Innovation of heterochromatin functions drives rapid evolution of essential ZAD-ZNF genes in *Drosophila*

Bhavatharini Kasinathan[1,2,3], Serafin U Colmenares III[4,5,6], Hannah McConnell[3], Janet M Young[3], Gary H Karpen[4,5,6], Harmit S Malik[3,7]*

[1]Medical Scientist Training Program, University of Washington School of Medicine, Seattle, United States; [2]Molecular and Cellular Biology Graduate program, University of Washington School of Medicine, Seattle, United States; [3]Division of Basic Sciences, Fred Hutchinson Cancer Research Center, Seattle, United States; [4]Biological Systems and Engineering Division, Lawrence Berkeley National Laboratory, Berkeley, United States; [5]Department of Molecular and Cell Biology, University of California at Berkeley, Berkeley, United States; [6]Innovative Genomics Institute, Berkeley, United States; [7]Howard Hughes Medical Institute, Fred Hutchinson Cancer Research Center, Seattle, United States

*For correspondence:
hsmalik@fhcrc.org

Competing interests: The authors declare that no competing interests exist.

Abstract Contrary to dogma, evolutionarily young and dynamic genes can encode essential functions. We find that evolutionarily dynamic *ZAD-ZNF* genes, which encode the most abundant class of insect transcription factors, are more likely to encode essential functions in *Drosophila melanogaster* than ancient, conserved *ZAD-ZNF* genes. We focus on the *Nicknack ZAD-ZNF* gene, which is evolutionarily young, poorly retained in *Drosophila* species, and evolves under strong positive selection. Yet we find that it is necessary for larval development in *D. melanogaster*. We show that *Nicknack* encodes a heterochromatin-localizing protein like its paralog *Oddjob*, also an evolutionarily dynamic yet essential *ZAD-ZNF* gene. We find that the divergent *D. simulans* Nicknack protein can still localize to *D. melanogaster* heterochromatin and rescue viability of female but not male *Nicknack*-null *D. melanogaster*. Our findings suggest that innovation for rapidly changing heterochromatin functions might generally explain the essentiality of many evolutionarily dynamic *ZAD-ZNF* genes in insects.

Introduction

Although organisms display enormous phenotypic diversity, their cellular organization and early development are highly conserved across broad taxonomic ranges (*Miklos and Rubin, 1996*). Such widespread conservation has led to a commonly held view that an ancient, conserved genetic architecture encodes fundamental biological functions, which appears to be largely borne out by comparative genomics (*Miklos and Rubin, 1996*; *Krylov et al., 2003*). However, recent studies demonstrated that 30% of 185 evolutionarily young genes in *D. melanogaster* (*Chen et al., 2010*) have acquired roles in development, cell biology, and reproduction that render them essential for viability or fertility (*Lee et al., 2017*; *Long et al., 2013*; *Ross et al., 2013*; *Lee et al., 2019*). Sometimes, evolutionary turnover of genes underlying essential cellular processes can occur, as seen in the evolution of kinetochore proteins (*Drinnenberg et al., 2016*). Even when they are retained over long evolutionary periods, genes encoding essential functions can evolve unexpectedly rapidly across plants and animals (*Malik and Henikoff, 2001*; *Talbert et al., 2004*). Thus, at least a subset

of essential functions is encoded by rapidly evolving genes or genes that are subject to high genetic turnover. The functional basis of this correlation, which runs counter to dogma, is unclear.

To study this unexpected class of rapidly evolving, essential genes further, we focused on the highly dynamic *ZAD-ZNF* gene family, which encodes several essential transcription factors. ZAD-ZNF proteins contain a conserved N-terminal ZAD (Zinc-finger-associated domain), a linker, and a C-terminal domain that includes tandem C2H2 zinc fingers (*Chung et al., 2002*; *Lespinet et al., 2002*). The ZAD facilitates protein-protein interactions but does not have DNA-binding ability, whereas the C2H2 zinc fingers often mediate sequence-specific DNA binding (*Jauch et al., 2003*). Unlike the ZAD and ZNF domains, which are homologous between different ZAD-ZNF proteins, the linker domains are highly variable across ZAD-ZNF proteins in both sequence and length and have no discernible structural motifs. *ZAD-ZNF* genes arose in the ancestor of vertebrates and arthropods, but dramatically expanded within insect lineages (*Chung et al., 2007*), becoming the most abundant class of TFs in many genomes, including in *D. melanogaster* (*Chung et al., 2002*). However, *ZAD-ZNF* gene repertoires can vary quite extensively across insect lineages via gene gains and losses (*Chung et al., 2002*; *Chung et al., 2007*); the causes and consequences of this gene dynamism are poorly understood.

Insect ZAD-ZNF proteins might act analogously to mammalian KRAB-Zinc finger (KZNFs), providing an explanation for their gene dynamism (*Chung et al., 2002*; *Chung et al., 2007*). *KZNF* genes arose in the ancestor of tetrapods and expanded through lineage-specific gene amplifications to become the most abundant class of transcription factors present in mammalian genomes (*Emerson and Thomas, 2009*; *Imbeault et al., 2017*). KZNF proteins contain an N-terminal KRAB (Krüppel-associated box) domain and C-terminal arrays of C2H2 zinc-finger domains that define their DNA-binding specificities. Many *KZNF* genes play a critical role in genome defense through recognition and repression of transposable elements (*Wolf and Goff, 2009*; *Rowe et al., 2010*). As a result, KZNF genes involved in genome defense are subject to rapid genetic turnover and positive selection of their DNA-binding domains (*Thomas and Schneider, 2011*). Similar functions might have driven the diversification of *ZAD-ZNF* genes in insects (*Chung et al., 2002*; *Chung et al., 2007*).

