomae et al., 2013). On polytene chromosomes, large genomic DNA structures that are isolated from endoreplicating tissue, the proteins HMR and LHR are additionally detected at several regions along the chromosome arms and at telomeres (Thomae et al., 2013). A common feature of these genomic regions is their association with heterochromatin features. Consistent with these findings, Thomae et al. uncovered that the HMR/LHR complex contains centromeric and pericentromeric factors apart from HP1a, such as Umbrea, NLP (nucleoplasmin-like protein), CENP-C (James et al., 1989; Padeken et al., 2013; Ross et al., 2013; Thomae et al., 2013; Vermaak and Malik, 2009) and the telomere protein Verrocchio (Raffa et al., 2010; Vedelek et al., 2015).

Both, telomeres and centromeres are heterochromatic structures with unique features and crucial cellular functions. Telomeres harbor a specialized type of heterochromatin that is composed of rapidly evolving DNA sequences and associated proteins (Anderson et al., 2009; Mefford and Trask, 2002; Raffa et al., 2011). A protective telomere cap structure, forming on this DNA, shields chromosome ends and prevents telomere fusion (Andreyeva et al., 2005; Raffa et al., 2011). The centromere is an assembly platform for kinetochore proteins, which facilitate accurate chromosome segregation during mitotic cell division. The centromere structure is characterized by conserved and unique epigenetic features: the histone H3 variant CID (centromere identifier in *Drosophila*) is present at the inner centromere region, which is embedded in densely packed heterochromatin (Allshire and Karpen, 2008). This centromere-

associated heterochromatin consists of long repetitive arrays and harbors selfish DNA sequences such as transposable elements (TEs) that are transcriptionally silenced (Birchler et al., 2000). The variation in copy number of such repetitive DNA is remarkably high and causes a broad variation in genome size among multicellular eukaryotes in general (Gregory, 2005) and among *Drosophila* species in particular (Bosco et al., 2007). On the one hand, the copy number of such DNA elements can rapidly change with processes that involve homology regions such as crossing over and duplication events (Charlesworth et al., 1994). On the other hand, selfish properties such as transposition for TEs (Hickey, 1982) and meiotic drive for satellite DNA sequences (Walker, 1971) can cause overrepresentation of repetitive DNA. In such a genetic conflict scenario, the action of selfish DNA on the host genome is detrimental to other genes, and therefore results in an arms race between selfish DNA and host defense mechanisms. The actions of selfish DNA have been proposed as an important driver of hybrid incompatibility (Brown and O'Neill, 2010; Crespi and Nosil, 2013; Johnson, 2010; Maheshwari and Barbash, 2011; Presgraves, 2010). As shown above, multiple *Drosophila* HI genes and gene products are associated with such genomic elements. **Figure 4** depicts the cytological localization of HMR/LHR to various heterochromatic regions that are subject to this postulated molecular arms race and suggests a relation between this localization and the observed HMR/LHR complex functions and phenotypes.

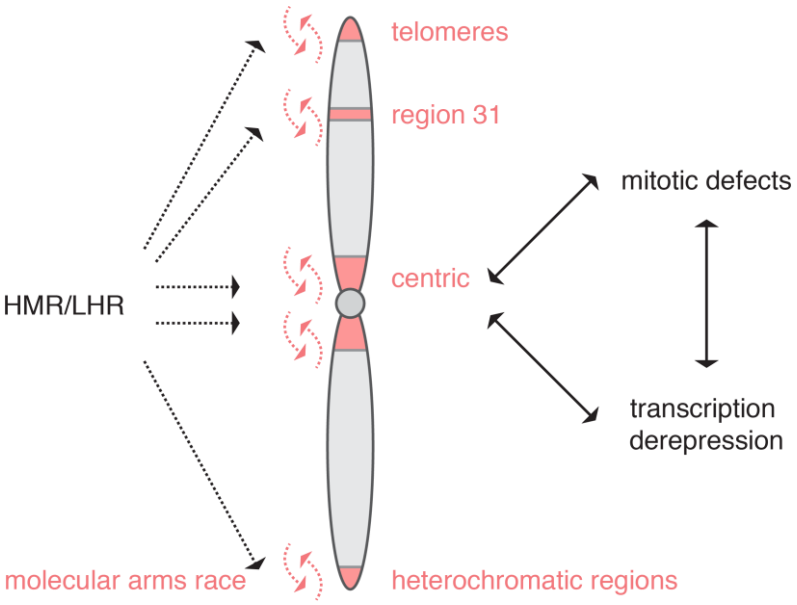

Figure 4. HMR and LHR localize to centromere-associated heterochromatin and are crucial for centromere function. Based on cytological staining, the HMR/LHR complex associates with various heterochromatic regions across the *Drosophila* genome including centromeric heterochromatin, telomeres and other distinct regions at the chromosome arms such as cytological region 31, all regions of a postulated arms race (see text). HMR's and LHR's localization is in accordance to their loss-of-function phenotypes. Mitotic defects and transcriptional derepression of underling repetitive elements and heterochromatic genes are consistent with impaired centromere function (see text).

1.7 HMR and LHR affect transcription of heterochromatic genes and repetitive DNA

The loss of HMR or LHR by RNAi knockdown in *D. melanogaster* cells results in an increased rate of mitotic defects indicating that the HMR/LHR complex does not only localize to the centromere but also has a critical function in chromosome segregation (Thomae et al., 2013). What is then the function of HMR and LHR at centromere regions? On the one hand, HMR and LHR were not identified in CID pull down (Barth et al., 2014) and the reduction of HMR and LHR in *D. melanogaster* does not affect the kinetochore structure or the incorporation of newly synthesized CID (Thomae et al., 2013). On the other hand, loss of HMR or LHR results in highly increased transcription of transposable elements, including those located in centromeric and heterochromatic regions. This was shown in cell lines using RNAi knockdown approaches (Thomae et al., 2013) and in ovaries of *Hmr* and *Lhr* mutant flies (Satyaki et al., 2014). In ovaries, HMR and LHR suppress satellite DNA transcription, repeats with germline-specific expression such as telomeric *HeT-A* and retrotransposon *copia* but also somatic expressed repeats such as *gypsy* (Satyaki et al.,

2014). In line with telomeric *HeT-A* derepression, *Hmr* and *Lhr* mutant flies also displayed higher copy number of *HeT-A* suggesting HMR/LHR being telomere cap proteins that regulate telomere length (Satyaki et al., 2014). However, the underlying mechanism of these observations is still poorly understood. Notably, also in hybrid male larvae where HMR and LHR proteins are present before the animal eventually dies, transposable elements are misregulated. The misregulation of repetitive DNA transcription could origin from defects in chromatin packaging, but could also be a secondary effect of hybrid developmental and physiological defects. Similar could be true for *Hmr* or *Lhr* loss-of-function mutant flies which display germline-associated phenotypes and chromosome segregation defects. Overall, HMR localizes to heterochromatin-associated sites in the genome, namely telomeric and centromeric regions, presumably via its interacting proteins, such as the heterochromatic proteins HP1a, SUVAR3-7 and the telomeric protein Verrocchio (**Figure 4**). Strikingly, these underling DNA sequence as well as their associated proteins are rapidly evolving (**Figure 4**). The role of HMR and LHR at such genomic sites is currently not well understood, but the loss of HMR or LHR results in transcription misregulation of coding and repetitive DNA at these sites (**Figure 4**).

1.8 Evolution of insulators and insulator proteins in *Drosophila*

Several genes associated with hybrid incompatibility are rapidly evolving. Further, a remarkably high portion of them is part of the MADF-BESS domain family in *Drosophila*, a protein family that successively gained new members during *Drosophila* evolution (Shukla et al., 2014). As described above, *Hmr* encodes four MADF domains and a BESS domain whereas *Lhr* lacks MADF domains but encodes a BESS domain (Brideau et al., 2006; Maheshwari et al., 2008). The BESS domain is often associated with the MADF domain (Bhaskar and Courey, 2002; Cutler et al., 1998; England et al., 1992) and is named after BEAF or BEAF-32, a protein that belongs to another highly remarkable group of proteins in the context of *Drosophila* evolution. These proteins emerged in *Drosophila* and are called insulator proteins (Heger et al., 2013). Insulator proteins bind to rapidly evolving DNA elements that adaptively change during evolution (Ni et al., 2012; Yang et al., 2012). Many transposons or transposon-derived sequences such as *gypsy* give rise to genomic insulator sites, regulatory elements across the genome that mediate chromatin organization into distinct domains and ensure gene transcription integrity.

While only one insulator protein, CCCTC-binding factor (CTCF), is known in vertebrates, arthropods have experienced a successive gain of insulator proteins during evolution (Heger et al., 2013; Pauli et al., 2016; Schoborg and Labrador, 2010). To date, *Drosophila* utilizes the largest known variety of insulator proteins such as CTCF, Boundary Element Associated Factor 32 (BEAF-32), Suppressor of Hairy wing [Su(Hw)], Modifier of mdg4 [Mod(mdg4)] and Centrosomal Protein 190 (CP190), which all affect nuclear architecture (Sexton et al., 2012). Different *Drosophila* species underwent multiple genomic rearrangements and transposon invasions (Bosco et al., 2007; Clark et al., 2007), which presumably resulted in an adaptive response of regulatory DNA elements to maintain spatial and temporal gene expression. For example, binding sites of the insulator proteins BEAF-32 and CTCF show a high degree of variability when compared among very closely related species. The gain of new insulator sites is an adaptive change in response to chromosome rearrangements as well as to the birth of genes and has a direct impact on species-specific transcription (Ni et al., 2012; Yang et al., 2012). This is very reminiscent of the situation for HMR, LHR, their interacting heterochromatin proteins and repetitive DNA elements described in the text before.

This work provides a novel link between HMR and insulator proteins. Using genome-wide mapping techniques, we found HMR localizing to discrete subsets of genomic insulator sites. One of them belongs to *gypsy* insulators that are marked by the presence of Su(Hw), Mod(mdg4) and CP190. The other subset of HMR binding sites associates with the insulator protein BEAF-32 and is exclusively found at the transcription start site of heterochromatic genes. Interestingly, such genes are transcriptionally affected in HMR and BEAF-32 mutants suggesting a functional relation between HMR, insulator proteins and heterochromatin. Further, these findings and the recurrent gain of insulator proteins in evolution suggest a putative role of insulator proteins and genome organization in the formation of species.

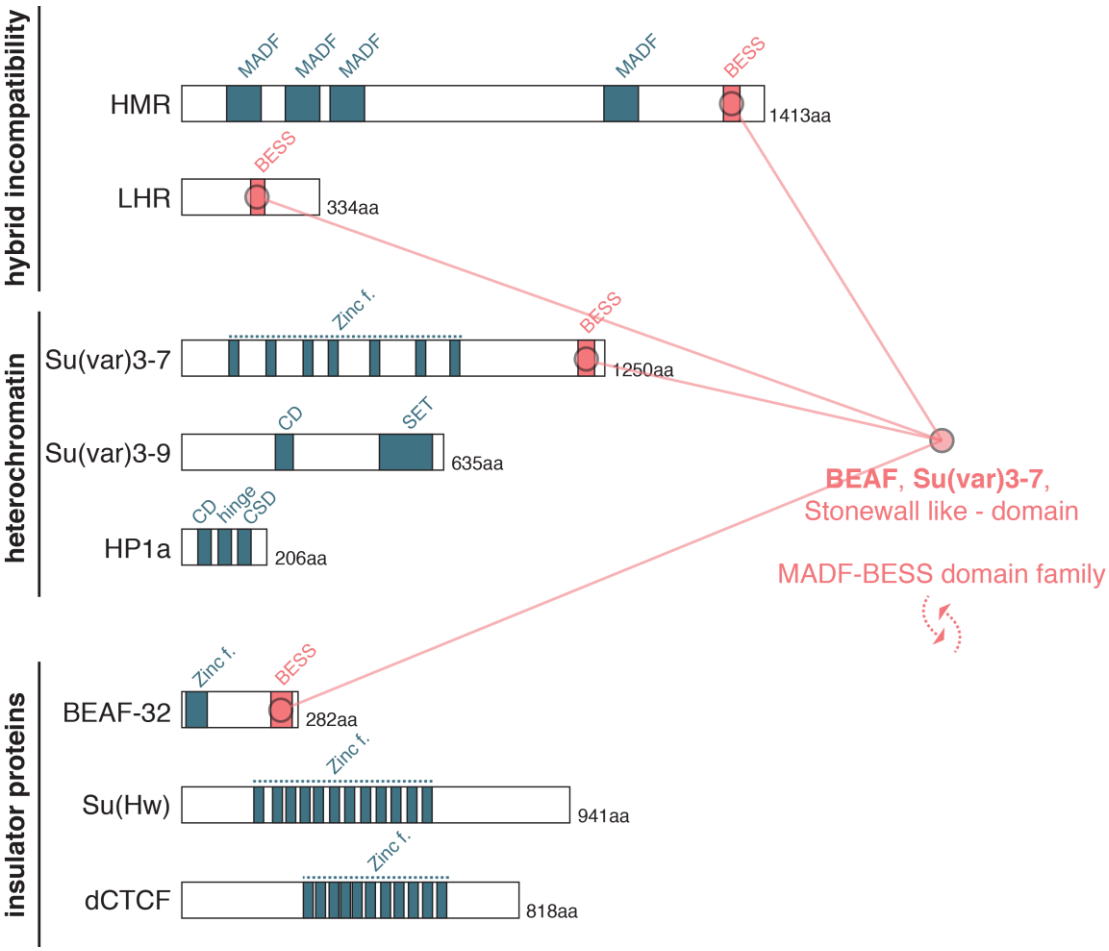

Figure 5. Hybrid incompatibility proteins, heterochromatin proteins and insulator proteins are interconnected by the BESS domain and partially subject to rapid evolution. HMR, LHR, Su(var)3-7 and BEAF-32 share the BESS domain. Insulator proteins and MADF-BESS domain family proteins successively gained new members during evolution (Heger et al., 2013; Shukla et al., 2014). In contrast to HP1a and Su(var)3-9, the BESS domain containing Su(var)3-7 is *Drosophila*-specific and evolved rapidly (Jaquet et al., 2006). Domain architecture obtained from: UniProt webpage, (Aagaard et al., 1999; Bhaskar and Courey, 2002; Brideau et al., 2006; Jaquet et al., 2002; Maksimenko and Georgiev, 2014).

1.9 Thesis aims

The *Drosophila* gene *Hmr* was among the first identified hybrid incompatibility genes and constitutes one of the most studied genes in the field of speciation. HMR is a chromatin component in *Drosophila* and forms a chromatin-associated complex with LHR and other proteins. What makes the study of HMR particularly interesting as a model in the field of chromatin and evolution is:

1st Understanding the driving forces that lead to positive selection of hybrid incompatibility genes during evolution.

2nd Understanding the reasons for this adaptive evolution being genetically incompatible with sibling species and detrimental in a hybrid background.

Both aspects are particularly related to the genomic localization of HMR. In *D. melanogaster* pure species, HMR localizes to centromeric heterochromatin and plays an important role in chromosome segregation and in the regulation of transposable elements and heterochromatic gene transcription. In hybrids, widespread mislocalization of HMR is suggested to cause lethality by a misregulation of de novo target loci. However, the molecular details of HMR binding sites and their putative role in HMR's function in pure species and in hybrid background remains to be elucidated. To gain new insights on the nature and biology of HMR binding sites we aimed to:

1st Identify and describe HMR's genomic binding sites on a molecular level in *D. melanogaster*.

2nd Dissect HMR's targeting mechanisms and recruitment to its binding sites in the *D. melanogaster* genome.

3rd Elaborate on HMR's function at its binding sites in the *D. melanogaster* genome.

To address these questions, we identify and describe the genome-wide binding properties of HMR in *D. melanogaster* Schneider S2 cells, an embryonic cell line. S2 cells are used as a popular model and provide genome-wide data sets of other chromatin features, which were generated by the modENCODE consortium. Further, this cell line is highly accessible to specific RNAi knockdown that allows specific and efficient reduction of proteins (Elbashir et al., 2001) and can be subjected to advanced genome editing (Bottcher et al., 2014; Kunzelmann et al., 2016).

To answer the first question we performed chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) on HMR using antibody directed against endogenous HMR generated prior to this study (Thomae et al., 2013). In collaboration

with Prof. Dr. Klaus Förstemann and his group (Biocenter LMU Munich), we generated an endogenously epitope-tagged HMR expressing cell line (Bottcher et al., 2014) to be used in ChIP with commercial antibodies. The derived HMR binding sites were characterized with respect to their underlying DNA sequence and binding overlap to other chromatin marks and chromatin-associated proteins.

To answer the second question, we took use of the RNAi knockdown strategy that allows efficient depletion of a target protein prior to ChIP. The knockdown of a putative HMR targeting factor followed by HMR ChIP allows assessing the specific effect of the knockdown on HMR binding relative to a negative control treatment. This strategy was combined with a detailed description and characterization of HMR binding sites (first question's answer) and tailored bioinformatics analysis.

To answer the third question, we took use of public data on gene transcription in *Hmr* mutant flies and other transcription datasets. By applying tailored bioinformatics analysis, we then related these data to the knowledge on the HMR binding sites.

Such molecular insights should be of general interest for the speciation research field that currently lacks a mechanistic understanding of speciation and its evolutionary driving forces. By describing HMR's genomic binding properties on a molecular level, we further promote the chromatin research field by characterizing a biologically relevant example for the deregulation of chromatin factors.

2 MATERIALS & METHODS

2.1 Materials

2.1.1 Cell lines

Cells used in this work were *D. melanogaster* S2-DRSC cells obtained from the DGRC (Schneider, 1972). For CRISPR/Cas9-based tagging of HMR, either S2-DRSC cells or S2-DRSC cells with myc-Cas9 integration, clone 9-4 (neomycin-resistant) and 5-3 (hygromycin-resistant) (Bottcher et al., 2014) were obtained from the lab of Prof. Dr. Klaus Förstemann. Cell lines generated with CRISPR/Cas9 system are listed in **Table 2**.

Table 2. List of cell lines generated with CRISPR/Cas9 system in this study.

Name	Description	ID and date of freezing
S2 HMR-Flag$_2$	Parental line: S2-DRSC U6-gRNA and pRB14 (myc-Cas9) transfected Population	p2-4, population (08/09/2014, 02/10/2014, 08/10/2014)
S2 HMR-GFP	Parental line: S2 myc-Cas9 (Bottcher et al., 2014) U6-gRNA Single clone	p4-2, clone #6 (08/09/2014, 02/10/2014)
S2 HMR-Strep$_2$	Parental line: S2-DRSC U6-gRNA and pRB14 (myc-Cas9) transfected Single clone	p2-5, clone #1 (08/09/2014, 02/10/214)
S2 HMR-Strep$_2$	Parental line: S2-DRSC U6-gRNA and pRB14 (myc-Cas9) transfected Single clone	p2-5, clone #4 (08/09/2014, 02/10/214)

2.1.2 Plasmids

Table 3. List of plasmids used in this study.

Plasmids	Description	Reference/Source
pRB14	Expression of myc-Cas9; used for CRISPR/Cas9 strategy	(Bottcher et al., 2014)
pRB17	Template for overlap-extension PCR to generate U6-gRNA; contains the U6-promotor with a T7 extension; used for CRISPR/Cas9 strategy	(Bottcher et al., 2014)
pMH3	Template for PCR to generate homologous recombination donors; containing the GFP-tag sequence and blasticidin resistance gene; used for CRISPR/Cas9 strategy (tagging, knockout blasticidin-2/2)	K. Förstemann
pMH4	Template for PCR to generate homologous recombination donors; containing the $Flag_2$-tag sequence and blasticidin resistance gene; used for CRISPR/Cas9 strategy (tagging)	(Bottcher et al., 2014)
pIW1	Template for PCR to generate homologous recombination donors; containing the $Strep_2$-tag sequence and blasticidin resistance gene; used for CRISPR/Cas9 strategy (tagging)	(Bottcher et al., 2014)
pKF277	Template for PCR to generate homologous recombination donors; containing blasticidin resistance gene; used for CRISPR/Cas9 strategy (tagging, knockout blasticidin-1-2)	K. Förstemann
pRB25	Template for PCR to generate homologous recombination donors; containing puromycin resistance gene; used for CRISPR/Cas9 strategy (tagging, knockout puromycin-1-2)	K. Förstemann
pSK23	Template for PCR to generate homologous recombination donors; containing puromycin resistance gene; used for CRISPR/Cas9 strategy (tagging, knockout puromycin-2-2)	K. Förstemann

2.1.3 Oligonucleotides

Oligonucleotides used in this work were obtained from Sigma-Aldrich (desalt, 100 µM in water). The sequences of oligonucleotides are listed in **Table 4**, **Table 5**, **Table 6**, **Table 7**. Oligonucleotides published in (Bottcher et al., 2014) were kindly provided by Prof. Dr. Klaus Förstemann.

Table 4. List of oligonucleotides used for CRISPR/Cas9-based tagging of HMR.

Name	Sequence (5'-3')	Reference/Source
***lig4* and *mus308* knockdown**		
lig4 RNAi for	TAATACGACTCACTATAGGGCCCAATGATCCAAAGTGTTTTTGCA	(Bottcher et al., 2014)
lig4 RNAi rev	TAATACGACTCACTATAGGGAAGTAGGATGCCTTCGCGA	(Bottcher et al., 2014)

mus308 RNAi for	TAATACGACTCACTATAGGGCTGGGACTCCACCGGAAAG	K. Förstemann
mus308 RNAi rev	TAATACGACTCACTATAGGGTACCGTCGCCGTCCAGTAATG	K. Förstemann
gRNA *in vitro* transcription template generation		
oligo scaffold	GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGC	(Bottcher et al., 2014)
oligo CRISPR target with T7 prom. (HMR tagging)	TAATACGACTCACTATAGCCACCGCCTTAGCTCTCGAAACTTTGTTTTAGAGCTA	Designed for this work; HMR-specific
oligo CRISPR target with T7 prom. (HMR N-term knockout)	CCTATTTTCAATTTAACGTCGTTGGAGCTATCAGGTGTCTGTTTAAGAGCTATGCTG	Designed for this work; HMR-specific
oligo CRISPR target with T7 prom. (HMR C-term knockout)	CCTATTTTCAATTTAACGTCGCCTTAGCTCTCGAAACTTTGTTTAAGAGCTATGCTG	Designed for this work; HMR-specific
oligo CRISPR target with T7 prom. (LHR N-term knockout)	CCTATTTTCAATTTAACGTCGAGTGGTATATTAAAACATAGTTTAAGAGCTATGCTG	Designed for this work; LHR-specific
oligo CRISPR target with T7 prom. (LHR C-term knockout)	CCTATTTTCAATTTAACGTCGGAAATATAAAATGCTATTGTTTAAGAGCTATGCTG	Designed for this work; LHR-specific
primer antis. scaffold	GCACCGACTCGGTGCCACT	(Bottcher et al., 2014)
U6-gRNA template generation by overlap-extension PCR		
U6-gRNA sense	GCTCACCTGTGATTGCTCCTAC	(Bottcher et al., 2014)
U6-gRNA antisense	GCTTATTCTCAAAAAAGCACCGACTCGGTGCCACT	(Bottcher et al., 2014)
Homologous recombination template generation (tagging)		
HMR sense	TGGGCCTACGCCGTCGGTAACTTGTCCACGGCCAGTCAGGATACACTGCTCGGCAAGATGACGCAGCTGTTCTCTAAATACGCCAAGGTCAATCCGCCACCGCCTGGATCTTCCGGATGGCTCGAG	Designed for this work; HMR-specific
HMR antisense	ACGGCGAAAGTTCTTACAGAGAATATGTATGACTAAACTACGTGTGCCAAAAGTTTCGAGAGGAAGTTCCTATTCTCTAGAAAGTATAGGAACTTCCATATG	Designed for this work; HMR-specific
Homologous recombination template generation (knockout)		
HMR N-term sense	CTCGACGGCTTGTGTGGGGGAAAGGCGCGCGTAGGTCAAGATTGGAGCTATCACGTGTCTAGGACGAAGTTCCTATACTTTCTAGAGAATAGGAACTTCCATATG	Designed for this work; HMR-specific

HMR C-term antisense	GGTACGGCGAAAGTTCTTACAGAGAATAT GTATGACTAAACTACGTGTGCCAAAAGTG TCGAGAGGAAGTTCCTATTCTCTAGAAAG TATAGGAACTTCCATATG	Designed for this work; HMR-specific
LHR N-term sense	TAGATTTTATTAAAGAAATTACCGTTAAGT GGTATATTAAAGCATACGGATGAATTAGT ACACAAGAAGTTCCTATACTTTCTAGAGA ATAGGAACTTCCATATG	Designed for this work; LHR-specific
LHR C-term antisense	CTCCTTGTTTGTTTAGTTATTAGTTCTTCG AGAATGCAAAGCAAGTGAAATATAAAATG TTATTTGAAGTTCCTATACTTTCTAGAGAA TAGGAACTTCCATATG	Designed for this work; LHR-specific
Split BlastR sense (#351)	ACAATCAACAGCATCCCCATCTC	K. Förstemann
Split BlastR antisense (#352)	TTCTCATTTCCGATCGCGACGATAC	K. Förstemann
Split PuroR sense (#781)	GGACGTTGGCTGCCGC	K. Förstemann
Split PuroR antisense (#782)	CCCCTGCTTCCACGCT	K. Förstemann
PCR analysis on genomic DNA for verification		
Hmr CDS sense	TATAAGCAGGTGAAGCCGAAC	Designed for this work; HMR-specific
Hmr downstream antisense	TGCCCTCATCGCTATCATTCTG	Designed for this work; HMR-specific
Copia antisense	GTAGGTTGAATAGTATATTCCAACAGCAT ATG	(Bottcher et al., 2014)

Table 5. List of oligonucleotides used for transcript analysis by cDNA-qPCR.

Name	Sequence (5'-3')	Publication
tub97EF for	GAGCAAGAACAGCAGCTACTTTGT	(Padeken et al., 2013)
tub97EF rev	CACCTTGACGTTGTTGGGAAT	(Padeken et al., 2013)
Hmr for	AATCGCTTGCGAAGAACACT	(Thomae et al., 2013)
Hmr rev	ACTGGCCGTGGACAAGTTAC	(Thomae et al., 2013)
Lhr for	CGCCAAGAGAAAGCTACTCG	(Thomae et al., 2013)
Lhr rev	CATGGCCGGACTGAGTAAAT	(Thomae et al., 2013)
HP1a for	AAGTCAGCCGCCTCCAAGAAGG	(Thomae et al., 2013)
HP1a rev	ATGGTGTCCTGCTCCGCATCTG	(Thomae et al., 2013)

Table 6. List of oligonucleotides used for ChIP-qPCR in the order applying to figures. Listed primers were designed in this study with Primer3. #: Primers are from (Wallace, 2010).

Name	Sequence (5'-3')
2L:302129-302248 for	CACAGCAACGAAGCTCTCTG
2L:302129-302248 rev	AGCATAGTGACCCGCATCTC
3R:23793216-23793267 for	GAGCAAGAACAGCAGCTACTTTGT
3R:23793216-23793267 rev	CACCTTGACGTTGTTGGGAAT
3RHet:2107224-2107336 for	AACCCTATCCAAATTTCGAACC
3RHet:2107224-2107336 rev	AGCCAAGATGAAGTCGATGC

4:855631-855744 for	TAAACTCAGCCCTGCATTCC
4:855631-855744 rev	GTGTTAAACCAATCCGAGACATC
2RHet:369982-370075 for	CATTTGACTTCTTCGACACGAC
2RHet:369982-370075 rev	GACACTGATTTACACAAAGCACAAC
2RHet:370407-370487 for	TGCATACCCTACAAATAGTTTTGC
2RHet:370407-370487 rev	TTGATCGGCTAAGTGAAGTGG
gypsy 5' for [#]	GGTTTCTCTAAAAAGTATGCAGC
gypsy 5' rev [#]	CTGGCCACGTAATAAGTGTGC
gtwin 5' for [#]	ATGAAGTCACTCGGCAACCT
gtwin 5' rev [#]	ACGCTTGGTAAAAGTATGCAATTG
3L:2244123-2244231 for	TTCCTGATACCAGGCGAAC
3L:2244123-2244231 rev	CAGTTCACTCCGCAGATACG
2L:22247197-22247280 for	CCGTACAATTTCCGAGCAG
2L:22247197-22247280 rev	GAAACTTGAAGAACCGATTGC

Table 7. List of primers used for RNAi experiments combined with ChIP.

Name	Sequence (5'-3')	Publication
CP190 RNAi for	TAATACGACTCACTATAGGGCCTGGCTG TGCCTGAGA	(Van Bortle et al., 2012)
CP190 RNAi rev	TAATACGACTCACTATAGGGCTGGTAGA CTTATGTCCGAAA	(Van Bortle et al., 2012)
CTCF RNAi for	TAATACGACTCACTATAGGGGAGCCCG ACATCAGTTCAAT	(Van Bortle et al., 2012)
CTCF RNAi rev	TAATACGACTCACTATAGGGGAGCACTT GAAGGATGGCTC	(Van Bortle et al., 2012)
GFP RNAi for	TTAATACGACTCACTATAGGGAGAACGT AAACGGCCACAAGTTCAGC	(Thomae et al., 2013)
GFP RNAi rev	TTAATACGACTCACTATAGGGAGATGCT GGTAGTGGTCGGCGAG	(Thomae et al., 2013)
GST RNAi for	TTAATACGACTCACTATAGGGAGAAGTT TGAATTGGGTTTGGAGTTTCC	(Thomae et al., 2013)
GST RNAi rev	TTAATACGACTCACTATAGGGAGATCGC CACCACCAAACGTGG	(Thomae et al., 2013)
HMR RNAi for	TTAATACGACTCACTATAGGGAGAGATG TGGAGGTCATAGAGAATCCGCCAATG	(Thomae et al., 2013)
HMR RNAi rev	TTAATACGACTCACTATAGGGAGAACCT TGTTGTGCAGGGAGTCCTCCGTC	(Thomae et al., 2013)
HP1a RNAi for	TAATACGACTCACTATAGGGAGATGGG CAAGAAAATCGACAACCCTGAG	(Thomae et al., 2013)
HP1a RNAi rev	TAATACGACTCACTATAGGGAGAGCGTC CTTGAGTTTCCTTGGCCTTG	(Thomae et al., 2013)

2.1.4 Antibodies and beads

Antibodies used in this work are listed in **Table 8**. The HMR- and LHR-specific rat monoclonal antibodies were generated by Dr. Elisabeth Kremmer at the Service Unit Monoclonal Antibodies at the Helmholtz-Zentrum-München and used in (Thomae et al., 2013). The HMR-specific antibodies were raised against an aminoterminal (N-terminal) fragment spanning amino acids 2-416 fused to an N-terminal Glutathione S-

transferase (GST) tag. The LHR-specific antibodies were raised against full length LHR N-terminally fused to the Maltose-Binding Protein (MBP). Beads used in this work for affinity purification are Protein A Sepharose 4 Fast Flow (GE Healthcare) and Protein G Sepharose 4 Fast Flow (GE Healthcare).

Table 8. List of antibodies used in this work.

Antibody	Host	Clonality	Reference/Provider	Application
Primary antibodies				
anti-Rat IgG (Fc)-unconj., MinX Hu	rabbit	polyclonal	RRID:AB2339804 Dianova, 312-005-046	IP: 10 µg per reaction
anti-CID 7A2	rat	monoclonal	Elisabeth Kremmer	IF: 1:200
anti-CP190 Bx63	rabbit	monoclonal	RRID:AB2615894 (Frasch et al., 1986)	WB: 1:3000
anti-CTCF N3	rabbit	polyclonal	RRID:AB2616159 (Bartkuhn et al., 2009)	WB: 1:3000 IF: 1:1000
anti-FLAG M2	mouse	monoclonal	RRID:AB262044 Sigma-Aldrich, F1804	WB: 1:10000 IP: 5 µL per reaction IF: 1:100
anti-FLAG M2	mouse	monoclonal	RRID:AB262044 Sigma-Aldrich, A2220	IP: 30 µL per reaction
anti-Histone H3	rabbit	polyclonal	RRID:AB302613 Abcam, ab1791	WB: 1:2000 IP: 2.5 µL per reaction
anti-H3K9me3	rabbit	polyclonal	RRID:AB2532132 Active Motif, 39161	WB: 1:2000 IP: 5 µL per reaction
anti-HMR 2C10	rat	monoclonal	RRID:AB2569849 (Thomae et al., 2013)	WB: 1:25 IP: 1 mL per reaction IF: 1:25
anti-HP1a C1A9	mouse	monoclonal	RRID:AB528276	WB: 1:25 IP: 300 µL per reaction
anti-LHR 12F4	rat	monoclonal	RRID:AB2569850 (Thomae et al., 2013)	WB: 1:25 IP: 1 mL per reaction IF: 1:25
anti-Tubulin	mouse	monoclonal	RRID:AB2241150 Abcam, ab44928	WB: 1:800
Secondary antibodies				
anti-mouse HRP-linked whole Ab	sheep	polyclonal	RRID: AB772210 GE Healthcare, NA931	WB: 1:5000
anti-rabbit HRP-linked whole Ab	donkey	polyclonal	RRID: AB772206 GE Healthcare, NA934	WB: 1:5000
anti-rat HRP-linked whole Ab	goat	polyclonal	RRID: AB772207 GE Healthcare, NA935	WB: 1:5000
anti-mouse Alexa Fluor conjugate	donkey	polyclonal	Jackson Immuno Research	IF: 1:300
anti-rabbit Alexa Fluor conjugate	donkey	polyclonal	Jackson Immuno Research	IF: 1:1000
anti-rat Alexa Fluor conjugate	donkey	polyclonal	Jackson Immuno Research	IF: 1:300

2.1.5 Kits, enzymes and markers

Table 9. List of kits, enzymes and markers used in this work.

Description	Supplier/Model
100 bp DNA ladder	New England Biolabs
1 kb DNA ladder	New England Biolabs
Agilent 2100 Bioanalyzer kits	Agilent DNA microfluidic chips Agilent DNA 1000 Reagents Agilent DNA 12000 Reagents
cDNA synthesis	Invitrogen SuperScript III First-Strand Kit
DNA purification kit	Qiagen QIAquick Gel Extraction Kit Sigma GenElute PCR Clean-Up Kit
ECL detection reagents	Bio-Rad Clarity™ Western ECL Substrate
Fluorometer reagents	Thermo Fischer Qubit dsDNA HS Assay Kit
In-vitro Transcription Kit	Ambion MEGAscript T7 Transcription Kit
	Ambion MEGAshortscript T7 Transcription Kit
Library preparation	NEBNext DNA Library Prep Diagenode MicroPlex Library Preparation Kit
Phusion HF DNA Polymerase	New England Biolabs
Protein marker	Peqlab peqGOLD Prestained Protein-Marker V
qPCR Reagents	Applied Biosystems Fast SYBR Green Master Mix
Ribonuclease A	Sigma, R4875
RNA isolation kit	QIAGEN RNeasy Mini Kit
Taq DNA Polymerase	New England Biolabs

2.1.6 Chemicals and consumables

Table 10. List of chemicals and consumables used in this work.

Description	Supplier
384 Well Lightcycler Plate	Sarstedt
384 Well Lightcycler Plate Sealing Tape, optical clear	Sarstedt
AFA Tubes (Tubes for Covaris S220 instrument)	Covaris S-Series Tube & Cap 12 x 24 mm
Agarose	Bio&Sell
AMPure XP	Agencourt
Aprotinin	Genaxxon bioscience
Bradford reagents	Bio-Rad Protein Assay
Bromophenol blue	Merck
Coomassie Brilliant Blue G250	Serva
Coverslip, Poly-L-Lysine coated	Sigma
DAPI	Life Technologies
DTT (1,4 Dithiothreitol)	Roth
DMSO (Dimethyl sulfoxide)	Sigma
DMP (Dimethyl pimelimidate)	Sigma
EDTA	VWR Prolabo
EGTA	Roth
Ethidium bromide	Merck
Falcon tubes 15 mL/ 50 mL	Sarstedt
Fetal calf serum	Sigma
Filter papers	Whatman

Folded filters	Schleicher & Schuell
Formaldehyde, methanol-stabilized	Sigma, 37% (w/v) solution
Formaldehyde, methanol-free	Thermo Scientific, 16% (w/v) solution
Glycerol	VWR
Glycine	Sigma
HEPES	Serva
HCl	VWR
Image-iT FX signal enhancer	Invitrogen
KH_2PO_4	Merck
KOH	Merck
KCl	Sigma
Leupeptin	Genaxxon bioscience
LiCl	VWR
Methanol	VWR
MG-132 (proteasome inhibitor)	Enzo Life Sciences
$MgCl_2$	VWR
Midori Green Direct	NIPPON Genetics
Na-azide (sodium azide)	Merck
Na-borate (sodium borate)	Sigma
Na_2HPO_4	Merck
NaCl	Serva
Na-DOC (Sodium deoxycholate)	Sigma
NaOH	Merck
NEM (N-ethylmaleimide, deubiquitinase inhibitor)	Thermo Fisher Scientific
Nitrocellulose membrane	Whatman Protan membrane BA85
Non-fat dry milk	Heirler
Normal goat serum	Dianova
NP-40	Perbio Science
Penicilin-Streptomycin	Gibco
Pepstatin	Genaxxon bioscience
Pipette, serological, 5 mL/ 10 mL/ 25 mL	Sarstedt
Pipettor filter tips	Biozym, Gilson
Pipettor tips	Brand, Sarstedt
PMSF (phenylmethanesulfonylfluoride)	Sigma
Protein A/ Protein G Sepharose beads	GE Healthcare
Roller bottles	Greiner Bio-One
Schneider's *Drosophila* Medium	Invitrogen
SDS (Sodium dodecyl sulfate)	Serva
SDS-PAGE gels	Expedeon RunBlue Precast Gels Serva SERVA*Gel* Precast Gels
Sealing foil	Brand PARAFILM
T75/ T175 Flasks	Greiner Bio-One
Tubes 1.5 mL	Greiner, Sarstedt
Tubes 1.5 mL, DNase-, Rnase free	Biozym
Tris (Tris(hydroxymethyl)-aminomethan)	Biozym
Triton X-100	Sigma
Tween 20	Sigma
Vectashield	Vector Labs
X-ray films	Fujifilm

2.1.7 Technical devices

Table 11. List of technical devices used in this work.

Description	Supplier/Model
-20 °C Freezer	Miele, Liebherr
-80 °C Freezer	GFL
4 °C Refrigerator	Liebherr
25 °C Incubator, roller bottles (cell culture)	Bellco-Tecnomara
26 °C Incubator (cell culture)	LMS
Bioanalyzer (platform for DNA analysis)	Agilent 2100 Bioanalyzer
Cell Counter	Casy cell counter, OMNI Life Science LUNA-II cell counter, Logos Biosystems
Centrifuges	Eppendorf Centrifuges 5424/ 5417C/ 5430R Hettich Rotina 46 Sigma 3-18 Thermo/Heraeus Pico 17
Chambers for running protein-gels	Invitrogen Novex Mini-Cell
Developer machine	Agfa Curix 60
Dounce homogenizer	B. Braun S fit pestle
Fluorometer	Thermo Fischer Qubit 3.0 Fluorometer
Hemocytometer	Hausser Scientific Bright-Line
Ice machine	Ziegra
Imaging system	Bio-Rad ChemiDoc Touch
Incubation shaker (Multitron)	Infors
Laminar-flow hood	CleanAir
Magnetic stirrer	Bachofer Ika-Combimag Reo
Microscope	Leica DMIL LED
Microwave	Daewoo
pH-meter	inoLab pH 720
Pipetboy	Brand accu-jet pro
Pipettors	Gilson
Power supply (run of protein-gels and blotting)	Bio-Rad PowerPac Basic
Precision scales	Kern ALS 120-4N Mettler Toledo
Quantitative Real-Time PCR instrument	Roche LightCycler 480 II
Rotator	Stuart SB3
Rolling station	IDL TRM-V
Scales	Kern PCB Sartorius Extend
Scanner	Epson Perfection V700 Photo
Shaker	Bachofer Vortex Genie neoLab DOS-10L
Sonicator	Covaris S220 Focused-ultrasonicator
Spectrophotometer	Biozym NanoDrop DeNovix DS-11 Peqlab NanoDrop ND-1000
Thermomixer	Eppendorf ThermoStat plus/ comfort
Vacuum concentrator	LaboGene MiniVac Systems
Water bath	B. Braun Thermomix 1420
Water filtering machine	Elga Purelab Ultra
Western transfer chambers	Bio-Rad Mini Trans-Blot Cell

2.1.8 Software

Table 12. List of software applications used in this work.