Despite their abundance, only a few *ZAD-ZNFs* have been functionally characterized, with studies restricted to *D. melanogaster*. Approximately half of all *ZAD-ZNF* genes are highly expressed in ovaries and early embryos (*Chung et al., 2002*), where they might play crucial roles for fertility or development. Biochemical characterization of 21 ZAD-ZNF proteins shows that they bind to unique DNA consensus sequences, with putative targets in the regulatory regions of specific target genes (*Krystel and Ayyanathan, 2013*). For example, the ZAD-ZNF protein M1BP (Motif 1 Binding Protein) binds core promoters and promotes the expression of numerous housekeeping genes (*Baumann et al., 2017*; *Li and Gilmour, 2013*). Similarly, *ZAD-ZNF* gene *Grauzone* promotes *cortex* expression and is necessary for meiotic progression during oogenesis (*Chen et al., 2000*; *Harms et al., 2000*; *Page and Orr-Weaver, 1996*). Finally *ZAD-ZNF* genes *Molting Defective, Ouija Board,* and *Séance* encode proteins that promote the transcription of heterochromatin-embedded genes, *Spookier* and *Neverland,* required for larval progression (*Uryu et al., 2018*). In contrast, other ZAD-ZNF proteins repress rather than drive transcription. For example, the ZAD-ZNF protein Odd-job localizes to pericentromeric heterochromatin and is required for gene silencing (position-effect variegation) (*Swenson et al., 2016*). Another ZAD-ZNF protein encoded by *CG17801* helps repress *HetA* and *Blood* transposable elements in the ovary (*Czech et al., 2013*).

Not all ZAD-ZNF functions are directly related to transcription. For example, the ZAD-ZNF proteins ZIPIC, Zw5, and Pita help organize chromatin architecture (*Gaszner et al., 1999*; *Zolotarev et al., 2016*) whereas the ZAD-ZNF protein Trade Embargo mediates the initiation of meiotic recombination during oogenesis (*Lake et al., 2011*). Finally, some ZAD-ZNFs might not function in the nucleus at all. Even though the Weckle ZAD-ZNF protein possesses zinc-finger domains, it localizes to the plasma membrane instead of nuclear chromatin, where it interacts with the Toll-MyD88 complex to help establish the anterior-posterior axis of the developing embryo (*Chen et al., 2006*).

The *ZAD-ZNF* gene family in *Drosophila* is ideal for studying the relationship between genetic innovation and essentiality because of its involvement with essential cellular processes despite rapid evolutionary dynamics. In order to study this relationship between rapid evolution and essentiality, we leveraged extensive phylogenomic and population genetics datasets in *Drosophila* species. We also took advantage of genome-wide screens for phenotype and tools for cytological and genetic

analyses in *D. melanogaster*. Using evolutionary analyses, we identified the *D. melanogaster ZAD-ZNF* genes that had been either subject to genetic turnover or positive selection. Although only a few *ZAD-ZNF* genes have undergone positive selection, we found that these genes are more likely to be required for viability or fertility in *D. melanogaster* than slowly-evolving *ZAD-ZNF* genes. We focused on the characterization of one of these positively-selected *ZAD-ZNF* genes: *Nicknack (CG17802, Nnk)*. We show that *Nicknack* is essential for larval development in *D. melanogaster* despite being evolutionarily young and differentially retained among *Drosophila* species. *Nicknack* belongs to a small cluster of *ZAD-ZNF* paralogs, the best characterized of which is *Oddjob (CG7357, Odj)* (*Swenson et al., 2016*). We show that both *Odj* and *Nnk* encode heterochromatin-localizing proteins. Although Odj broadly localizes to heterochromatin, we found that Nnk predominantly localizes to discrete foci within heterochromatin. Surprisingly, despite a strong signature of positive selection, we found that the protein encoded by the divergent *D. simulans Nicknack* ortholog can still localize to heterochromatin in *D. melanogaster* cells. Furthermore, *D. simulans Nicknack* can significantly rescue the viability of *D. melanogaster Nicknack-null* females but is unable to rescue *Nicknack-null* males. Based on our functional and cytological analyses, we conclude that rapidly changing requirements for heterochromatin function likely drove essential innovation of *ZAD-ZNF* genes such as *Nicknack* and *Oddjob* in *Drosophila*.

Results

ZAD-ZNFs genes are dynamic and diverse in *Drosophila*

We searched Flybase to identify all genes in *D. melanogaster* that encode a ZAD domain (PF07776; Pfam database, Pfam.org). We found 91 *ZAD-ZNF* genes distributed across Chromosomes 2, 3, and X. 37 *ZAD-ZNF* genes occur in 13 gene clusters, containing two or more tandemly-arrayed *ZAD-ZNF* genes. Of these, many *ZAD-ZNF* genes share intron/exon structures with their neighbors and likely arose via segmental duplication. In contrast, seven *ZAD-ZNF* genes (*CG3032, CG4318, CG9215, CG44002, CG17361, CG7963, CG17359*) lack introns found in their closest relatives; we infer that these genes were likely born via retrotransposition.

Although ZAD-ZNF proteins are defined as having both a ZAD and a ZNF domain, further analysis of ZAD-containing proteins using NCBI's Conserved Domain Database revealed significant variation in the ZNF domains. We found that ZAD-containing proteins have an average of six C2H2 domains. However, some ZAD-containing proteins have no C2H2 domains at all (*dlip, dbr, CG15435, CG31109, CG31457*), whereas others contain up to 23 C2H2 domains (*CG11902*). In addition to the ZAD in the N-terminus, 13 ZAD-ZNF proteins possess another N-terminal domain, such as a Sodium-Calcium exchanger domain (*CG12391*) or ASF1 histone chaperone domain (*CG10321*). The functional significance of these additional domains is unknown.

Next, we surveyed *ZAD-ZNF* genes in all 12 previously sequenced and annotated *Drosophila* genomes. These species represent a range of evolutionary divergence from *D. melanogaster*, from just a few million years (*e.g., D. simulans*), to more than 40 million years (*e.g., D. virilis*) (*Drosophila 12 Genomes Consortium, 2007*). Our analysis reveals a surprisingly wide range in number of *ZAD-ZNF* genes across different *Drosophila* species genomes (*Figure 1*). For example, we found that the *D. persimilis* genome encodes 130 *ZAD-ZNF* genes whereas the *D. willistoni* genome encodes only 75 *ZAD-ZNF* genes (*Supplementary file 1*). Such analyses are dependent on the state of completion and annotation of individual *Drosophila* species' genomes. Thus, this number may be an underestimate. To complement these analyses, we assessed the apparent age of each *D. melanogaster ZAD-ZNF* gene by examining the presence of orthologs in the 12 annotated genomes of *Drosophila* species (flybase.org). We found that 74 of 91 *D. melanogaster ZAD-ZNF* genes arose in or prior to the common ancestor of all *Drosophila* species, 40 million years ago. Of these 74 *ZAD-ZNF* genes, 61 have been preserved over *Drosophila* evolution, whereas 13 have been lost in at least one lineage or species. We estimate that 17 *ZAD-ZNF* genes arose via gene duplication after the last common ancestor of *D. melanogaster* and *D. virilis*. At least 3 *ZAD-ZNF* genes found in *D. melanogaster* (*CG4318, CG17612, neu2*) originated via gene duplication less than 10 million years ago. Our findings complement previous large-scale surveys that identified rapid changes in *ZAD-ZNF* gene repertoires within insect genomes (*Chung et al., 2002*; *Lespinet et al., 2002*; *Chung et al., 2007*).
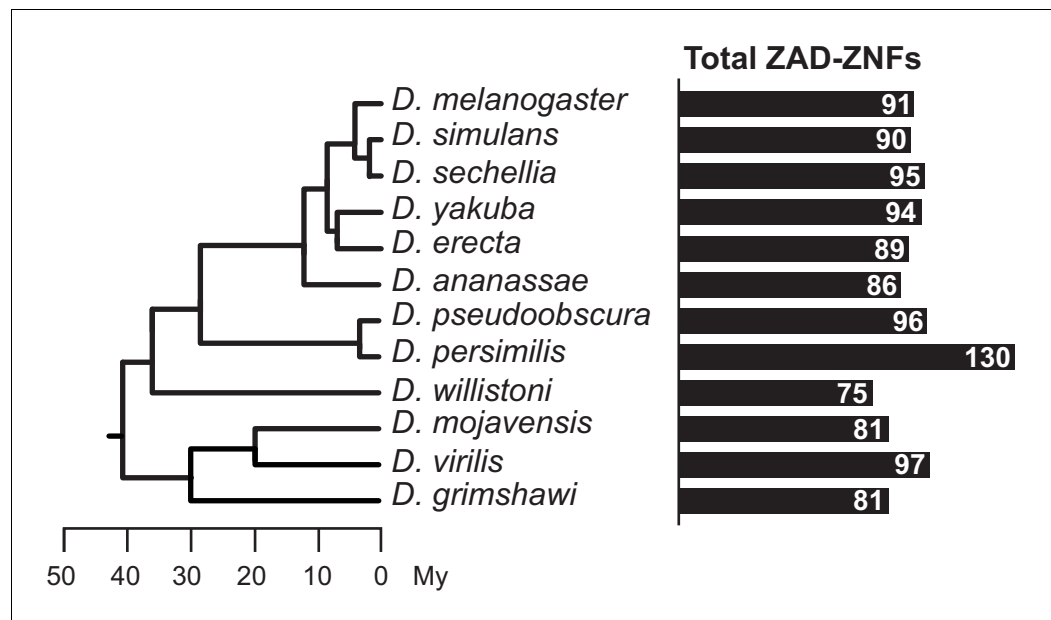
Figure 1. Total number of *ZAD-ZNF* genes across *Drosophila*. Phylogeny of 12 *Drosophila* genomes with a scale bar showing approximate divergence times (**Drosophila 12 Genomes Consortium, 2007**). The number of *ZAD*-containing genes in each *Drosophila* species genome (**Supplementary file 1**) is indicated by black bars.

Rapidly evolving *ZAD-ZNF* genes are frequently essential in *D. melanogaster*

These rapid changes in *ZAD-ZNF* gene repertoires suggested that selection might favor their genetic innovation. We investigated whether evolutionary retention is a predictor of essentiality. We took advantage of the fact that knockdown or knockout phenotypes have been characterized for almost all *D. melanogaster genes.* Indeed, 85 of 91 *D. melanogaster ZAD-ZNF* genes have associated phenotypic outcomes (summarized in **Supplementary file 1**). Of these, knockdown or knockout of 22 *ZAD-ZNF* genes showed complete lethality or sterility in *D. melanogaster*, whereas the other 63 were determined not to be essential.

Of the 61 *ZAD-ZNF* genes with orthologs retained in all 12 annotated *Drosophila* species' genomes, we found that 14 are essential for either fertility or viability in *D. melanogaster*, whereas 42 genes are not (five genes have no phenotypic data available). In comparison, we found that 8 of the 30 genes not universally conserved in *Drosophila* species are essential, whereas 21 are not (one has no phenotypic data available). Thus, somewhat surprisingly, we find that genes not globally retained over *Drosophila* evolution are just as likely to encode a necessary function in *D. melanogaster* as genes that have been strictly maintained over 40 million years of *Drosophila* evolution (8:21 versus 14:42, p=0.8, two-tailed Fisher's exact test) (**Table 1**). These findings further support the idea that gene families subject to rapid evolutionary turnover may become involved in essential functions.