Device/Application	Software
Genome-wide binding profile visualization	BioViz Integrated Genome Browser
Genome-wide data analysis	R studio and others, see section 2.2.9 Data analysis
Image processing	Adobe Photoshop Adobe Illustrator Bio-Rad Image Lab ImageJ
Office and general data analysis	Microsoft Word Microsoft Excel Microsoft PowerPoint
Primer design	Primer3 (Rozen and Skaletsky, 2000) UCSC *in silico* PCR (Jim Kent)
Quantitative DNA fragment size analysis	Agilent 2100 Expert software
qPCR	Roche LightCycler 480 SW 1.5
Sequence alignments	FlyBase Basic Local Alignment Search Tool Softonic Serial Cloner 2-6-1

2.1.9 Datasets

Table 13. List of publically available genome-wide datasets used in this work.

Type	Target, sample	System	Access	Publication
ChIP-chip	BEAF-32	*D.mel* S2 cells	GEO GSE32815	(Riddle et al., 2011)
ChIP-seq	CP190	*D.mel* S2 cells	GEO GSE32815	(Ong et al., 2013)
ChIP-seq	CTCF	*D.mel* S2 cells	GEO GSE32815	(Ong et al., 2013)
ChIP-seq	H3K27me3	*D.mel* S2 cells	GEO GSE27111	(Negre et al., 2011)
ChIP-seq	Mod(mdg4)	*D.mel* S2 cells	GEO GSE32815	(Ong et al., 2013)
ChIP-seq	Su(Hw)	*D.mel* S2 cells	GEO GSE32815	(Ong et al., 2013)
RNA-chip	untreated	*D.mel* S2 cells	GEO GSE46020	(Rus et al., 2013)
RNA-chip	BEAF-32 mutant	*D.mel* larvae wing imaginal disc	GEO GSE36736	(Gurudatta et al., 2012)
RNA-seq	BEAF-32 knockdown	*D.mel* S2 cells	GEO GSE57168	(Lhoumaud et al., 2014)
RNA-seq	HMR mutant	*D.mel* larvae	BioProject PRJNA236022	(Satyaki et al., 2014)
RNA-seq	HMR mutant	*D.mel* ovaries	BioProject PRJNA236022	(Satyaki et al., 2014)

2.2 Methods

2.2.1 Cell culture and RNAi

Cell culture and passaging

D. melanogaster S2-DRSC cells were grown at 26 °C in Schneider's *Drosophila* medium supplemented with 10 % fetal calf serum and antibiotics (100 units/mL penicillin and 100 µg/mL streptomycin) using 25 cm^2, 75 cm^2 or 175 cm^2 corning flask with 5 mL, 15 mL or 36 mL medium volume respectively. Cells were kept at a density of 0.5-7x10^6 cells/mL.

Cell counting and harvesting

Cell density was determined using a Hemocytometer, Casy cell counter or LUNA-II cell counter following suppliers' instructions. For harvesting, cells were resuspended, transferred to falcon tube and spun down at 160 g for 5 min.

Cell long-term storage

For long-term storage of cells, cells were harvested in logarithmic growth phase and resuspended in freezing solution (50 % fetal calf serum, 40 % Schneider's *Drosophila* medium, 10 % DMSO). To freeze slowly, 1 mL resuspension (containing approx. 5x10^7 cells) was placed in cryovial in freezing rack with isopropanol over night at -80 °C and transferred to liquid nitrogen for long-term storage afterwards. To thaw quickly, cells were thawed by resuspending in 15 mL supplemented Schneider's *Drosophila* medium and seeded in 75 cm^2 corning flask. After cells settled down, medium was renewed and cells were cultured as described above.

Gene knockdown by RNAi

For RNAi experiments, cells were seeded to a density of 1x10^6 cells/mL one day before dsRNA treatment. For dsRNA treatment, medium was removed and replaced with serum-free medium containing 10 µg/mL dsRNA (e.g. 175 cm^2 corning flask: 120 µg dsRNA in 12 mL medium). After 1 hr of incubation, serum-containing medium was supplied (e.g. 175 cm^2 corning flask: 24 mL supplemented medium). Cells were split 1:1 after three days and were harvested seven days after dsRNA treatment (**Figure 9**).

The dsRNA was prepared using the MEGAScript T7 Transcription Kit according to manufacturer's instructions. DsRNA was synthesized in 40 µL total volume from

300 ng gel-purified *in vitro* transcription template. *In vitro* transcription was performed overnight at 37 °C, followed by 15 min DNase treatment and products isolated by LiCl precipitation. Primers for generating *in vitro* transcription template are listed in **Table 7** and were verified for dsRNA sequence specificity by using E-RNAi (http://www.dkfz.de/signaling/e-rnai3/) (Horn and Boutros, 2010). Knockdown efficiency was tested on the protein level by Western Blot analysis on whole cell lysate and partially tested on the transcript level by reverse transcription followed by quantitative real-time PCR.

2.2.2 Genome editing using CRISPR/Cas9

The clustered, regularly interspaced, short palindromic repeats (CRISPR)/Cas9 system from *Streptococcus pyogenes* has been used for sequence-specific targeting of the Cas9 nuclease. Cas9 induces DNA double strand breaks at genomic sites, which are defined by complementary sequence-specific guide RNA (gRNA). Providing PCR-based homologous recombination donors and suppressing the alternative non-homologous end joining (NHEJ) DNA repair pathway, enables the genomic introduction of protein tags and resistance cassette (**Figure 6 A**) or gene knockout by substituting the gene body with a resistance cassette (**Figure 6 B**).

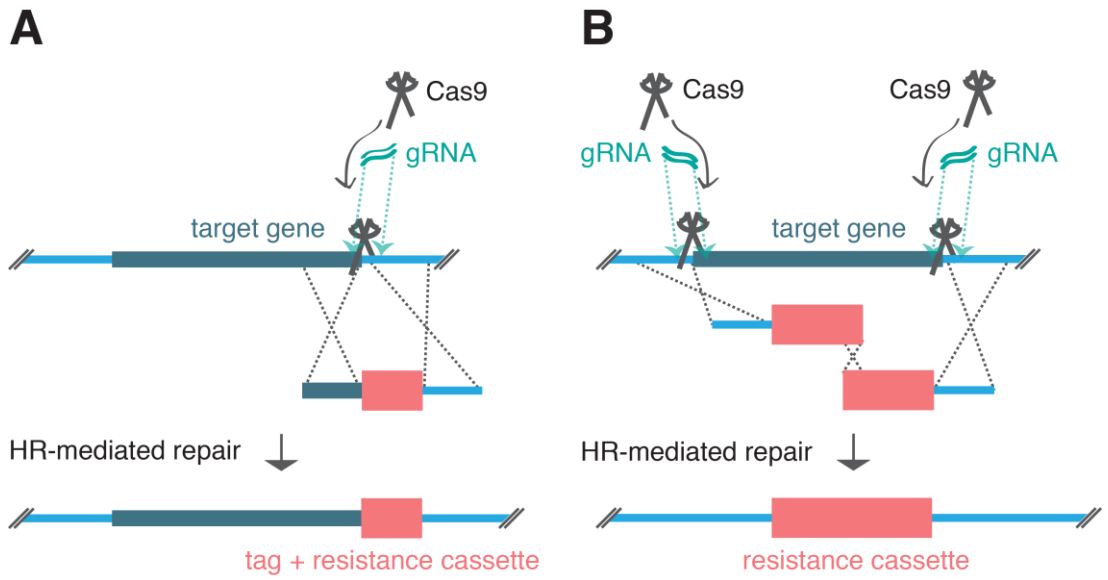

Figure 6. Experimental strategy for endogenous epitope tagging and gene knockout using CRISPR/Cas9 system. (A) The Cas9 nuclease is directed to genomic sites by sequence-specific guide RNA (gRNA) and induces DNA double strand break at the gene body 3' end. The DNA defect is repaired by homologous recombination (HR). By providing an artificial HR donor, the epitope tag and resistance cassette are integrated. The functional resistance cassette allows for later positive selection (see also **Figure 7**). **(B)** DNA double strand breaks are induced at the gene body 5' end and 3' end. By providing two artificial homologous recombination donors that complement for a functional resistance cassette, the gene is replaced by a resistance cassette that allows for later positive selection.

Gene knockout strategies and endogenous tagging of *Hmr* in S2 cells were performed in collaboration with Prof. Dr. Klaus Förstemann and were performed exactly as described in (Bottcher et al., 2014; Kunzelmann et al., 2016). For endogenous tagging, we provided gRNA by *in vitro* transcription or by transfecting U6-gRNA DNA template (**Figure 7**, **Figure 11**). For gene knockout, we provided gRNA by transfecting U6-gRNA DNA template and aimed to promote HR-mediated repair by *mus308* and *lig4* RNAi and performed optional double selection (blasticidin and puromycin or blasticidin only). An experimental setup overview and reagents are shown in **Figure 7**, **Figure 8** and **Table 4**. The cell lines generated in this way are listed in **Table 2**.

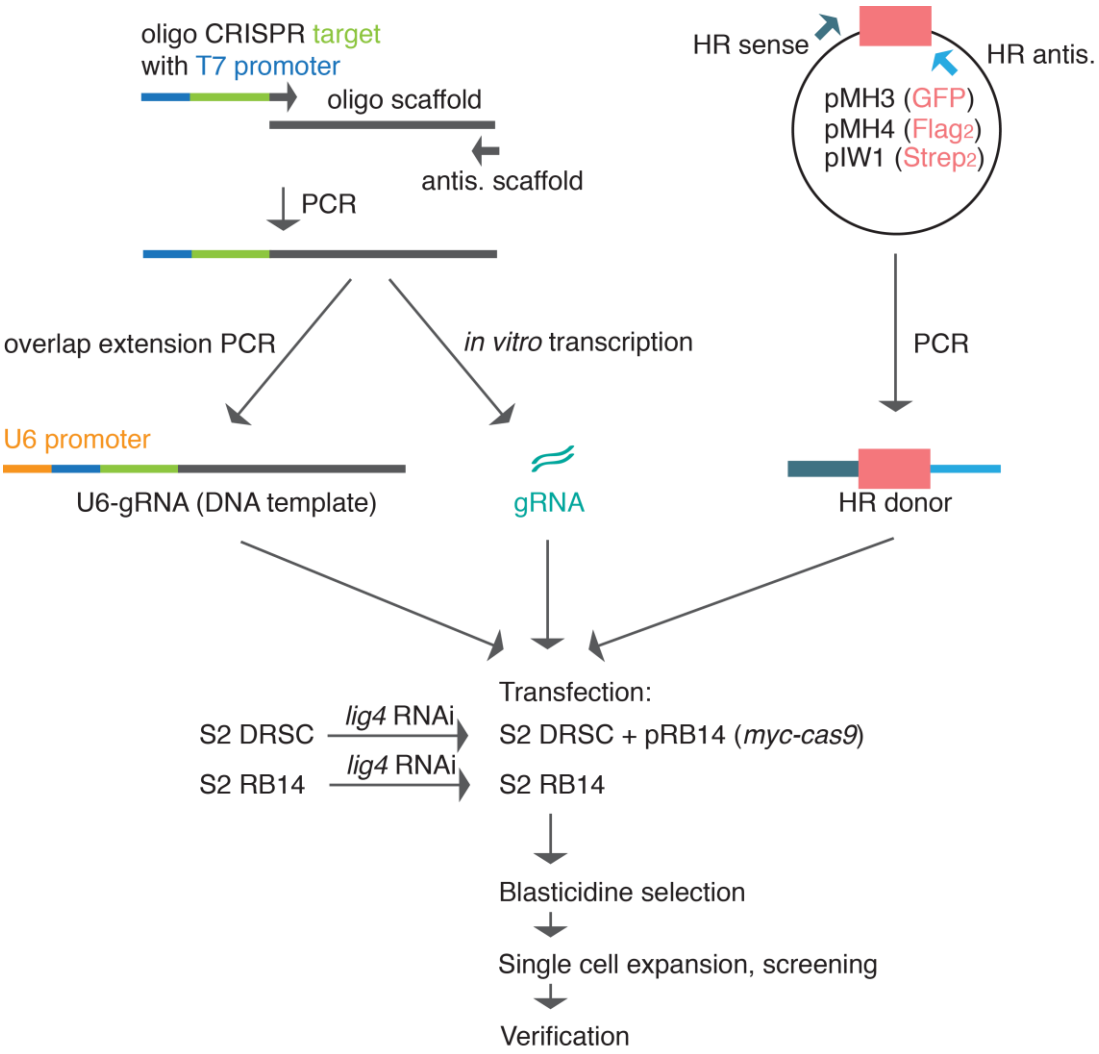

Figure 7. Reagents and workflow for endogenous epitope tagging using CRISPR/Cas9 system. (Top left and center) A crucial step in CRISPR/Cas9-mediated genome editing is the sequence-specific targeting of the nuclease enzyme Cas9 by site-specific guide RNA (gRNA). gRNA was either transfected after *in vitro* transcription or was transfected as DNA template for later *in vivo* host transcription. For T7 promoter driven *in vitro* transcription, template DNA was generated by PCR using site-specific primer pairs. For host *in vivo* transcription, the gRNA sequence is under control of a host U6 promoter fused to the initial PCR product by overlap extension PCR. **(Top right)** Artificial homologous recombination donor fragments coding for GFP, $Flag_2$ or $Strep_2$ and a resistance cassette were generated by PCR using site-specific primer pairs. Prior to transfection, the NHEJ pathway, an alternative mechanism to homologous recombination repair, was suppressed by knockdown of *lig4*. Workflow according to (Bottcher et al., 2014).

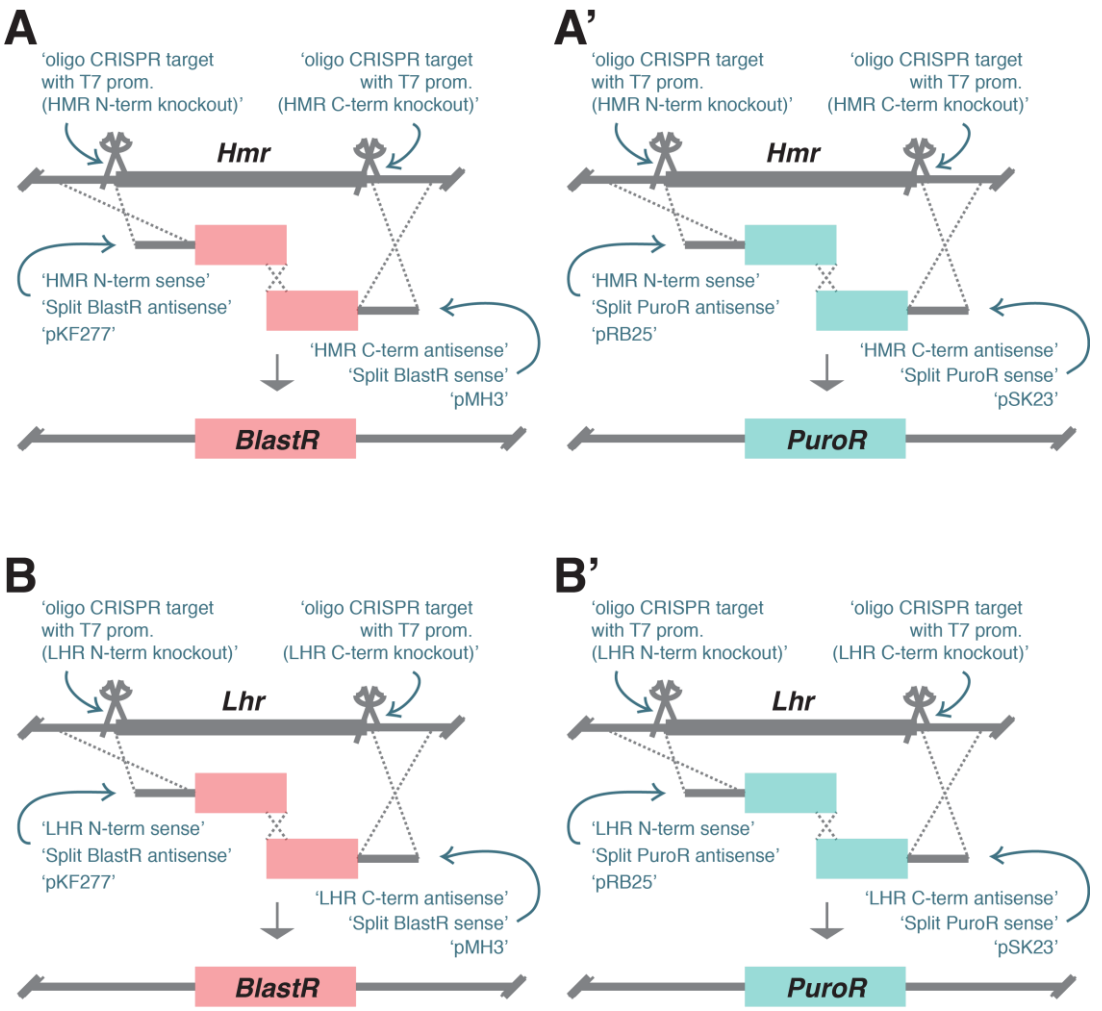

Figure 8. Reagents for gene knockout using CRISPR/Cas9 system. (A and A') Reagents for *Hmr* knockout. **(B and B')** Reagents for *Lhr* knockout. The gene body is cleaved at 5' end and 3' end. Homologous recombination donors complement for a functional resistance cassette, blasticidin resistance gene (*BlastR*) and puromycin resistance gene (*PuroR*) that allows for later positive selection. gRNA are provided as U6-gRNA DNA template (see **Figure 7**). Workflow according to (Bottcher et al., 2014) with additional knockdown of *mus308* and optional double selection (blasticidin and puromycin or blasticidin only).

2.2.3 Protein methods

Cell lysis

Cells were harvested and resuspended in RIPA (140 mM NaCl, 10 mM Tris HCl pH 7.3, 1 mM EDTA pH 8.0, 1 % Triton X100, 0.1 % SDS, 0.1 % Na-DOC), typically $1x10^6$ cells in 200 µL RIPA, supplemented with protease inhibitors and benzonase, vortexed and kept on ice for about 30 min. The lysate was boiled in Laemmli and subjected directly to further analysis or stored at -20 °C.

Nuclear extract preparation

Following steps were all performed on ice, cold room and using ice-cold reagents. All centrifugation steps were performed in 4 °C cooled centrifuges. Cells were harvested and washed twice with PBS (2.7 mM KCl, 136 mM NaCl, 4 mM Na_2HPO_4, 1.7 mM KH_2PO_4). The pure cell volume (PCV) was estimated and cells resuspended in 3 volumes PCV of hypotonic Buffer A (10 mM HEPES pH 7.6, 15 mM KCl, 2 mM $MgCl_2$, 0.1 mM EDTA pH 8.0). Cells were incubated for 30 min for swelling and homogenized using a Dounce homogenizer. Afterwards, nuclei were stabilized by adding 1/10 volume Buffer B (50mM HEPES pH 7.6, 1 M KCl, 30 mM $MgCl_2$, 0.1 mM EDTA pH 8.0) and spun down for 25 min at 8500 g. After removal of supernatant, the nuclear pellet volume (NPV) was estimated and nuclear pellet resuspended in 3 NPV of Buffer A : Buffer B mix (volume ratio 9 : 1). Nuclear proteins were isolated by adding 1/10 NPV of 4 M $(NH_4)_2SO_4$ pH 7.9 and incubation on rotating wheel for 25 min. To remove unsoluble cellular debris, sample was ultracentrifuged for 90 min at 100000 g. Supernatant was recovered and determined for volume. Proteins were precipitated from supernatant by adding 0.3 g of solid $(NH_4)_2SO_4$ per 1 mL supernatant volume under continuous stirring. After 15 min of incubation time, precipitate was spun down for 35 min at 15000g, dissolved in 0.2 NPV of Buffer C (25 mM HEPES pH 7.6, 150 mM KCl, 12.5 mM MgCl2, 0.1 mM EDTA pH 8.0, 10 % glycerol) and dialyzed against three times against 1 L of Buffer C (supplemented with protease inhibitors) for 80 min each. Remaining precipitates were spun down for 15 min at 14000 g. The supernatant was subjected for Bradford assay and quickly frozen in liquid nitrogen and stored at -80 °C.

Bradford assay

Protein quantification was performed with Bradford assay (Bradford, 1976) using Bio-Rad Protein Assay according to manufacturer's instructions.

Affinity Purification

Affinity Purifications were essentially performed as described for immunoprecipitation reaction in ChIP. Instead of using sheared chromatin, nuclear extract was used as starting material and instead of RIPA Buffer, BC150 Buffer (25 mM HEPES pH 7.6, 150 mM KCl, 12.5 mM $MgCl_2$, 0.1 mM EDTA pH 8.0) was used as immunoprecipitation buffer and washing buffer. Nuclear extract was diluted in BC 150 Buffer and subjected to immunoprecipitation. 400 µg of diluted nuclear extract were used per immunoprecipitation reaction.

After immunoprecipitation, beads, input and supernatants (optional) were boiled in Laemmli buffer and subjected directly to further analysis or stored at -20 °C.

SDS-Polyacrylamid-Gelelectrophoresis (SDS-PAGE)

SDS-PAGE was used to separate proteins according to their electrophoretic mobility (Laemmli, 1970). The samples were boiled in Laemmli (5xLaemmli: 250 mM Tris HCl pH 6.8, 10 % SDS, 500 mM DTT, 25 % Glycerol, 0.5 % Bromophenol blue) at 95 °C for 10 min. Optionally, the heating time was increased: 65 °C for 25 min (first step) and 95 °C for 5 min (second step) to ensure complete reduction of antibodies. For work with crosslinked material it was heated 65 °C for 2 hrs (first step) and 95 °C for 10 min (second step) to reverse crosslinks.

Gels were purchased from Expedeon (RunBlue Precast Gels) and Serva (SERVAGel Precast Gels) and protocol performed as described in manufacturer's instructions.

All samples were spun down at maximal speed for 10 min and supernatant was used for analysis. Electrophoresis was typically performed at 180 V. Protein standards were purchased from Peqlab and used according to manufacturer's instructions. For visualization after separation, gels were either subjected to Coomassie-staining or Western Blot analysis.

Coomassie-staining

Coomassie-staining was used to visualize proteins on SDS-PAGE gels and was performed as described in (Bramhall et al., 1969).

Western Blot analysis

Western Blot analysis was used to specifically visualize proteins upon SDS-PAGE by antibody-based detection. The following protocol is based on (Towbin and Gordon, 1984). After SDS-PAGE, the gel and other components were assembled in the

transfer apparatus in the following order: case (black side), sponge, two whatman papers, SDS-PAGE gel, nitrocellulose membrane, two whatman papers, sponge, case (clear side). The components were soaked in Western Blot Buffer (20 mM Tris HCl, 192 mM glycine, 20 % methanol) before. The transfer was performed at constant voltage of 50 V for 4 hrs (max 150 mA). After transfer, the membrane was washed with PBS and blocked for 30 min (up to overnight) in 5 % non-fat dry milk dissolved in PBS. The membrane was incubated with primary antibody in 1 % non-fat dry milk dissolved in PBS-Tween 20 (PBS-T; final concentration of Tween 20: 0.02 - 0.1 %) at 4 °C overnight. The membrane was washed twice for 10 min in PBS-T and incubated with secondary antibody for 1 hr at room temperature. Afterwards, membrane was washed twice in PBS-T for 10 min each and subjected to chemiluminescence detection using Bio-Rad Clarity™ Western ECL Substrate according to manufacturer's instructions. Chemiluminescence signals were detected by exposing to X-ray films or by using Bio-Rad ChemiDoc Touch imaging system.

2.2.4 Nucleic Acids methods

Polymerase Chain Reaction

Polymerase Chain Reaction (PCR) (Mullis et al., 1986) was used for DNA amplification, generation of *in vitro* transcription templates and other primer-sequence-directed DNA products (e.g. overlap extension PCR for CRISPR/Cas9 strategy). If not stated otherwise, *Taq* DNA Polymerase or Phusion HF DNA Polymerase and other reaction components from New England Biolabs were used according to manufacturer's instructions.

PCR products were analyzed with agarose gel electrophoresis, restriction digest (New England Biolabs reagents and protocols) or sequenced using GATC or MWG sequencing service. For DNA purification from PCR reaction, Sigma GenElute PCR Clean-Up Kit was used according to manufacturer's instructions.

RNA-isolation and cDNA synthesis

For transcription analysis, cDNA was prepared and subjected to qPCR. Total RNA was isolated from 5-6x10^6 cells using Qiagen RNeasy Mini Kit according to manufacturer's instructions and eluted in 30 µL. 500-1000 ng of total RNA were treated with DNase at 37 °C for 1 hr and cDNAs were synthesized by using SuperScript III First-Strand Kit with random hexamers primers according to

manufacturer's instructions. Upon synthesis, cDNA was treated with RNAse H at 37 °C for 20 min.

Preparative and analytical agarose gel electrophoresis

Agarose gel electrophoresis was performed to analyze and separate DNA fragments according to their size (Aaij and Borst, 1972). 1-2 % agarose gel was prepared by dissolving agarose in TBE buffer (90 mM Tris, 90 mM Boric acid, 2 mM EDTA) by boiling in the microwave. Loading dye and DNA ladder from New England Biolabs were used according to manufacturer's instructions. For visualization of DNA under UV light (254 nm), either the agarose gel was supplied with ethidiumbromide (1 µg/ml) or the DNA sample was supplied with Midori Green Direct according to manufacturer's instructions. Electrophoresis was performed with up to 10 V/cm gel length.

For DNA purification from gel, DNA band was excised from the gel with scalpel and purified using Qiagen QIAquick Gel Extraction Kit according to manufacturer's instructions.

2.2.5 Immunofluorescence imaging

Immunolocalization

Immunofluorescence analysis was performed to characterize HMR's and LHR's localization manner in CRISPR/Cas9 system generated *Schneider* S2 cell lines expressing endogenously epitope-tagged HMR. Further, we used immunofluorescence to investigate localization changes of HMR, LHR and CTCF after RNAi knockdown treatment in S2 cells.

For analysis, cells were subjected on a Poly-L-Lysine coated coverslip and settled for 20 min in a humidified chamber (150 µL of cell suspension with approx. 5-6x10^6 cells/mL per coverslip). Coverslips were rinsed in a 12-well plate for 5 min with PBS and cells fixed for 12 min at room temperature in PBS containing 3.7 % formaldehyde and 0.3 % Triton X-100. After fixation, cells were immediately washed with PBS twice to remove residual formaldehyde before permeabilization. Cells were permeabilized for 6 min on ice with ice cold PBS containing 0.25 % Triton X-100 and washed with PBS twice afterwards. Then, coverslips were transferred to humidified chamber. Permeabilized material was blocked for 45 min at room temperature with Image-iT FX signal enhancer (100 µL per coverslip). Primary antibodies were diluted in PBS containing 5 % normal goat serum, spun down at max. speed for 5 min and

supernatant added to coverslip (100 µL per coverslip) and incubated for 2 hr at room temperature. Coverslips were washed in a 12-well plate for 5 min with PBS containing 0.1 % Triton X-100 and with PBS for 5 min afterwards. Fluorophore coupled-secondary antibodies were diluted in PBS containing 5 % normal goat serum, spun down at max. speed for 5 min and supernatant added to coverslip (100 µL per coverslip) and incubated for 1 hr at room temperature. Coverslips were rinsed in a 12-well plate twice with PBS containing 0.1 % Triton X-100 and were washed with PBS for 5 min afterwards. After washing, coverslips were transferred to humidified chamber and incubated for 6 min with PBS containing 200 ng/mL DAPI, washed twice in a 12-well plate with PBS, mounted in Vectashield (4.5 µL per coverslip) and sealed with nail polish. Afterwards, coverslips could be stored at 4 °C protected from light.

Confocal microscopy imaging

Images were acquired using a Leica TCS SP5 confocal microscope with a 63x glycerol immersion objective (NA = 1.3). Z stacks were deconvolved using the Huygens Essential Software (SVI). Other image processing steps were performed with ImageJ.

2.2.6 Chromatin immunoprecipitation (ChIP)

Chromatin immunoprecipitation (ChIP) experiments were performed to analyze the genomic localization of HMR in S2 cells. The experiment requires chromatin crosslinking, isolation and preparation steps, immunoprecipitation, DNA isolation and DNA analysis. The basic ChIP procedure is outlined in **Figure 9 A**. Within this work, multiple steps in the protocol were adjusted and optimized. The protocol is based on the modENCODE ChIP-seq protocol, provided by the lab of Prof. Dr. Gary Karpen (Lawrence Berkeley National Laboratory, CA) (Landt et al., 2012).

If not stated otherwise, chromatin was crosslinked and isolated from $200x10^6$ cells in 50 mL tubes scale and aliquoted for long-term storage and further processing in 1.5 mL tubes scale (chromatin isolated from $50x10^6$ cells per tube). Typically, chromatin prepared from $15\text{-}40x10^6$ cells was used per immunoprecipitation reaction. For nuclear lysis and following steps, all buffers were supplemented with inhibitors Aprotinin, Pepstatin, Leupeptin, PMSF and MG-132.

Fixation, lysis and chromatin preparation

Cells were harvested in logarithmic growth phase, washed once in PBS (2.7 mM KCl, 136 mM NaCl, 4 mM Na_2HPO_4, 1.7 mM KH_2PO_4, tempered, 25 °C), resuspended in PBS (tempered, 25 °C) and crosslinked with 1 % formaldehyde (methanol-stabilized, 37 % stock solution) for 5 min at 25 °C. Crosslinking reaction was stopped with 12 % glycine (1.5 M stock solution), mixing and incubating on ice for 5 min. Following steps were all performed on ice, cold room and using ice-cold reagents. All centrifugation steps were performed in 4 °C cooled centrifuges. The cells were washed once in PBS and lysed for 10 min in Buffer A (10 mM HEPES pH 7.6, 10 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.25 % Trition X-100). The crosslinked chromatin was spun down for 2 min at 1500 g and resuspended in Buffer B (10 mM HEPES pH 7.6, 100 mM NaCl, 1 mM EDTA pH 8.0, 0.5 mM EGTA pH 8.0, 0.01 % Triton X-100). Chromatin was aliquoted in 1.5 mL reaction tubes, spun down for 2 min at 1500 g and the chromatin pellet frozen in liquid nitrogen and stored at -80 °C.

The crosslinked chromatin obtained from previous step was thawed at 4 °C for 10 min and washed with 1 mL TE Buffer (10 mM Tris-Cl pH 7.3, 1 mM EDTA pH 8.0) to remove residual Buffer B. Chromatin was spun down for 2 min at 1500 g and resuspended in 1 mL TE Buffer. Suspension was transferred to an AFA Tube (Covaris S-Series Tube & Cap 12 x 24 mm) and SDS stock solution was added to the desired final concentration of 0.1 %.

For chromatin AFA shearing, Covaris S220 Focused-ultrasonicator device was used with default settings except for the following parameters:

Duty cycle	5 %
Peak incident power	140 W
Cycles per burst	200
Processing time	flexible, range from 5 min to 60 min
Water bath temperature	4 °C
Water level (RUN)	level 10

After shearing, the buffer composition was adjusted to 1 % Triton X-100, 0.1 % Na-DOC and 140 mM NaCl using stock solutions. PMSF and MG-132 was supplemented and rotated for 10 min. The samples were centrifuged for 20 min at maximal speed to pellet insoluble material. The supernatant was transferred into a fresh tube with

equilibrated Protein A/G Sepharose beads (30 µL beads per 500 µL supernatant) and incubated for 1 hr on rotating wheel. After pre-clearing, the samples were centrifuged for 20 min at maximal speed and supernatant subjected to immunoprecipitation reactions. 100 µL of supernatant (soluble sheared chromatin prepared from $2.5x10^6$ cells) was kept as input sample.

Immunoprecipitation from sheared chromatin

Immunoprecipitation was performed using rat, mouse and rabbit antibodies non-covalently coupled to Protein A/G Sepharose as solid phase. In immunoprecipitations with rat antibodies, a bridging antibody from rabbit, anti-Rat IgG (Dianova, 312-005-046), was used. Antibodies from other species were directly coupled to beads. 30 µL equilibrated beads were used per immunoprecipitation reaction. Prior to immunoprecipitation, antibodies were incubated with beads overnight on rotating wheel – if required anti-Rat IgG was coupled to beads beforehand for 1 hr at room temperature – and surplus antibody removed by washing once with RIPA buffer afterwards.

300-800 µL of soluble sheared chromatin (prepared from $15\text{-}40x10^6$ cells) were incubated overnight with the antibody-coupled beads constantly rotating at 4 °C. The beads were spun down for 1 min at 500 g and washed 5 times with 500 µL RIPA Buffer for 5 min constantly rotating. In last washing step, the beads and associated chromatin were transferred to fresh 1.5 mL tubes. After washing steps, the samples were processed as described in the following section.

DNA isolation

Beads with bound chromatin fragments were spun down, and resuspended in 100 µL RIPA Buffer. Input and ChIP samples were treated with RNase A (final concentration of 0.2 mg/mL) for 30 min at 37 °C. Afterwards, SDS stock solution was added to 0.5 % final concentration and samples treated with Proteinase K (final concentration of 1 mg/mL) for 2 hrs at 56 °C for protein digestion. Crosslinks were reversed by overnight incubation at 65 °C.

The DNA fragments were purified using the Sigma GenElute PCR Clean-Up Kit according to the manufacturer's instructions. The purified DNA was eluted in 50 µL ddH_2O or elution buffer and DNA concentration measured with NanoDrop Spectrophotometer (only suited for input sample) or Qubit Fluorometer (suited for input and ChIP sample) according to manufacturer's instructions. Quantitative DNA

fragment size analysis (without PCR-based amplification only suited for input sample) was performed using the Agilent 2100 Bioanalyzer with the Agilent DNA 1000 and DNA 12000 Kits and the Agilent 2100 Expert software according to the manufacturer's instructions.

For sequence enrichment analysis, DNA was subjected to qPCR (typically input DNA diluted 1:1000, ChIP DNA diluted 1:9 used) or library preparation and sequencing (typically 5-20 ng used).

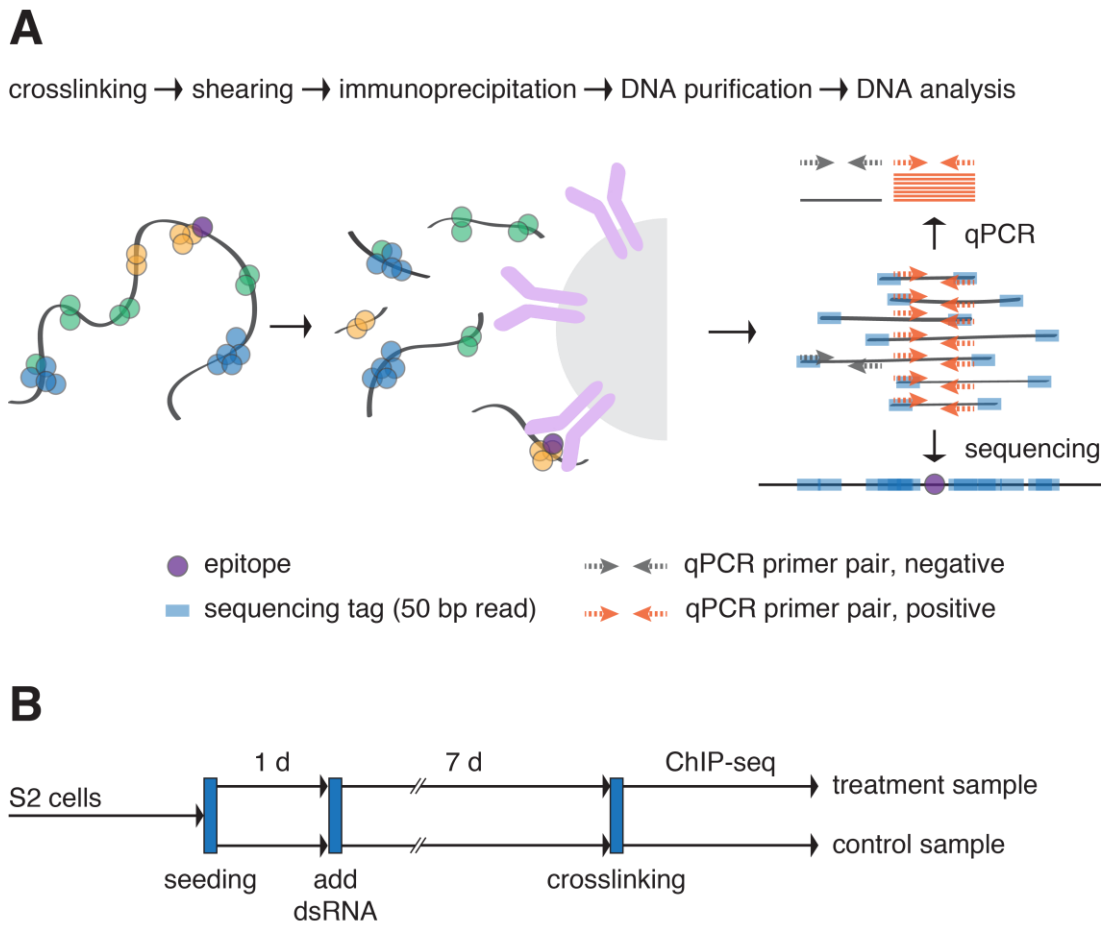

Figure 9. Experimental approach for ChIP combined with RNAi knockdown (A) ChIP experiment overview. Chromatin ImmunoPrecipitation (ChIP) can be used to identify protein-binding sites or to assay the enrichment and enrichment-changes at known binding sites. After crosslinking of intact cells, chromatin is isolated and sheared into small fragments of approximately 200 bp in length. The material is subjected to an IP reaction that enriches for chromatin fragments bound or crosslinked to the protein of interest (purple). After purification, DNA can be subjected to high-throughput sequencing or assayed by quantitative real-time PCR. **(B)** Experimental setup for RNAi followed by ChIP. *Drosophila* S2 cells were treated with dsRNA. RNAi lasted for 7 days and included treatment sample and a negative control sample with unspecific dsRNA. After crosslinking and ChIP-seq, the samples were compared to obtain RNAi-specific differences.

2.2.7 Quantitative real-time PCR

Quantitative real-time PCR (qPCR) was used for relative quantification of DNA obtained from ChIP reaction or from reverse transcription reaction. The following sections refer to ChIP but also apply to DNA obtained from reverse transcription reaction. Reverse transcription-specific differences are highlighted in the section *Transcription analysis using quantitative real-time PCR*.

Primer design

Oligonucleotides used for quantitative real-time PCR are listed in **Table 5** and **Table 6** and were designed by using Primer3 (http://biotools.umassmed.edu/bioapps/primer3_www.cgi) (Rozen and Skaletsky, 2000), with default settings except for the following parameters:

Product size:	Min: 90, Opt: 120, Max: 140
Primer Tm:	Min: 58, Opt: 60, Max: 61
Max Tm Difference:	1.0
Primer GC%:	Min: 50, Max: 60
Max Poly-X:	3
CG Clamp:	1

Primer sequences were verified for specificity using UCSC *in silico* PCR (https://genome.ucsc.edu/cgi-bin/hgPcr).

qPCR performance and analysis

Quantitative real-time PCR was performed using the Applied Biosystems Fast SYBR Green Master Mix, Sarstedt 384 Well Lightcycler Plates and the real-time PCR system Roche LightCycler 480 II according to the manufacturer's instructions. All reactions were performed in triplicates. Fluorescence from the amplification reaction is proportional to the amount of amplified product and allows the determination of the starting template amount. The cycle number at which the fluorescent signal is detected is called the threshold cycle number (Ct).

Reaction mixture:

5 µl	2x Applied Biosystems Fast SYBR Green Master Mix
1 µl	Primer forward, 3 mM
1 µl	Primer reverse, 3 mM
1 µl	ddH_2O
2 µl	DNA template, diluted

qPCR program:

Mode	Temperature	Duration	Heating Rate	Cycles
Polymerase Activation	95 °C	20 s	4.4 °C/s	1
Quantification[#]	95 °C	3 s	4.8 °C/s	45
	60 °C	30 s	2.2 °C/s	
Melting curves[Ω]	95 °C	-	0.11 °C/s	1

#. Single acquisition. Ω. Multiple Acquisitions (5 per 1 °C)

For each primer pair a standard curve was generated using a 10-fold serial dilution of ChIP input DNA. The starting quantity of template ranged from 5 ng to 0.005 ng. The Ct values were plotted against the log of the starting quantity for each dilution. The primer efficiency E for each primer pair is calculated from the slope of the standard curve:

$$E = 10^{-1/slope} \quad \textbf{(Equation 1)}$$

(with: *E = primer efficiency; slope = slope of the standard curve generated by using a 10-fold serial dilution of ChIP input DNA*)

Melting-curve analysis was performed in order to confirm that the primer pairs and reaction conditions only yield the specific PCR product. The presence of unspecific product as for example caused by primer-dimer formation, gives one or more additional melting curves. Calculation of primer efficiency as well as melting curve analysis was performed using Roche LightCycler 480 software.

Based on Ct values, it is possible to determine the DNA amount in any given sample. The ChIP input sample represents the total amount of DNA used in the ChIP, while the ChIP sample represents the amount of DNA specifically bound by the immunoprecipitated protein. The efficiency of ChIP is expressed as % input, which is calculated as the ratio of the amount of DNA obtained from the ChIP over that

measured in the ChIP input sample, see **Equation 2-4**. For calculation of the fold enrichment, the % input value of the target ChIP is divided by the % input value of the control ChIP, see **Equation 5**. In order to evaluate whether the binding of the protein of interest is specific to a given locus (locus-specific fold enrichment), the fold enrichment at the *target locus* in the target ChIP is divided by the fold enrichment at the negative control locus (*tub97EF*), see **Equation 6**. The analysis of qPCR data was performed using Microsoft Excel.

$$\Delta Ct = Ct_{ChIP\ input} - Ct_{ChIP} \quad \textbf{(Equation 2)}$$

$$DNA_{rel} = E^{\Delta Ct} \quad \textbf{(Equation 3)}$$

(with: *DNA_{rel} = relative concentration of DNA in sample; E = primer efficiency*)

$$\%input = DNA_{rel} \times \frac{d(ChIP)}{d(ChIP\ input)} \times \frac{v(ChIP\ input)}{v(ChIP)} \times 100 \quad \textbf{(Equation 4)}$$

(with: *%input = DNA amount in ChIP as percentage of ChIP input taking into account primer efficiency; qPCR template dilution factors and volume factors; DNA_{rel} = relative concentration of DNA in sample; d(ChIP) = dilution factor of ChIP sample; d(ChIP input) = dilution factor of ChIP input sample; v(ChIP input) = volume factor of ChIP input sample; v(ChIP) = volume factor of ChIP sample*)

$$fold\ enichment = \frac{\%input\ (target\ ChIP, target\ locus)}{\%input\ (control\ ChIP, target\ locus)} \quad \textbf{(Equation 5)}$$

(with: *%input = DNA amount in ChIP as percentage of ChIP input taking into account primer efficiency; qPCR template dilution factors and volume factors*)

$$specific\ fold\ enr. = \frac{fold\ enr.\ (target\ ChIP, target\ locus)}{fold\ enr.\ (target\ ChIP, control\ locus)} \quad \textbf{(Equation 6)}$$

Transcription analysis using quantitative real-time PCR

Based on Ct values it is possible to determine the change of transcript levels in a tester treatment relative to a calibrator treatment (control treatment, here GFP RNAi or GST RNAi) and to a reference gene (here *tub97EF*), see **Equation 7-9**. The analysis of qPCR data was performed using Microsoft Excel.

$$\Delta Ct_{normalized} = Ct_{target\ gene} - Ct_{reference\ gene} \quad \textbf{(Equation 7)}$$

$$\Delta\Delta Ct = \Delta Ct_{normalized,tester} - \Delta Ct_{normalized,calibrator} \quad \textbf{(Equation 8)}$$

$$Expression\ ratio = E^{\Delta\Delta Ct} \quad \textbf{(Equation 9)}$$

(with: *E = primer efficiency*)

2.2.8 ChIP DNA library preparation and next-generation sequencing

Next-generation sequencing was used for genome-wide analysis of DNA obtained from ChIP reactions. For sequencing, all libraries were prepared using MicroPlex (Diagenode) or NEBNext (NEB) Library Preparation kit according to manufacturer's instructions. In brief, DNA fragment ends were repaired, ligated to adaptor elements, amplified by PCR and purified for sequencing. If required, DNA was concentrated using vacuum concentrator. All libraries were prepared without size-selection. DNA was purified using Agencourt AMPure XP reagents, and concentration was measured with Qubit Fluorometer according to manufacturer's instructions.

Quantitative DNA fragment size analysis was performed using the Agilent 2100 Bioanalyzer with the Agilent DNA 1000 and DNA 12000 Kits and the Agilent 2100 Expert software according to the manufacturer's instructions.

Sequencing was performed in the group of Dr. Helmut Blum, the Genomics unit of LAFUGA at the LMU Gene Center with the help and advice of Dr. Stefan Krebs. Samples were sequenced single-end, 50 bp on Illumina HiSeq1500 sequencer. Samples description and the obtained read numbers are listed in **Table 14**.

Table 14. List of ChIP-seq samples and the obtained read numbers used in this work.

Cell line, treatment	Antibody	Sample name	ID	Index$^{\Omega}$	unique read number
S2, untreated	HMR 2C10	HMR_1	TG7	#3	7321851
S2, untreated	HP1a C1A9	HP1a_1	TG9	#5	9615590
S2, untreated	FLAG M2	IgG_1	TG6	#2	8941371
S2, untreated	LHR 12F4	LHR_1	TG8	#4	7931351
S2, untreated	Input	Input_1	TG5	#1	10427301
S2, untreated	HMR 2C10	HMR_2	TG45	#8	11790768
S2, untreated	HP1a C1A9	HP1a_2	TG47	#6	3793451
S2, untreated	FLAG M2	IgG_2	TG46	#7	17068323
S2, untreated	Input	Input_2	TG44	#1	14681351
S2, untreated	HMR 2C10	HMR_3	TG37	#2	11727510
S2, untreated	Input	Input_3	TG36	#1	14734701
S2, untreated	HP1a C1A9	HP1a_4	TG4	#10	13525876
S2, untreated	Input	Input_4	TG1	#7	21798351
S2, *Hmr*-Flag$_2$, untreated	HMR 2C10	HMR_5	TG12	#8	11786439
S2, *Hmr*-Flag$_2$, untreated	FLAG M2	FLAG_5	TG11	#7	16721688
S2, *Hmr*-Flag$_2$, untreated	Input	Input_5	TG10	#6	14215373
S2, Ctrl RNAi	HMR 2C10	HMR_CtrlRNAi_1	TG19	#7	8637377
S2, Ctrl RNAi	HP1a C1A9	HP1a_CtrlRNAi_1	TG16	#4	765717
S2, Ctrl RNAi	Histone H3	H3_CtrlRNAi_1	TG27	#10	13497422
S2, Ctrl RNAi	H3K9me3	H3K9me3_CtrlRNAi_1	TG33	#4	9610889
S2, Ctrl RNAi	Input	Input_CtrlRNAi_1	TG13	#1	14443029
S2, HMR RNAi	HMR 2C10	HMR_HMRRNAi_1	TG20	#8	11868648
S2, HMR RNAi	HP1a C1A9	HP1a_HMRRNAi_1	TG17	#5	7419588
S2, HMR RNAi	Histone H3	H3_HMRRNAi_1	TG28	#11	10562127
S2, HMR RNAi	H3K9me3	H3K9me3_HMRRNAi_1	TG34	#5	8553421
S2, HMR RNAi	Input	Input_HMRRNAi_1	TG14	#2	14445163
S2, CP190 RNAi	HMR 2C10	HMR_CP190RNAi_1	TG21	#9	10825131
S2, CP190 RNAi	HP1a C1A9	HP1a_CP190RNAi_1	TG18	#6	10031543
S2, CP190 RNAi	Histone H3	H3_CP190RNAi_1	TG29	#12	13085885
S2, CP190 RNAi	H3K9me3	H3K9me3_CP190RNAi_1	TG35	#6	11137815
S2, CP190 RNAi	Input	Input_CP190RNAi_1	TG15	#3	15526600
S2, Ctrl RNAi	HMR 2C10	HMR_CtrlRNAi_2	TG52	#9	14947403
S2, Ctrl RNAi	Input	Input_CtrlRNAi_2	TG48	#2	11772305
S2, HMR RNAi	HMR 2C10	HMR_HMRRNAi_2	TG53	#10	15528028
S2, HMR RNAi	Input	Input_HMRRNAi_2	TG49	#3	16101359
S2, CP190 RNAi	HMR 2C10	HMR_CP190RNAi_2	TG54	#11	15719248
S2, CP190 RNAi	Input	Input_CP190RNAi_2	TG50	#4	1024172
S2, CTCF RNAi	HMR 2C10	HMR_CTCFRNAi_2	TG55	#12	17217080
S2, CTCF RNAi	Input	Input_CTCFRNAi_2	TG51	#5	15926054
S2, Ctrl RNAi	Histone H3	H3_CtrlRNAi_2	TG59	#4	1231249
S2, Ctrl RNAi	Input	Input_CtrlRNAi_2	TG56	#1	10756706
S2, HMR RNAi	Histone H3	H3_HMRRNAi_2	TG60	#5	8934748
S2, HMR RNAi	Input	Input_HMRRNAi_2	TG57	#2	8693426
S2, CP190 RNAi	Histone H3	H3_CP190RNAi_2	TG61	#6	11286236
S2, CP190 RNAi	Input	Input_CP190RNAi_2	TG58	#3	12157132
S2, CTCF RNAi	Histone H3	H3_CTCFRNAi_2	TG63	#12	11511532
S2, CTCF RNAi	Input	Input_CTCFRNAi_2	TG62	#11	13177107

Ω. Sanger index sequences used for multiplexing: #1: ATCACG, #2: CGATGT, #3: TTAGGC, #4: TGACCA, #5: ACAGTG, #6: GCCAAT, #7: CAGATC, #8: ACTTGA, #9: GATCAG, #10: TAGCTT, #11: GGCTAC, #12: CTTGTA

2.2.9 Data analysis

The raw reads were aligned to the *D. melanogaster* genome assembly (UCSC dm3) using Bowtie (version 2.2.6) (Langmead et al., 2009) and excluding chromosome Uextra. Only uniquely mapped reads are kept using samtools (version 1.2) (Li et al., 2009). The raw read quality was accessed using FASTQC (version 11.5) (Andrews, 2010) and reads filtering was performed using FastX (version 0.0.13) (Hannon, 2010). Sequencing tracks of both fold enrichment and log (of base 2) transformation with parameter settings –m FE and –m logLR –p 0.00001 were generated using MACS (version 2.1.1) (Zhang et al., 2008), which were then visualized using IGB (Nicol et al., 2009) and IGV (Thorvaldsdottir et al., 2013) genome viewers. Peak calling was performed using HOMER 4.8 with parameter settings -style factor -size 200 -fragLength 200 -inputFragLength 200 (Heinz et al., 2010). Motif search and peak annotation were performed using ChIPseeks implementation of HOMER (Chen et al., 2014).

For downstream analysis, peaks identified in two out of three biological replicates were taken. Downstream analysis steps were performed using Python and R and parts of data preprocessing was done using ChipPeakAnno (Zhu et al., 2010). For the clustering of HMR peaks according to adjacent HP1a ChIP signals, three clusters were generated with K-means algorithm (MacQueen, 1967).

For repeat analysis, reads from ChIP-seq experiments were mapped to RepBase version 19.10 (Bao et al., 2015) using bowtie (Langmead et al., 2009). Only unique reads were kept for analysis. For each repetitive element log (of base 2) fold change was calculated. For the read density tracks, deepTools (version 2.3.3.5) (Ramirez et al., 2016) with parameter sets --ratio ratio --pseudocount=1 was utilized to normalize against the control.

To examine ChIP signals across species, genome positions of interest were lifted into the sibling species genome using UCSC liftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver). We used liftOver to convert genome coordinates between *D. melanogaster* dm3 and *D. simulans* droSim1 assemblies with setting “Minimum ratio of bases that must remap: 0.1” which is the minimum number of bases that must be directly aligned in gapless blocks from the first genome to the second.

3 RESULTS

The gene pair *Hmr* and *Lhr* is one of the most studied HI gene pairs. However, the molecular details of HMR's localization, HMR's targeting to its genomic binding sites and HMR's actions on the *D. melanogaster* pure species genome are not well understood. This work for the first time describes the localization of HMR to chromatin on a genome-wide scale and provides a novel link between HMR and genomic insulator sites.

3.1 HMR and LHR interact and mutually affect their protein level

The HI genes *Hmr* and *Lhr* encode the proteins HMR and LHR. These proteins physically interact and stabilize each other in a common complex. This was demonstrated in L2-4 cells and flies (Satyaki et al., 2014; Thomae et al., 2013). It even cumulated in a model proposing a protein level-dependent mechanism for hybrid incompatibility (Thomae et al., 2013). Therefore, we were aiming to verify these findings in the S2 cell system that we use in this study.

To demonstrate the physical interaction of HMR and LHR in S2 cells, we purified HMR and LHR from *D. melanogaster* S2 cell nuclear extract with anti-HMR and anti-LHR antibodies and analyzed the immunoprecipitates by Western Blot. Analogous to previous findings, LHR co-purifies with HMR and *vice versa* (**Figure 10 A**).

To demonstrate the mutual dependency of HMR and LHR in S2 cells, we applied HMR- and LHR-directed knockdown by RNAi and analyzed the resulting protein levels of HMR and LHR in cell lysates by Western Blot. RNAi knockdown takes use of the cell-owned RNAi pathway that regulates cellular transcript levels. Here, the treatment with specific double-strand RNA (dsRNA) efficiently reduces the transcripts of the desired target gene, which eventually results in a reduction of the translated protein (Elbashir et al., 2001). Analogous to previous findings, the protein levels of HMR and LHR are mutually dependent (**Figure 10 B**). In HMR knockdown, the LHR protein is depleted together with HMR and *vice versa*. This mutual dependency of the protein levels could reflect a contribution to synthesis or stability of the respective complex partner. Therefore, we further analyzed the *Hmr* mRNA and *Lhr* mRNA levels in HMR and LHR knockdown. While the *Hmr* mRNA levels are not affected in LHR knockdown and the *Lhr* mRNA levels are not affected in

HMR knockdown, the protein levels of HMR and LHR are highly dependent on each other (**Figure 10 C**). Similar was reported for components of the polycomb PRC2 complex (Tan et al., 2007), Histone Methly Transferases (Fritsch et al., 2010) and the pair JIL-1/PW53 (Dr. Catherine Regnard, personal communication). Altogether, these results indicate that HMR and LHR stabilize each other in a common complex.

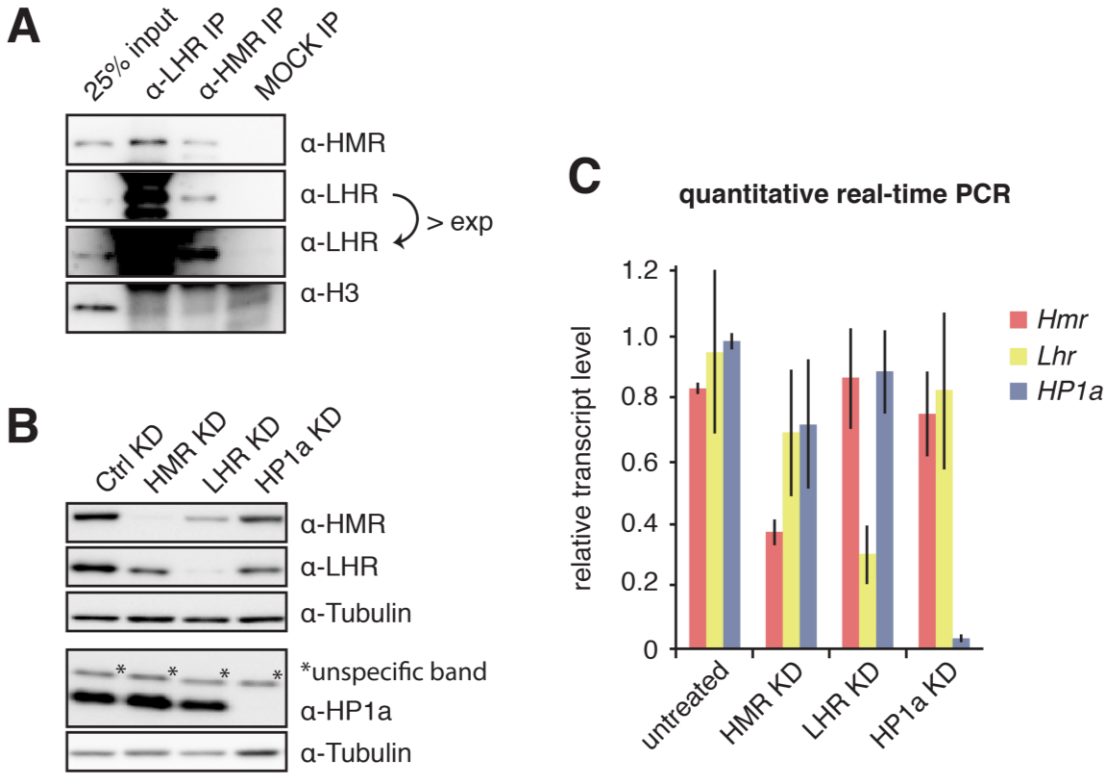

Figure 10. HMR and LHR coimmunoprecipitate and stabilize each other posttranscriptionally. (A) Western Blot of Coimmunoprecipitates from S2 cell nuclear extract using anti-HMR and anti-LHR antibody shows interaction of HMR and LHR. Rabbit anti-Rat IgG served as negative control (MOCK). **(B)** Western Blot of S2 cell lysates after specific and control RNAi-mediated knockdown (KD) shows mutual protein level dependency of HMR and LHR. **(C)** *Hmr* and *Lhr* transcript levels assayed by qPCR after RNAi treatment. Shown are transcript levels relative to control RNAi treatment after internal normalization against *tub97EF*. Error bars indicate SD of five biological replicates.

3.2 Endogenous epitope tagging of HMR using CRISPR/Cas9 system

3.2.1 Experimental strategy for endogenous epitope tagging

The major aim of this work was to identify genomic binding sites of HMR using chromatin immunoprecipitation. The quality of such an experiment is highly dependent on the antibody specificity and the enrichment by affinity purification (Landt et al., 2012; Orlando, 2000). For HMR complex immunoprecipitation, Dr. Andreas Thomae generated stable cell lines expressing epitope-tagged HMR (Thomae et al., 2013). The drawback of these cell lines is that HMR is not under control of the endogenous promoter. This could result in increased HMR protein dosage. As HMR protein levels are crucial for HMR's genomic localization and function in pure species (Thomae et al., 2013), we decided to generate cells expressing endogenously epitope-tagged HMR. For this purpose, I applied the CRISPR/Cas9 system that was adapted for targeted integration of chromosomal fragments in eukaryotic genomes. The experiments were performed in collaboration with Prof. Dr. Klaus Förstemann, who adapted and established this approach for *Drosophila* (Bottcher et al., 2014; Kunzelmann et al., 2016).

The key step in CRISPR/Cas9 genome editing is a sequence-specific targeting of the nuclease enzyme Cas9. In this way, DNA double strand breaks at defined genomic positions are generated. The specificity of the cleavage is ensured by a sequence-specific guide RNA that directs the Cas9 enzyme to its genomic target sites. The induced DNA double strand breaks are repaired by homologous recombination (HR), an alternative mechanism to the non-homologous end-joining (NHEJ) repair. NHEJ can be suppressed by knockdown of *lig4* or *mus308*, two crucial components of the NHEJ pathway (Bottcher et al., 2014; Kunzelmann et al., 2016). By providing artificial HR donor fragments, an epitope tag and resistance cassette can be placed at the 3' end of the gene body upstream of the 3' UTR resulting in a C-terminally tagged protein (MATERIALS & METHODS, **Figure 6**, **Figure 7**). After transfection, cells were selected for the presence of the resistance cassette and screened for site-specific integration.

3.2.2 Verification of cell lines generated by CRISPR/Cas9 system

In this work, I generated HMR-Flag$_2$, HMR-Strep$_2$ and HMR-GFP expressing S2 cells. HMR-Flag$_2$ expressing cells were derived from a uniformly tagged cell population with almost 100 % tagging efficiency. HMR-Strep$_2$ and HMR-GFP expressing cells were derived from single clones according to (Bottcher et al., 2014). Site-specific integration of the HR donor into the genomic landing site was tested by PCR on genomic DNA and analysis of the resulting PCR products (**Figure 11**). The first PCR primer pair targets only tagged allele (**Figure 11 A**, purple triangles) whereas the second PCR primer pair targets wild type genome sequence (**Figure 11 A**, green triangles) and results in a short favored PCR product for untagged alleles and in a long unfavored PCR product for tagged alleles. Therefore, the first primer pair is suited to test site-specific integration, whereas the second primer pair ensures high integration frequency across the allele homologs and across the cell population.

The HMR-Flag$_2$ expressing cells, obtained after resistance cassette-based selection, showed a very high site-specific integration frequency (**Figure 11 B**) and were directly used for further analysis. HMR-Strep$_2$ and HMR-GFP expressing cells showed site-specific integration events but also the presence of untagged wild type alleles (**Figure 11 C**). Therefore, we derived single clones from HMR-Strep$_2$ and HMR-GFP expressing cell populations. Clone 1 and clone 4 from HMR-Strep$_2$ and clone 6 from HMR-GFP expressing cells display site-specific integration at all *Hmr* gene copies.

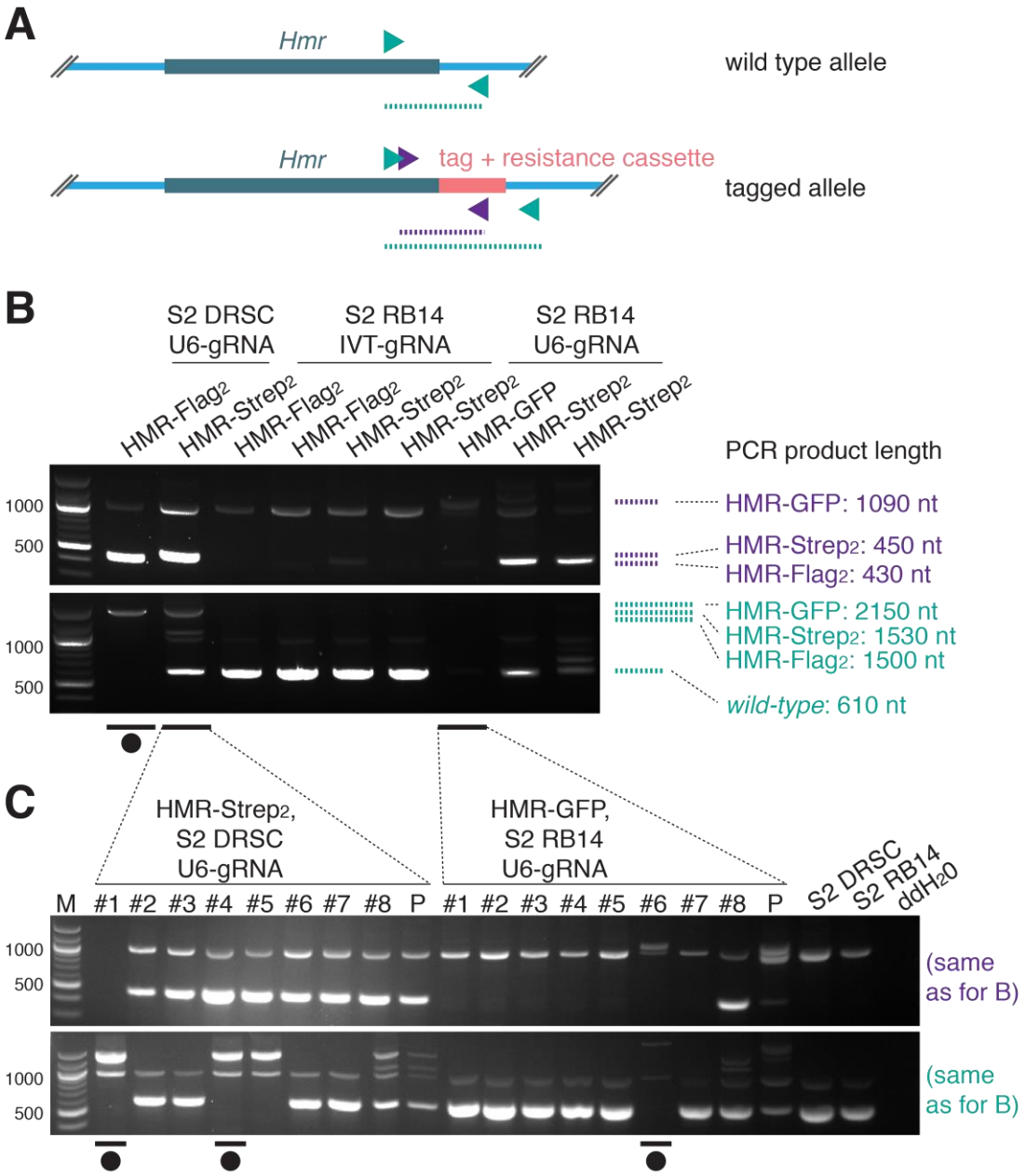

Figure 11. Verification of CRISPR/Cas9 system-based generated cell lines by PCR on genomic DNA. **(A)** PCR primer pair (purple and green triangles) locations and their expected products (dashed lines) for wild type *Hmr* allele and tagged *Hmr* allele. Site-specific integration is tested with a primer pair located upstream and downstream of the homology region. **(B)** PCR analysis on genomic DNA from cell population after resistance cassette based selection to verify site-specific integration at the *Hmr* locus. The expected product sizes are indicated on the right. **(C)** PCR analysis as described in (B) on genomic DNA from single cells (#) and their parental population (P). Successfully verified samples are marked with a filled circle.

3.2.3 Characterization of HMR-Flag$_2$ expressing cells

Among the cell lines generated by CRISPR/Cas9, colleagues and I used HMR-Flag$_2$ expressing S2 cells for further analysis. For initial characterization, I analyzed HMR's protein level and immunohistological localization. The HMR-Flag$_2$ fusion protein is expressed to protein levels comparable with wild type cells based on Western Blot analysis on cell lysates (**Figure 12 A**). Nevertheless, HMR-Flag$_2$ levels are slightly higher than HMR wild type levels. This might be due to a loss of the *Hmr* 3' untranslated region (UTR) and its potential regulatory function on HMR. In HMR-Flag$_2$ cells, the tag and the resistance cassette are placed at the 3' end of the gene body, which results in a loss of the 3' UTR in the *Hmr* gene transcripts. Further, I monitored the cellular localization of HMR-Flag$_2$ by immunofluorescence microscopy. The HMR-Flag$_2$ fusion protein exhibits a prominent localization to the centromere, here indicated by the overlap with the centromeric histone H3 variant Cid (**Figure 12 B**, white arrows). The centromeric localization of HMR in *D. melanogaster* S2 cells was described previously (Thomae et al., 2013).

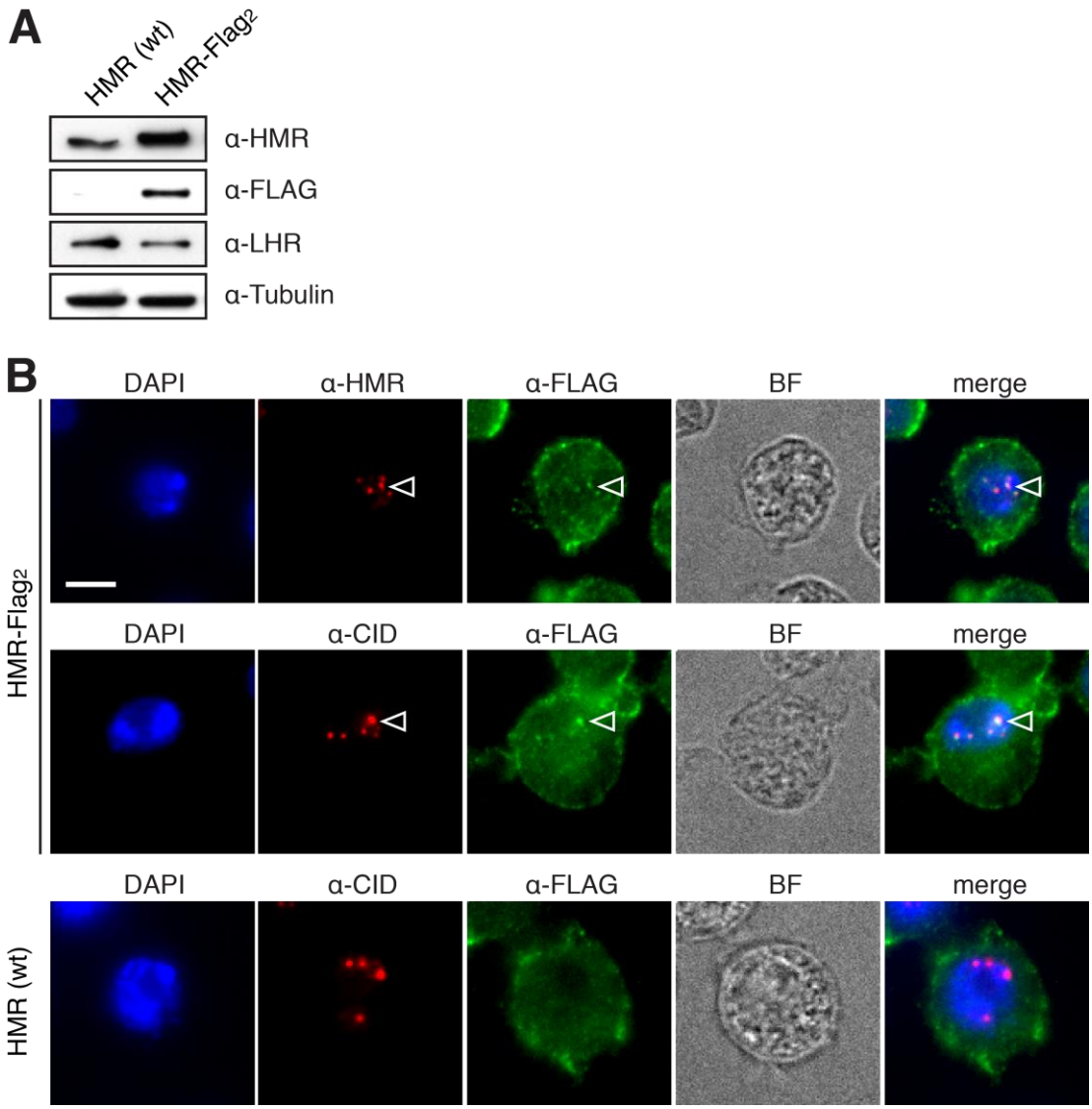

Figure 12. Characterization of HMR-Flag$_2$ expressing cells generated by CRISPR/Cas9. **(A)** Western Blot analysis on cell lysates to assay HMR protein level. LHR and Tubulin protein detection served as control. **(B)** Immunofluorescence microscopy on cells co-stained for HMR and FLAG (top row, HMR-Flag$_2$) or CID and FLAG (middle row, HMR-Flag$_2$ and bottom row, wild type). White arrows indicate signal overlap of HMR, CID and FLAG in HMR-Flag$_2$ cells. Scale bar represents 5 µm.

3.3 Genome-wide binding map of HMR in *D. melanogaster*

Previous immunohistological studies demonstrated that HMR localizes to centromere or centromere-proximal regions in diploid cells (Maheshwari and Barbash, 2012; Thomae et al., 2013). Further, HMR localizes to telomeres and distinct regions across the chromosome arms on polytene chromosomes (Thomae et al., 2013). However, a high resolution binding map of HMR that allows dissecting the underlying binding site features, such as DNA sequence and co-binding proteins, was lacking so far.

To access genomic binding of HMR in high resolution, we applied chromatin immunoprecipitation (ChIP) in *D. melanogaster* S2 cells. In brief, ChIP involves covalent *in vivo* crosslinking of proteins and DNA, chromatin fragmentation, immunoprecipitation of the target protein and associated chromatin, isolation and analysis of the enriched DNA (MATERIALS & METHODS, **Figure 9**).

3.3.1 ChIP-seq of the HMR/LHR complex

The antibody used in ChIP is crucial for the experimental outcome and interpretation (Landt et al., 2012; Orlando, 2000). We applied a highly specific anti-HMR (Thomae et al., 2013) and anti-FLAG antibody in combination with HMR-$Flag_2$ expressing cells generated with CRISPR/Cas9 system to target HMR. Further, we applied an anti-LHR antibody (Thomae et al., 2013) to target the HMR interacting protein LHR (**Figure 13 A**). Notably, all these antibodies are monoclonal and therefore target a single epitope present in the peptide they were raised against. The anti-HMR antibody was raised against the HMR N-terminus and therefore targets HMR and HMR-$Flag_2$ (**Figure 13 A**). The anti-FLAG antibody targets the C-terminal double FLAG tag of HMR-$Flag_2$ but serves as a powerful negative control when applying in HMR expressing wild type cells where the FLAG epitope is absent (**Figure 13 A**).

HMR ChIP and FLAG ChIP were performed in wild type and HMR-$Flag_2$ cells. LHR ChIP was performed in wild type cells. Different antibodies that target HMR/LHR in ChIP give a similar signal pattern distribution across genomic regions but are not entirely overlapping (**Figure 13 B**). As expected, all ChIP signals are absent in the negative control (**Figure 13 B**).

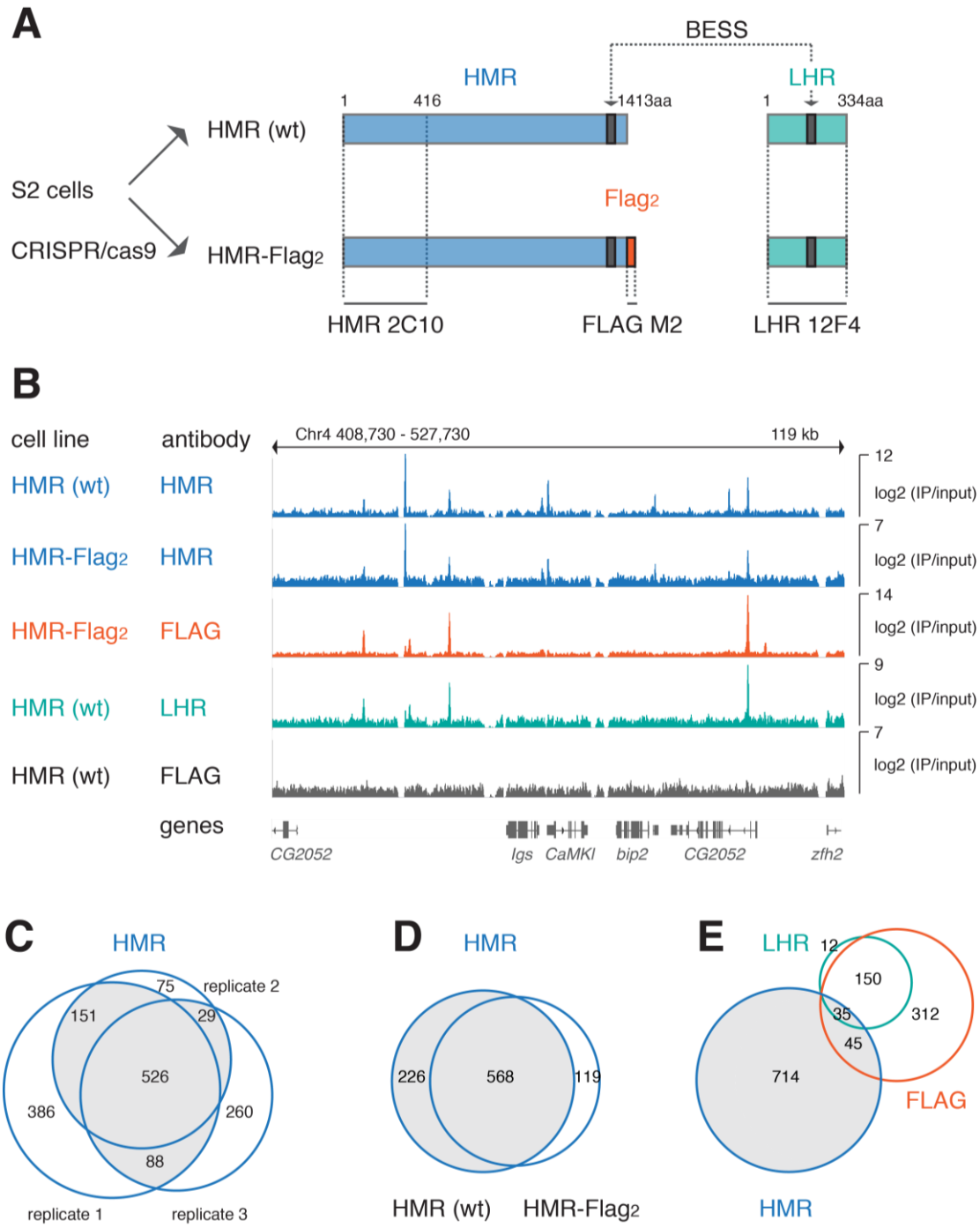

Figure 13. ChIP-seq of the HMR/LHR complex. (A) Monoclonal antibodies used in ChIP and the peptide locations where the antibodies were raised against. Wild type (wt) S2 cells and HMR-Flag$_2$ expressing cells generated with CRISPR/Cas9 system were used. Anti-HMR antibody was raised against HMR N-terminus (Thomae et al., 2013) whereas anti-FLAG antibody targets the C-terminal double FLAG tag of HMR-Flag$_2$ and serves as negative control in wt cells. Anti-LHR antibody was raised against whole LHR (Thomae et al., 2013). HMR and LHR presumably interact via BESS domain located at HMR C-terminus (Brideau et al., 2006) **(B)** Genome browser view of signals from ChIP reactions targeting HMR/LHR complex. Used cell line and antibody are stated. Control is anti-FLAG in HMR (wt) expressing cells. **(C)** Venn diagram of HMR peaks identified in three independent biological replicates.

Peaks identified in at least two out of three are highlighted and used for downstream analysis. **(D)** Venn diagram of HMR peaks identified in HMR (wt) and HMR-$Flag_2$ expressing cells. **(E)** Venn diagram of HMR, FLAG (targeting HMR) and LHR peaks.

Using the anti-HMR antibody, we derived a set of 794 HMR binding sites, present in at least two out of three independent biological replicates from HMR expressing wild type cells (**Figure 13 C**). These binding sites also largely overlap with the anti-HMR ChIP in HMR-$Flag_2$ expressing cells (**Figure 13 D**). However, when applying the anti-FLAG antibody in HMR-$Flag_2$ expressing cells, we derived a set of peaks that shares only 80 out of 542 with anti-HMR antibody (**Figure 13 E**). The results from (**Figure 13 D**) demonstrate that the differences among these peak sets are not due to differences between the HMR and the HMR-$Flag_2$ protein. Therefore, we conclude that the differences of ChIP peak sets are due to the ChIP antibody and not due to intrinsic HMR and HMR-$Flag_2$ properties. Assuming that the antibodies recognize single epitopes in the crosslinked HMR/LHR complex *in vivo* structure, both, anti-FLAG and anti-LHR antibody, target the C-terminal region of HMR: anti-FLAG antibody targets the C-terminal tag, anti-LHR antibody targets LHR which interacts with HMR's C-terminal BESS domain (Thomae et al., 2013). In accordance with these findings, the peak set derived with anti-LHR antibody is highly reminiscent of the anti-FLAG ChIP in HMR-$Flag_2$ expressing cells (185 out of 197 overlapping, **Figure 13 E**). A heatmap-based genome-wide analysis of the corresponding ChIP-seq profiles along the anti-HMR and anti-FLAG ChIP peaks confirms this (**Figure 14**).