Gene duplication and loss is only one form of genetic innovation. We also analyzed the *D. melanogaster ZAD-ZNF* genes for signatures of recent positive selection. We performed McDonald-Kreitman tests of *ZAD-ZNF* orthologs found in both *D. melanogaster* and the closely-related species, *D. simulans*. The McDonald-Kreitman test assesses sequence diversity within a species versus divergence between species by comparing the ratio of non-synonymous (amino-acid altering, or replacement) to synonymous substitutions fixed during the divergence of the two species (Dn: Ds), to that of non-synonymous to synonymous polymorphisms within a species (Pn: Ps) (**McDonald and Kreitman, 1991**). The Pn: Ps ratio is a proxy for functional constraint acting on a gene within species and is expected to be similar to Dn: Ds between species under the null hypothesis. However, a higher than expected number of fixed non-synonymous changes would indicate the action of adaptive evolution during species divergence (**McDonald and Kreitman, 1991**).

Table 1. Links between *ZAD-ZNF* gene dynamism and essentility.
(A) *ZAD-ZNF* genes are just as likely to be essential whether or not they are conserved across all 12 *Drosophila* species. Note that five genes for which phenotypic data is not available were not included in these analyses. (B) Positively-selected *ZAD-ZNF*s (via McDonald-Kreitman test, *Table 2*) are more likely to be essential than *ZAD-ZNF*s that have not evolved under positive selection. *D. simulans* lacks one of the 91 *D. melanogaster* genes and no data on essentiality is available for six genes. p-values were calculated by two-tailed Fisher's exact test.

Conservation of *ZAD-ZNF*s is not correlated with essentiality in *Drosophila*

ZAD-ZNFs	Essential	Not essential	p-value
conserved across *Drosophila*	14	42	-
not conserved across *Drosophila*	8	21	0.80

Positively selected *ZAD-ZNF*s are more likely to be essential than *ZAD-ZNF*s not under positive selection in *D. melanogaster*

ZAD-ZNFs	Essential	Not essential	p-value
not under positive selection	14	59	-
positive selection	8	4	0.0016

We took advantage of previous efforts that sequenced the genomes of hundreds of *D. melanogaster* strains and a reference *D. simulans* strain (*Lack et al., 2016*; *Langley et al., 2012*) to perform the McDonald-Kreitman test using the Popfly server (popfly.uab.cat) (*Hervas et al., 2017*). Of the 91 *D. melanogaster ZAD-ZNF* genes, only *CG2202* is absent in *D. simulans*. We found that 12 out of the remaining 90 *ZAD-ZNF*s show evidence for recent adaptive evolution, *i.e.*, have an excess of fixed non-synonymous changes (Dn) (*Table 2*). Using *D. yakuba* as an outgroup species, we also polarized fixed differences between *D. melanogaster* and *D. simulans* to assess whether the lineage leading to *D. melanogaster* showed evidence of positive selection in these 12 *ZAD-ZNF* genes. We found that 5 out of 12 genes showed evidence of positive selection with this polarized McDonald-Kreitman test (*Table 2*).

Subsequently, we used a domain-restricted analysis to define which of the three protein domains (ZAD, linker, or ZNF) were subject to positive selection. Although we found evidence of domain-specific positive selection in eight cases (two in the case of the polarized McDonald-Kreitman test), we were unable to define the domain subject to positive selection in four cases, due to small numbers of intra-species polymorphisms (*Table 2*). One gene (*CG7386*) showed signatures of adaptive evolution in its ZAD domain, whereas three genes (*CG2712, CG7386, CG10321*) showed evidence of adaptive evolution in the ZNF domain. Unexpectedly, we found that the linker domain has evolved under positive selection in six *ZAD-ZNF* genes. The biochemical function of these linker domains is largely unexplored as they lack predicted structural motifs (*Chung et al., 2002*; *Chung et al., 2007*), are highly variable in length and sequence even between orthologs, and are predicted to encode intrinsically disordered domains (*Ishida and Kinoshita, 2007*). Our findings implicate the poorly characterized linker region in mediating the adaptive potential of several *ZAD-ZNF* genes.

Using this signature of positive selection in a subset of *ZAD-ZNF* genes, we evaluated whether positively selected genes encode essential functions. Intriguingly, we found that 8 of 12 *ZAD-ZNF* genes that have evolved under positive selection are essential, whereas only 14 of the remaining 73 genes that have been phenotypically assayed are essential (six have no phenotypic data available). Thus, contrary to the dogma, we find that *ZAD-ZNF* genes that are subject to positive selection are more likely to be essential for viability or fertility (8:4 versus 14:59, two-tailed Fisher's exact test, p=0.0016) (*Table 1*). Our findings not only imply that several essential *ZAD-ZNF* genes evolve rapidly, but also raise the possibility that rapid evolution of some *ZAD-ZNF* genes might be critical for organismal viability.

Table 2. Positively selected *ZAD-ZNF*s.

Summary statistics for the McDonald-Kreitman (MK) test (with statistically significant values at p<0.05). The number of polymorphisms within *D. melanogaster* at synonymous (Ps) and non-synonymous (Pn) sites are compared to the number of all fixed synonymous (Ds) and non-synonymous (Dn) changes between *D. melanogaster* and *D. simulans*. A polarized MK test only considers synonymous (Ds) and non-synonymous (Dn) changes that were fixed along the lineage leading to *D. melanogaster*. MK test statistics were calculated either for the whole gene or individual domains. A neutrality index (N.I.) of <1 suggests an excess of fixed non-synonymous changes between species (i.e., positive selection). Phenotypes of these genes are taken from sources that are listed in ***Supplementary file 1***.

ZAD-ZNF gene (mutant phenotype)	Region	Alignment length (codons)	McDonald-Kreitman test (unpolarized)						Polarized changes (*D. melanogaster*)				Polarized changes (*D. simulans*)		Total polarized Dn	Total polarized Ds
			p-value	Dn	Ds	Pn	Ps	NI	p-value	Dn	Ds	NI	Dn	Ds		
Nnk (CG17802) (lethal)	full-length	448	**0.007**	52	24	14	20	0.323	**0.041**	17	8	0.329	26	16	43	24
	ZAD	71	0.119	6	2	2	4	0.167	0.058	3	0	0.000	3	2	6	2
	linker	216	**0.019**	34	10	9	10	0.265	0.070	11	3	0.245	15	7	26	10
	C2H2	136	0.091	6	11	0	6	0.000	0.155	2	5	0.000	4	6	6	11
Odj (CG7357) (lethal)	full-length	438	**0.000**	40	32	5	25	0.160	**0.002**	18	15	0.167	16	17	34	32
	ZAD	73	0.197	4	4	0	2	0.000	0.248	1	1	0.000	3	3	4	4
	linker	138	0.053	18	9	4	8	0.250	0.309	5	4	0.400	9	5	14	9
	C2H2	135	0.099	2	7	0	11	0.000		0	4		2	3	2	7
Trem (CG4413) (female sterile)	full-length	442	**0.015**	31	30	3	14	0.207	0.335	6	13	0.464	21	17	27	30
	ZAD	77	0.571	2	6	0	1	0.000		0	1		2	5	2	6
	linker	192	**0.036**	18	13	2	8	0.181	0.407	4	7	0.438	13	6	17	13
	C2H2	133	0.239	3	10	0	5	0.000	0.338	1	5	0.000	2	5	3	10
Zw5 (CG2711) (lethal)	full-length	590	**0.004**	55	51	6	22	0.253	0.120	15	23	0.418	28	28	43	51
	ZAD	78	0.125	3	4	0	4	0.000	0.285	1	3	0.000	1	1	2	4
	linker	229	0.251	32	17	3	4	0.398	0.867	7	8	0.857	16	9	23	17
	C2H2	217	0.060	12	25	0	8	0.000	0.159	3	11	0.000	7	14	10	25
CG2712 (viable)	full-length	524	**0.003**	70	36	15	24	0.321	0.210	21	19	0.565	37	15	58	34
	ZAD	80	0.161	10	6	3	6	0.300	0.569	2	2	0.500	7	4	9	6
	linker	231	0.089	45	14	10	8	0.389	0.106	16	4	0.313	18	8	34	12
	C2H2	190	**0.038**	15	14	2	10	0.187	0.759	3	11	0.733	12	3	15	14
D19B (CG10270) (viable)	full-length	774	**0.005**	11	42	3	66	0.174	**0.049**	4	20	0.227	5	21	9	41
	ZAD	72	0.121	2	4	0	6	0.000	0.134	1	2	0.000	1	2	2	4
	linker	171	0.182	6	12	2	13	0.308	0.347	0	6		4	5	4	11
	C2H2	257		0	6	0	25			0	3		0	3	0	6
	C2H2-2	78		0	2	0	7			0	2		0	0	0	2

*Table 2 continued on next page*

Table 2 continued

ZAD-ZNF gene (mutant phenotype)	Region	Alignment length (codons)	McDonald-Kreitman test (unpolarized)						Polarized changes (*D. melanogaster*)				Polarized changes (*D. simulans*)		Total polarized Dn	Total polarized Ds
			p-value	Dn	Ds	Pn	Ps	NI	p-value	Dn	Ds	NI	Dn	Ds		
CG7386 (lethal)	full-length	688	0.002	78	38	28	36	0.379	0.117	26	18	0.538	47	19	73	37
	ZAD	72	0.858	3	6	3	5	1.200	0.506	1	4	2.400	1	2	2	6
	linker	116	0.118	12	5	13	1	5.417	0.133	4	2	6.500	7	3	11	5
	C2H2-1	21		0	0	0	1			0	0		0	0	0	0
	C2H2-2	163	0.005	14	7	5	16	0.156	0.639	2	4	0.625	12	3	14	7
	C2H2-3	78	0.098	2	6	3	1	9.000	0.540	1	1	3.000	0	5	1	6
CG17359 (lethal)	full-length	344	0.000	43	17	3	14	0.085	0.000	18	6	0.071	20	11	38	17
	ZAD	82	0.157	3	1	1	3	0.111	0.540	1	1	0.333	1	0	2	1
	linker	134	0.000	28	4	1	5	0.029	0.001	9	0	0.000	15	4	24	4
	C2H2	105	0.064	10	10	0	4	0.000	0.040	6	4	0.000	4	6	10	10
wek (CG4148) (lethal; maternal effect lethal)	full-length	473	0.001	29	42	3	30	0.145	0.596	3	19	0.633	16	23	19	42
	ZAD	70	0.206	2	3	0	3	0.000		0	3		1	0	1	3
	linker	194	0.012	22	19	3	14	0.185	0.456	3	7	0.500	11	12	14	19
	C2H2	157	0.074	5	17	0	12	0.000		0	6		4	11	4	17
CG10321 (viable)	full-length	835	0.001	28	50	9	63	0.255	0.007	11	20	0.260	14	29	25	49
	ZAD	69		0	2	0	4			0	0		0	2	0	2
	linker	546	0.008	24	38	8	42	0.302	0.072	8	15	0.357	13	22	21	37
	C2H2	146	0.024	3	7	0	15	0.000	0.010	2	3	0.000	1	4	3	7
mld (CG34100) (lethal)	full-length	2021	0.020	84	70	37	57	0.541	0.328	32	36	0.730	37	32	69	68
	ZAD	70		0	0	0	0			0	0		0	0	0	0
	linker	1143	0.282	63	51	30	34	0.714	0.824	24	25	0.919	29	24	53	49
	linker-2	278	0.075	6	8	2	13	0.205	0.292	2	4	0.308	2	4	4	8
	C2H2-1	81	0.237	4	5	0	2	0.000	0.290	2	3	0.000	2	2	4	5
	C2H2-2	110	0.427	1	1	1	4	0.250	0.121	1	0	0.000	0	1	1	1
CG4282 (viable)	full-length	652	0.030	13	32	2	25	0.197	0.061	6	16	0.213	8	15	14	31
	ZAD	75	0.053	2	0	1	4	0.000	0.121	1	0	0.000	1	0	2	0
	linker	248	0.252	9	18	1	7	0.286	0.433	4	11	0.393	6	6	10	17
	C2H2	290		0	11	0	14			0	3		0	8	0	11