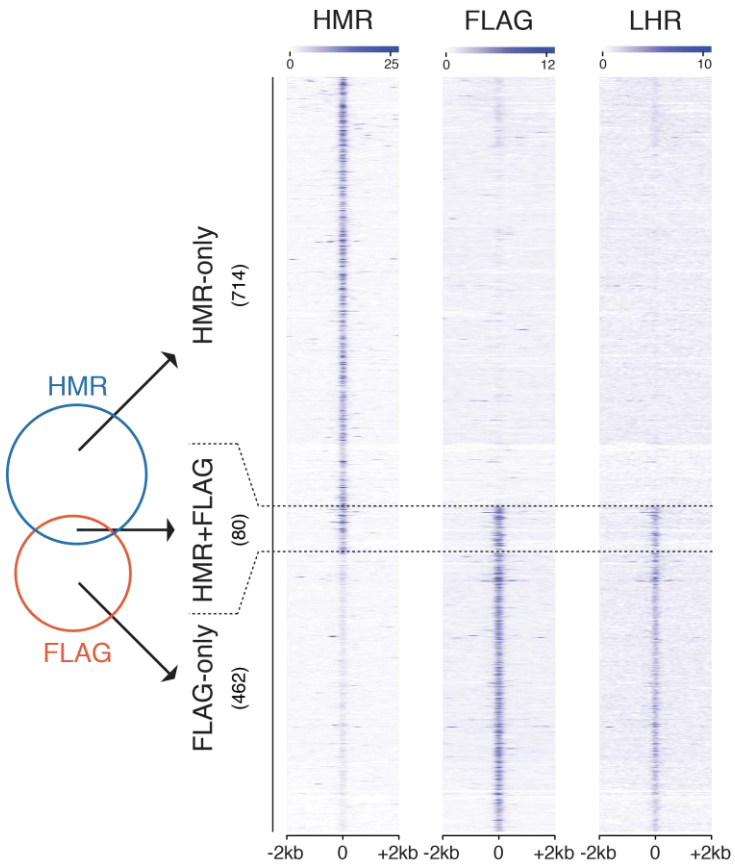

Figure 14. ChIP-seq profiles of anti-FLAG (HMR-Flag$_2$ expressing cells) and anti-LHR antibody are distinct from anti-HMR antibody. Heatmaps of signals of anti-FLAG (HMR-Flag$_2$ expressing cells), anti-LHR and anti-HMR ChIP. Genomic regions are centered on anti-HMR ChIP peaks (HMR-only and HMR+FLAG) and on anti-FLAG ChIP peaks (FLAG-only).

3.3.2 HMR binding site verification for genome-wide analysis

Different HMR ChIP data derived with different monoclonal antibodies share only a subset of peaks. To verify our ChIP results and to decide which peak set to use for subsequent genome-wide analysis, we compared the genomic location of HMR ChIP peaks with the ChIP signals obtained in the negative control IgG ChIP. First, we selected four positive and two negative control loci in ChIP-qPCR. We observe enrichment of HMR at the tested HMR binding sites compared to negative control reactions (**Figure 15**).

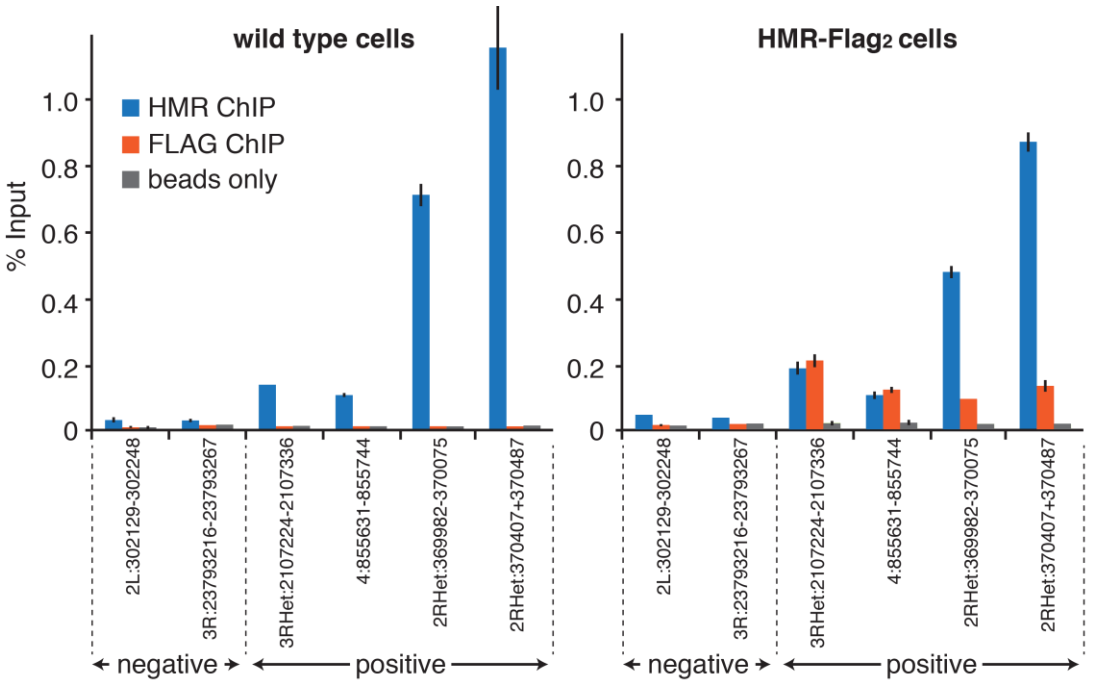

Figure 15. ChIP-qPCR showing HMR enriched at HMR binding sites identified in ChIP-seq compared to negative control regions. HMR ChIP is enriched for binding sites in wild type and HMR-Flag$_2$ expressing cells. In contrast, FLAG ChIP is enriched for HMR binding sites only in HMR-Flag$_2$ expressing cells but not in wild type cells lacking FLAG tag epitope. Data are represented as mean ± SD of three technical replicates.

Second, we compared the binding sites on a genome-wide scale in ChIP-seq (**Figure 16**). Here, two different antibodies (anti-HMR and anti-FLAG) were applied to the same chromatin from wild type cells. Almost none of the HMR binding sites were enriched in the negative anti-FLAG control IgG (**Figure 16 A, A'**). Further, we compared the genomic location of FLAG ChIP peaks from HMR-Flag$_2$ expressing cells with the ChIP signals obtained in the negative control IgG. Here, in contrast to (**Figure 16 A, A'**), the same antibody (anti-FLAG) was applied to two different chromatin types (HMR-Flag$_2$ expressing cells and wild type cells). Again, almost none of the HMR binding sites were enriched in the negative control (**Figure 16 B, B'**).

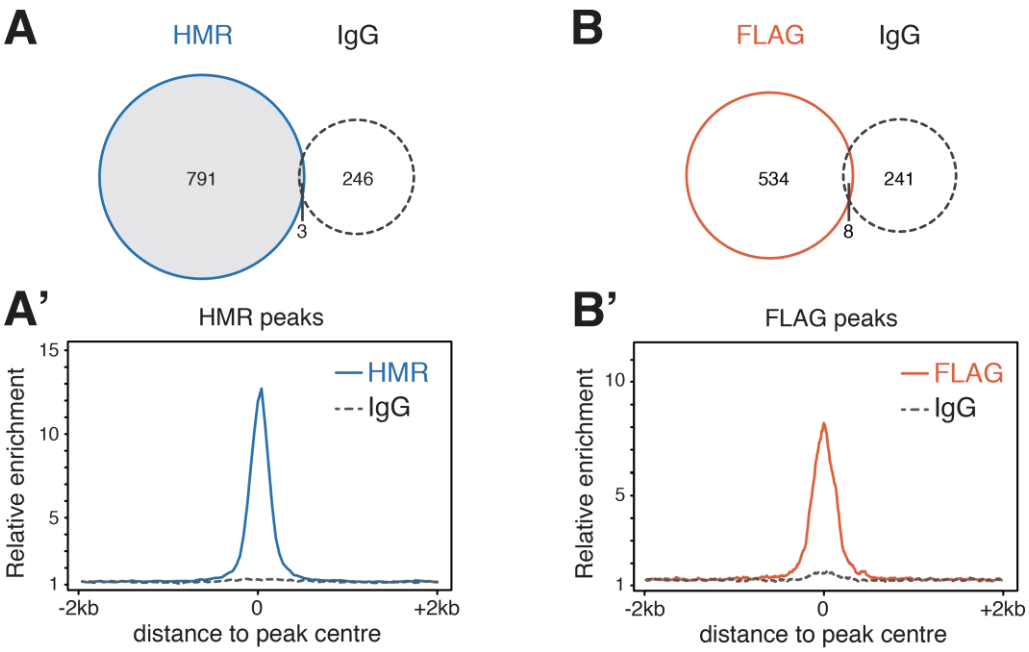

Figure 16. HMR and FLAG ChIP signals are not overlapping with negative control IgG ChIP signals. The FLAG ChIP in HMR-Flag$_2$ cells (orange) targets HMR whereas FLAG ChIP in wt cells (grey) serves as negative control IgG **(A)** Venn diagram of HMR and control IgG peaks (pool of peaks from two independent biological replicates). **(A')** Composite plot of HMR and control IgG ChIP signals at genomic HMR peak positions. **(B)** Venn diagram of FLAG and control IgG peaks. **(B')** Composite plot of FLAG and control IgG ChIP signals at genomic FLAG peak positions.

After comparing with negative control IgG ChIP, we compared the identified binding sites with the recently published collection of phantom peaks, genomic regions which are prone to generate unspecific signals in ChIP experiments (Jain et al., 2015). Interestingly, the ChIP peaks derived from anti-FLAG and anti-LHR antibody are largely overlapping with phantom peak regions, whereas the anti-HMR antibody-based peaks are not (**Figure 17**). This finding does not mean that anti-LHR and anti-FLAG ChIP profiles are necessarily incorrect. In fact, the phantom peak positions share various features that are essential for HMR's binding properties such as the presence of insulator proteins and their corresponding DNA binding motifs as demonstrated and discussed later in this work (Jain et al., 2015). However, in order to rule out potential unspecific ChIP signals, we decided to proceed analysis exclusively with the peak set derived from the anti-HMR antibody which contains a subset of the anti-FLAG and anti-LHR antibody-based ChIP peaks (**Figure 13 E**).

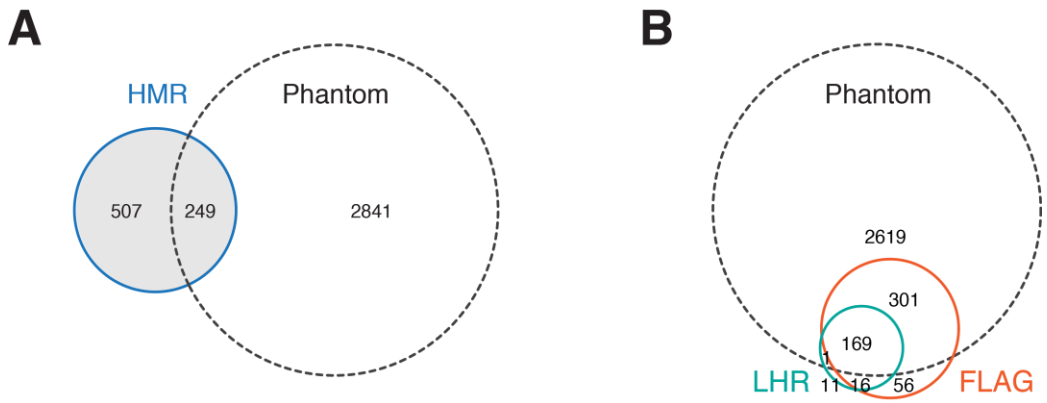

Figure 17. HMR, FLAG and LHR ChIP peak overlap with genomic phantom peak positions that can give raise to false positive ChIP signals. Genomic phantom peak positions used for analysis were described for S2 cells in (Jain et al., 2015). **(A)** Venn diagram of HMR and phantom peaks. **(B)** Venn diagram of FLAG, LHR and phantom peaks.

To further validate HMR's genome-wide binding data, we applied HMR-directed knockdown by RNAi, assayed the resulting protein levels of HMR in cell lysates by Western Blot as described before (**Figure 10 B**) and subjected cells to HMR ChIP. Next, we monitored the genome-wide changes in enrichment against negative control knockdown cells (**Figure 18**, also **Figure 19 C**). Though, HMR knockdown results in a genome-wide reduction of HMR binding and showed a substantial effect on many HMR binding sites, a subset of HMR binding sites remains enriched in HMR knockdown (**Figure 18**). Such knockdown resistant binding sites were described for other chromatin-associated factors before and can be interpreted as high-affinity binding sites (Schwartz et al., 2012) or assigned to incomplete removal of the target protein. The HMR ChIP reaction supernatants are not depleted for the HMR epitope after ChIP in untreated cells (**Figure S1**) which means that the antibody amount in ChIP is limiting to the yield of ChIP DNA rather than the epitope amount. In case the epitope is more abundant than the epitope-recognizing antibody, the detectable effect of an epitope targeting knockdown can decrease.

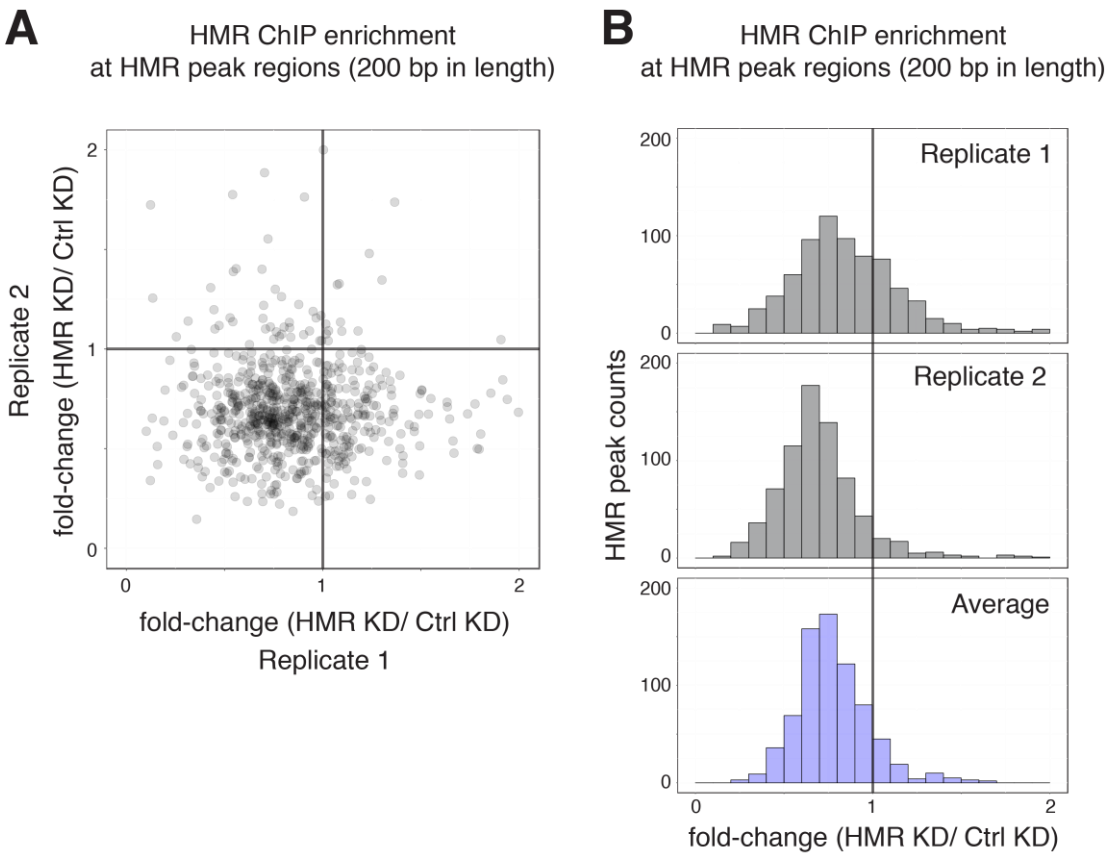

Figure 18. Changes in HMR ChIP enrichment after HMR knockdown versus GST control knockdown. (A) Scatter plot showing the fold changes of normalized HMR ChIP tag number mapped to a 200 bp HMR peak region in two biological replicates. Peak regions with less than 50 aligned tags were excluded from the analysis. **(B)** Histograms showing the frequency of HMR peaks that display HMR ChIP signal reduction after HMR knockdown.

3.3.3 HMR ChIP-seq data resembles prior immunohistological studies

Using different antibodies and RNAi knockdown strategy in ChIP, we derived a reliable set of 794 HMR binding sites. To compare this high-resolution HMR binding information with prior localization studies on HMR, we monitored HMR binding site distribution and density across the *D. melanogaster* 2nd chromosome (**Figure 19 A**). Strikingly, HMR binding sites are enriched proximate to centromere and cytological region 31. HMR was described to bind these two genomic regions in independent immunohistological experiments before (Maheshwari and Barbash, 2012; Thomae et al., 2013). Furthermore, region 31 contains eight bands of HP1a antibody staining (Fanti et al., 2003). The anti-HP1a ChIP in our experiments (**Figure 19 C**) is in agreement with this finding. In line with HP1a ChIP-seq data obtained from the modENCODE consortium (Kharchenko et al., 2011), we observe HP1a enriched at pericentromeric heterochromatin together with HMR when spotting the

heterochromatin-euchromatin border on the 2nd chromosome (**Figure 19 B**). Even though HMR and HP1a were reported to physically interact, HMR binding sites do not completely cover HP1a chromatin domains but rather mark their edges (**Figure 19 C**).

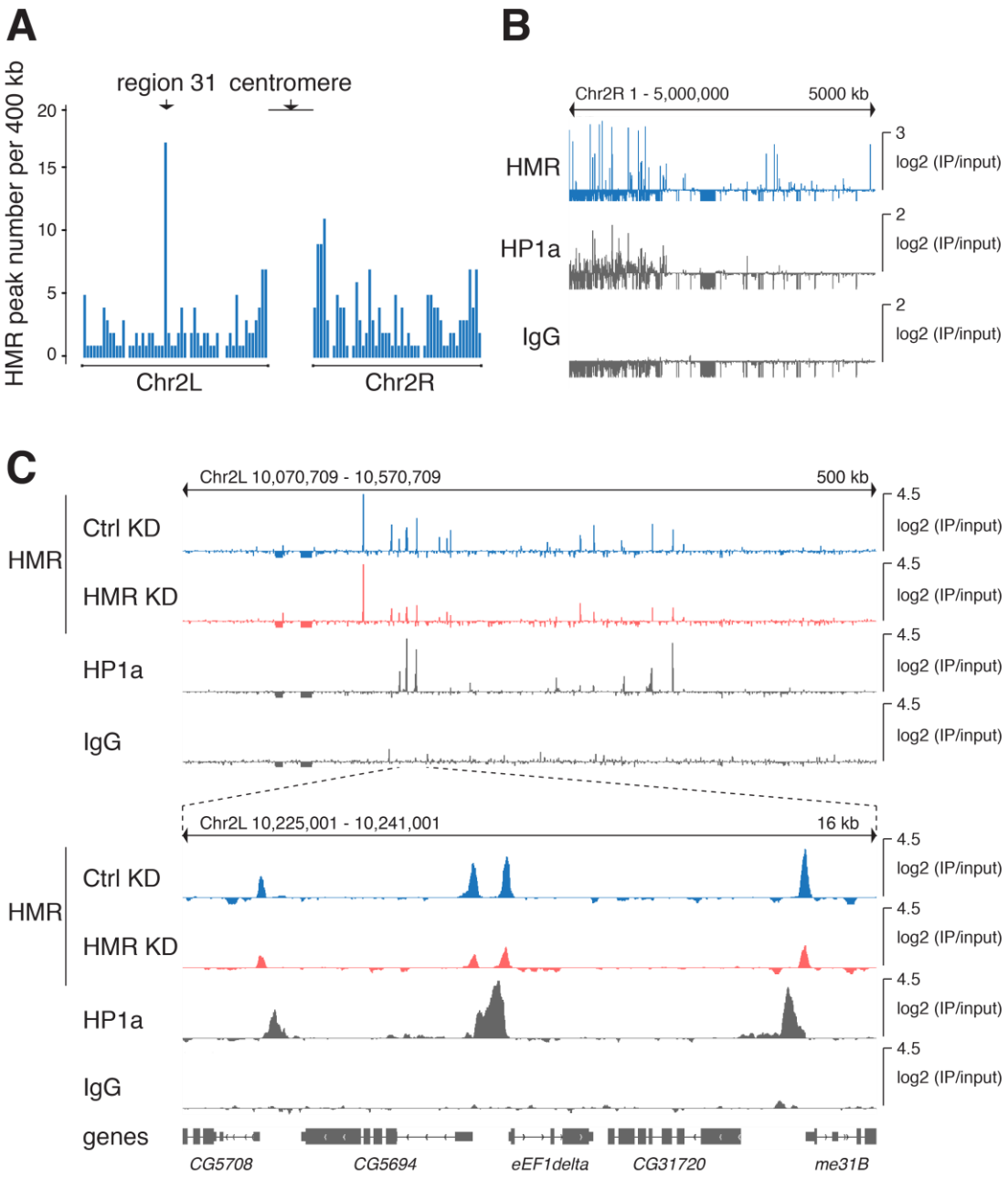

Figure 19. HMR binding at pericentromeric and other heterochromatic regions (A) Histogram showing HMR peak density across left arm (2L) and right arm (2R) of the 2nd chromosome. Cytological region 31 and centromere-proximal regions are enriched for HMR binding sites. **(B)** Genome browser view of HMR, HP1a and control IgG ChIP signals at 2nd chromosome centromere-proximal region. **(C)** Genome browser view of HMR, HP1a and control IgG ChIP signals at region 31. HMR ChIP signals are displayed for control knockdown and HMR knockdown.

In summary, HMR is shown to be enriched at heterochromatic genome regions by independent techniques in different biological material. First, this provides evidence for a high reliability of our ChIP data and, second, it suggests that HMR localization is at least partially consistent between cell types.

3.4 HMR binding sites largely overlap with genomic insulators

Chromatin-associated proteins can distribute along chromatin and assemble in domains such as HP1a. This results in a dispersed ChIP signal. Alternatively, proteins bind sequence-specific such as transcription factors that give rise to a sharp ChIP signal peak. The genomic tracks (**Figure 19 C**) and composite plots of HMR ChIP signals at HMR binding sites (**Figure 16 A', B'**) revealed sharp ChIP signals with approximately 200 nucleotides in width, suggesting HMR being part of chromatin residing complexes with well-defined genomic binding positions.

To identify potential targeting factors of HMR, we performed DNA sequence motif analysis with HOMER. Strikingly, the HMR binding sites are enriched for DNA sequence motifs that associate with the insulator DNA binding proteins BEAF-32 (27 %) and Su(Hw) (26 % and 17 %) (**Figure 20 A**). Both proteins can bind DNA sequence-specific by zinc-finger DNA-binding domains and are associated with other functional insulator complex components such as CP190 and Mod(mdg4) (Gause et al., 2001; Ghosh et al., 2001; Pai et al., 2004). We used insulator protein ChIP data derived from S2 cells (Ong et al., 2013; Riddle et al., 2011) and compared them with the HMR binding profile. Indeed, HMR and insulator complex components show a strong binding overlap (**Figure 20 B**).

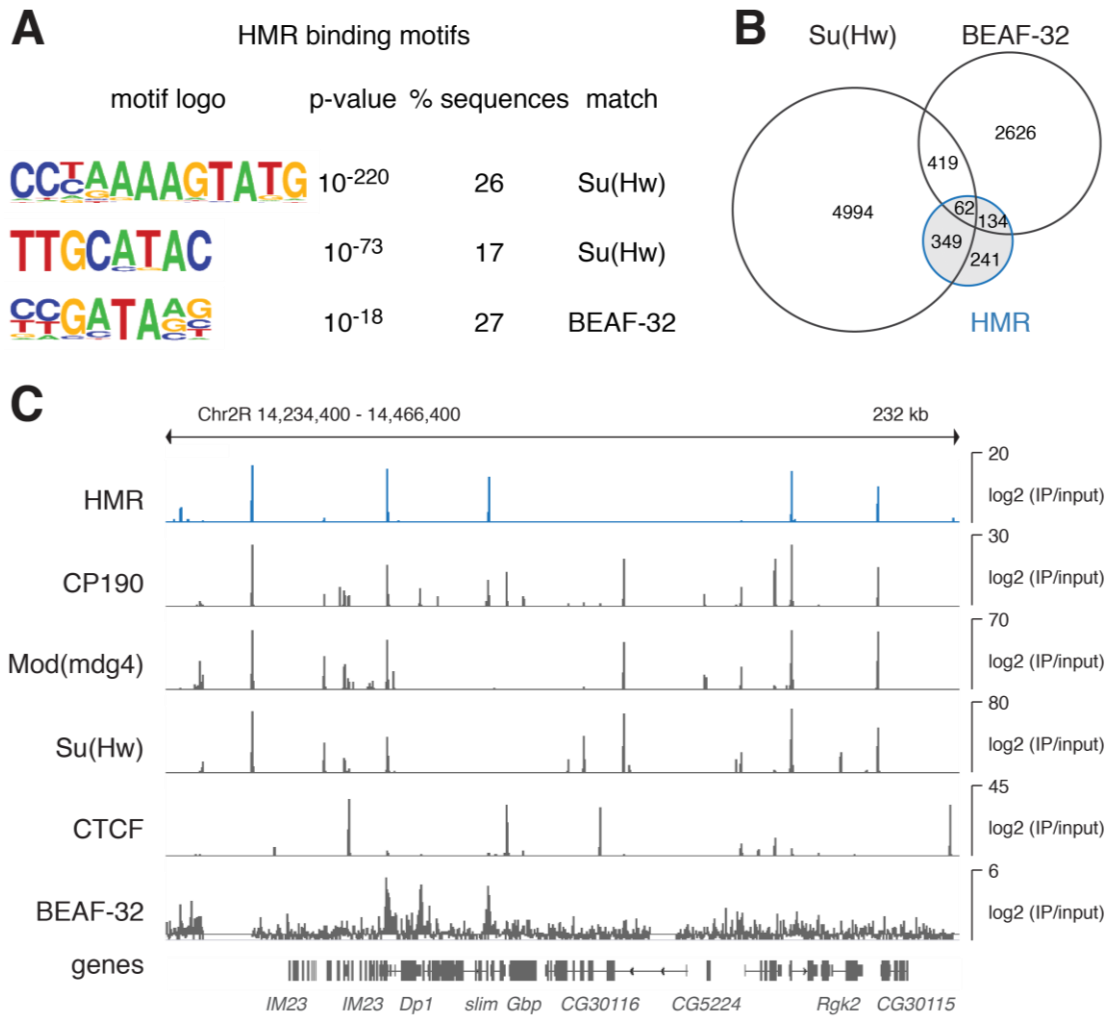

Figure 20. HMR localizes to genomic binding sites of the insulator proteins Su(Hw) and BEAF-32. (A) Sequence motifs identified within HMR peak regions using HOMER motif search. Motif logo, p-value of enrichment, percentage of regions with this motif and putative binding factors are indicated. **(B)** Venn diagram of HMR, Su(Hw) (Ong et al., 2013) and BEAF-32 (Riddle et al., 2011) ChIP peaks. **(C)** Genome browser view of HMR, the insulator proteins CP190, Mod(mdg4), Su(Hw), CTCF (Ong et al., 2013) and BEAF-32 (Riddle et al., 2011) ChIP signals shows binding overlap of HMR with various genomic insulator sites.

Genomic insulator sites are occupied by different combinations of insulator complex components (Schwartz et al., 2012). Judging from our ChIP data, HMR binds to various of these insulator complexes (**Figure 20 C**). HMR is particularly enriched at the well characterized class of *gypsy* insulators. *Gypsy* insulators are composed of Su(Hw) binding motifs, and several proteins, including Su(Hw), Mod(mdg4), and CP190 (Georgiev and Gerasimova, 1989; Gerasimova et al., 1995; Pai et al., 2004; Parkhurst et al., 1988). Approximately half of the HMR binding sites are associated with *gypsy* insulator (**Figure 21 A**). The first identified *gypsy* insulator site is named

1A-2, due to its cytological location, and is located between the *yellow* and *achaete* genes (Gerasimova et al., 1995). HMR also binds to this region (**Figure 21 B**).

Overall, in accordance with BEAF-32 and Su(Hw) binding motif enrichment, HMR binding sites are mainly associated with BEAF-32-bound insulator or Su(Hw)-bound *gypsy* insulator sites.

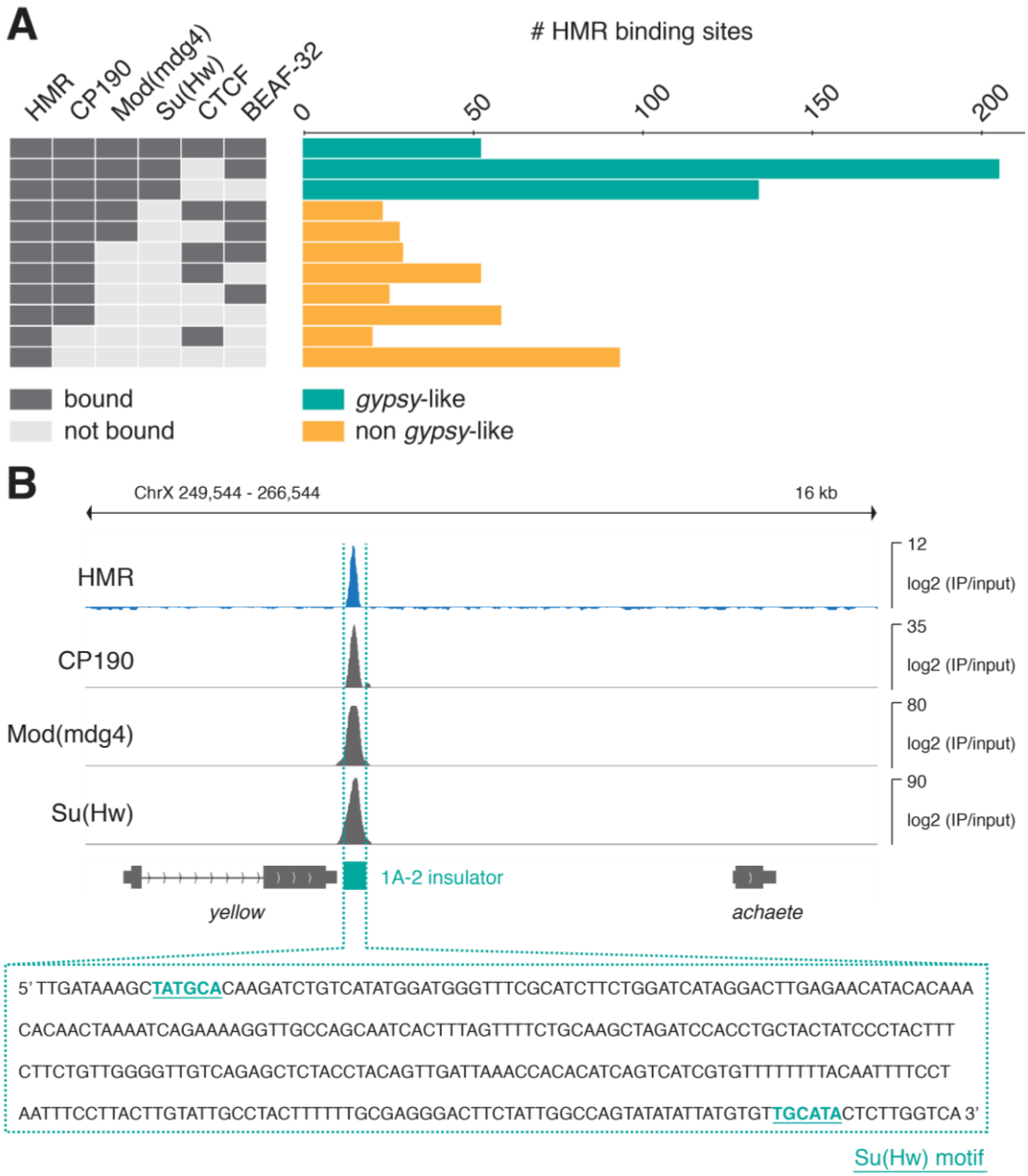

Figure 21. HMR localizes to *gypsy* and non-*gypsy* insulator sites. (A) Combinatorial binding pattern of HMR with insulator proteins. ChIP peak overlap of HMR with CP190, Mod(mdg4), Su(Hw), CTCF (Ong et al., 2013) and BEAF-32 (Riddle et al., 2011). HMR peaks assigned to the Su(Hw)-containing *gypsy* groups are depicted in green, non-*gypsy* groups are depicted in orange. Combinations with less than ten HMR peaks are not shown. **(B)** Genome browser view of ChIP signals of HMR and the *gypsy* insulator proteins CP190, Mod(mdg4) and Su(Hw) (Ong et al., 2013) at the *gypsy* insulator 1A-2. Su(Hw) binding motifs identified in the underlying DNA sequence are depicted in green.

3.5 HMR localizes to the insulator of *gypsy* and *gypsy-twin* retrotransposons

Genome-wide analysis of HMR binding sites revealed an extensive localization of HMR to genomic insulator sites, in particular to the *gypsy* insulator that is derived from *gypsy* transposons. Further, HMR interacts with HP1a and affects repetitive DNA transcription (Satyaki et al., 2014; Thomae et al., 2013).

To investigate, whether HMR localizes to repetitive DNA, we assayed the enrichment of repeats by mapping HMR, HP1a and Su(Hw) ChIP-seq reads to the RepBase repeat database (Bao et al., 2015). HP1a is a heterochromatin component covering several repetitive DNA elements. Indeed, HP1a ChIP is enriched for the centromeric heterochromatin-associated *Dodeca* satellite (DMSAT6) (Abad et al., 1992) and the transposable elements *Rt1a* and *Rt1b* (DMRT1A, DMRT1B) (**Figure 22 A**) (Kaminker et al., 2002). These repeats however, are neither enriched in HMR ChIP nor in Su(Hw) ChIP. Instead, HMR ChIP and Su(Hw) ChIP are both enriched for the long terminal repeat (LTR) retrotransposon *gypsy*, and *gypsy-twin* (*gtwin)* (**Figure 22 A'**, **A''**). *Gtwin* is a *gypsy*-related sequence (Bowen and McDonald, 2001; Kotnova et al., 2005) and both retroelements are suppressed in transposition by the *flamenco* locus in *Drosophila* (Kotnova et al., 2005; Prud'homme et al., 1995; Razorenova et al., 2001). HMR's localization to the *gypsy* and *gtwin* 5' end is in good agreement with the strong binding correlation of HMR with insulator proteins. HMR, Su(Hw), Mod(mdg4) and CP190 show almost identical binding pattern across *gypsy* and *gtwin* and are specifically enriched at the LTR's 5' region (**Figure 22 B**, **B'**). This is the region where the *gypsy* insulator is located. The *gypsy* insulator is composed of a 340-bp sequence containing 12 Su(Hw) binding motifs bound by Su(Hw) and other protein complex components (**Figure 22 B**, **B'**).

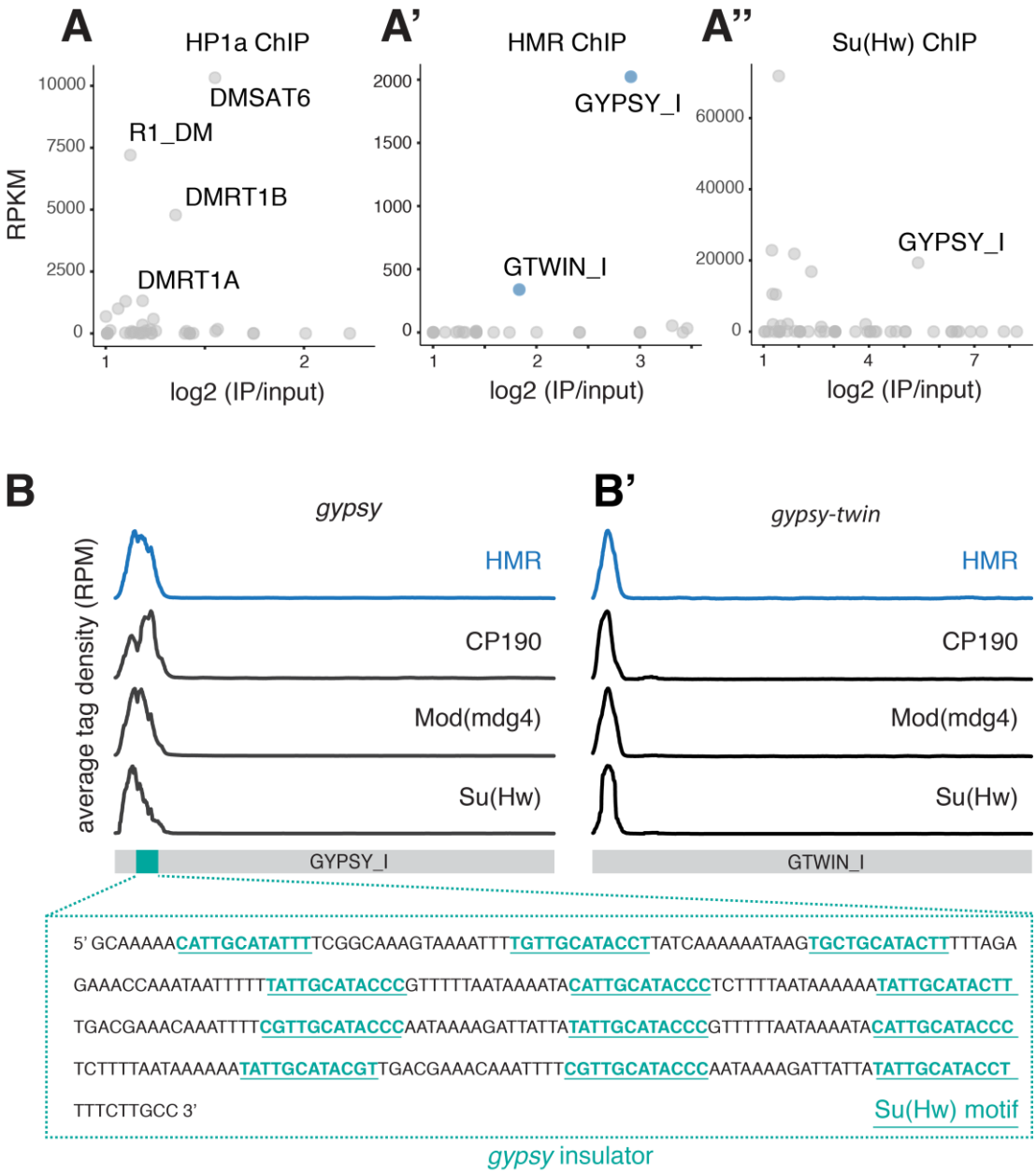

Figure 22. HMR localizes with *gypsy* insulator to *gypsy* and *gtwin* 5' region. (A, A', A'') ChIP tag enrichment of HP1a, HMR and Su(Hw) ChIP (Ong et al., 2013) at repetitive DNA elements from RepBase. The log2-fold enrichment over input is plotted against the RPKM of an individual repeat sequence. Only repetitive elements that are more than 2-fold enriched are displayed. **(B, B')** ChIP tag distribution of HMR, CP190, Mod(mdg4), Su(Hw) ChIP (Ong et al., 2013) across the *gypsy* (GYPSY_I) and *gtwin* (GTWIN_I) retrotransposon sequence. HMR is enriched with *gypsy* insulator proteins at the *gypsy* insulator sequence (green).

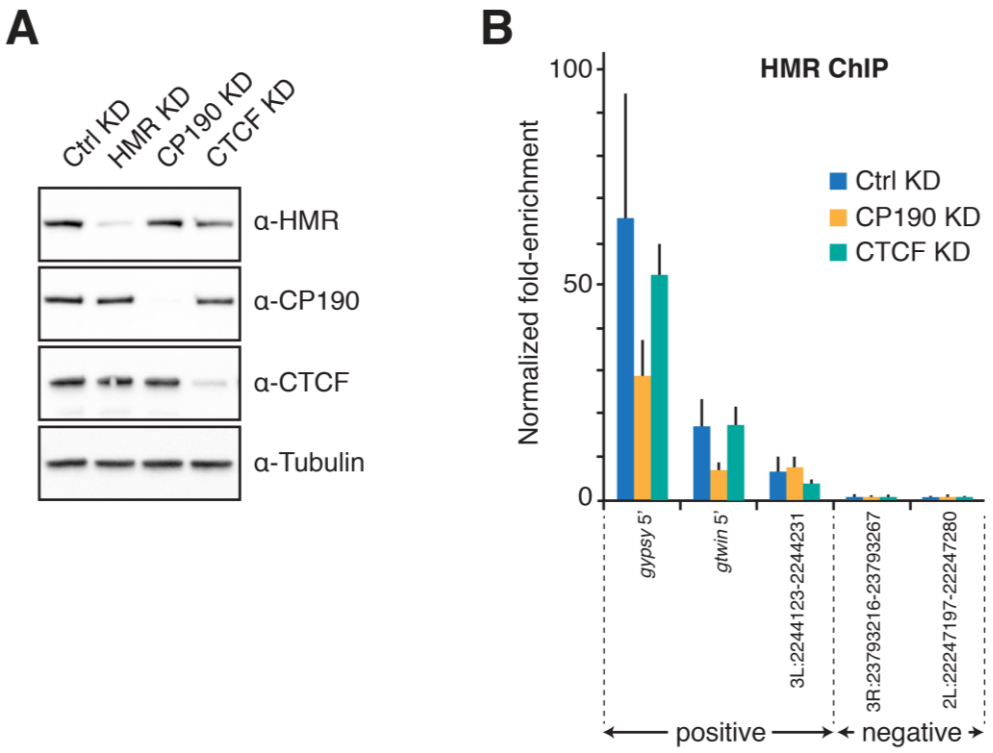

Figure 23. HMR localizes to *gypsy* and *gtwin* 5' region dependent on CP190. (A) Western Blot analysis on cell lysates to assay protein levels after HMR, CP190 and CTCF knockdown (KD). Tubulin protein detection served as control. **(B)** Normalized fold-enrichment of HMR after CP190 KD, CTCF KD at *gypsy* and *gtwin* 5' region, a HMR binding site without CP190 (positive) and two negative control regions (negative). Data are represented as mean ± SD of three independent biological replicates.