The *Oddjob-Nicknack* cluster of *ZAD-ZNF* genes evolves dynamically in *Drosophila*

To further investigate the biological basis for the correlation between positive selection and gene essentiality, we decided to focus on one cluster of *ZAD-ZNF* genes on chromosome 3 of *D. melanogaster* (*Figure 2A*). This cluster of five genes contains two of the eight positively-selected, essential genes in *D. melanogaster*: *Oddjob* (*Odj, CG7357*) and *CG17802*; both also show evidence of positive selection in the polarized McDonald-Kreitman test (*Table 2*). *Odj is the* best-characterized gene in this cluster. The Oddjob protein interacts with Heterochromatin Protein 1a (HP1a),
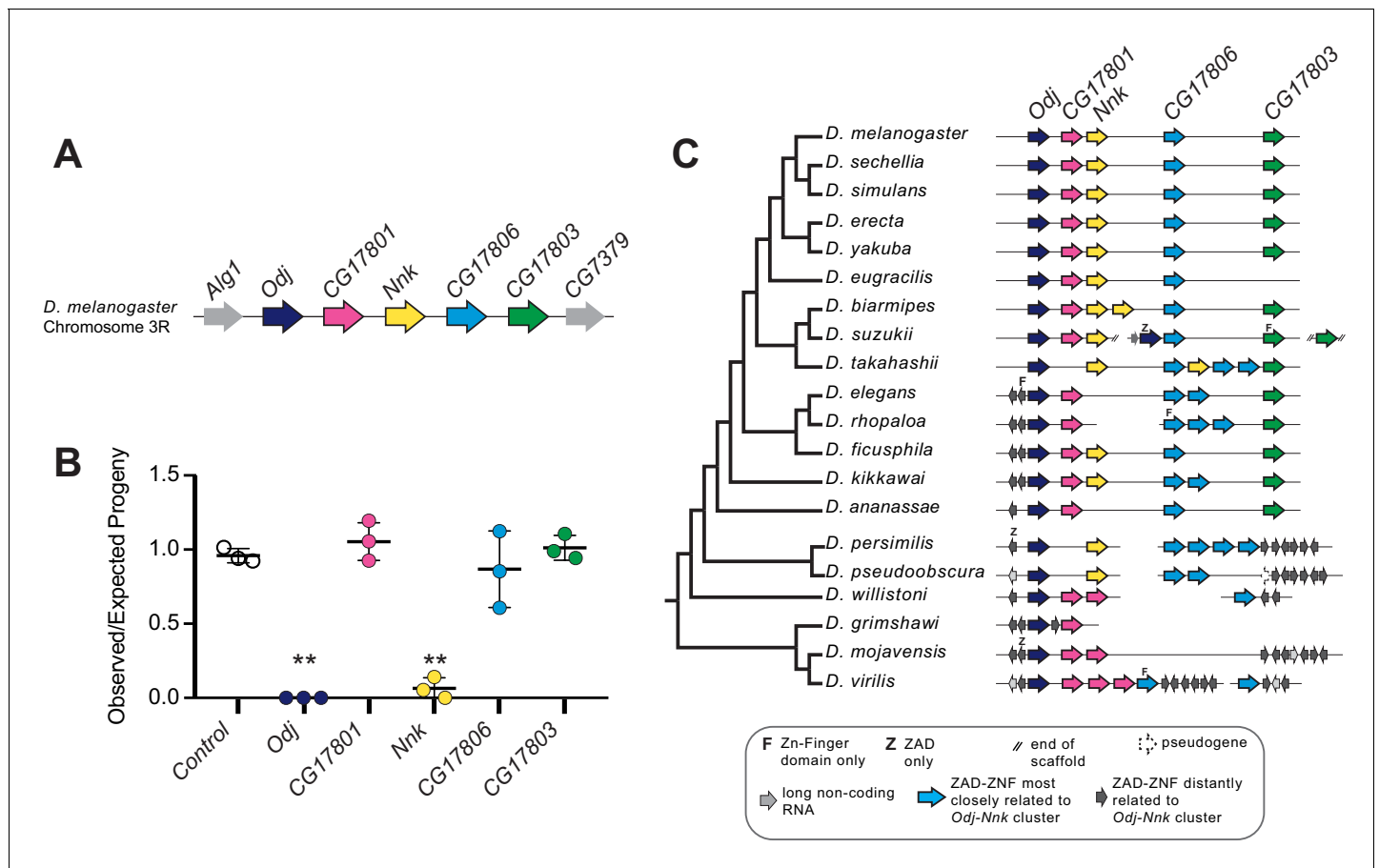
Figure 2. *Nicknack* and *Oddjob* are essential for viability. (**A**) A schematic of the *Oddjob-Nicknack* cluster of *ZAD-ZNF* genes in *D. melanogaster.* (**B**) Viability of adult flies ubiquitously knocked down for each *Odj-Nnk* cluster member. The vertical axis shows a ratio of the observed/expected number of knockdown progeny per cross; each cross is represented by a point. Ubiquitous knockdown of either *Odj* or *Nnk* greatly reduces adult viability. Horizontal bars represent the mean and one standard deviation. We compared controls and *ZAD-ZNF* knockdowns using a two-tailed Student's t-test; ** denotes p-value<0.01. (**C**) Phylogenomic analysis of the *Odj-Nnk* cluster shows that *Oddjob, CG17801* and *CG17806* all date back to the origin of the *Drosophila* genus. However, *Oddjob* is the only gene within the cluster retained in all queried *Drosophila* species; *CG17801* has been lost within the *obscura* clade and in *D. takahashii*, whereas *CG17806* has undergone numerous gains and losses. *CG17803* arose in the ancestor of *D. melanogaster* and *D. ananassae* but was lost in *D. eugracilis. Nicknack* arose in the ancestor of *D. melanogaster* and *D. pseudoobscura*, duplicated in *D. biarmipes* and *D. takahashii,* but was independently lost in *D. ananassae* and the *D. elegans/D. rhopaloa* lineage.

The online version of this article includes the following figure supplement(s) for figure 2:

Figure supplement 1. Phylogenetic analysis of all genes found within the *Odj-Nnk ZAD-ZNF* cluster across 20 *Drosophila* species based on a multiple alignment of their ZNF domains (**Supplementary file 2**).

dynamically localizes to heterochromatin, and is required for heterochromatin-mediated gene silencing, or position-effect variegation (**Swenson et al., 2016**). Knockdown or over-expression of *Odj* is lethal in *D. melanogaster*, although the functional basis for this lethality is still unknown (**Schertel et al., 2015**). In keeping with the *Oddjob* nomenclature theme of 'James Bond henchmen,' and the mutant phenotype (described below), we renamed *CG17802* as *Nicknack*, or *Nnk*. This cluster of *ZAD-ZNF* genes also includes three genes that do not evolve under positive selection (*CG17801, CG17806, and CG17803*) in *D. melanogaster* (**Figure 2A**). Although previous studies found that the knockdown of *CG17801* in the germline led to de-repression of HetA and Blood transposable elements in the ovary, *CG17801* knockdown did not significantly impair fertility or viability (**Czech et al., 2013**). The other two *ZAD-ZNF* genes in this cluster (*CG17803, CG17806*) have not been previously investigated.

We evaluated the functional roles of all five *ZAD-ZNF* genes in the *Odj-Nnk* cluster by generating knockdowns using an Actin5C-Gal4 driver and RNAi constructs specific to each gene (*Figure 2B*) to generate ubiquitous knockdown of individual genes in the resulting zygotes. Consistent with previous results (*Schertel et al., 2015*), we found that the ubiquitous knockdown of *Odj* or *Nnk* is lethal. In contrast, we found that knockdowns of *CG17801*, *CG17803* and *CG17806* are viable (*Figure 2B*). Based on these results, we conclude that the two positively-selected members of this *ZAD-ZNF* cluster – *Odj* and *Nnk* – are both essential, confirming the unexpected correlation we previously observed between positive selection and gene essentiality in the *ZAD-ZNF* gene family (*Table 1B*).

To gain deeper insight into the evolutionary dynamics of the *Odj-Nnk* cluster, we identified orthologs of these genes using reciprocal TBLASTN searches with each of the five *D. melanogaster* genes as queries. We searched both the originally-sequenced, well-annotated 12 *Drosophila* genomes (*Drosophila 12 Genomes Consortium, 2007*), as well as eight additional genomes that were subsequently sequenced to sample the melanogaster group within the *Sophophora* subgenus more densely (*Chen et al., 2014*). In several cases, we were not able to confidently assign the TBLASTN hits to orthologous groups because they matched closely to more than one *D. melanogaster* gene, or there were several putative hits within a single genome, or because the hit contained only the ZAD or ZNF domain.

To more rigorously identify orthologs, we conducted phylogenetic analyses (*Figure 2—figure supplement 1*) using a multiple alignment of the ZNF domain (*Supplementary file 2*). Our phylogenomic analyses reveal that *Odj-Nnk* cluster evolution was highly dynamic during the evolution of the *Drosophila* genus (summarized in *Figure 2C*). Although *Oddjob* orthologs are present throughout 40 million years of *Drosophila* evolution, no other member of the *D. melanogaster Odj-Nnk* cluster is universally present in *Drosophila* species. In addition to *Odj*, *CG17801* and *CG17806* also date back prior to the origin of *Drosophila* but unlike *Odj*, *CG17801* and *CG17806* have since been lost in some species (*Figure 2C*). While *CG17801* has been lost in the *obscura* group,*CG17806* underwent multiple independent duplication and loss events. *CG17803* arose in the ancestor of *D. melanogaster* and *D. ananassae* and underwent two independent losses. Finally, *Nnk* appears to have arisen in the ancestor of *D. melanogaster* and *D. pseudoobscura* (~30 mya, *Figure 1*), and later experienced multiple independent duplications and losses (*Figure 2C*). We note that our estimates for the age of *Nnk* are higher than those reported previously (*Chen et al., 2010*), likely because the prior estimate was based on fewer sequenced species. Based on these analyses, we conclude that, despite its essential function in *D. melanogaster*, *Nnk* is not universally conserved in *Drosophila* species.

Nicknack is an essential ZAD-ZNF gene in D. melanogaster

Nnk has a dramatic evolutionary history: young evolutionary age, differential retention, positive selection. Yet, it serves an essential function in *D. melanogaster* based on our and other previous analyses. However, this claim of essentiality has been challenged by two previous findings. First, the screen that originally identified *Nnk* as an essential gene used the KK RNA-interference (RNAi) collection from the Vienna *Drosophila* Stock Center (VDRC) (*Figure 3A*; *Chen et al., 2010*). Many lines in this 'KK' collection were later found to harbor a second-site mutation that caused lethality as a result of ectopic *tiptop* expression when crossed to *GAL4*-driver lines (*Green et al., 2014*; *Vissers et al., 2016*). As a result of its dependence on the KK lines, the claim of *Nnk* essentiality remained ambiguous. A second *Nnk* mutant allele, created by CRISPR-Cas9-mediated mutagenesis (hereafter referred to as *CRISPR-null*), had a four base pair deletion within the coding sequence of the gene that created a frameshift and a premature stop codon within the linker domain of *Nnk* (*Kondo et al., 2017*; *Figure 3A*). Although this deletion was homozygous lethal, it was unexpectedly viable when paired with a deficiency covering this *ZAD-ZNF* cluster (*Kondo et al., 2017*), again challenging the result that *Nnk* is an essential gene.