To test whether HMR binding to the *gypsy* insulator depends on the insulator complex, we performed HMR ChIP after knockdown of CP190 and CTCF (**Figure 23 A**) and probed immunoprecipitated DNA by qPCR (**Figure 23 B**). CP190 is required for *gypsy* insulator complex recruitment and maintenance whereas the insulator protein CTCF is not required (Pai et al., 2004; Schwartz et al., 2012). After CP190 knockdown, HMR binding is reduced at *gypsy* and *gtwin* 5' region, the regions where the insulator complex is located. A CTCF knockdown does not affect the enrichment at these sites (**Figure 23 B**). This finding suggests that HMR's binding to *gypsy* insulators is dependent on CP190.

In summary, HMR binds to the *gypsy* and *gypsy-twin* elements. HMR's binding to these sites dependents on the presence of CP190 or the *gypsy* insulator complex in general.

3.6 HMR's localization to genomic insulator sites depends on the presence of the insulator protein complex

HMR ChIP-qPCR analysis after knockdown experiments demonstrated that HMR's binding to the *gypsy* insulator complex at the *gypsy* and *gtwin* 5' region is dependent on CP190. To test whether CP190 also impacts HMR binding on a genome-wide scale, we performed HMR ChIP-seq after CP190 knockdown (analogous shown in **Figure 23 A**). We analyzed the enrichment changes against control knockdown by sequencing and group-wise comparison (**Figure 24 A**). According to literature, we expect the *gypsy*-like group to be affected by CP190 knockdown, whereas the non *gypsy*-like group to be not affected (**Figure 24 B**) (Schwartz et al., 2012). After CP190 knockdown, HMR binding is only reduced in the *gypsy*-like group (**Figure 24 C**). To control for specificity, we subjected HMR ChIP after HMR knockdown to the same analysis. Here, HMR binding is affected equally in both groups (**Figure 24 D**). A quantification of the effect on HMR binding by assessing the aligned ChIP tags to HMR binding sites confirmed specificity (**Figure 25**).

The insulator protein CP190 serves as an adaptor protein for large insulator complex structures that are placed in nucleosome depleted regions of the genome (Bartkuhn et al., 2009). As nucleosome occupancy can serve as a proxy for insulator complex integrity at these sites (Schwartz et al., 2012), we wondered whether the CP190 knockdown affects nucleosome occupancy. To access nucleosome occupancy, we performed Histone H3 ChIP. The H3 ChIP signal is increased in the *gypsy*-like group of HMR binding sites after CP190 knockdown, but not in the non *gypsy*-like group (**Figure 24 C**). After HMR knockdown, none of the groups are affected in H3 binding (**Figure 24 D**). On the one hand, this demonstrates the importance of CP190 for *gypsy* insulator complex integrity. The loss of HMR binding at these sites is accompanied by the loss of the *gypsy* insulator complex. On the other hand, these findings suggest that HMR does not fulfill an essential role in maintaining *gypsy* insulator complex integrity.

Overall, these results demonstrate the important role of insulator complexes in mediating HMR's binding to chromatin.

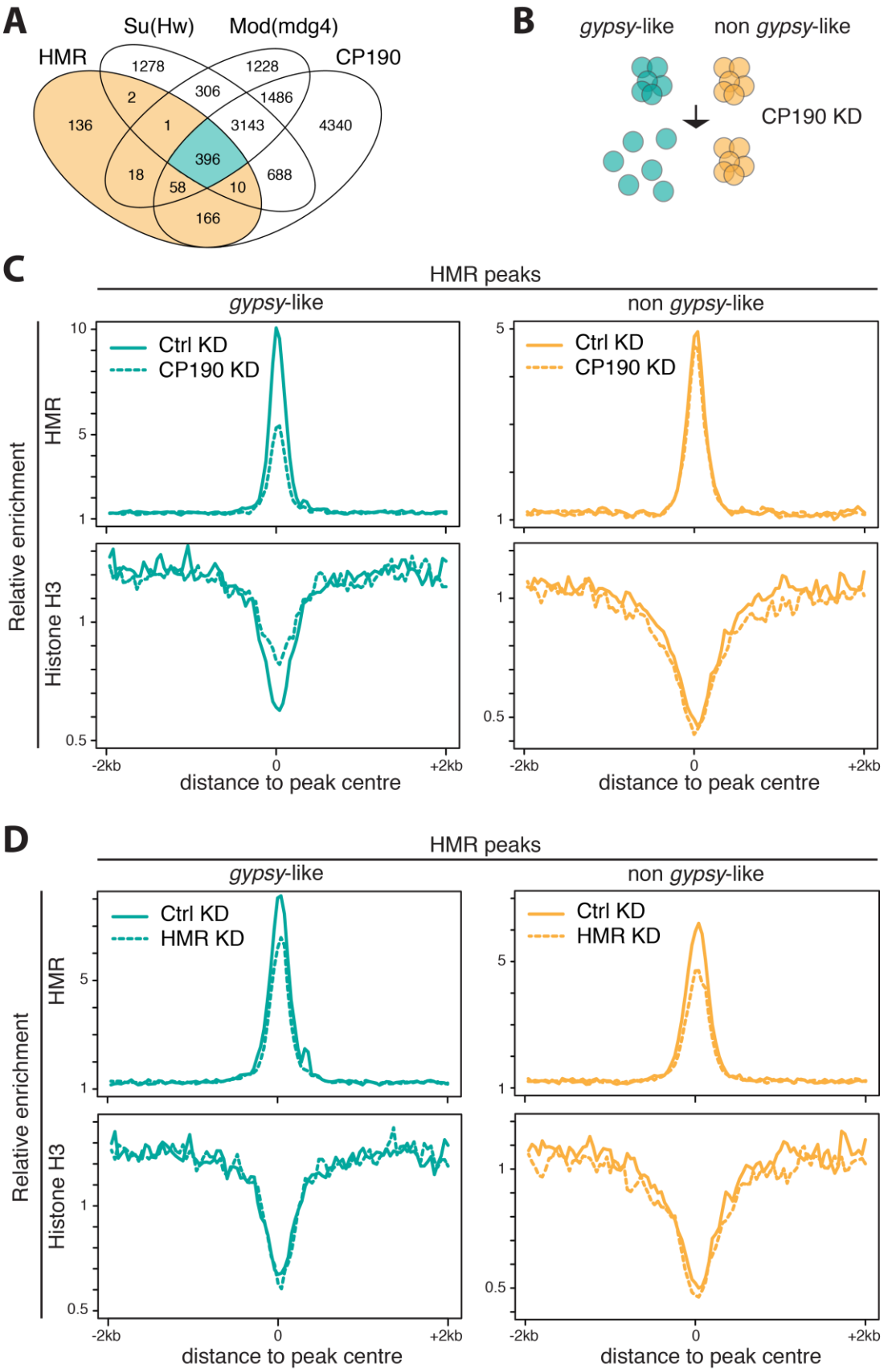

Figure 24. Insulator protein complexes are crucial for HMR's genome-wide localization. **(A)** Venn diagram of HMR ChIP peaks with ChIP peaks of the *gypsy* insulator proteins CP190, Mod(mdg4) and Su(Hw) (Ong et al., 2013). HMR binding sites are grouped in *gypsy*-like group (green) and non *gypsy*-like group (orange). **(B)** CP190 serves as an adaptor

protein for *gypsy* insulator complexes and is crucial for complex stability at chromatin (Schwartz et al., 2012). **(C, D)** Composite plot of HMR and Histone H3 ChIP signals at HMR binding sites after CP190 (C) and HMR knockdown (D). Monitored groups are defined in (A).

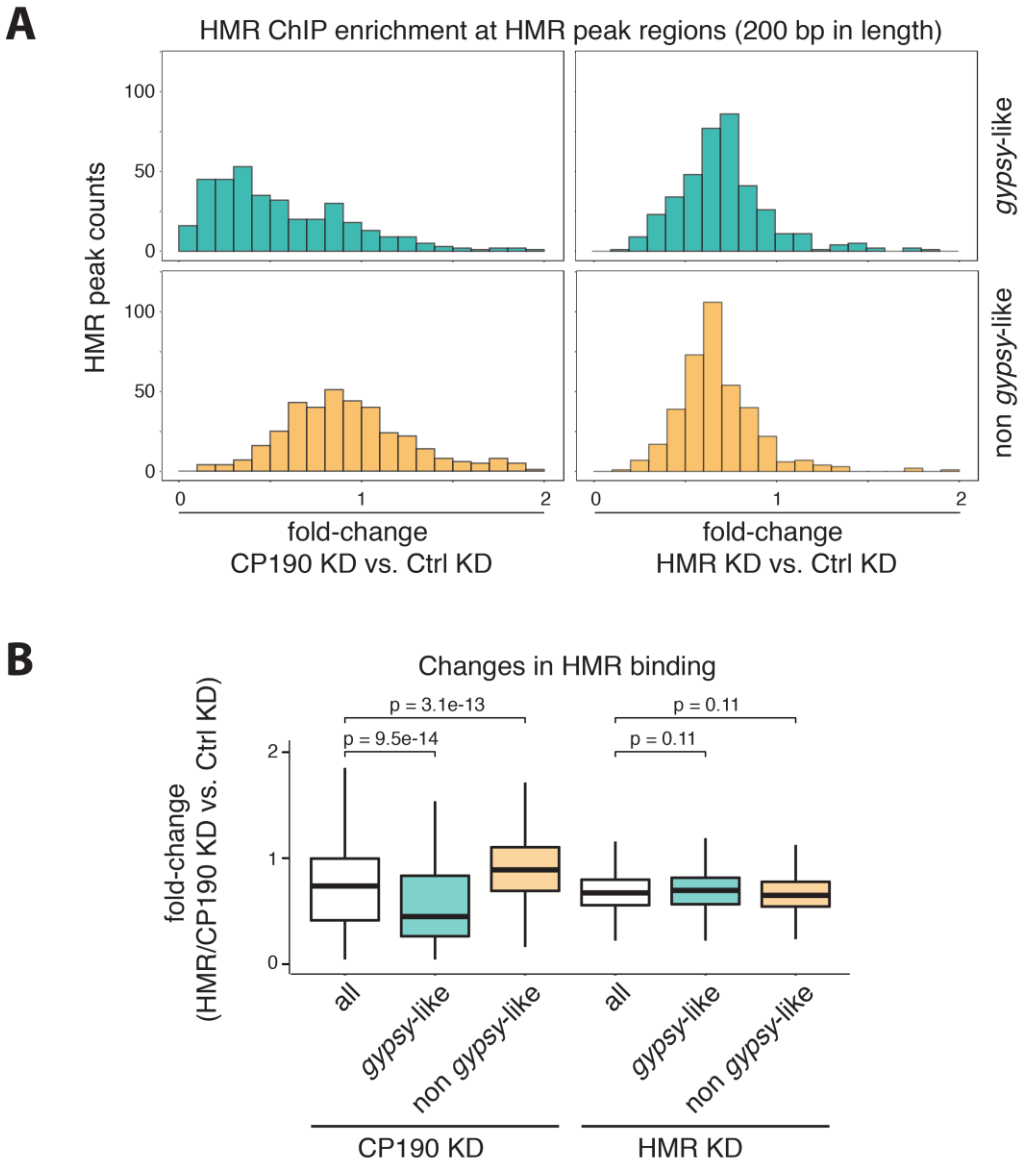

Figure 25. Insulator protein complexes are crucial for HMR's genome-wide localization - Quantification. (A) Histograms showing the frequency of HMR peaks that display HMR ChIP signal reduction after CP190 and HMR knockdown in *gypsy*-like and non *gypsy*-like HMR peak regions. Analyzed are the fold changes of input-normalized HMR ChIP tag number mapped to a 200 bp HMR peak region. Peak regions with less than 50 aligned tags were excluded from the analysis. **(B)** Analysis of data shown in (A). Box plots represent fold-change of HMR ChIP signal. Significance of difference was estimated with p-values calculated with Wilcoxon rank sum test (Wilcoxon, 1946). For both box plots, the box represents the interval that contains the central 50 % of the data with the line indicating the median. The length of the whiskers is 1.5 times the interquartile distance (IQD).

3.7 HMR does not localize to insulator bodies and HMR's localization to centromere-proximal regions is independent from CP190

In cytological staining, nuclear *gypsy* insulator complexes concentrate in structures, so called insulator bodies, which associate with the nuclear lamina (Byrd and Corces, 2003). The formation of insulator bodies correlates with *gypsy* insulator function and requires the *gypsy* insulator proteins Su(Hw), CP190 and Mod(mdg4) (Capelson and Corces, 2005, 2006). The *gypsy* insulator proteins and also CTCF are present in *gypsy* insulator bodies. It was shown that CTCF and Mod(mdg4) fail to form insulator bodies in the absence of CP190 (Gerasimova et al., 2007). We aimed to investigate whether HMR also localizes to such insulator bodies and whether insulator bodies are dependent on HMR. To do so, we monitored the cellular localization of HMR, LHR and CTCF after HMR, LHR and CP190 RNAi knockdown by immunofluorescence microscopy. RNAi efficiency was tested by probing protein levels in cell lysates (analogous shown in **Figure 23 A**). The CP190 knockdown serves as a positive control. The loss of CP190 results in a loss of insulator bodies marked by CTCF whereas HMR or LHR loss does not affect insulator bodies' integrity (**Figure 26**). HMR and LHR localize to distinct cellular foci which were identified as centromeres or centromere-proximal regions in independent experiments (**Figure 12 B** and (Thomae et al., 2013)) and are not overlapping with CTCF signals in immunofluorescence. In addition, the cytological localization of HMR and LHR is mutually dependent but is not affected by CP190 knockdown. This indicates that HMR's localization to centromere-proximal regions is independent of CP190.

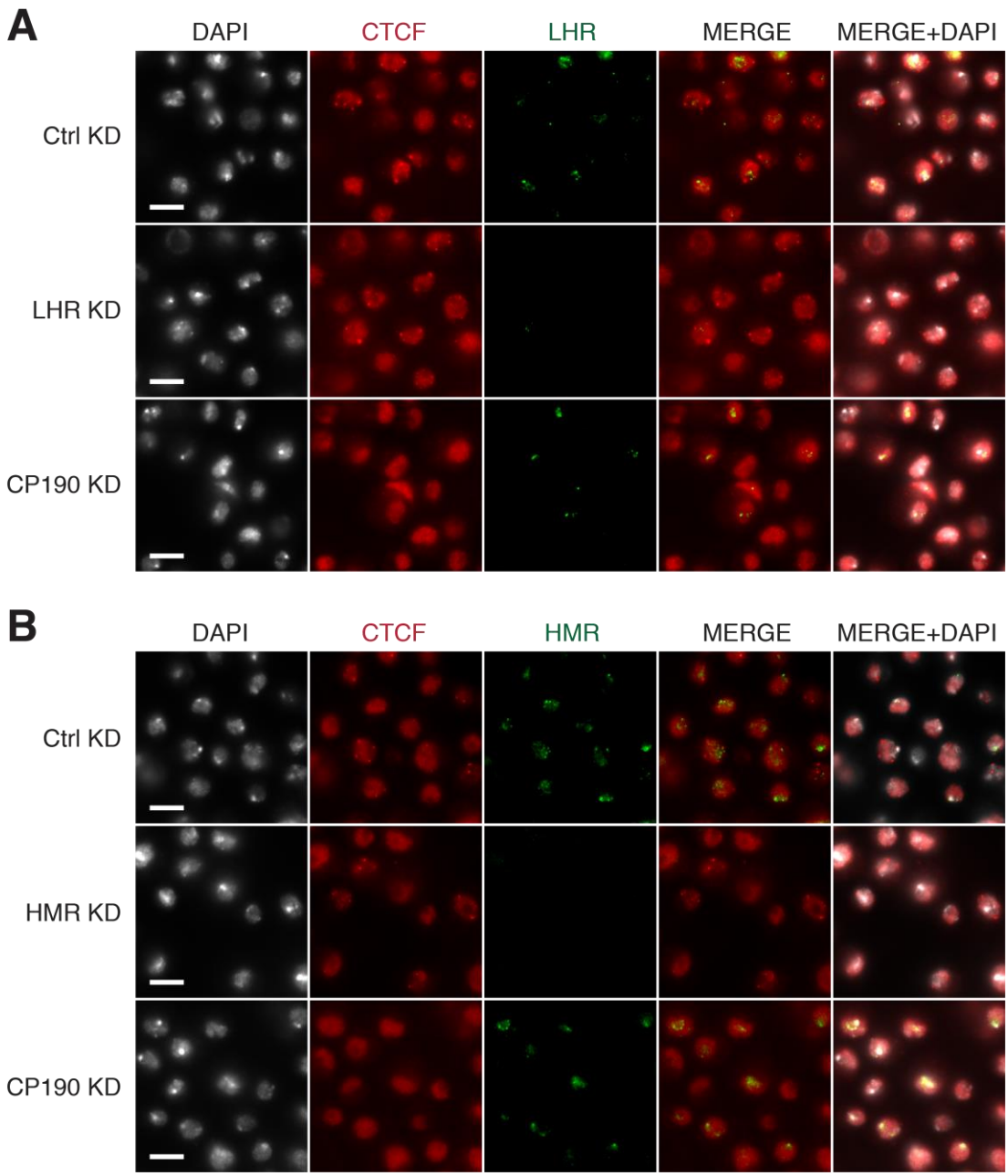

Figure 26. HMR's and LHR's localization to centromere-proximal regions is CP190-independent. (A) Immunolocalization of LHR and CTCF in LHR knockdown and CP190 knockdown S2 cells. LHR signal is lost in LHR knockdown, CTCF-stained insulator bodies are lost in CP190 knockdown. **(B)** Immunolocalization of HMR and CTCF in HMR knockdown and CP190 knockdown S2 cells. HMR signal is lost in HMR knockdown, CTCF-stained insulator bodies are lost in CP190 knockdown. Neither HMR, nor LHR immunolocalization are affected in CP190 knockdown. Scale bars represent 5 μm. The experiment was performed in two independent biological replicates (one replicate is shown).

3.8 HMR associates with chromatin state borders at active genes

The vast majority of HMR binding sites overlap with genomic insulator sites. Insulators have been reported to coincide with transcription start sites (TSS) and chromatin domain borders (Cuddapah et al., 2009; Guelen et al., 2008; Schwartz et al., 2012; Sexton et al., 2012). Based on the combinatorial pattern of 18 histone modifications, the *Drosophila* S2 cell genome was partitioned into nine chromatin states (Kharchenko et al., 2011).

To investigate whether HMR binding sites mark chromatin domain borders at TSS, we used this data set and monitored the annotated states (color representation as described in (Kharchenko et al., 2011)) relative to the position of HMR peaks. We sorted the genomic regions according to non-bordering (annotation is identical up- and downstream of peak summit) or bordering (annotation is different up- and downstream of peak summit). Within these groups, we sorted the genomic regions, first, according to the states upstream of the HMR peak summit and, second, according to the annotation downstream of the HMR peak summit (**Figure 27 A**).

A major fraction of the HMR binding sites is found at transcriptionally active TSS (red, state 1) and intronic regions (brown, state 3). Strikingly, approximately half of the HMR binding sites border chromatin states (**Figure 27 A**). At these borders, HMR is predominantly associated to transcriptionally active TSS (red, state 1) or separates domains from extensive silent domains (white, state 9). An enlarged view on a bordering subset, aligned with our HP1a and HMR ChIP data (**Figure 27 A**), demonstrates reliability and mutual comparability of our ChIP data and the chromatin states classification. HMR signal is distinctly placed at domain borders, whereas HP1a signal resembles heterochromatin regions (dark blue, state 7). When we applied the 5-state annotation, based on Filion et al. (Filion et al., 2010), we could confirm the classification into the distinct chromatin states. However, when applying this annotation, we could not dissect for non-bordering and bordering sets, presumably due to limited resolution (**Figure 27 A**).

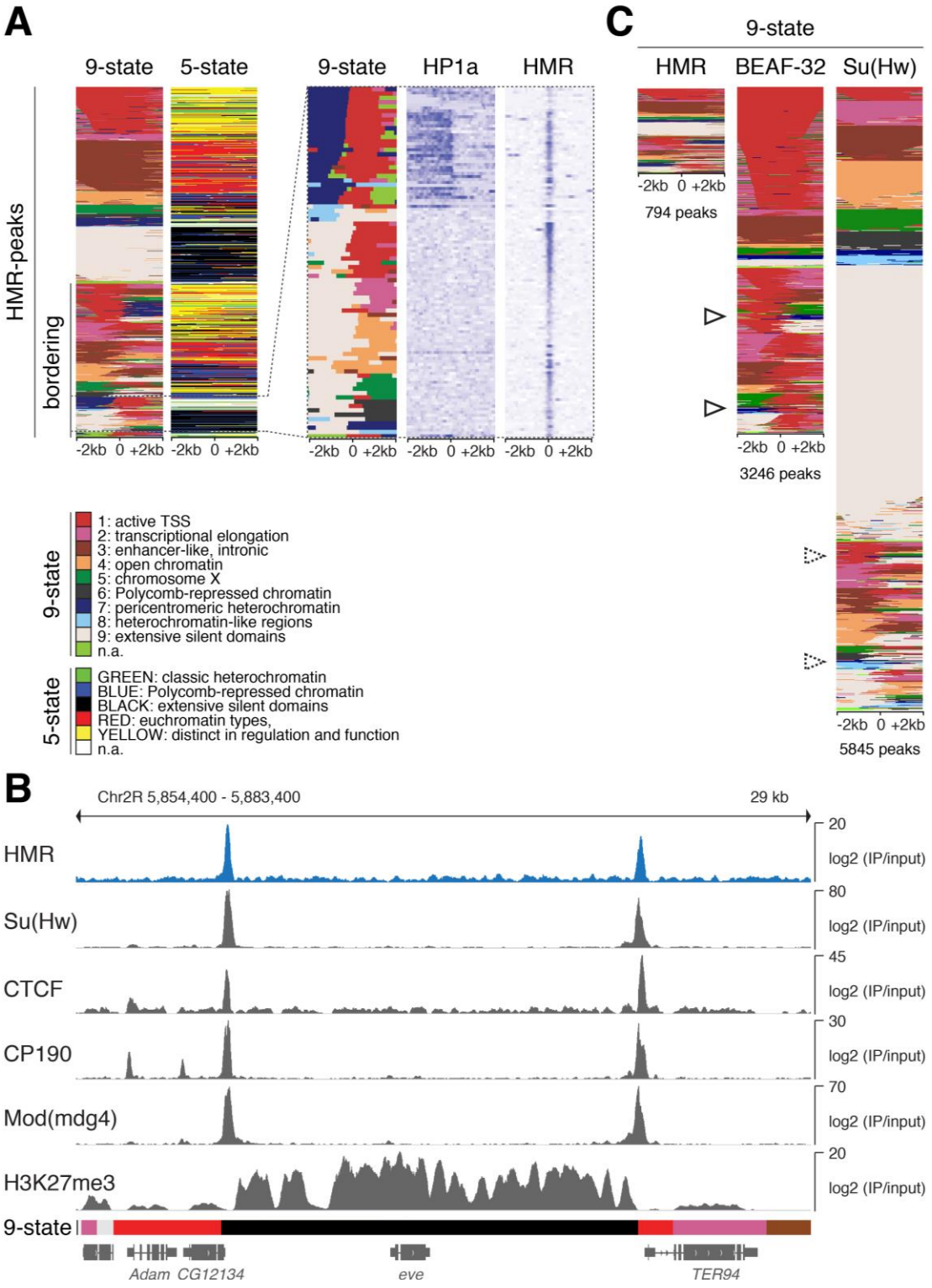

Figure 27. HMR and the associated insulator proteins BEAF-32 and Su(Hw) border chromatin domains at active TSS. (A) Chromatin state annotation at HMR binding sites. Genomic sites are clustered according to the spatial pattern of chromatin states upstream and downstream of the HMR binding site. Each row represents a HMR binding site. Chromatin state annotation is taken from (Kharchenko et al., 2011) (9-state) and (Filion et al., 2010) (5-state). Zoom-in additionally shows heatmaps of HMR and HP1a ChIP signals. **(B)** Genome browser view of HMR, Su(Hw), CTCF, CP190, Mod(mdg4) (Ong et al., 2013) and Histone mark H3K27me3 (Negre et al., 2011) ChIP signals at *even-skipped* (*eve*) gene locus. **(C)**

Same as (A) but showing BEAF-32 and Su(Hw) binding sites (Ong et al., 2013) in addition to HMR binding sites. HMR and BEAF-32 border heterochromatin states at TSS whereas Su(Hw) does not (depicted with arrows).

Insulator proteins have been shown to mark H3K27me3 chromatin domains and to be functionally relevant at these sites by demarcating chromatin domain borders (Bartkuhn et al., 2009; Cuddapah et al., 2009; Negre et al., 2010; Van Bortle et al., 2012). We also find HMR associated to these genomic sites (**Figure 27 B**). The *gypsy* insulator can modulate these chromatin structures and block the spreading of the H3K27me3 histone mark (Comet et al., 2011; Kahn et al., 2006). This restrictive role on H3K27me3 mark spreading has been confirmed in cultured cells, but only for domain borders adjacent to non-active genes. At the borders adjacent to active genes, insulators seem to be dispensable for H3K27me3 domain maintenance (Schwartz et al., 2012).

The genome-wide analysis of HMR binding sites showed that HMR overlaps with Su(Hw) and BEAF-32 binding sites. To dissect insulator- and HMR-specific differences, we next applied the same analysis on Su(Hw) and BEAF-32 binding sites. Interestingly, Su(Hw), BEAF-32 and HMR show distinct properties in chromatin domain bordering. HMR and BEAF-32 specifically border heterochromatin whereas Su(Hw) does not (**Figure 27 C**, arrow).

In summary, HMR borders chromatin states. We supplement this finding with two independent chromatin state annotations and our own ChIP data on the heterochromatin mark HP1a. HP1a chromatin domain borders associate to HMR and BEAF-32 binding sites.

3.9 HMR borders HP1a domains at active promoters together with BEAF-32

HMR and the *gypsy* insulator DNA binding protein Su(Hw) share genomic binding sites and mark chromatin domain borders. Focusing on HMR- and Su(Hw)-specific differences, we realized that HMR borders HP1a chromatin regions whereas Su(Hw) does not. In fact, there is a second class of HMR binding sites that is Su(Hw)-independent but associates with HP1a and BEAF-32 (**Figure 28 A**).

To investigate these differences genome-wide, we sorted the HMR binding sites according to the presence of HP1a and compared the two classes of HMR binding sites against each other. Class 1 associates with HP1a and borders HP1a chromatin at

the TSS of actively transcribed genes (124 out of 794) whereas class 2 contains the major fraction of HMR binding sites (670 out of 794) but does not associate with HP1a (**Figure 28 A**). When we compared class 1 and class 2 for their distribution across the genome, we found class 1 enriched at region 31, centromere-proximal regions and the 4th chromosome, whereas class 2 binding sites are rather equally distributed (**Figure 28 B**).

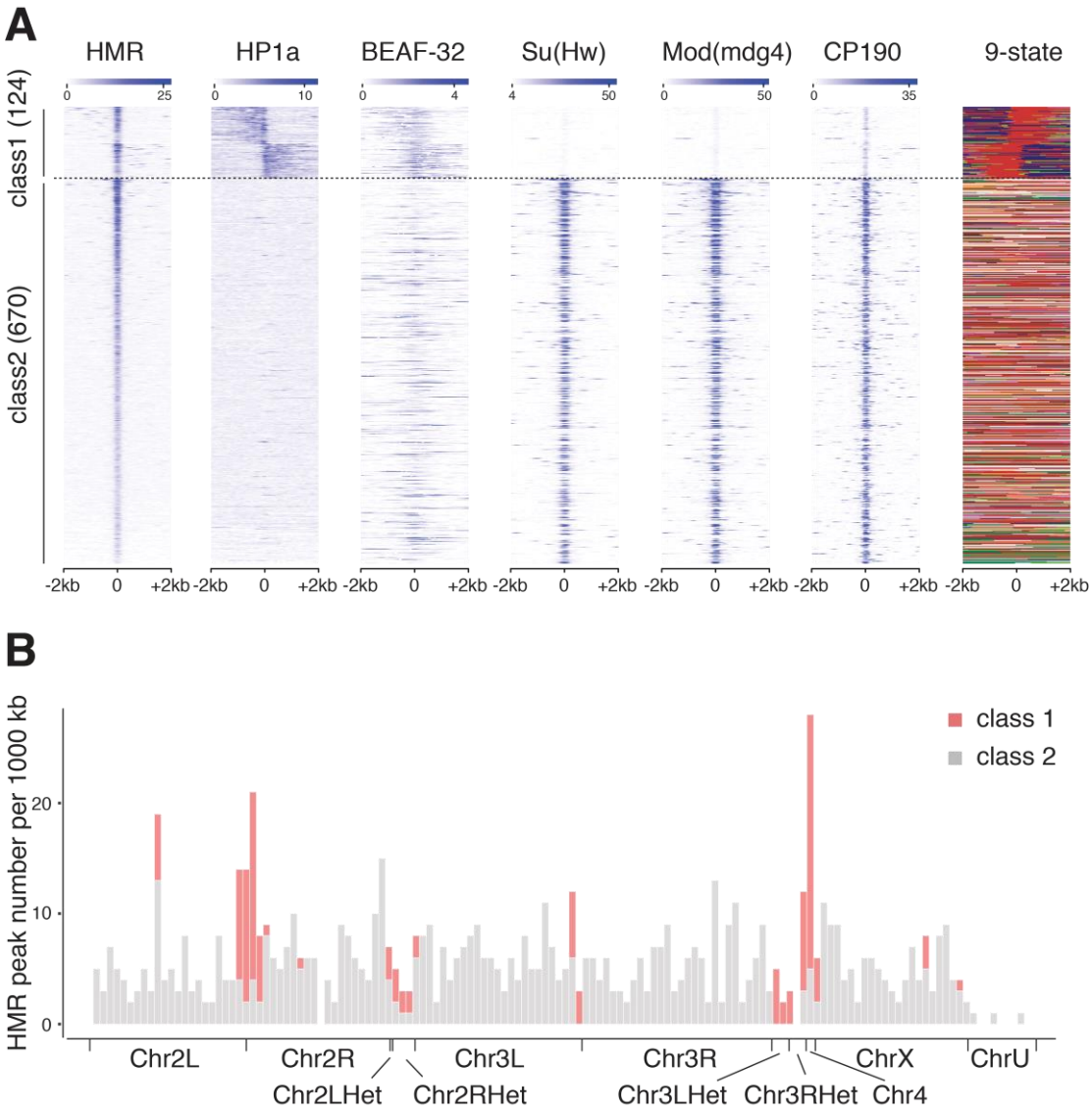

Figure 28. HMR borders heterochromatin at centromere-proximal regions, region 31 and the 4th chromosome. (A) Heatmaps of HMR, the heterochromatin marks HP1a and the insulator proteins BEAF-32 (Riddle et al., 2011), Su(Hw), Mod(mdg4) and CP190 (Ong et al., 2013) ChIP signals. Genomic regions are centered on HMR binding sites, clustered according to adjacent HP1a signals and sorted by HMR signal intensity. **(B)** Histogram showing HMR peak density across the annotated *D. melanogaster* genome. Class 1 HMR binding sites are enriched at region 31, centromere-proximal regions and the 4th chromosome.

We compared the class 1 and class 2 binding sites for their ChIP signal overlap with insulator proteins. The HP1a-associated HMR binding sites (class 1) are essentially lacking Su(Hw) and other *gypsy* insulator proteins but are enriched for BEAF-32 binding sites (**Figure 29 A**). Same is true for the underlying consensus sequences identified by HOMER. The BEAF-32 binding motif is predominantly present in class 1 whereas Su(Hw) binding motifs are predominantly present in class 2 (**Figure 29 B**).

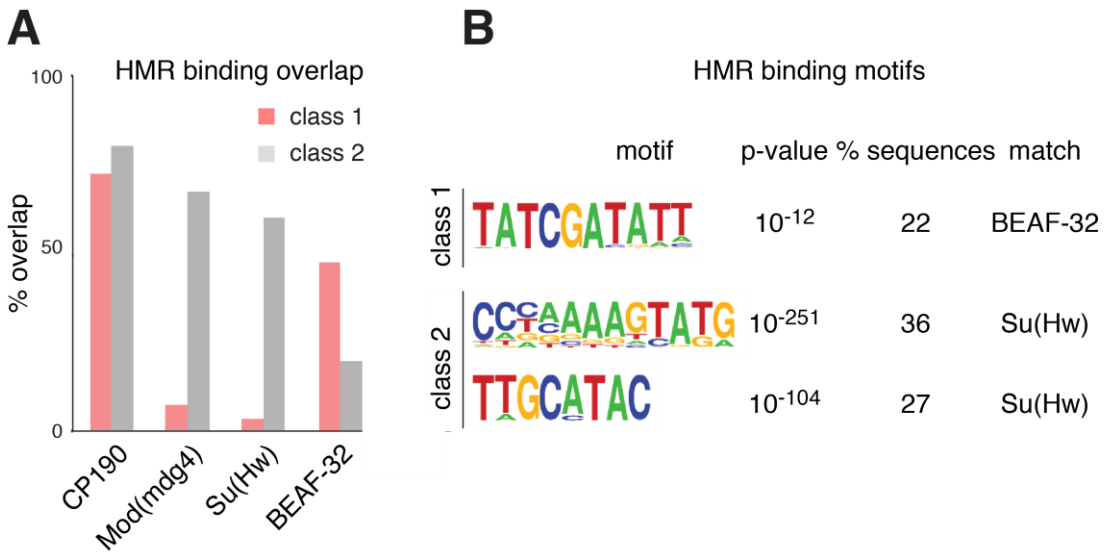

Figure 29. HMR overlaps with BEAF-32 binding sites at heterochromatin domain borders. (A) ChIP peak overlap of the insulator proteins CP190, Mod(mdg4), Su(Hw) (Ong et al., 2013) and BEAF-32 (Riddle et al., 2011) with class 1 and class 2 HMR binding sites. **(B)** Sequence motifs identified within class 1 and class 2 HMR peak regions using HOMER motif search. Motif logo, p-value of enrichment, percentage of regions with this motif and putative binding factors are indicated.

Class 1 and class 2 are also different when investigating their annotation. According to HOMER peak annotation, almost all of the HP1a-associated HMR binding sites (class 1) are in close proximity to transcription start sites (TSS) (90 %), whereas class 2 binding sites are annotated as various genomic elements (**Figure 30 A**). For class 1, HMR binds to very narrow regions placed between the HP1a-marked regions and the gene body. To gather information on the transcriptional activity and characteristics of these genes, we monitored transcript levels of the HMR-bound genes. Analysis of RNA-seq based transcription profiling from *Drosophila* S2 cells demonstrates that the genes bound by HMR are actively transcribed (**Figure 30 B**). In agreement with that, HMR binding correlates with the acetyltransferase CREB binding protein (dCBP/p300), H3K27ac and H2BK5ac when subjecting HMR peak

positions to i-cisTarget web tool (Herrmann et al., 2012) (data not shown). In *Drosophila*, the two histone marks correlate well together and likely give redundant information (Karlic et al., 2010). H3K27ac is thought to be produced by dCBP/p300, which is present at enhancers and promoters (Tie et al., 2009).

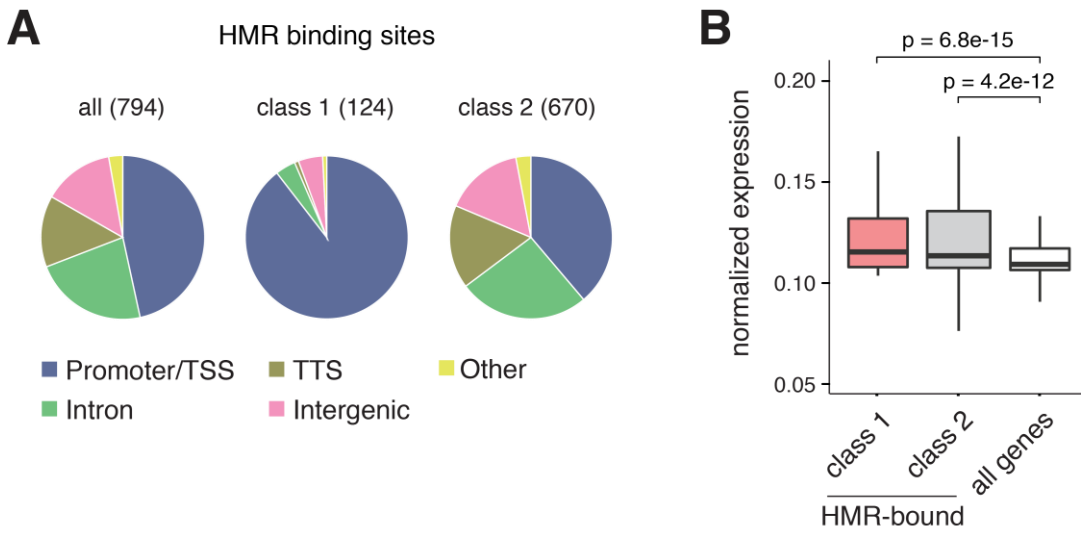

Figure 30. HMR borders heterochromatin domains at active gene promoters (A) Distribution of class 1 and class 2 HMR binding sites for various genomic annotations assigned by HOMER. **(B)** Box plot comparing normalized expression level of all genes to HMR-bound genes (promoter/TSS annotated, see (A)). S2 cell mRNA levels were used according to (Rus et al., 2013). Significance of difference was estimated with p-values calculated with Wilcoxon signed rank test (Wilcoxon, 1946). For both box plots, the box represents the interval that contains the central 50 % of the data with the line indicating the median. The length of the whiskers is 1.5 times the interquartile distance (IQD).

Overall, these results demonstrate the presence of two distinct HMR binding site groups when comparing to HP1a. First, HMR can associate to *gypsy* insulator sites, which are not associated to HP1a chromatin and distributed throughout the genome. Second, HMR can associate to BEAF-32 binding sites which border HP1a chromatin at the TSS of actively transcribed genes.

3.10 HMR promotes transcription at HP1a domain borders

We identified a group of HMR binding sites, which border heterochromatin domains and overlap with BEAF-32 binding sites. At these sites, HMR is predominantly enriched near transcription start sites of highly expressed genes. To monitor a potential role of HMR in transcription of protein coding genes, we took use of transcription analysis data derived from *Hmr* mutant larvae and ovary tissue (Satyaki et al., 2014). In these studies, HMR was reported to impact preferentially heterochromatic gene transcription (Satyaki et al., 2014). In order to assess HMR binding site-specific differences, we compared the transcript changes in class 1 and in class 2. We selected the TSS-annotated HMR binding sites and analyzed the associated gene transcription. In *Hmr* mutant tissue, transcript levels of class 1-associated genes are reduced compared to wild type (**Figure 31**), indicating that HMR promotes gene transcription at HP1a domain borders.

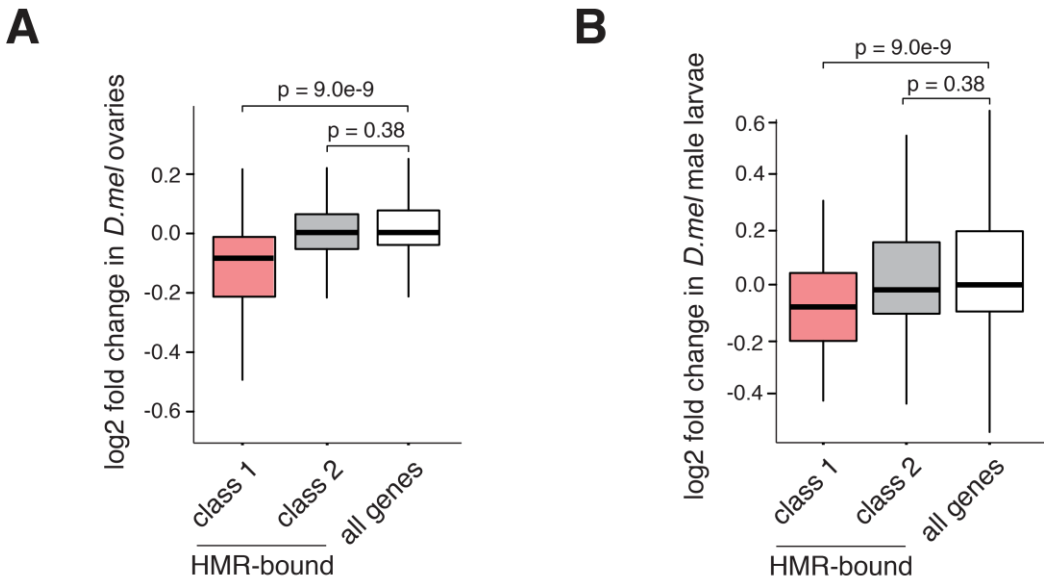

Figure 31. HMR enhances gene transcription at HP1a domain borders (class 1 HMR binding sites). Box plot showing the log2 fold change of protein coding gene transcripts of all analyzed genes and HMR-bound genes (promoter/TSS annotated) in class 1 and in class 2. The RNA-seq data comes from experiments done in *D. melanogaster* ovaries **(A)** and *D. melanogaster* larvae **(B)** (Satyaki et al., 2014). Significance of difference was estimated with p-values calculated with Wilcoxon rank sum test (Wilcoxon, 1946). For both box plots, the box represents the interval that contains the central 50 % of the data with the line indicating the median. The length of the whiskers is 1.5 times the interquartile distance (IQD).

Given the extensive binding overlap between HMR and BEAF-32 at the transcriptionally affected class 1 HMR binding sites, we next asked whether BEAF-32 has an impact on the transcription of these genes, too. We reanalyzed transcription analysis data of BEAF-32 mutant larvae wing imaginal disc tissue and S2 cell transcription after BEAF-32 knockdown with respect to HMR binding sites (**Figure 32**, similar as in **Figure 31**). In addition to class 1 and class 2 HMR binding site-associated genes, we monitored control groups of BEAF-32-bound and Su(Hw)-bound genes. Comparing these groups among each other, we can conclude that the loss of HMR and BEAF-32 result in similar transcriptional changes for class 1-associated genes and genes bound by HMR and BEAF-32 (independent from the classification given in **Figure 28**) (**Figure 32**). Notably, the effect on transcription in S2 cells is substantially lower than in the other monitored systems (**Figure 32**). Even though HMR and BEAF-32 bind to the affected regions in S2 cells, their function at such genomic sites might be rather relevant in more specific tissue or distinct times during development.

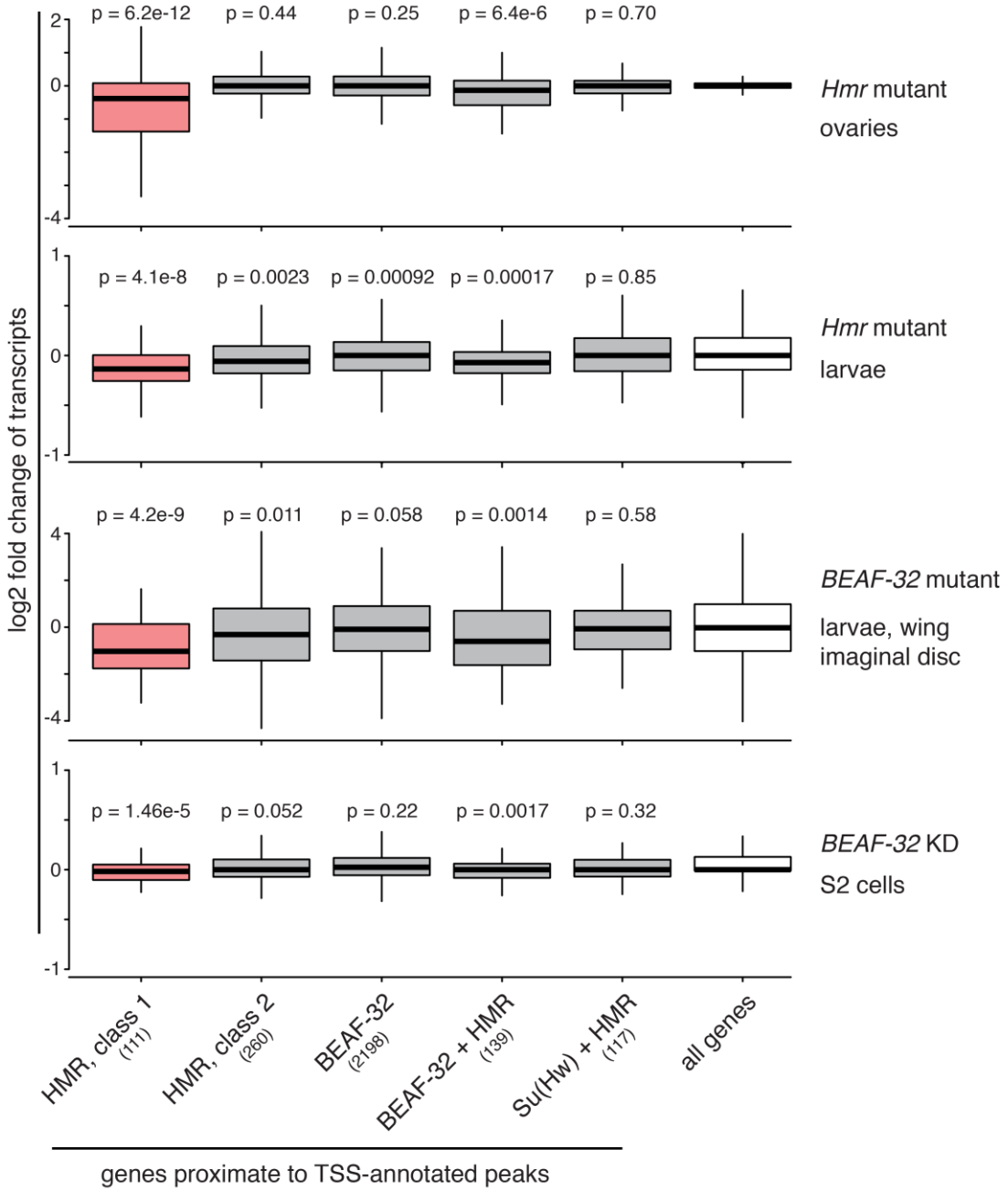

Figure 32. BEAF-32 and HMR enhance gene transcription at shared binding sites and class 1 HMR binding sites. Box plot showing the log2 fold change of protein coding gene transcripts of analyzed genes in proximity to TSS-annotated HMR, BEAF-32 and/or Su(Hw) binding sites. The RNA-seq data comes from publically available data sets (MATERIALS & METHODS, **Table 13**) (Gurudatta et al., 2012; Lhoumaud et al., 2014; Satyaki et al., 2014). Significance of difference was estimated with p-values calculated with Wilcoxon rank sum test (Wilcoxon, 1946). The box represents the interval that contains the central 50 % of the data with the line indicating the median. The length of the whiskers is 1.5 times the interquartile distance (IQD).

3.11 HMR at heterochromatin borders does not affect heterochromatin maintenance and localizes CP190-independent

We identified HMR being very precisely located between HP1a domain and the body of genes. The transcription of such genes seems to be promoted by HMR. The role of HP1a on gene transcription is described recently to be very diverse (reviewed in (Yasuhara and Wakimoto, 2006)). From a naive point of view, HP1a can either enhance the transcription of associated genes or can repress transcription by spreading into the active gene body. Given the strong association of HMR to the HP1a chromatin border, we wondered whether HMR has a functional role in maintaining HP1a chromatin integrity or in preventing HP1a spreading. To investigate the effect of HMR on HP1a and the associated histone mark H3K9me3, we performed HP1a and H3K9me3 ChIP after HMR knockdown. First, we analyzed cell lysates after HMR knockdown in Western Blot. HMR loss affects neither the cellular levels of HP1a nor the cellular levels of H3K9me3 (**Figure 33 A**). To test whether HMR loss affects their localization manner and enrichment, we monitored HP1a and H3K9me3 ChIP signal in class 1. Reorientation of class 1 genomic regions (**Figure 33 B**) allows displaying the HMR-marked boundary of heterochromatin and active gene body. Interestingly, neither HP1a ChIP signal, nor H3K9me3 ChIP signal are changed after HMR knockdown (**Figure 33 C**). This finding indicates that HMR neither maintains heterochromatin integrity at these sites, nor prevents the spreading of heterochromatin into the active gene body. Interestingly, the ChIP signal pattern of HP1a and H3K9me3 differs at these sites. The HP1a signal covers HMR binding sites whereas H3K9me3 signal is present beyond the HMR binding sites. These sites are depleted for nucleosomes, a feature that was associated with insulator complex structures (**Figure 24 C**, **D**) (Bartkuhn et al., 2009). The loss of CP190 results in increased nucleosome occupancy and loss of HMR binding at Su(Hw)-bound *gypsy* insulator sites (**Figure 24 C**). Both observations are in agreement with CP190's importance for *gypsy* insulator complex formation (Schwartz et al., 2012). In contrary, class 1 HMR binding sites are essentially lacking Su(Hw) but are enriched for BEAF-32. To assess CP190's role at these sites, we assayed HMR and Histone H3 ChIP signal after CP190 knockdown (**Figure 33 D**). In contrast to HMR's *gypsy* insulator binding sites (**Figure 24**), class 1 HMR binding sites seem to be CP190-independent. CP190 knockdown neither reduces HMR binding, nor affects nucleosome occupancy at these sites (**Figure 33 D**). This observation is in agreement

with previous studies on the binding dependency of BEAF-32 and CP190. Both proteins share numerous binding sites, but, in contrast to *gypsy* insulator complexes, BEAF-32 binds independent of CP190 (Schwartz et al., 2012).

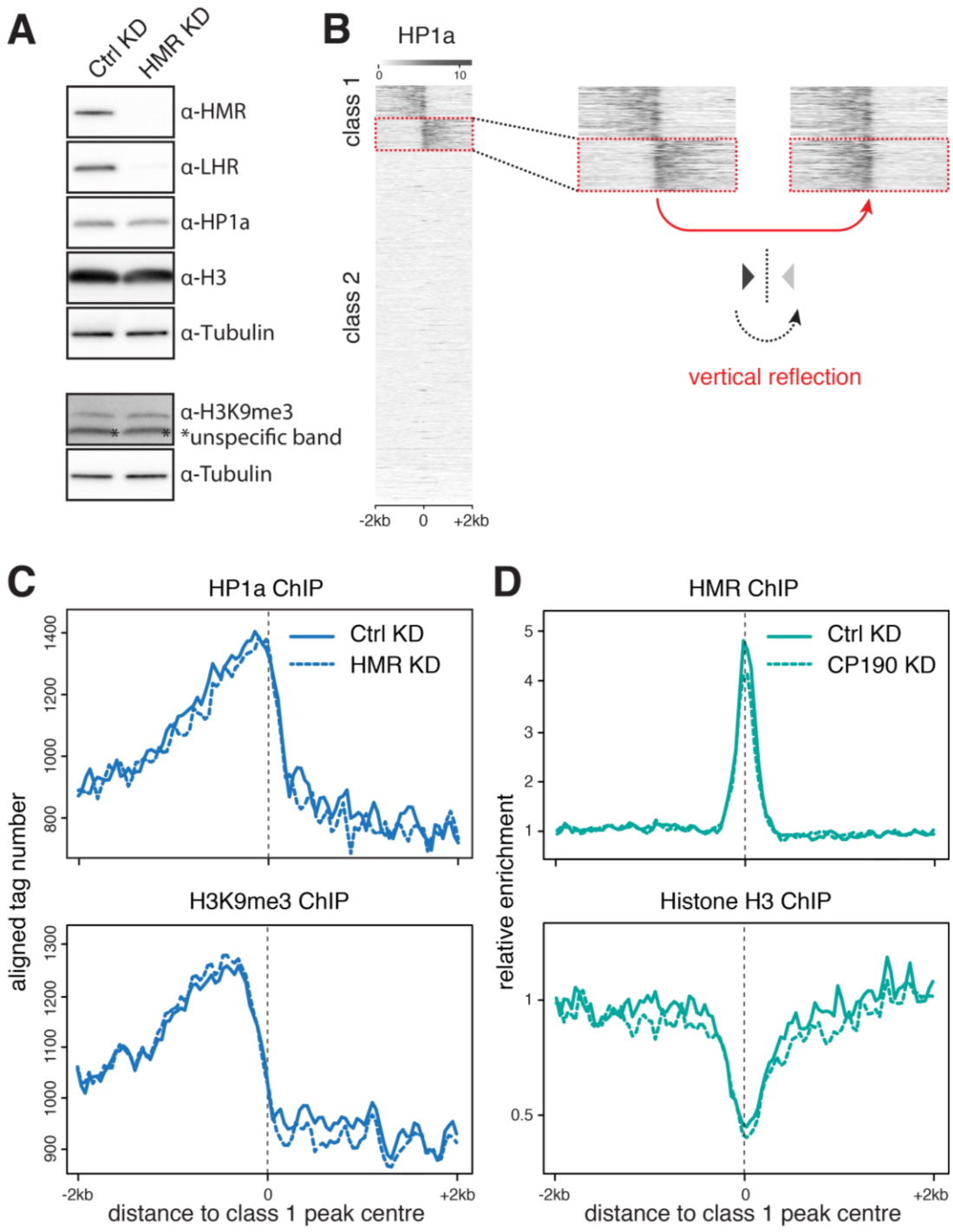

Figure 33. HMR in class 1 HMR binding sites does not affect heterochromatin maintenance and seems to localize CP190-independent. (A) Western Blot analysis on cell lysates to assay protein levels after HMR knockdown. Tubulin protein detection served as control. **(B)** Line out for class 1 HMR binding sites orientation for analysis of heterochromatin after HMR knockdown. Bottom cluster of HMR binding sites within class 1 is vertically reflected. **(C)** Composite plots of HP1a and H3K9me3 ChIP signals at class 1 HMR binding sites after HMR knockdown. Orientation of genomic locations as described in (B) **(D)** Composite plots of HMR and Histone H3 ChIP signals at class 1 HMR binding sites after HMR knockdown. Orientation of genomic locations as described in (B).

4 DISCUSSION

The protein HMR causes reproductive isolation of two closely related *Drosophila* species. HMR in *D. melanogaster* was intensively studied over the last decades and served as the most popular example of a speciation protein that associates with chromatin. However, a detailed understanding of HMR's binding to the genome and the principles underlying HMR targeting to its binding sites was lacking. To the best of our knowledge this study presents the first genome-wide high resolution map of HMR's binding across the *D. melanogaster* genome and provides a novel link between HMR and genomic insulator sites, a class of regulatory elements that rapidly evolved in *Drosophila* and gave rise to several *Drosophila*-specific proteins.

4.1 HMR localizes to *gypsy* insulators and BEAF-32 insulators at heterochromatic regions

Using anti-HMR antibody in chromatin-immunoprecipitation sequencing (ChIP-seq) experiments in *D. melanogaster* Schneider S2 cells, we demonstrate an extensive localization of HMR to two groups of genomic insulators. The first group, *gypsy* insulators, associates with Su(Hw) and other *gypsy* insulator complex components and locates along euchromatic chromosome arms. The second group, associates with the insulator protein BEAF-32 and locates at heterochromatic regions (illustrated in **Figure 34**). This second group is particularly interesting, as HMR, LHR and HP1a physically interact and were shown to localize at the same genomic binding sites (Alekseyenko et al., 2014; Brideau et al., 2006; Greil et al., 2007; Satyaki et al., 2014; Thomae et al., 2013). In our experimental setup, HMR locates to transcription start sites at the boundaries between HP1a-marked heterochromatin and active genes together with BEAF-32. These binding sites are embedded within larger heterochromatic regions such as pericentromeric regions, region 31 and the 4th chromosome (illustrated in **Figure 34**). Notably, we do not detect HMR as a constitutive component of heterochromatin such as HP1a or the histone mark H3K9me3.

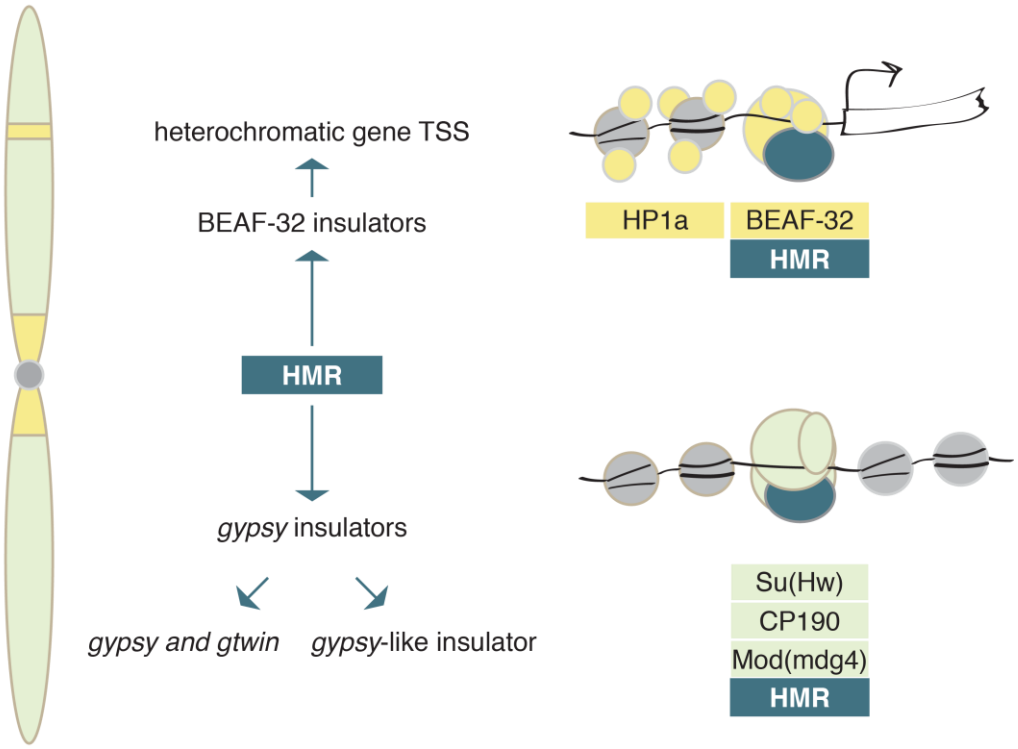

Figure 34. HMR localizes to euchromatic *gypsy* (green) and heterochromatic BEAF-32 insulators (yellow). HMR-bound *gypsy* insulators localize along chromosome arm regions and at the 5' end of the repetitive elements *gypsy* and *gtwin*. HMR-bound BEAF-32 insulators localize at heterochromatic regions, primarily centromere-proximal regions, region 31 and the 4th chromosome. HMR borders HP1a-marked heterochromatin domains together with BEAF-32 at the transcription start site (TSS) of actively transcribed heterochromatic genes.

To the best of our knowledge, this study for the first time acquired high-resolution binding data of HMR under endogenous expression. The difference between this study and prior studies could be due to technical limitations such as mapping resolution or spatial and temporal HMR expression differences. Cytological staining of polytene chromosomes identified HMR and its interaction partner LHR at heterochromatic regions across chromosome arms, in particular at cytological region 31 (Thomae et al., 2013). We found this region particularly enriched for the BEAF-32-associated group of HMR binding sites. Further, we found HMR binding sites enriched at pericentromeric heterochromatin, an observation that resembles the centromere-proximal binding of HMR and LHR (Thomae et al., 2013). However, recent studies also report HMR localizing with repeats such as *2L3L*, *Dodeca* and *GA*-rich satellites (Satyaki et al., 2014), a finding that we cannot confirm with our ChIP experiments. ChIP data interpretation relies on the mapping of short sequences to the genome. In a conventional ChIP-seq analysis that only analyzes uniquely mapped reads, *Drosophila* centromere core regions and other highly repetitive arrays of

genomic DNA are excluded (Sun et al., 2003; Sun et al., 1997). To address this limitation, we performed repeat analysis for HMR and HP1a ChIP and confirmed that HP1a is associated to repeats such as satellite DNA. Consistent with our data on uniquely mapped regions across the chromosome arms (illustrated in **Figure 34**), we find HMR not associated to satellite DNA but to the repetitive elements *gypsy* and *gtwin*, the only repeats that harbor an insulator binding site within their sequence. The convergent results from cytological staining on *Drosophila* polytene chromosomes (Brideau et al., 2006; Thomae et al., 2013) and the ChIP-seq data from this study indicate that HMR's binding is at least partially conserved. However, we cannot rule out that HMR's binding is spatially and temporally regulated and varies over cell cycle, the tissue and the developmental stages. In fact, HMR displays a tissue-specific mRNA (Graveley et al., 2011) expression and a tissue-specific localization in cytological staining (Dr. Andreas Thomae, unpublished). Further, there is increasing evidence for a temporal regulation of HMR over the cell cycle. HMR's binding to centromere-proximal regions is lost in mitosis (Thomae et al., 2013) and potentially linked to a proteasome-mediated degradation of HMR upon mitotic nuclear envelope breakdown. These cell-to-cell differences are not addressable in cell population-based analysis such as ChIP-seq. A comprehensive understanding of HMR's association to chromatin during development, and within different cell types and tissues is still lacking and is going to be a highly interesting aspect for future research.

Our findings further indicate that HMR does not directly bind to DNA. Even though, HMR's MADF domains were associated with direct DNA binding activity (Barbash et al., 2003; Bhaskar and Courey, 2002), we found HMR's binding sites mostly associated with DNA binding motifs assigned to the direct DNA binding activity of Su(Hw) or BEAF-32. The postulated direct DNA binding activity of HMR could be tested using an *in vitro* genome-wide DNA binding assay (Villa et al., 2016). How well such assay reflects the *in vivo* binding of the dosage-dependent HMR remains to be tested.

4.2 HMR's genomic localization depends on insulator complexes

Apart from descriptive binding information on HMR, detailed information on HMR's recruitment to chromatin was lacking. Prior studies showed that the binding of the HMR interaction partner LHR is dependent on HMR but not *vice versa* (Satyaki et al., 2014; Thomae et al., 2013). Further, LHR was reported to localize dependent on HP1a (Greil et al., 2007). However, in the current study as well as in former studies HMR, LHR and HP1a protein levels were shown to be mutually dependent which makes such studies difficult (Satyaki et al., 2014; Thomae et al., 2013). Here, we took advantage of the S2 cell system being suitable for highly efficient RNAi-mediated protein knockdown strategies to test the targeting mechanism of HMR to its genomic binding sites. By efficiently depleting CP190, a crucial component of the *gypsy* insulator complex (Schwartz et al., 2012), we could demonstrate that the binding of HMR to genomic *gypsy* insulators depends on the integrity of the residing insulator complex. One possibility is that HMR interacts physically with one or more components of the *gypsy* insulator complex. Immunoprecipitation experiments on CP190 in embryonic nuclear extract (Moshkovich et al., 2011) identified HMR as a prominent interactor of CP190 (Dr. Elisa Lei, personal communication). Further, Su(Hw) copurified with HMR and LHR while Mod(mdg4) copurified with HMR (Thomae et al., 2013). However, experiments performed by immunoprecipitating endogenously expressed HMR and LHR could not finally confirm the reported interactions (**Figure S2**). To which extent HMR and insulator complex components directly interact remains to be tested. Apart from a direct recruitment of HMR by the insulator complex, it is possible that HMR is directed to its genomic sites by a chromatin feature that is exposed in the presence of the insulator complex. Insulator complexes localize in nucleosome free regions (Bartkuhn et al., 2009; Schwartz et al., 2012). The loss of the insulator complex results in increased nucleosome occupancy at HMR binding sites, which in turn could result in masking some other HMR binding feature. In contrast to the *gypsy* insulator, CP190 loss did not impair the binding of HMR to BEAF-32-associated insulators. It remains to be addressed what recruits HMR to the BEAF-32-associated heterochromatic sites in the genome. At this point, we speculate that HMR is recruited to heterochromatin by the BEAF-32 insulator complex, analogous to the CP190-dependent *gypsy* insulator complex along the chromosome arms. In contrast to *gypsy* insulators, however, BEAF-32 insulators act independently of CP190 (Schwartz et al., 2012). In agreement with that, CP190

knockdown disrupts cellular *gypsy* insulator structures and HMR's binding to genomic *gypsy* insulators but does not disturb HMR's binding to BEAF-32 insulators in ChIP and HMR's centromeric localization in immunofluorescence. However, in contrast to HP1a and *gypsy* insulator complex components, BEAF-32 was not enriched in HMR or LHR immunoprecipitations (Thomae et al., 2013). As HMR and LHR interact with HP1a (Alekseyenko et al., 2014; Brideau et al., 2006; Greil et al., 2007; Satyaki et al., 2014; Thomae et al., 2013), it is also possible that HMR gets recruited to these genomic binding sites by HP1a. Paradoxically, HMR is only detected at HP1a heterochromatin boundaries. A similar pattern was reported for the heterochromatin boundary-factor Epe1 in yeast (Braun et al., 2011). Epe1 is recruited to heterochromatin by physical interactions but is then actively removed from the heterochromatin body by specific polyubiquitylation and degradation and persists only at the boundaries (Braun et al., 2011). HMR localization seems to rather happen downstream of HP1a. This is further supported by the fact that HMR knockdown is not affecting HP1a and H3K9me3 marks, also not at HMR binding sites. However, testing a potential recruitment of HMR by HP1a with HMR ChIP after HP1a knockdown could be difficult as cellular HMR and LHR protein levels are sensitive to HP1a. In such an approach, it has to be clarified whether changes in HMR binding are specific. This can be done by comparing HP1a-associated HMR binding sites against an internal control group which is not associated to HP1a.

4.3 Is HMR a functional insulator complex component?

In the previous sections, localization and recruitment of HMR to its genomic binding sites were discussed. Even though, the HMR binding sites identified in this study share features with insulator protein binding sites, this does not necessarily link HMR to insulator complex function. Whether HMR plays a role at its genomic binding sites in the maintenance or establishment of insulator complexes is not known. CP190 serves as an adaptor protein for *gypsy* insulator complexes, dramatically changes nucleosome occupancy and insulator bodies built by CTCF and *gypsy* (Bartkuhn et al., 2009; Gerasimova et al., 2007; Schwartz et al., 2012). In contrast to that, HMR loss neither influences nucleosome occupancy, nor insulator bodies. Importantly, HMR localizes only to a subset of the identified insulator sites, indicating that HMR acts downstream of general insulator complex components and is not responsible for their maintenance.

Even though HMR's binding seems to depend on the insulator complexes and not *vice versa*, it is possible that HMR establishes insulator complexes *de novo*. This could be mediated by the interaction of HMR with insulator complex components. In an experiment where HMR was artificially targeted between an enhancer and a reporter gene promoter, HMR strongly interfered with the reporter gene expression (Thomae et al., 2013). This property is highly reminiscent of the described enhancer blocking activity of insulator proteins (reviewed in (Gaszner and Felsenfeld, 2006; Schwartz and Cavalli, 2017; West et al., 2002)) and potentially involves the establishment of an insulator complex structure. Such putative insulator recruitment function of HMR could be highly relevant in hybrids, where HMR binding sites are gained (Thomae et al., 2013). HMR's gained binding sites might then serve as recruitment platforms for novel unspecific insulator sites.

In summary, we found HMR not being associated to classical heterochromatin regions but localizing to genomic insulator sites. We could demonstrate that HMR's binding to euchromatic *gypsy* insulators depends on the presence of the *gypsy* insulator complex but not *vice versa*. The targeting mechanism of HMR to BEAF-32 insulators at heterochromatic genes remains to be solved. However, this group of HMR binding sites resembles previous cytological studies that describe HMR being associated to several heterochromatic regions.

4.4 HMR mediates the expression of heterochromatic genes – potentially in concert with BEAF-32

The high density of HMR binding sites at pericentromeric heterochromatin correlates well with the strong colocalization of HMR and the centromeric H3 variant CID in immunolocalization experiments (Thomae et al., 2013). However, the fact that the purification of chromatin containing the centromeric H3 variant CID did not identify HMR argues against HMR being a *bona fide* component of the centromere core region (Barth et al., 2014). Notably, the copies of *gypsy* are located at centromeric and/or pericentromeric regions (Heredia et al., 2004). However, CP190 knockdown that abolishes HMR binding to the 5' *gypsy* insulator does not affect the centromeric localization of HMR, indicating HMR being recruited to the centromere in a different manner. Further, centromeric satellite DNA is lacking in HMR ChIP. We therefore assume that the binding of HMR and BEAF-32 to the TSS of actively transcribed heterochromatic genes result in HMR's cytological localization to the centromere.

Interestingly, HMR loss and BEAF-32 loss result in phenotypes that are similar and linked to centromeric function. In S2 cells, either the knockdown of BEAF-32 or the knockdown of HMR causes mitotic chromosome segregation defects (Emberly et al., 2008; Thomae et al., 2013). Further, flies that carry a mutation in *Hmr* or flies in which BEAF-32 is only contributed maternally both display reduced female fertility (Aruna et al., 2009; Roy et al., 2007). In contrast to BEAF-32, the *gypsy* insulator proteins Mod(mdg4) and CP190 are not required for oogenesis (Baxley et al., 2011; Chodagam et al., 2005). This raises a hypothesis in which the lack of HMR binding to BEAF-32-associated heterochromatic regions is responsible for the female sterility phenotype observed in *Hmr* mutant flies. But how can we explain these phenotypes on a mechanistic level that involves HMR's and BEAF-32's action on chromatin?

It was suggested that the BEAF-32-associated phenotype is caused by the impact of BEAF-32 on the transcription of genes involved in cell cycle regulation (Emberly et al., 2008). Such connection to cell cycle regulator genes was not observed for the genes misregulated in *Hmr* mutant flies (Satyaki et al., 2014). However, our analysis on published transcription data indicates that loss of either HMR or BEAF-32 result in a similar effect on heterochromatic gene transcription. Whether these changes in transcription are due to direct transcriptional regulation or whether these are secondary effects due to the mitotic defect phenotype or an overall altered chromatin structure remains to be solved. Importantly, also Su(Hw) is required for female germline development (Baxley et al., 2011; Klug et al., 1968; Parkhurst et al., 1988). This function however, is independent from the other *gypsy* insulator proteins CP190 and Mod(mdg4). It was shown that Su(Hw) can serve as a transcriptional regulator which has a direct impact on its germline-associated phenotype (Soshnev et al., 2013). Even though a critical Su(Hw)-regulated target gene, *Rbp9* (Soshnev et al., 2013), was not affected in *Hmr* mutant tissue (Satyaki et al., 2014), it demonstrates that a direct regulation of few target genes might also cause HMR and BEAF-32 mutant phenotypes.

The mechanisms by which BEAF-32 affects gene transcription are currently not well understood, but BEAF-32 is often found near paused RNA Polymerase II (Jiang et al., 2009). There is growing evidence that BEAF-32 fulfills a role in gene expression by regulating RNA Polymerase II promoter proximal pausing (Duarte et al., 2016; Li and Gilmour, 2011) (Prof. John T. Lis, personal communication). Additionally, BEAF-32 localizes not only to the promoter of heterochromatic genes but also, together with MOF, upstream of MSL-bound genes (Philip et al., 2012). This is not the case for the

insulator proteins CP190, Su(Hw) and CTCF (Philip et al., 2012). MOF and the MSL complex are part of the dosage compensation system that equalizes X-linked gene expression between males and females (reviewed in (Conrad and Akhtar, 2012; Straub and Becker, 2007)). Even though the lack of BEAF-32 was not found to affect dosage compensation, it disrupts male X chromosome morphology (Roy et al., 2007). Interestingly, male X chromosome morphology is also highly sensitive to Su(var)3-7 and HP1a protein levels (Spierer et al., 2008; Spierer et al., 2005). Both proteins are interactors of HMR (Thomae et al., 2013). Apart from its impact on male X chromosome morphology, the heterochromatic factor Su(var)3-7 is required for oogenesis and female fertility (Basquin et al., 2014). It is tempting to speculate that the morphologic phenotypes of HMR, LHR, BEAF-32 and Su(var)3-7 are not due to site-specific actions but rather due to severe changes in chromatin structure and integrity. Heterochromatin is crucial in mediating chromosome segregation and telomere protection (Allshire and Karpen, 2008; Andreyeva et al., 2005; Raffa et al., 2011). Therefore, HMR-related phenotypes, such as mitotic defects in cultured cells and proliferative fly tissue and telomere lengthening (Satyaki et al., 2014), point towards a functional role of HMR in setting up heterochromatin. Transcription analysis in *Hmr* mutant flies revealed a strong misregulation of heterochromatic genes as well as repeat sequences (Satyaki et al., 2014). The heterochromatic genes affected by HMR are present at centromere-proximal regions and the 4th chromosome, which are both repeat-rich heterochromatic regions (Lohe et al., 1993; Sun et al., 2003; Sun et al., 1997). It is possible that HMR specifically functions on these genes to ensure their expression in a repeat-rich and transcriptionally silent environment. In turn, this gene expression could ensure the surrounding heterochromatin integrity (Yasuhara and Wakimoto, 2006). Mechanistically, HMR could serve as an insulator barrier together with BEAF-32 in antagonizing the spreading of repressive heterochromatin marks (Gaszner and Felsenfeld, 2006; Schwartz and Cavalli, 2017; West et al., 2002).
We can relate the set of misregulated heterochromatic genes well to our HMR binding data. However, we do not find HMR directly binding to misregulated repeats apart from *gypsy* and *gtwin*. This indicates that the effect of HMR to repeat transcription is not mediated by direct binding of HMR. The misregulation of TEs is a stress-response and therefore often associated with stress-related phenotypes such as mitotic defect (Dr. Severine Chambeyron, personal communication). Therefore, the TE misregulation upon HMR loss is not necessarily a direct effect of HMR but could also be a consequence of the mitotic defect phenotype. To shed light on this, we aimed to

test a putative role of HMR on heterochromatin by monitoring HP1a and H3K9me3 binding after HMR knockdown but could not detect any effect of HMR on heterochromatin maintenance. Whether HMR has a direct role in TE repression by establishing or maintaining heterochromatin remains to be tested in future studies. A functional role of HMR in heterochromatin establishment and maintenance could be strongly dependent on development timing and development-specific expression of HMR. The concepts of heterochromatin initiation and maintenance were dissected in *S. pombe* first (Grewal, 2010; Moazed, 2009) but similar is true for *Drosophila*. For example, the *Drosophila* PIWI is crucial for the initiation but not for the maintenance of H3K9me3 mark across repeats in ovaries. This temporal-specific effect of PIWI on H3K9me3 was demonstrated by inducing PIWI knockdown at various developmental stages and monitoring H3K9me3 by ChIP (Dr. Severine Chambeyron, personal communication).

Overall, it is tempting to speculate that HMR is part of a functional protein network together with LHR, Su(var)3-7 and the insulator protein BEAF-32, in setting up chromatin morphology and heterochromatin in particular by maintaining gene expression in a repeat-rich environment.

4.5 How do our findings help in understanding HMR's gain-of-function in hybrids?

Hybrid flies lack imaginal discs and suffer from impaired cell proliferation in mitotic tissues such as the larval brain (Bolkan et al., 2007). HMR and LHR were shown to localize to centromeres and pericentromeric heterochromatin, where they contribute to chromosome segregation in mitosis and to the suppression of transposable elements (Satyaki et al., 2014; Thomae et al., 2013). More recently, it was shown that knockdown of the hybrid incompatibility factor $gfzf_{sim}$ restores cell proliferation in hybrid male larval brain and rescues hybrid lethality (Phadnis et al., 2015). These findings implicate that the reason for hybrid male lethality are cell proliferation defects. In such a scenario, the combination of HMR_{mel} and LHR_{sim} in hybrids causes dysfunction at centromeres or pericentromeric heterochromatin, which then triggers a mitotic cell cycle checkpoint set by gfzf (Phadnis et al., 2015).

How does a better understanding of HMR's binding properties in pure species help in dissecting HMR's deleterious function in hybrids? Importantly, the pure species function is distinct from the detrimental gain-of-function in a hybrid background. As

discussed in the introduction section, HMR is not essential in the pure species background but gains a lethal function in the hybrid background. Polytene chromosomes from hybrid male larvae show a widespread mislocalization of HMR across the genome (Thomae et al., 2013), indicating that genomic localization matters for *D.mel/D.sim* hybrid incompatibility. Thomae et al. could mirror such mislocalization by HMR overexpression and proposed that the gain of HMR binding sites is due to increased HMR and LHR protein levels in hybrids (Thomae et al., 2013).

In this first scenario, HMR's function is conserved across species and differences between species consist primarily in HMR and LHR protein levels. First, amino acid residues that are critical for hybrid incompatibility are conserved across species, and second, *Hmr* orthologs from *D. simulans* and *D. mauritiana* can partially rescue female fertility defects in *Hmr* mutant *D. melanogaster* flies (Aruna et al., 2009; Satyaki et al., 2014). Purely protein dosage-driven changes in localization and function were described for other chromatin components including the heterochromatic Su(var) proteins (Eissenberg et al., 1990; Eissenberg et al., 1992; Locke et al., 1988; Schotta et al., 2002; Spierer et al., 2008). The characteristics of HMR binding sites gained in hybrids are not known. It is possible that HMR localizes to other genomic insulator binding sites, to heterochromatic regions or to other, unrelated genomic regions. At these sites, HMR could have the potential to fulfill its pure species function at a non-designated site in the genome. This could be recruitment of an insulator complex accompanied by chromosome conformational changes, gained enhancer-blocking or barrier activity, establishment of heterochromatin, bidirectional transcription or others. Overall, the range of possibilities is broad and highly speculative.

In a second scenario, hybrid incompatibility is caused by a species-specific function of HMR. HMR orthologues exhibit different properties in a hybrid background. This is indicated by the fact that *D.mel/D.sim* hybrid males rescued by Hmr_{mel} mutation are not killed by *Hmr* orthologs from sibling species (Barbash et al., 2004). Notably, the pure species function that can be complemented with sibling species orthologues (Aruna et al., 2009; Satyaki et al., 2014) is distinct from the detrimental function in hybrids. Even though Thomae et al. concluded that protein interactions are rather similar for HMR and LHR orthologues, the interaction between LHR and Su(Hw), for example, is *D. melanogaster*-specific (Thomae et al., 2013). Such differences might

be sustainable in the pure species background but detrimental in the hybrid background.

Taking both options into account, it could also be that intrinsic HMR properties are crucial in defining HMR protein levels in a hybrid background. Such could be the presence or absence of posttranslational modifications on residues that could result in stabilization or destabilization of the protein or severe posttranslational processing such as enzymatic cleavage accompanied with degradation.

In any case, it is likely that the integrity of heterochromatin is not preserved under these conditions. HMR mutations, HMR knockdown, HMR overexpression and hybrids, all are accompanied by mitotic defects and misregulation of heterochromatic genes and transposable elements (Satyaki et al., 2014; Thomae et al., 2013). Even though we did not observe HMR binding extensively to repeats, we could clearly see that HMR-bound heterochromatic genes are affected in HMR and BEAF-32 mutant flies. Whether the transcriptional changes are cause or consequence of the mitotic defect phenotypes associated with these mutations remains to be solved. Further, transcriptional changes can be attributed to delay in development, differences in tissue types, or impaired heterochromatin. Species-specific heterochromatin can cause mitotic chromosome segregation defects (Ferree and Barbash, 2009) as pericentromeric heterochromatin ensures centromere and kinetochore integrity (reviewed in (McKinley and Cheeseman, 2016)). In addition, heterochromatin mediates chromosome segregation for the X and the 4^{th} chromosome in *D. melanogaster* by the formation of heterochromatic threads in mitosis and meiosis (Baumann et al., 2007; Chan et al., 2007; Dernburg et al., 1996; Hughes et al., 2009; Karpen et al., 1996; Theurkauf and Hawley, 1992). A failure of pairing heterochromatic regions results in a failure of chromosome segregation and mitotic arrest. The *359-bp* satellite repeats on the X chromosome, for example, were shown to cause widespread mitotic defects and hybrid lethality from *D. simulans* mothers and *D. melanogaster* fathers (Ferree and Barbash, 2009).

In this respect, it is also notable that hybrid lethality depends on the *D. melanogaster* X chromosome, as male hybrids that carry a *D. simulans* X chromosome and an autosomal copy of Hmr_{mel} survive (Barbash, 2010; Hutter et al., 1990). The *D. melanogaster* dosage compensation system that upregulates expression on the male X chromosome (reviewed in (Conrad and Akhtar, 2012; Straub and Becker, 2007) shows signs of positive selection and was proposed to cause hybrid lethality by species-specific divergence of the involved components (Pal-Bhadra et al., 2004;

Rodriguez et al., 2007). Using mutants of four main *D. melanogaster* dosage compensation genes, Barbash genetically tested whether failure in dosage compensation decreases hybrid male viability and concluded that it was rather increased (Barbash, 2010). However, it was also shown that the chromatin structure of the male X chromosome is particularly dependent on the dosage of the heterochromatic proteins Su(var)3-7 and HP1a (Spierer et al., 2008; Spierer et al., 2005). Both proteins interact with HMR and LHR (Thomae et al., 2013). Given the additional link of BEAF-32 to dosage compensation, it could well be that these pathways are highly interconnected, for instance by a mass action-based model in which the protein dosage of the corresponding factors is essential (Locke et al., 1988). Based on the law of mass action, a dosage-dependent influence was, for example, proposed for the Su(var) proteins and their extent on the assembly of heterochromatin (Locke et al., 1988). Altogether, it is important to note that hybrid incompatibility cannot be nailed down to *Hmr*, *Lhr*, or other HI genes alone but involves the interplay of such with a species-specific cellular background.

4.6 HMR's divergent evolution could have been triggered by changes in repeat copy number

How hybrid incompatibility genes evolve and what drives their rapid evolution is a long standing question in the field of evolutionary research. Heterochromatin was considered as one of the key components in prior models that propose a molecular arms race between fast evolving sequences such as short satellite repeats, transposable elements and their corresponding regulatory chromatin surrounding which results in adaptive changes (Brown and O'Neill, 2010; Crespi and Nosil, 2013; Johnson, 2010; Maheshwari and Barbash, 2011; Presgraves, 2010). In the case of *Drosophila*, *D. simulans* has a smaller genome, with about 4-fold less satellite DNA (Bosco et al., 2007) and less transposable elements (Dowsett and Young, 1982; Lerat et al., 2011; Vieira and Biemont, 2004; Vieira et al., 2012) than its sibling species *D. melanogaster*. *Gypsy* and *gtwin* are unique among the transposable elements as they encode full-length envelope proteins in various *Drosophila* species. These species range from very close *D. melanogaster* sibling species to more distant siblings such as *D. virilis* (Ludwig and Loreto, 2007; Mejlumian et al., 2002). This finding and the phylogeny of *gypsy* and *gtwin* elements across various host species support a

hypothesis by which their copies are able to spread horizontally between sexually isolated species (Heredia et al., 2004; Ludwig and Loreto, 2007).

However, two findings do not match with a direct molecular arms race scenario between HMR and TEs. First, we did not find HMR bound to transposable elements apart from *gypsy* and *gtwin*. Second, most of the transposable elements which are misregulated in *Hmr* mutant flies are transpositionally inactive in the *D. melanogaster* species (Kofler et al., 2012; Satyaki et al., 2014).

Alternatively, HMR could have evolved to ensure the expression of heterochromatic genes surrounded by repetitive DNA. There is growing evidence that heterochromatin and its components are important for the transcription of genes embedded in repetitive DNA (Yasuhara and Wakimoto, 2006). Class 1 HMR binding sites associate with heterochromatin and transcriptionally active genes. We therefore want to propose a model where HMR serves as an activating factor to keep these genes actively transcribed under the evolutionary pressure of a gained number of repetitive DNA (illustrated in **Figure 35 A**). At least two evidences support this model. First, the expression of such genes is affected in *Hmr* mutant flies and also in BEAF-32 mutant flies. Second, these regions are subject to repetitive DNA evolution and substantially differ in *D. melanogaster* and *D. simulans* (Bosco et al., 2007; Dowsett and Young, 1982; Lerat et al., 2011; Vieira and Biemont, 2004; Vieira et al., 2012). *D. melanogaster* exhibits a higher expression of HMR and has a higher amount of repetitive DNA then *D. simulans* where HMR is virtually absent (Thomae et al., 2013) and the amount of repetitive DNA is lower. What triggered the evolution of HMR with respect to changes in repetitive DNA elements if HMR is not capable to bind such sequences? The heterochromatic protein Su(var)3-7 is capable to bind short repeats directly (Cleard et al., 1997; Delattre et al., 2000). Su(var)3-7 evolved fast and is, in contrast to HP1a and Su(var)3-9, *Drosophila*-specific (Jaquet et al., 2006), same as HMR. According to that, we speculate that Su(var)3-7 served as a 'molecular sensor' for changes in repetitive DNA. In turn, this resulted in an adaptive change for HMR and LHR, both interacting with Su(var)3-7 (illustrated in **Figure 35 A**) (Thomae et al., 2013). Notably, the cellular function of Su(var) proteins, as well as the ones of HMR and LHR are highly dependent on the protein's concentration (Eissenberg et al., 1990; Eissenberg et al., 1992; Locke et al., 1988; Satyaki et al., 2014; Schotta et al., 2002; Thomae et al., 2013). In that way, changes in repeat copy number could have spurred a cascade of mass action-based changes. From these changes, HMR could have gained a function in activating heterochromatic gene

transcription or, more general, supported adaptive chromatin changes due to increased repeat copy number (illustrated in **Figure 35 A**). Such scenario would constitute an example of how a chromatin-related molecular arms race can drive the evolution of hybrid incompatibility (illustrated in **Figure 3**, **Figure 35**).

4.7 Did HMR and BEAF-32 coevolve to ensure gene expression under the selective pressure of genome sequence changes?

The class 1 HMR binding sites fit a model that proposes adaptive changes in HMR due to changes in repeat copy number. A preliminary analysis on the evolutionary conservation of HMR binding sites suggest that HMR class 1 binding site regions are indeed less conserved than other HMR or insulator protein binding sites (**Figure S3**). As pointed out in prior sections, HMR and BEAF-32 might act in concert at these genomic sites to ensure heterochromatic gene transcription.

Did HMR, BEAF-32 and other heterochromatic components evolve interdependently in fulfilling their function? HMR, LHR and Su(var)3-7, the proteins involved in the proposed model of molecular arms race, are *Drosophila* lineage-specific. Notably, phylogenetic analysis demonstrated that non-CTCF insulator proteins such as Su(Hw), CP190, Mod(mdg4) and BEAF-32 are limited to arthropods and were successively gained during evolution (Heger et al., 2013; Schoborg and Labrador, 2010). Further, it reveals that BEAF-32 is exclusively *Drosophila* lineage-specific, same as HMR, LHR and Su(var)3-7 (Pauli et al., 2016). It is striking that most of these proteins belong to the MADF-BESS domain family. This is an evolutionary young protein family and has only few members in *Drosophila* which are presumably derived from a common ancestor gene (Shukla et al., 2014). HMR, LHR, Su(var)3-7 and BEAF-32 all govern MADF and/or BESS domains. Further, isoforms of the *Drosophila* protein dADD1, an orthologue to the ADD domain of human ATRX, gained MADF domains suggesting a divergence of the corresponding gene in *Drosophila* (Lopez-Falcon et al., 2014) (Dr. Addie Kolybaba-Steward, Dr. Anne-Kathrin Classen, personal communication). This is particularly interesting as the human ATRX is involved in heterochromatin formation and transposon silencing (Bassett et al., 2008; Groh and Schotta, 2017; Sadic et al., 2015). Apart from that, dADD1 interacts with HP1a (Alekseyenko et al., 2014; Lopez-Falcon et al., 2014). Also another HI protein, which is Overdrive, belongs to the MADF-BESS domain family. Altogether, heterochromatin components (Su(var)3-7, dADD1), insulator

protein BEAF-32 and HI proteins (HMR, LHR, Overdrive) are originating from the MADF-BESS domain family and show signs of rapid evolution. As discussed before, most of these proteins are interconnected, either functionally or by biochemical interaction, often via their BESS domain. We therefore speculate that, in a first step, these proteins coevolved, and, in a second step, adaptively evolved in a molecular arms race with repetitive DNA (illustrated in **Figure 35**). Shared domain architecture is not limited to MADF and BESS. Interestingly, also the insulator protein Mod(mdg4) and the hybrid incompatibility protein gfzf share a rarely occurring FLYWCH zinc finger domain (Phadnis et al., 2015). How exactly the birth of these genes and their gain or loss of function in diverse *Drosophila* species occurred during evolution remains to be solved. On the one hand, HMR is present in *D. melanogaster* and *D. simulans* and HMR orthologues can substitute sibling species HMR function (Aruna et al., 2009; Satyaki et al., 2014). On the other hand, HMR expression levels are substantially higher in *D. melanogaster* (Thomae et al., 2013). Whether HMR gained a species-specific function in *D. melanogaster* or lost an ancestral function in *D. simulans* is unclear.

Further, it is possible that HMR acquired multiple functions that relate to heterochromatin and genomic insulators. HMR's class 1 binding sites at heterochromatic genes could be linked to evolutionary changes in repeat copy number, heterochromatic gene transcription and female fertility (see above), whereas HMR's function at *gypsy* insulator sites could have been gained another time in evolution. Notably, heterochromatin factor Su(var)3-7 and insulator proteins share another feature: they bind to rapidly changing DNA sequence and adaptively change their genomic binding sites whereas their binding motifs and their DNA binding domains are conserved (Ni et al., 2012; Yang et al., 2012). Insulator proteins presumably have a role in coordinating genome organization and function during evolution. Studies that come to this conclusion were done on the *Drosophila* insulator proteins CTCF and BEAF-32, comparing binding sites in the species *D. melanogaster*, *D. simulans*, *D. yakuba* and *D. pseudoobscura* (Ni et al., 2012; Yang et al., 2012). In this four *Drosophila* species, the insulators binding motifs are virtually identical. However, the number and genomic position of insulator binding sites diverged over time. Newly evolved binding sites of CTCF and BEAF-32 are correlating with the occurrence of new genes, changes in gene locations, genome rearrangements and genome size changes during *Drosophila* evolution (Ni et al., 2012; Yang et al., 2012). These findings suggest a strong positive selection of newly

gained insulator binding sites during evolution. In preliminary analysis, we observe that HMR indeed binds together with insulator proteins in regions at the genome that are less conserved across *D. melanogaster* and *D. simulans* than regions bound by insulator proteins in the absence of HMR (**Figure S3**). Whether HMR fulfills a function at these sites and which fraction of the HMR binding regions (gene body, regulatory regions, neighboring repeats) differ between species, however, is unknown. A very interesting example of a new born *D. melanogaster*-specific gene bound by HMR (**Figure S4**) is *sphinx,* a non-coding RNA gene that is involved in regulating *D. melanogaster* male courtship behavior (Chen et al., 2011). It could well be that the birth and regulation of genes such as *sphinx,* that affect phenotypic traits, constitute a critical step in prezygotic reproductive isolation.

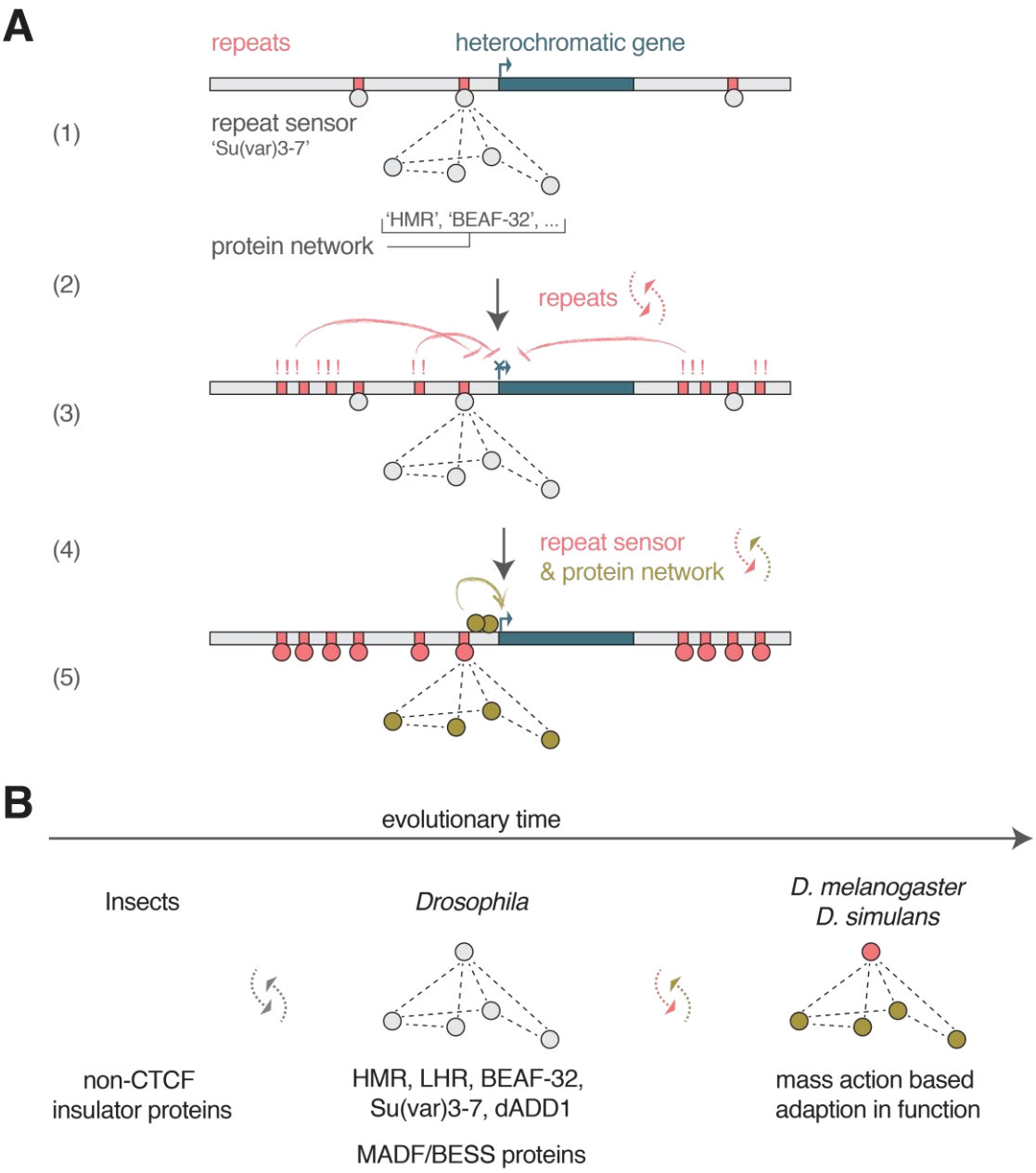

Figure 35. Model of how repetitive DNA and insulator proteins result in adaptive evolution and species-specific function of HMR. (A) (1) Active transcription of heterochromatic genes surrounded by repeats which are sensed and regulated by a mass action-based protein network that consists in Su(var) proteins, HMR, LHR and insulator proteins. (2, 3) Increase in repeat copy number antagonizes heterochromatic gene transcription and therefore (4) triggers mass action-based functional adaption in the protein network. (5) HMR and BEAF-32 ensure heterochromatic gene transcription. **(B)** Successive gain of insulator proteins in insects during evolution. The MADF-BESS domain family gives rise to a mass action-based protein network that is involved in the regulation of transposable elements and adaptively changed due to changes in repeat copy number.

SUPPLEMENTAL DATA

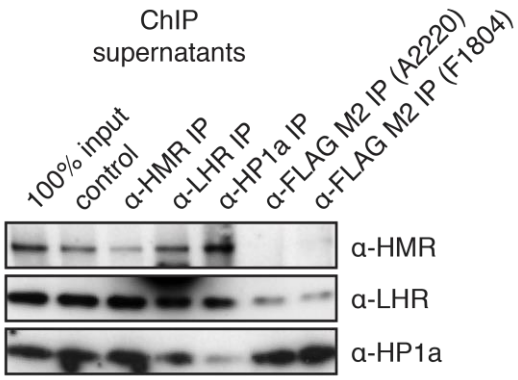

Figure S1. Western Blot of CoImmunoprecipitates from HMR-Flag$_2$ expressing S2 cell cross linked chromatin using anti-HMR, anti-LHR, anti-HP1a and anti-FLAG (A2220 and F1804) antibody. Supernatant of IP reactions was loaded after reverse crosslinking. Co-depletion of HMR and LHR in anti-FLAG IP and partial depletion of HMR in anti-HMR IP and HP1a in anti-HP1a IP. Protein A/G sepharose beads served as negative control.

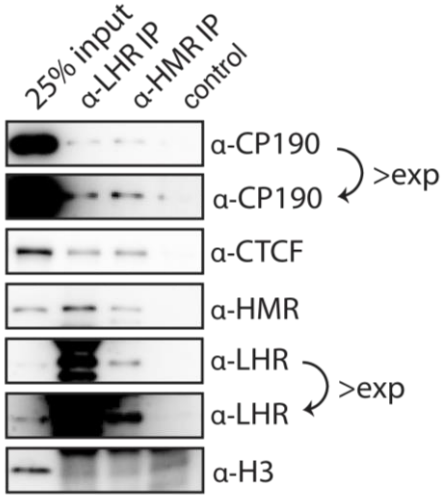

Figure S2. Western Blot of CoImmunoprecipitates from S2 cell nuclear extract using anti-HMR and anti-LHR antibody shows interaction of HMR, LHR and the insulator proteins CP190 and CTCF. Protein A/G sepharose beads served as negative control.

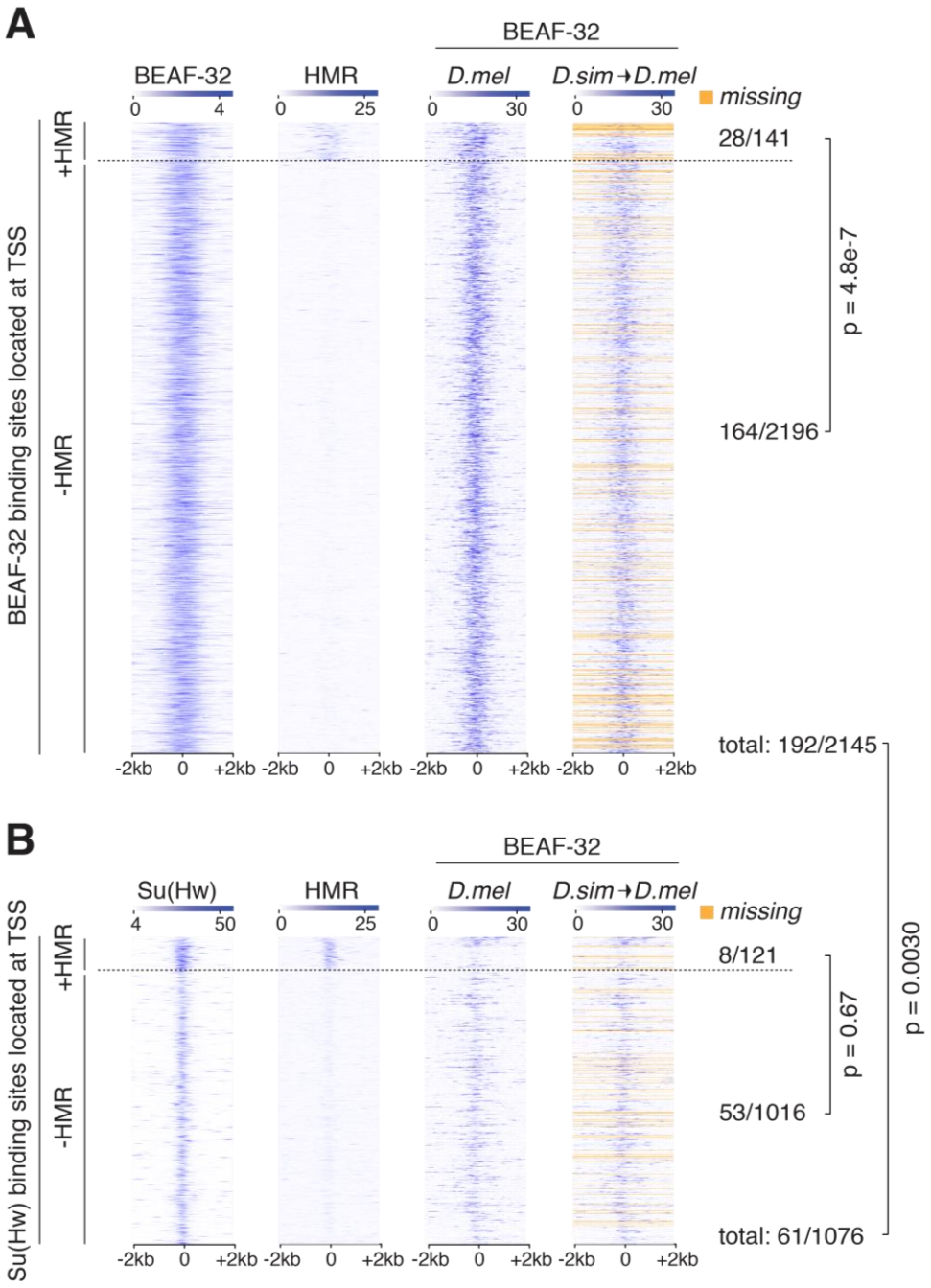

Figure S3. HMR binding sites are enriched for fast evolving BEAF-32 binding sites (part1/2) (see following page).

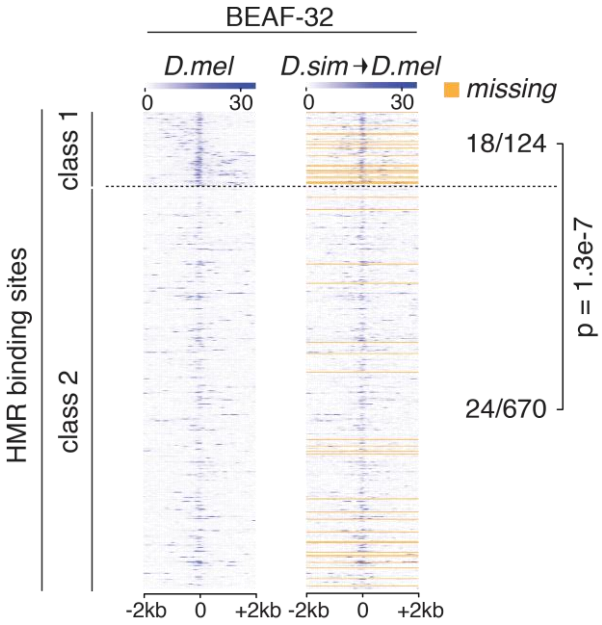

Figure S3. HMR binding sites are enriched for fast evolving BEAF-32 binding sites (part2/2) Heatmaps of BEAF-32 ChIP signals derived from *D. melanogaster* (*D.mel*) and *D. simulans* (*D.sim*) embryos (Yang et al., 2012). Signals are centered around TSS-annotated BEAF-32 binding sites **(part 1/2 A)**, TSS-annotated Su(Hw) binding sites **(part 1/2 B)** and HMR binding sites **(part 2/2)**, clustered according to adjacent HP1a signals and sorted by HMR intensity. To display *D.sim*-derived BEAF-32 ChIP signals according to the homologous *D.mel* genome position, *D.mel* genomic positions were lifted into the *D.sim* genome using liftOver. In case of a missing homologous sequence, the genome position is marked in *orange*. The significance of difference between the proportion of missing genome positions in class 1 and class 2 was estimated with p-values calculated with proportional test.

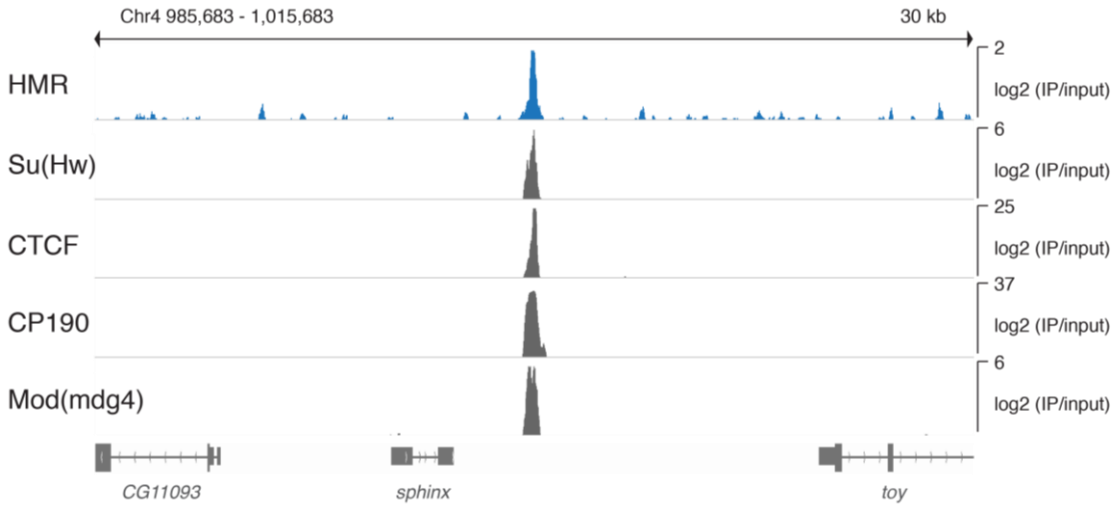

Figure S4. Genome browser view of HMR, the insulator proteins CP190, Mod(mdg4), Su(Hw) and CTCF (Ong et al., 2013) in proximity to the *D. melanogaster*-specific gene *sphinx* (Chen et al., 2011).

ABBREVIATIONS

aa	amino acid
Ab	Antibody
AFA	Adaptive Focused Acoustics
ATRX	Alpha-thalassemia X-linked mental retardation
BEAF-32	Boundary Element Associated Factor 32
BESS	BEAF, Su-Var(3-7), Stonewall-like
BF	Bright field
bp	Basepair(s)
ChIP	Chromatin immunoprecipitation
CENP	Centromeric protein
CID	Centromer identifier in *Drosophila*
CRISPR	clustered, regularly interspaced, short palindromic repeats
CP190	Centrosomal Protein 190
CD	chromo domain
CSD	chromo-shadow domain
Ctrl	Control
DAPI	4',6-diamidino-2-phenylindole
DCC	Dosage compensation complex
DGRC	*Drosophila* genomics resource center
CBP	CREB-binding protein
Ct	threshold cycle number
CTCF	CCCTC-binding factor
D. mel	*Drosophila melanogaster*
D. sim	*Drosophila simulans*
Da	Dalton
dADD1	*Drosophila* ADD1
DGRC	*Drosophila* genomics resource center
DNA	Deoxyribonucleic acid
dsRNA	Double stranded RNA
ECL	enhanced chemiluminescent
eve	*even-skipped*
for	forward
g	standard gravity
gfzf	*glutathione-S-transferase-containing FLYWCH zinc finger protein*
gRNA	guide RNA
GFP	Green fluorescent protein
GST	Glutathione S-transferase
gtwin	*gypsy-twin*
HI	Hybrid Incompatibility
Hmr	*Hybrid male rescue*
Hmr_{mel}	*Drosophila melanogaster Hmr*
Hmr_{sim}	*Drosophila simulans Hmr*
hr	Hour
HRP	Horseradish Peroxidase
HP1a	Heterochromatin Protein 1a
HR	Homologous recombination
IF	Immunofluorescence
IgG	immunoglobulin G
IP	Immunoprecipitation
IQD	Interquartile distance

IVT	*in vitro* transcription
kb	Kilobase(s)
kDa	Kilodalton
Lhr	*Lethal hybrid rescue*
Lhr_{mel}	*Drosophila melanogaster Lhr*
Lhr_{sim}	*Drosophila simulans Lhr*
LTR	long terminal repeat
MADF	Myb/SANT-like in ADF1
MBP	Maltose-Binding Protein
min	Minute(s)
mL	Milliliter
Mod(mdg4)	Modifier of mdg4
modENCODE	MODel organism ENCyclopedia Of DNA Elements
MOF	Males absent on the first
mRNA	messenger RNA
MSL	Male specific lethal
NA	numerical aperture
NHEJ	non-homologous end joining
NLP	nucleoplasmin-like protein
NPC	nuclear pore complex
NPV	nuclear pellet volume
nt	nucleotide(s)
OdsH	*Odysseus-site homeobox*
Ovd	*Overdrive*
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate-buffered saline
PCV	pure cell volume
PRC	Polycomb Repressive Complex
Prdm9	PR domain-containing protein 9
qPCR	Quantitative real-time Polymerase Chain Reaction
rev	reverse
RIPA	Radioimmunoprecipitation assay
RNA	Ribonucleic acid
RNAi	RNA interference
RPM	Reads Per Million mapped reads
s	second(s)
SD	standard deviation
SDS	Sodium dodecyl sulfate
Su(Hw)	Suppressor of Hairy wing
Su(var)205	Suppressor of variegation 205
TE	transposable element
TSS	Transcription start site
TTS	Transcription termination site
UCSC	University of California, Santa Cruz
UTR	Untranslated region
UV	Ultraviolet
WB	Western Blot
wt	wild type
Zhr	*Zygotic hybrid rescue*

REFERENCES

Aagaard, L., Laible, G., Selenko, P., Schmid, M., Dorn, R., Schotta, G., Kuhfittig, S., Wolf, A., Lebersorger, A., Singh, P.B., *et al.* (1999). Functional mammalian homologues of the Drosophila PEV-modifier Su(var)3-9 encode centromere-associated proteins which complex with the heterochromatin component M31. EMBO J *18*, 1923-1938.

Aaij, C., and Borst, P. (1972). The gel electrophoresis of DNA. Biochim Biophys Acta *269*, 192-200.

Aasland, R., Stewart, A.F., and Gibson, T. (1996). The SANT domain: a putative DNA-binding domain in the SWI-SNF and ADA complexes, the transcriptional co-repressor N-CoR and TFIIIB. Trends Biochem Sci *21*, 87-88.

Abad, J.P., Carmena, M., Baars, S., Saunders, R.D., Glover, D.M., Ludena, P., Sentis, C., Tyler-Smith, C., and Villasante, A. (1992). Dodeca satellite: a conserved G+C-rich satellite from the centromeric heterochromatin of Drosophila melanogaster. Proc Natl Acad Sci U S A *89*, 4663-4667.

Alekseyenko, A.A., Gorchakov, A.A., Zee, B.M., Fuchs, S.M., Kharchenko, P.V., and Kuroda, M.I. (2014). Heterochromatin-associated interactions of Drosophila HP1a with dADD1, HIPP1, and repetitive RNAs. Genes Dev *28*, 1445-1460.

Allshire, R.C., and Karpen, G.H. (2008). Epigenetic regulation of centromeric chromatin: old dogs, new tricks? Nat Rev Genet *9*, 923-937.

Anderson, J.A., Gilliland, W.D., and Langley, C.H. (2009). Molecular population genetics and evolution of Drosophila meiosis genes. Genetics *181*, 177-185.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data (http://www.bioinformatics.babraham.ac.uk/projects/fastqc).

Andreyeva, E.N., Belyaeva, E.S., Semeshin, V.F., Pokholkova, G.V., and Zhimulev, I.F. (2005). Three distinct chromatin domains in telomere ends of polytene chromosomes in Drosophila melanogaster Tel mutants. J Cell Sci *118*, 5465-5477.

Aruna, S., Flores, H.A., and Barbash, D.A. (2009). Reduced fertility of Drosophila melanogaster hybrid male rescue (Hmr) mutant females is partially complemented by Hmr orthologs from sibling species. Genetics *181*, 1437-1450.

Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. Nature *410*, 120-124.

Bao, W., Kojima, K.K., and Kohany, O. (2015). Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA *6*, 11.

Barbash, D.A. (2010). Genetic testing of the hypothesis that hybrid male lethality results from a failure in dosage compensation. Genetics *184*, 313-316.

Barbash, D.A., Awadalla, P., and Tarone, A.M. (2004). Functional divergence caused by ancient positive selection of a Drosophila hybrid incompatibility locus. PLoS Biol *2*, e142.

Barbash, D.A., Siino, D.F., Tarone, A.M., and Roote, J. (2003). A rapidly evolving MYB-related protein causes species isolation in Drosophila. Proc Natl Acad Sci U S A *100*, 5302-5307.

Barth, T.K., Schade, G.O., Schmidt, A., Vetter, I., Wirth, M., Heun, P., Thomae, A.W., and Imhof, A. (2014). Identification of novel Drosophila centromere-associated proteins. Proteomics *14*, 2167-2178.

Bartkuhn, M., Straub, T., Herold, M., Herrmann, M., Rathke, C., Saumweber, H., Gilfillan, G.D., Becker, P.B., and Renkawitz, R. (2009). Active promoters and insulators are marked by the centrosomal protein 190. EMBO J *28*, 877-888.

Basquin, D., Spierer, A., Begeot, F., Koryakov, D.E., Todeschini, A.L., Ronsseray, S., Vieira, C., Spierer, P., and Delattre, M. (2014). The Drosophila Su(var)3-7 gene is

required for oogenesis and female fertility, genetically interacts with piwi and aubergine, but impacts only weakly transposon silencing. PLoS One *9*, e96802.
Bassett, A.R., Cooper, S.E., Ragab, A., and Travers, A.A. (2008). The chromatin remodelling factor dATRX is involved in heterochromatin formation. PLoS One *3*, e2099.
Baumann, C., Korner, R., Hofmann, K., and Nigg, E.A. (2007). PICH, a centromere-associated SNF2 family ATPase, is regulated by Plk1 and required for the spindle checkpoint. Cell *128*, 101-114.
Baxley, R.M., Soshnev, A.A., Koryakov, D.E., Zhimulev, I.F., and Geyer, P.K. (2011). The role of the Suppressor of Hairy-wing insulator protein in Drosophila oogenesis. Dev Biol *356*, 398-410.
Bayes, J.J., and Malik, H.S. (2009). Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. Science *326*, 1538-1541.
Bhaskar, V., and Courey, A.J. (2002). The MADF-BESS domain factor Dip3 potentiates synergistic activation by Dorsal and Twist. Gene *299*, 173-184.
Biggs, W.H., 3rd, Zavitz, K.H., Dickson, B., van der Straten, A., Brunner, D., Hafen, E., and Zipursky, S.L. (1994). The Drosophila rolled locus encodes a MAP kinase required in the sevenless signal transduction pathway. EMBO J *13*, 1628-1635.
Birchler, J.A., Bhadra, M.P., and Bhadra, U. (2000). Making noise about silence: repression of repeated genes in animals. Curr Opin Genet Dev *10*, 211-216.
Bohmdorfer, G., and Wierzbicki, A.T. (2015). Control of Chromatin Structure by Long Noncoding RNA. Trends Cell Biol *25*, 623-632.
Bolkan, B.J., Booker, R., Goldberg, M.L., and Barbash, D.A. (2007). Developmental and cell cycle progression defects in Drosophila hybrid males. Genetics *177*, 2233-2241.
Bosco, G., Campbell, P., Leiva-Neto, J.T., and Markow, T.A. (2007). Analysis of Drosophila species genome size and satellite DNA content reveals significant differences among strains as well as between species. Genetics *177*, 1277-1290.
Bottcher, R., Hollmann, M., Merk, K., Nitschko, V., Obermaier, C., Philippou-Massier, J., Wieland, I., Gaul, U., and Forstemann, K. (2014). Efficient chromosomal gene modification with CRISPR/cas9 and PCR-based homologous recombination donors in cultured Drosophila cells. Nucleic Acids Res *42*, e89.
Bowen, N.J., and McDonald, J.F. (2001). Drosophila euchromatic LTR retrotransposons are much younger than the host species in which they reside. Genome Res *11*, 1527-1540.
Bradford, M.M. (1976). A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Anal Biochem *72*, 248-254.
Bramhall, S., Noack, N., Wu, M., and Loewenberg, J.R. (1969). A simple colorimetric method for determination of protein. Anal Biochem *31*, 146-148.
Brasher, S.V., Smith, B.O., Fogh, R.H., Nietlispach, D., Thiru, A., Nielsen, P.R., Broadhurst, R.W., Ball, L.J., Murzina, N.V., and Laue, E.D. (2000). The structure of mouse HP1 suggests a unique mode of single peptide recognition by the shadow chromo domain dimer. EMBO J *19*, 1587-1597.
Braun, S., Garcia, J.F., Rowley, M., Rougemaille, M., Shankar, S., and Madhani, H.D. (2011). The Cul4-Ddb1(Cdt)(2) ubiquitin ligase inhibits invasion of a boundary-associated antisilencing factor into heterochromatin. Cell *144*, 41-54.
Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X., and Barbash, D.A. (2006). Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. Science *314*, 1292-1295.
Brown, J.D., and O'Neill, R.J. (2010). Chromosomes, conflict, and epigenetics: chromosomal speciation revisited. Annu Rev Genomics Hum Genet *11*, 291-316.

Byrd, K., and Corces, V.G. (2003). Visualization of chromatin domains created by the gypsy insulator of Drosophila. J Cell Biol *162*, 565-574.
Capelson, M., and Corces, V.G. (2005). The ubiquitin ligase dTopors directs the nuclear organization of a chromatin insulator. Mol Cell *20*, 105-116.
Capelson, M., and Corces, V.G. (2006). SUMO conjugation attenuates the activity of the gypsy chromatin insulator. EMBO J *25*, 1906-1914.
Capelson, M., Liang, Y., Schulte, R., Mair, W., Wagner, U., and Hetzer, M.W. (2010). Chromatin-bound nuclear pore components regulate gene expression in higher eukaryotes. Cell *140*, 372-383.
Chan, K.L., North, P.S., and Hickson, I.D. (2007). BLM is required for faithful chromosome segregation and its localization defines a class of ultrafine anaphase bridges. EMBO J *26*, 3397-3409.
Charlesworth, B., Sniegowski, P., and Stephan, W. (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. Nature *371*, 215-220.
Chen, T.W., Li, H.P., Lee, C.C., Gan, R.C., Huang, P.J., Wu, T.H., Lee, C.Y., Chang, Y.F., and Tang, P. (2014). ChIPseek, a web-based analysis tool for ChIP data. BMC Genomics *15*, 539.
Chen, Y., Dai, H., Chen, S., Zhang, L., and Long, M. (2011). Highly tissue specific expression of Sphinx supports its male courtship related role in Drosophila melanogaster. PLoS One *6*, e18853.
Chodagam, S., Royou, A., Whitfield, W., Karess, R., and Raff, J.W. (2005). The centrosomal protein CP190 regulates myosin function during early Drosophila development. Curr Biol *15*, 1308-1313.
Clark, A.G., Eisen, M.B., Smith, D.R., Bergman, C.M., Oliver, B., Markow, T.A., Kaufman, T.C., Kellis, M., Gelbart, W., Iyer, V.N., *et al.* (2007). Evolution of genes and genomes on the Drosophila phylogeny. Nature *450*, 203-218.
Cleard, F., Delattre, M., and Spierer, P. (1997). SU(VAR)3-7, a Drosophila heterochromatin-associated protein and companion of HP1 in the genomic silencing of position-effect variegation. EMBO J *16*, 5280-5288.
Comet, I., Schuettengruber, B., Sexton, T., and Cavalli, G. (2011). A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. Proc Natl Acad Sci U S A *108*, 2294-2299.
Conrad, T., and Akhtar, A. (2012). Dosage compensation in Drosophila melanogaster: epigenetic fine-tuning of chromosome-wide transcription. Nat Rev Genet *13*, 123-134.
Coyne, J., and Orr, H. (2004). Speciation (Sinauer Associates).
Crespi, B., and Nosil, P. (2013). Conflictual speciation: species formation via genomic conflict. Trends Ecol Evol *28*, 48-57.
Cuddapah, S., Jothi, R., Schones, D.E., Roh, T.Y., Cui, K., and Zhao, K. (2009). Global analysis of the insulator binding protein CTCF in chromatin barrier regions reveals demarcation of active and repressive domains. Genome Res *19*, 24-32.
Cutler, G., Perry, K.M., and Tjian, R. (1998). Adf-1 is a nonmodular transcription factor that contains a TAF-binding Myb-like motif. Mol Cell Biol *18*, 2252-2261.
Danzer, J.R., and Wallrath, L.L. (2004). Mechanisms of HP1-mediated gene silencing in Drosophila. Development *131*, 3571-3580.
Darwin, C. (1859). On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. (London: John Murray Albemarle Street).
Delattre, M., Spierer, A., Tonka, C.H., and Spierer, P. (2000). The genomic silencing of position-effect variegation in Drosophila melanogaster: interaction between the

heterochromatin-associated proteins Su(var)3-7 and HP1. J Cell Sci *113 Pt 23*, 4253-4261.
Dernburg, A.F., Sedat, J.W., and Hawley, R.S. (1996). Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell *86*, 135-146.
Devlin, R.H., Bingham, B., and Wakimoto, B.T. (1990). The organization and expression of the light gene, a heterochromatic gene of Drosophila melanogaster. Genetics *125*, 129-140.
Dillon, N., and Festenstein, R. (2002). Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. Trends Genet *18*, 252-258.
Dobzhansky, T. (1936). Studies on Hybrid Sterility. II. Localization of Sterility Factors in Drosophila Pseudoobscura Hybrids. Genetics *21*, 113-135.
Dowsett, A.P., and Young, M.W. (1982). Differing levels of dispersed repetitive DNA among closely related species of Drosophila. Proc Natl Acad Sci U S A *79*, 4570-4574.
Duarte, F.M., Fuda, N.J., Mahat, D.B., Core, L.J., Guertin, M.J., and Lis, J.T. (2016). Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes Dev *30*, 1731-1746.
Ebert, A., Schotta, G., Lein, S., Kubicek, S., Krauss, V., Jenuwein, T., and Reuter, G. (2004). Su(var) genes regulate the balance between euchromatin and heterochromatin in Drosophila. Genes Dev *18*, 2973-2983.
Eissenberg, J.C., James, T.C., Foster-Hartnett, D.M., Hartnett, T., Ngan, V., and Elgin, S.C. (1990). Mutation in a heterochromatin-specific chromosomal protein is associated with suppression of position-effect variegation in Drosophila melanogaster. Proc Natl Acad Sci U S A *87*, 9923-9927.
Eissenberg, J.C., Morris, G.D., Reuter, G., and Hartnett, T. (1992). The heterochromatin-associated protein HP-1 is an essential protein in Drosophila with dosage-dependent effects on position-effect variegation. Genetics *131*, 345-352.
Eissenberg, J.C., and Reuter, G. (2009). Cellular mechanism for targeting heterochromatin formation in Drosophila. Int Rev Cell Mol Biol *273*, 1-47.
Elbashir, S.M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001). Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. Nature *411*, 494-498.
Emberly, E., Blattes, R., Schuettengruber, B., Hennion, M., Jiang, N., Hart, C.M., Kas, E., and Cuvier, O. (2008). BEAF regulates cell-cycle genes through the controlled deposition of H3K9 methylation marks into its conserved dual-core binding sites. PLoS Biol *6*, 2896-2910.
England, B.P., Admon, A., and Tjian, R. (1992). Cloning of Drosophila transcription factor Adf-1 reveals homology to Myb oncoproteins. Proc Natl Acad Sci U S A *89*, 683-687.
Fanti, L., Berloco, M., Piacentini, L., and Pimpinelli, S. (2003). Chromosomal distribution of heterochromatin protein 1 (HP1) in Drosophila: a cytological map of euchromatic HP1 binding sites. Genetica *117*, 135-147.
Ferree, P.M., and Barbash, D.A. (2009). Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. PLoS Biol *7*, e1000234.
Filion, G.J., van Bemmel, J.G., Braunschweig, U., Talhout, W., Kind, J., Ward, L.D., Brugman, W., de Castro, I.J., Kerkhoven, R.M., Bussemaker, H.J., *et al.* (2010). Systematic protein location mapping reveals five principal chromatin types in Drosophila cells. Cell *143*, 212-224.

Frasch, M., Glover, D.M., and Saumweber, H. (1986). Nuclear antigens follow different pathways into daughter nuclei during mitosis in early Drosophila embryos. J Cell Sci *82*, 155-172.
Fritsch, L., Robin, P., Mathieu, J.R., Souidi, M., Hinaux, H., Rougeulle, C., Harel-Bellan, A., Ameyar-Zazoua, M., and Ait-Si-Ali, S. (2010). A subset of the histone H3 lysine 9 methyltransferases Suv39h1, G9a, GLP, and SETDB1 participate in a multimeric complex. Mol Cell *37*, 46-56.
Gaszner, M., and Felsenfeld, G. (2006). Insulators: exploiting transcriptional and epigenetic mechanisms. Nat Rev Genet *7*, 703-713.
Gause, M., Morcillo, P., and Dorsett, D. (2001). Insulation of enhancer-promoter communication by a gypsy transposon insert in the Drosophila cut gene: cooperation between suppressor of hairy-wing and modifier of mdg4 proteins. Mol Cell Biol *21*, 4807-4817.
Georgiev, P.G., and Gerasimova, T.I. (1989). Novel genes influencing the expression of the yellow locus and mdg4 (gypsy) in Drosophila melanogaster. Mol Gen Genet *220*, 121-126.
Gerasimova, T.I., Gdula, D.A., Gerasimov, D.V., Simonova, O., and Corces, V.G. (1995). A Drosophila protein that imparts directionality on a chromatin insulator is an enhancer of position-effect variegation. Cell *82*, 587-597.
Gerasimova, T.I., Lei, E.P., Bushey, A.M., and Corces, V.G. (2007). Coordinated control of dCTCF and gypsy chromatin insulators in Drosophila. Mol Cell *28*, 761-772.
Ghosh, D., Gerasimova, T.I., and Corces, V.G. (2001). Interactions between the Su(Hw) and Mod(mdg4) proteins required for gypsy insulator function. EMBO J *20*, 2518-2527.
Granzotto, A., Lopes, F.R., Lerat, E., Vieira, C., and Carareto, C.M. (2009). The evolutionary dynamics of the Helena retrotransposon revealed by sequenced Drosophila genomes. BMC Evol Biol *9*, 174.
Graveley, B.R., Brooks, A.N., Carlson, J.W., Duff, M.O., Landolin, J.M., Yang, L., Artieri, C.G., van Baren, M.J., Boley, N., Booth, B.W., *et al.* (2011). The developmental transcriptome of Drosophila melanogaster. Nature *471*, 473-479.
Gregory, T.R. (2005). Synergy between sequence and size in large-scale genomics. Nat Rev Genet *6*, 699-708.
Greil, F., de Wit, E., Bussemaker, H.J., and van Steensel, B. (2007). HP1 controls genomic targeting of four novel heterochromatin proteins in Drosophila. EMBO J *26*, 741-751.
Greil, F., van der Kraan, I., Delrow, J., Smothers, J.F., de Wit, E., Bussemaker, H.J., van Driel, R., Henikoff, S., and van Steensel, B. (2003). Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. Genes Dev *17*, 2825-2838.
Grewal, S.I. (2010). RNAi-dependent formation of heterochromatin and its diverse functions. Curr Opin Genet Dev *20*, 134-141.
Grewal, S.I., and Rice, J.C. (2004). Regulation of heterochromatin by histone methylation and small RNAs. Curr Opin Cell Biol *16*, 230-238.
Groh, S., and Schotta, G. (2017). Silencing of endogenous retroviruses by heterochromatin. Cell Mol Life Sci.
Grossman, E., Medalia, O., and Zwerger, M. (2012). Functional architecture of the nuclear pore complex. Annu Rev Biophys *41*, 557-584.
Guelen, L., Pagie, L., Brasset, E., Meuleman, W., Faza, M.B., Talhout, W., Eussen, B.H., de Klein, A., Wessels, L., de Laat, W., *et al.* (2008). Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. Nature *453*, 948-951.

Gurudatta, B.V., Ramos, E., and Corces, V.G. (2012). The BEAF insulator regulates genes involved in cell polarity and neoplastic growth. Dev Biol *369*, 124-132.
Hannon, G.J. (2010). G. J. Fastx-toolkit. FASTQ/A short-reads pre-processing tools. (http//hannonlab. cshl. edu/fastx_ toolkit (2010)).
Heger, P., George, R., and Wiehe, T. (2013). Successive gain of insulator proteins in arthropod evolution. Evolution *67*, 2945-2956.
Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell *38*, 576-589.
Heitz, E. (1930). Der Bau der somatischen Kerne von Drosophila melanogaster. Z indukt Abstammungs- Vererbungsl *54*, 248-249.
Heredia, F., Loreto, E.L., and Valente, V.L. (2004). Complex evolution of gypsy in Drosophilid species. Mol Biol Evol *21*, 1831-1842.
Herrmann, C., Van de Sande, B., Potier, D., and Aerts, S. (2012). i-cisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. Nucleic Acids Res *40*, e114.
Hickey, D.A. (1982). Selfish DNA: a sexually-transmitted nuclear parasite. Genetics *101*, 519-531.
Holoch, D., and Moazed, D. (2015). RNA-mediated epigenetic regulation of gene expression. Nat Rev Genet *16*, 71-84.
Horn, T., and Boutros, M. (2010). E-RNAi: a web application for the multi-species design of RNAi reagents--2010 update. Nucleic Acids Res *38*, W332-339.
Hughes, S.E., Gilliland, W.D., Cotitta, J.L., Takeo, S., Collins, K.A., and Hawley, R.S. (2009). Heterochromatic threads connect oscillating chromosomes during prometaphase I in Drosophila oocytes. PLoS Genet *5*, e1000348.
Hutter, P., Roote, J., and Ashburner, M. (1990). A genetic basis for the inviability of hybrids between sibling species of Drosophila. Genetics *124*, 909-920.
Jacobs, S.A., and Khorasanizadeh, S. (2002). Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. Science *295*, 2080-2083.
Jacobs, S.A., Taverna, S.D., Zhang, Y., Briggs, S.D., Li, J., Eissenberg, J.C., Allis, C.D., and Khorasanizadeh, S. (2001). Specificity of the HP1 chromo domain for the methylated N-terminus of histone H3. EMBO J *20*, 5232-5241.
Jain, D., Baldi, S., Zabel, A., Straub, T., and Becker, P.B. (2015). Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments. Nucleic Acids Res *43*, 6959-6968.
James, T.C., Eissenberg, J.C., Craig, C., Dietrich, V., Hobson, A., and Elgin, S.C. (1989). Distribution patterns of HP1, a heterochromatin-associated nonhistone chromosomal protein of Drosophila. Eur J Cell Biol *50*, 170-180.
Jaquet, Y., Delattre, M., Montoya-Burgos, J., Spierer, A., and Spierer, P. (2006). Conserved domains control heterochromatin localization and silencing properties of SU(VAR)3-7. Chromosoma *115*, 139-150.
Jaquet, Y., Delattre, M., Spierer, A., and Spierer, P. (2002). Functional dissection of the Drosophila modifier of variegation Su(var)3-7. Development *129*, 3975-3982.
Jiang, N., Emberly, E., Cuvier, O., and Hart, C.M. (2009). Genome-wide mapping of boundary element-associated factor (BEAF) binding sites in Drosophila melanogaster links BEAF to transcription. Mol Cell Biol *29*, 3556-3568.
Johnson, N.A. (2010). Hybrid incompatibility genes: remnants of a genomic battlefield? Trends Genet *26*, 317-325.
Kahn, T.G., Schwartz, Y.B., Dellino, G.I., and Pirrotta, V. (2006). Polycomb complexes and the propagation of the methylation mark at the Drosophila ubx gene. J Biol Chem *281*, 29064-29075.

Kalverda, B., and Fornerod, M. (2010). Characterization of genome-nucleoporin interactions in Drosophila links chromatin insulators to the nuclear pore complex. Cell Cycle *9*, 4812-4817.
Kaminker, J.S., Bergman, C.M., Kronmiller, B., Carlson, J., Svirskas, R., Patel, S., Frise, E., Wheeler, D.A., Lewis, S.E., Rubin, G.M., *et al.* (2002). The transposable elements of the Drosophila melanogaster euchromatin: a genomics perspective. Genome Biol *3*, RESEARCH0084.
Karlic, R., Chung, H.R., Lasserre, J., Vlahovicek, K., and Vingron, M. (2010). Histone modification levels are predictive for gene expression. Proc Natl Acad Sci U S A *107*, 2926-2931.
Karpen, G.H., Le, M.H., and Le, H. (1996). Centric heterochromatin and the efficiency of achiasmate disjunction in Drosophila female meiosis. Science *273*, 118-122.
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.
Klug, W.S., Bodenstein, D., and King, R.C. (1968). Oogenesis in the suppressor of hairy-wing mutant of Drosophila melanogaster. I. Phenotypic characterization and transplantation experiments. J Exp Zool *167*, 151-156.
Kofler, R., Betancourt, A.J., and Schlotterer, C. (2012). Sequencing of pooled DNA samples (Pool-Seq) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. PLoS Genet *8*, e1002487.
Kotnova, A.P., Karpova, N.N., Feoktistova, M.A., Liubomirskaia, N.V., Kim, A.I., and Il'in Iu, V. (2005). [Retrotransposon gtwin: structural analysis and distribution in Drosophila melanogaster strains]. Genetika *41*, 23-29.
Kunzelmann, S., Böttcher, R., Schmidts I., and Förstemann, K. (2016). A Comprehensive Toolbox for Genome Editing in Cultured Drosophila melanogaster Cells. G3 (Bethesda) *6*, 1777-1785.
Lachner, M., O'Carroll, D., Rea, S., Mechtler, K., and Jenuwein, T. (2001). Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. Nature *410*, 116-120.
Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.
Landt, S.G., Marinov, G.K., Kundaje, A., Kheradpour, P., Pauli, F., Batzoglou, S., Bernstein, B.E., Bickel, P., Brown, J.B., Cayting, P., *et al.* (2012). ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Genome Res *22*, 1813-1831.
Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol *10*, R25.
Le, M.H., Duricka, D., and Karpen, G.H. (1995). Islands of complex DNA are widespread in Drosophila centric heterochromatin. Genetics *141*, 283-303.
Lee, H.Y., Chou, J.Y., Cheong, L., Chang, N.H., Yang, S.Y., and Leu, J.Y. (2008). Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. Cell *135*, 1065-1073.
Lerat, E., Burlet, N., Biemont, C., and Vieira, C. (2011). Comparative analysis of transposable elements in the melanogaster subgroup sequenced genomes. Gene *473*, 100-109.
Lhoumaud, P., Hennion, M., Gamot, A., Cuddapah, S., Queille, S., Liang, J., Micas, G., Morillon, P., Urbach, S., Bouchez, O., *et al.* (2014). Insulators recruit histone

methyltransferase dMes4 to regulate chromatin of flanking genes. EMBO J *33*, 1599-1613.
Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R., and Genome Project Data Processing, S. (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics *25*, 2078-2079.
Li, J., and Gilmour, D.S. (2011). Promoter proximal pausing and the control of gene expression. Curr Opin Genet Dev *21*, 231-235.
Li, Y., Danzer, J.R., Alvarez, P., Belmont, A.S., and Wallrath, L.L. (2003). Effects of tethering HP1 to euchromatic regions of the Drosophila genome. Development *130*, 1817-1824.
Liang, Y., and Hetzer, M.W. (2011). Functional interactions between nucleoporins and chromatin. Curr Opin Cell Biol *23*, 65-70.
Locke, J., Kotarski, M.A., and Tartof, K.D. (1988). Dosage-dependent modifiers of position effect variegation in Drosophila and a mass action model that explains their effect. Genetics *120*, 181-198.
Lohe, A.R., Hilliker, A.J., and Roberts, P.A. (1993). Mapping simple repeated DNA sequences in heterochromatin of Drosophila melanogaster. Genetics *134*, 1149-1174.
Lopez-Falcon, B., Meyer-Nava, S., Hernandez-Rodriguez, B., Campos, A., Montero, D., Rudino, E., Vazquez, M., Zurita, M., and Valadez-Graham, V. (2014). Characterization of the Drosophila group ortholog to the amino-terminus of the alpha-thalassemia and mental retardation X-Linked (ATRX) vertebrate protein. PLoS One *9*, e113182.
Ludwig, A., and Loreto, E.L. (2007). Evolutionary pattern of the gtwin retrotransposon in the Drosophila melanogaster subgroup. Genetica *130*, 161-168.
MacQueen, J. (1967). Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, volume 1: Statistics. Some methods for classification and analysis of multivariate observations. (University of California Press).
Maheshwari, S., and Barbash, D.A. (2011). The genetics of hybrid incompatibilities. Annu Rev Genet *45*, 331-355.
Maheshwari, S., and Barbash, D.A. (2012). Cis-by-Trans regulatory divergence causes the asymmetric lethal effects of an ancestral hybrid incompatibility gene. PLoS Genet *8*, e1002597.
Maheshwari, S., Wang, J., and Barbash, D.A. (2008). Recurrent positive selection of the Drosophila hybrid incompatibility gene Hmr. Mol Biol Evol *25*, 2421-2430.
Maksimenko, O., and Georgiev, P. (2014). Mechanisms and proteins involved in long-distance interactions. Front Genet *5*, 28.
Mayr, E. (1942). Systematics and the Origin of Species (New York: Columbia University Press).
McKinley, K.L., and Cheeseman, I.M. (2016). The molecular basis for centromere identity and function. Nat Rev Mol Cell Biol *17*, 16-29.
Mefford, H.C., and Trask, B.J. (2002). The complex structure and dynamic evolution of human subtelomeres. Nat Rev Genet *3*, 91-102.
Mejlumian, L., Pelisson, A., Bucheton, A., and Terzian, C. (2002). Comparative and functional studies of Drosophila species invasion by the gypsy endogenous retrovirus. Genetics *160*, 201-209.
Meller, V.H., Joshi, S.S., and Deshpande, N. (2015). Modulation of Chromatin by Noncoding RNA. Annu Rev Genet *49*, 673-695.
Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C., and Forejt, J. (2009). A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science *323*, 373-375.
Moazed, D. (2009). Small RNAs in transcriptional gene silencing and genome defence. Nature *457*, 413-420.

Moshkovich, N., Nisha, P., Boyle, P.J., Thompson, B.A., Dale, R.K., and Lei, E.P. (2011). RNAi-independent role for Argonaute2 in CTCF/CP190 chromatin insulator function. Genes Dev *25*, 1686-1701.
Muller, H.J., and Pontecorvo, G. (1942). Recessive genes causing interspecific sterility and other disharmonies between Drosophila melanogaster and simulans. Genetics *27*, 157.
Mullis, K., Faloona, F., Scharf, S., Saiki, R., Horn, G., and Erlich, H. (1986). Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol *51 Pt 1*, 263-273.
Murzina, N., Verreault, A., Laue, E., and Stillman, B. (1999). Heterochromatin dynamics in mouse cells: interaction between chromatin assembly factor 1 and HP1 proteins. Mol Cell *4*, 529-540.
Negre, 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.
Negre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G., Feng, X., Ahmad, K., Russell, S., White, R.A., *et al.* (2010). A comprehensive map of insulator elements for the Drosophila genome. PLoS Genet *6*, e1000814.
Ni, X., Zhang, Y.E., Negre, N., Chen, S., Long, M., and White, K.P. (2012). Adaptive evolution and the birth of CTCF binding sites in the Drosophila genome. PLoS Biol *10*, e1001420.
Nicol, J.W., Helt, G.A., Blanchard, S.G., Jr., Raja, A., and Loraine, A.E. (2009). The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets. Bioinformatics *25*, 2730-2731.
Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated at lysine 9. Nature *416*, 103-107.
Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S., and Ponting, C.P. (2009). Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. PLoS Genet *5*, e1000753.
Ong, C.T., Van Bortle, K., Ramos, E., and Corces, V.G. (2013). Poly(ADP-ribosyl)ation regulates insulator function and intrachromosomal interactions in Drosophila. Cell *155*, 148-159.
Orlando, V. (2000). Mapping chromosomal proteins in vivo by formaldehyde-crosslinked-chromatin immunoprecipitation. Trends Biochem Sci *25*, 99-104.
Padeken, J., Mendiburo, M.J., Chlamydas, S., Schwarz, H.J., Kremmer, E., and Heun, P. (2013). The nucleoplasmin homolog NLP mediates centromere clustering and anchoring to the nucleolus. Mol Cell *50*, 236-249.
Pai, C.Y., Lei, E.P., Ghosh, D., and Corces, V.G. (2004). The centrosomal protein CP190 is a component of the gypsy chromatin insulator. Mol Cell *16*, 737-748.
Pal-Bhadra, M., Leibovitch, B.A., Gandhi, S.G., Chikka, M.R., Bhadra, U., Birchler, J.A., and Elgin, S.C. (2004). Heterochromatic silencing and HP1 localization in Drosophila are dependent on the RNAi machinery. Science *303*, 669-672.
Parkhurst, S.M., Harrison, D.A., Remington, M.P., Spana, C., Kelley, R.L., Coyne, R.S., and Corces, V.G. (1988). The Drosophila su(Hw) gene, which controls the phenotypic effect of the gypsy transposable element, encodes a putative DNA-binding protein. Genes Dev *2*, 1205-1215.
Pauli, T., Vedder, L., Dowling, D., Petersen, M., Meusemann, K., Donath, A., Peters, R.S., Podsiadlowski, L., Mayer, C., Liu, S., *et al.* (2016). Transcriptomic data from panarthropods shed new light on the evolution of insulator binding proteins in insects : Insect insulator proteins. BMC Genomics *17*, 861.

Phadnis, N., Baker, E.P., Cooper, J.C., Frizzell, K.A., Hsieh, E., de la Cruz, A.F., Shendure, J., Kitzman, J.O., and Malik, H.S. (2015). An essential cell cycle regulation gene causes hybrid inviability in Drosophila. Science *350*, 1552-1555.
Phadnis, N., and Orr, H.A. (2009). A single gene causes both male sterility and segregation distortion in Drosophila hybrids. Science *323*, 376-379.
Philip, P., Pettersson, F., and Stenberg, P. (2012). Sequence signatures involved in targeting the Male-Specific Lethal complex to X-chromosomal genes in Drosophila melanogaster. BMC Genomics *13*, 97.
Pollard, T.D., Earnshaw, W.C., and Lippincott-Schwartz, J. (2007). Cell Biology, 2nd Edition (Saunders).
Presgraves, D.C. (2010). The molecular evolutionary basis of species formation. Nat Rev Genet *11*, 175-180.
Presgraves, D.C., Balagopalan, L., Abmayr, S.M., and Orr, H.A. (2003). Adaptive evolution drives divergence of a hybrid inviability gene between two species of Drosophila. Nature *423*, 715-719.
Price, T. (2008). Speciation in Birds (Roberts & Company).
Prud'homme, N., Gans, M., Masson, M., Terzian, C., and Bucheton, A. (1995). Flamenco, a gene controlling the gypsy retrovirus of Drosophila melanogaster. Genetics *139*, 697-711.
Raffa, G.D., Ciapponi, L., Cenci, G., and Gatti, M. (2011). Terminin: a protein complex that mediates epigenetic maintenance of Drosophila telomeres. Nucleus *2*, 383-391.
Raffa, G.D., Raimondo, D., Sorino, C., Cugusi, S., Cenci, G., Cacchione, S., Gatti, M., and Ciapponi, L. (2010). Verrocchio, a Drosophila OB fold-containing protein, is a component of the terminin telomere-capping complex. Genes Dev *24*, 1596-1601.
Ramirez, F., Ryan, D.P., Gruning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dundar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res *44*, W160-165.
Razorenova, O.V., Karpova, N.N., Smirnova Iu, B., Kusulidu, L.K., Reneva, N.K., Subocheva, E.A., Kim, A.I., Liubomirskaia, N.V., and Il'in Iu, V. (2001). [Interlineage distribution and characteristics of the structure of two subfamilies of Drosophila melanogaster MDG4 (gypsy) retrotransposon]. Genetika *37*, 175-182.
Richards, E.J., and Elgin, S.C. (2002). Epigenetic codes for heterochromatin formation and silencing: rounding up the usual suspects. Cell *108*, 489-500.
Riddle, N.C., Minoda, A., Kharchenko, P.V., Alekseyenko, A.A., Schwartz, Y.B., Tolstorukov, M.Y., Gorchakov, A.A., Jaffe, J.D., Kennedy, C., Linder-Basso, D., *et al.* (2011). Plasticity in patterns of histone modifications and chromosomal proteins in Drosophila heterochromatin. Genome Res *21*, 147-163.
Rodriguez, M.A., Vermaak, D., Bayes, J.J., and Malik, H.S. (2007). Species-specific positive selection of the male-specific lethal complex that participates in dosage compensation in Drosophila. Proc Natl Acad Sci U S A *104*, 15412-15417.
Ross, B.D., Rosin, L., Thomae, A.W., Hiatt, M.A., Vermaak, D., de la Cruz, A.F., Imhof, A., Mellone, B.G., and Malik, H.S. (2013). Stepwise evolution of essential centromere function in a Drosophila neogene. Science *340*, 1211-1214.
Roy, S., Gilbert, M.K., and Hart, C.M. (2007). Characterization of BEAF mutations isolated by homologous recombination in Drosophila. Genetics *176*, 801-813.
Rozen, S., and Skaletsky, H. (2000). Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol *132*, 365-386.
Rus, F., Flatt, T., Tong, M., Aggarwal, K., Okuda, K., Kleino, A., Yates, E., Tatar, M., and Silverman, N. (2013). Ecdysone triggered PGRP-LC expression controls Drosophila innate immunity. EMBO J *32*, 1626-1638.

Ryan, K.J., and Wente, S.R. (2000). The nuclear pore complex: a protein machine bridging the nucleus and cytoplasm. Curr Opin Cell Biol *12*, 361-371.
Sadic, D., Schmidt, K., Groh, S., Kondofersky, I., Ellwart, J., Fuchs, C., Theis, F.J., and Schotta, G. (2015). Atrx promotes heterochromatin formation at retrotransposons. EMBO Rep *16*, 836-850.
Satyaki, P.R., Cuykendall, T.N., Wei, K.H., Brideau, N.J., Kwak, H., Aruna, S., Ferree, P.M., Ji, S., and Barbash, D.A. (2014). The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS Genet *10*, e1004240.
Sawamura, K. (2012). Chromatin evolution and molecular drive in speciation. Int J Evol Biol *2012*, 301894.
Sawamura, K., Maehara, K., Mashino, S., Kagesawa, T., Kajiwara, M., Matsuno, K., Takahashi, A., and Takano-Shimizu, T. (2010). Introgression of Drosophila simulans nuclear pore protein 160 in Drosophila melanogaster alone does not cause inviability but does cause female sterility. Genetics *186*, 669-676.
Sawamura, K., and Yamamoto, M.T. (1997). Characterization of a reproductive isolation gene, zygotic hybrid rescue, of Drosophila melanogaster by using minichromosomes. Heredity *79*, 97-103.
Sawamura, K., Yamamoto, M.T., and Watanabe, T.K. (1993). Hybrid lethal systems in the Drosophila melanogaster species complex. II. The Zygotic hybrid rescue (Zhr) gene of D. melanogaster. Genetics *133*, 307-313.
Schneider, I. (1972). Cell lines derived from late embryonic stages of Drosophila melanogaster. J Embryol Exp Morphol *27*, 353-365.
Schoborg, T.A., and Labrador, M. (2010). The phylogenetic distribution of non-CTCF insulator proteins is limited to insects and reveals that BEAF-32 is Drosophila lineage specific. J Mol Evol *70*, 74-84.
Schotta, G., Ebert, A., Krauss, V., Fischer, A., Hoffmann, J., Rea, S., Jenuwein, T., Dorn, R., and Reuter, G. (2002). Central role of Drosophila SU(VAR)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. EMBO J *21*, 1121-1131.
Schotta, G., and Reuter, G. (2000). Controlled expression of tagged proteins in Drosophila using a new modular P-element vector system. Mol Gen Genet *262*, 916-920.
Schwartz, Y.B., and Cavalli, G. (2017). Three-Dimensional Genome Organization and Function in Drosophila. Genetics *205*, 5-24.
Schwartz, Y.B., Linder-Basso, D., Kharchenko, P.V., Tolstorukov, M.Y., Kim, M., Li, H.B., Gorchakov, A.A., Minoda, A., Shanower, G., Alekseyenko, A.A., *et al.* (2012). Nature and function of insulator protein binding sites in the Drosophila genome. Genome Res *22*, 2188-2198.
Seehausen, O., Butlin, R.K., Keller, I., Wagner, C.E., Boughman, J.W., Hohenlohe, P.A., Peichel, C.L., Saetre, G.P., Bank, C., Brannstrom, A., *et al.* (2014). Genomics and the origin of species. Nat Rev Genet *15*, 176-192.
Seetharam, A.S., and Stuart, G.W. (2013). Whole genome phylogeny for 21 Drosophila species using predicted 2b-RAD fragments. PeerJ *1*, e226.
Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M., Parrinello, H., Tanay, A., and Cavalli, G. (2012). Three-dimensional folding and functional organization principles of the Drosophila genome. Cell *148*, 458-472.
Shukla, V., Habib, F., Kulkarni, A., and Ratnaparkhi, G.S. (2014). Gene duplication, lineage-specific expansion, and subfunctionalization in the MADF-BESS family patterns the Drosophila wing hinge. Genetics *196*, 481-496.
Singh, P.B., and Georgatos, S.D. (2002). HP1: facts, open questions, and speculation. J Struct Biol *140*, 10-16.

Soshnev, A.A., Baxley, R.M., Manak, J.R., Tan, K., and Geyer, P.K. (2013). The insulator protein Suppressor of Hairy-wing is an essential transcriptional repressor in the Drosophila ovary. Development *140*, 3613-3623.
Spierer, A., Begeot, F., Spierer, P., and Delattre, M. (2008). SU(VAR)3-7 links heterochromatin and dosage compensation in Drosophila. PLoS Genet *4*, e1000066.
Spierer, A., Seum, C., Delattre, M., and Spierer, P. (2005). Loss of the modifiers of variegation Su(var)3-7 or HP1 impacts male X polytene chromosome morphology and dosage compensation. J Cell Sci *118*, 5047-5057.
Spofford, J.B. (1976). Position-effect variegation in Drosophila. In The genetics and biology of Drosophila, vol 1c, M. Ashburner, and E. Novitski, eds. (New York: Academic Press), pp. 955–1018.
Stewart, M.D., Li, J., and Wong, J. (2005). Relationship between histone H3 lysine 9 methylation, transcription repression, and heterochromatin protein 1 recruitment. Mol Cell Biol *25*, 2525-2538.
Straub, T., and Becker, P.B. (2007). Dosage compensation: the beginning and end of generalization. Nat Rev Genet *8*, 47-57.
Sturtevant, A.H. (1920). Genetic Studies on DROSOPHILA SIMULANS. I. Introduction. Hybrids with DROSOPHILA MELANOGASTER. Genetics *5*, 488-500.
Sun, S., Ting, C.T., and Wu, C.I. (2004). The normal function of a speciation gene, Odysseus, and its hybrid sterility effect. Science *305*, 81-83.
Sun, X., Le, H.D., Wahlstrom, J.M., and Karpen, G.H. (2003). Sequence analysis of a functional Drosophila centromere. Genome Res *13*, 182-194.
Sun, X., Wahlstrom, J., and Karpen, G. (1997). Molecular structure of a functional Drosophila centromere. Cell *91*, 1007-1019.
Tan, J., Yang, X., Zhuang, L., Jiang, X., Chen, W., Lee, P.L., Karuturi, R.K., Tan, P.B., Liu, E.T., and Yu, Q. (2007). Pharmacologic disruption of Polycomb-repressive complex 2-mediated gene repression selectively induces apoptosis in cancer cells. Genes Dev *21*, 1050-1063.
Tang, S., and Presgraves, D.C. (2009). Evolution of the Drosophila nuclear pore complex results in multiple hybrid incompatibilities. Science *323*, 779-782.
Theurkauf, W.E., and Hawley, R.S. (1992). Meiotic spindle assembly in Drosophila females: behavior of nonexchange chromosomes and the effects of mutations in the nod kinesin-like protein. J Cell Biol *116*, 1167-1180.
Thomae, A.W., Schade, G.O., Padeken, J., Borath, M., Vetter, I., Kremmer, E., Heun, P., and Imhof, A. (2013). A pair of centromeric proteins mediates reproductive isolation in Drosophila species. Dev Cell *27*, 412-424.
Thomas, J.H., Emerson, R.O., and Shendure, J. (2009). Extraordinary molecular evolution in the PRDM9 fertility gene. PLoS One *4*, e8505.
Thorvaldsdottir, H., Robinson, J.T., and Mesirov, J.P. (2013). Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform *14*, 178-192.
Tie, F., Banerjee, R., Stratton, C.A., Prasad-Sinha, J., Stepanik, V., Zlobin, A., Diaz, M.O., Scacheri, P.C., and Harte, P.J. (2009). CBP-mediated acetylation of histone H3 lysine 27 antagonizes Drosophila Polycomb silencing. Development *136*, 3131-3141.
Ting, C.T., Tsaur, S.C., Wu, M.L., and Wu, C.I. (1998). A rapidly evolving homeobox at the site of a hybrid sterility gene. Science *282*, 1501-1504.
Towbin, H., and Gordon, J. (1984). Immunoblotting and dot immunobinding--current status and outlook. J Immunol Methods *72*, 313-340.
Van Bortle, K., Ramos, E., Takenaka, N., Yang, J., Wahi, J.E., and Corces, V.G. (2012). Drosophila CTCF tandemly aligns with other insulator proteins at the borders of H3K27me3 domains. Genome Res *22*, 2176-2187.

Vasu, S.K., and Forbes, D.J. (2001). Nuclear pores and nuclear assembly. Curr Opin Cell Biol *13*, 363-375.
Vedelek, B., Blastyak, A., and Boros, I.M. (2015). Cross-Species Interaction between Rapidly Evolving Telomere-Specific Drosophila Proteins. PLoS One *10*, e0142771.
Vermaak, D., and Malik, H.S. (2009). Multiple roles for heterochromatin protein 1 genes in Drosophila. Annu Rev Genet *43*, 467-492.
Vieira, C., and Biemont, C. (2004). Transposable element dynamics in two sibling species: Drosophila melanogaster and Drosophila simulans. Genetica *120*, 115-123.
Vieira, C., Fablet, M., Lerat, E., Boulesteix, M., Rebollo, R., Burlet, N., Akkouche, A., Hubert, B., Mortada, H., and Biemont, C. (2012). A comparative analysis of the amounts and dynamics of transposable elements in natural populations of Drosophila melanogaster and Drosophila simulans. J Environ Radioact *113*, 83-86.
Villa, R., Schauer, T., Smialowski, P., Straub, T., and Becker, P.B. (2016). PionX sites mark the X chromosome for dosage compensation. Nature *537*, 244-248.
Walker, P.M. (1971). Origin of satellite DNA. Nature *229*, 306-308.
Wallace, H.A. (2010). The role of the Suppressor of Hairy-wing insulator protein in chromatin organization and expression of transposable elements in Drosophila melanogaster (University of Tennessee).
Watanabe, T.K., and Kawanishi, M. (1979). Mating preference and the direction of evolution in Drosophila. Science *205*, 906-907.
Wei, K.H., Clark, A.G., and Barbash, D.A. (2014). Limited gene misregulation is exacerbated by allele-specific upregulation in lethal hybrids between Drosophila melanogaster and Drosophila simulans. Mol Biol Evol *31*, 1767-1778.
West, A.G., Gaszner, M., and Felsenfeld, G. (2002). Insulators: many functions, many mechanisms. Genes Dev *16*, 271-288.
Wilcoxon, F. (1946). Individual comparisons of grouped data by ranking methods. J Econ Entomol *39*, 269.
Wu, C.I., and Ting, C.T. (2004). Genes and speciation. Nat Rev Genet *5*, 114-122.
Yang, J., Ramos, E., and Corces, V.G. (2012). The BEAF-32 insulator coordinates genome organization and function during the evolution of Drosophila species. Genome Res *22*, 2199-2207.
Yang, Q., Rout, M.P., and Akey, C.W. (1998). Three-dimensional architecture of the isolated yeast nuclear pore complex: functional and evolutionary implications. Mol Cell *1*, 223-234.
Yasuhara, J.C., and Wakimoto, B.T. (2006). Oxymoron no more: the expanding world of heterochromatic genes. Trends Genet *22*, 330-338.
Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., *et al.* (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol *9*, R137.
Zhu, L.J., Gazin, C., Lawson, N.D., Pages, H., Lin, S.M., Lapointe, D.S., and Green, M.R. (2010). ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. BMC Bioinformatics *11*, 237.

DANKSAGUNG

An erster Stelle möchte ich Prof. Dr. Axel Imhof für die Aufnahme in sein Team für meine Masterarbeit und die anschließende Doktorarbeit danken. In den letzten Jahren hast Du mich auf jede erdenkliche Art unterstützt und gefördert, mir dabei gleichzeitig aber auch sehr viel Freiraum ermöglicht, so dass ich mich während der Zeit in Deinem Team persönlich und beruflich enorm weiterentwickeln konnte. Deine stets offene Tür, den ehrlichen und wertschätzenden Austausch auf Augenhöhe mit Dir, Dein stetes Interesse und Vermittlung neuer Ansätze und kreativer Ideen für mein Projekt habe ich sehr geschätzt. Auch die wissenschaftlichen und nicht-wissenschaftlichen Aktivitäten außerhalb der „eigenen vier Wände“, Konferenzen und Vorträge, das Science Camp, Skifahren und Hüttenbesuche mit der Gruppe und anderes sind nicht selbstverständlich und haben mich über die Zeit stets weiter motiviert.

Außerdem möchte ich Prof. Dr. Peter Becker dafür danken, dass ich die Zeit an seinem Institut verbringen konnte, in der Schillerstraße und ab 2015 im neuen BMC. Die kollegiale und freundschaftliche Atmosphäre, die Du im Lehrstuhl geschaffen hast, hat mich sehr positiv geprägt und inspiriert. Ein großes Dankeschön auch für Deinen Einsatz in allen Belangen der Doktoranden und Doktorandinnen, die ich im Molekularrat vertreten durfte. Ich habe die gemeinsamen Seminare im Institut, die Retreat-Fahrten, Chromatin Days und Laborausflüge sehr genossen! Danke auch an Edith Müller und Caroline Brieger für Eure unbürokratische und schnelle Unterstützung im Sekretariat.

I would like to thank the members of my Thesis Advisory Committee, Prof. Dr. Klaus Förstemann, Prof. Dr. Nicolas Gompel, Prof. Dr. Axel Imhof and Dr. Tobias Straub for critical and highly fruitful discussions on my project and their time in supporting my work and ideas.

The scientific process was strongly supported by Bo Sun and Dr. Pawel Smialowski who provided outstanding bioinformatics data analysis and interpretation. Thank you both for all your help and support! Your participation and contribution to the project was highly motivating and highly relevant for the project process and conclusions. You helped me in sharpening my bioinformatics mind and it was a pleasure working together with you! The hands-on work at the bench was strongly supported by Ülkü Uzun, who joined me as a summer student and helped me as a very talented and motivated student not only in establishing new technical procedures but also

challenged my mind by asking the right questions. Thank you Ülkü! Further, I want to thank Prof. Dr. Klaus Förstemann and his lab for the highly fruitful and important work and advice on the CRISPR/Cas9 genome editing and the scientific discussions we got involved in. In addition, I want to thank Dr. Stefan Krebs and the lab of Dr. Helmut Blum for outstanding service in high-throughput sequencing.

I would like to thank my current and former colleagues Dr. Andreas Thomae, Bo Sun, Andrea Lukacs, Natalia Kochanova and Ülkü Uzun, all working on the field of speciation. I enjoyed our in-deep discussions in the so called 'speciation meetings' and your continuous support, ideas and help as colleagues and friends!

I further want to thank all my other colleagues and friends of the Imhof team in particular Moritz, Irene, Magdalini, Simone, Miriam, Vio, Jianhua, Ignasi, Shibo, Andreas, Pierre and Marc. Thank you all for your daily support and help in all aspects whether it was a critical comment in the group meeting, helping out with reagents and protocols or an interesting scientific discussion outside of the lab. Thanks to my former mentor in the lab, Dr. Andreas Thomae for introducing me to the field of speciation, educating my mind for independent science and overwhelming support in getting started in the project.

You and all the members of the whole department, past and present, in particular the people from the 'nerd lab' and the 'octopi' made my PhD life a time that I will look back with pleasure. I enjoyed our discussions on science, even at very late hours, spontaneous lab dinners, hiking, skiing and so much more with you!

I would also like to thank the IRTG Program Coordinator Dr. Elizabeth Schroeder-Reiter for her huge commitment, Dr. Jörn Böke and the members of the PhD training. The programs activities and workshops were amazing and provided an outstanding platform for mutual scientific exchange and personal development.

Zuletzt gebührt ein sehr großer Dank meinen Eltern und meiner Schwester, für alles was sie mir ermöglicht haben und immer noch ermöglichen. Der Weg hierher wurde vor allem auch durch Eure fortwährende und kompromisslose Unterstützung und Wertschätzung geebnet!