Given these ambiguous results and the importance of *Nnk* for our claims of *ZAD-ZNF* essentiality despite evolutionary innovation, we re-investigated whether *Nnk* is essential for viability in *D. melanogaster*. We found several lines of evidence to support the conclusion that it is indeed essential (*Figure 3B–D*). First, we began by validating the RNAi line used in the original study (*Chen et al., 2010*), which first identified *Nnk* as a young, essential gene. We found that the *Nnk* VDRC RNAi line does not have an insertion upstream of the *tiptop* gene, which is associated with lethality in other KK lines (*Green et al., 2014*; *Vissers et al., 2016*). Second, we were able to rescue this lethality via
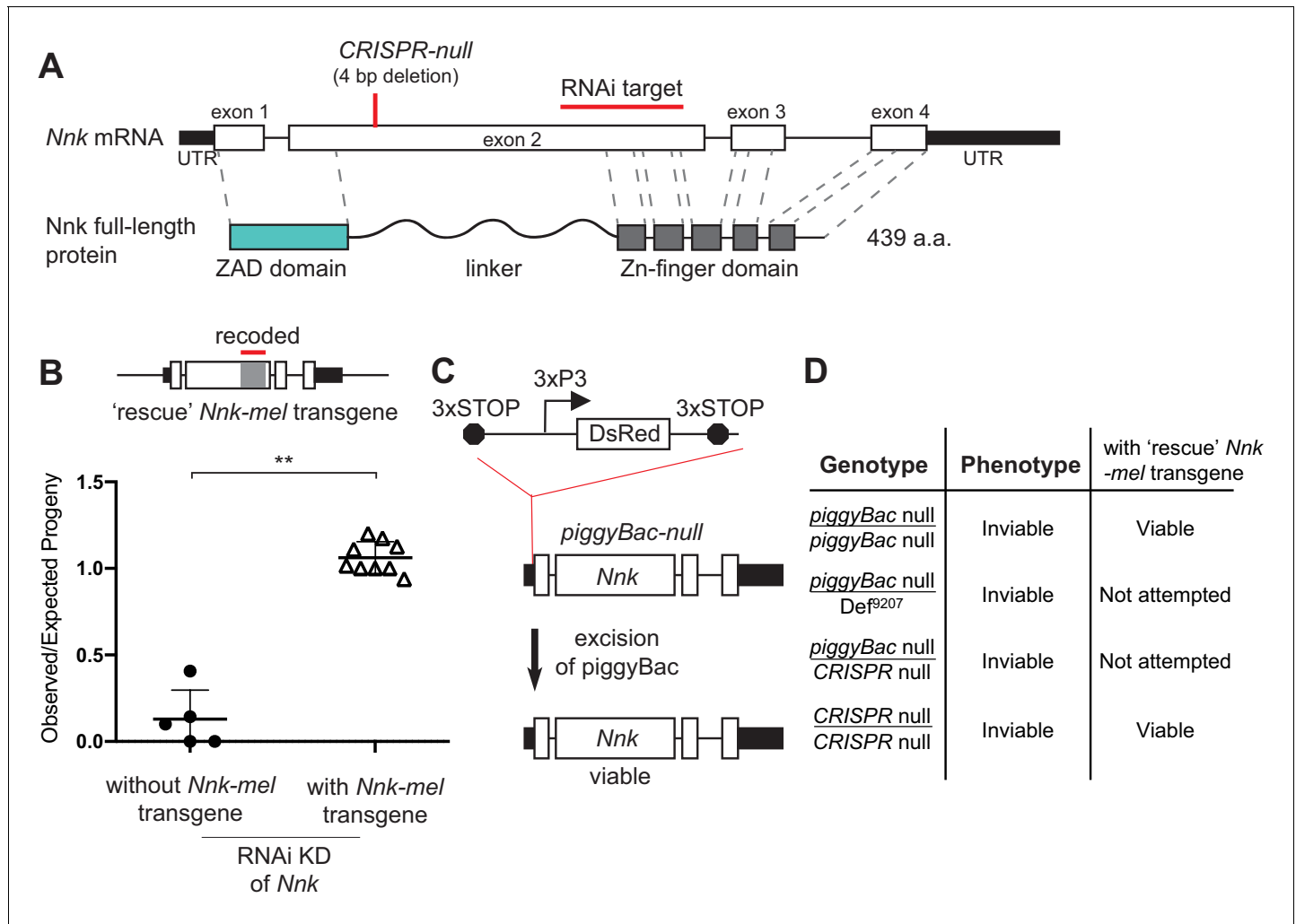
Figure 3. *Nicknack* RNAi and mutant alleles. (A) Schematic of *Nnk* gene. The *Nnk* gene contains four coding exons and encodes a full-length protein containing ZAD, linker, and ZNF domains. The CRISPR-null is a 4 bp deletion within exon 2 (*Kondo et al., 2017*): this deletion creates a frameshift and a premature stop codon within the linker region. The VDRC RNAi KK line creates double-stranded RNA corresponding to a region of exon 2 (schematized). (B) Schematic of *Nnk-rescue* transgene design, which contains the genomic region 1 kb upstream of the *Nnk* start codon and 1 kb downstream of the stop codon. The region targeted by the RNAi construct (highlighted) was recoded by synonymous mutations in the *Nnk*-rescue transgene to make it resistant to RNAi. Upon ubiquitous knockdown of *Nnk* using Act5c-GAL4 driven RNAi, adult viability is very low, but is significantly restored by the RNAi-resistant *Nnk*-rescue transgene. Horizontal bars represent the mean and error bars represent one standard deviation of replicates of single-pair matings for each cross. Vertical axis represents the observed/expected number of progeny from each cross. Asterisks indicate p<0.01, two-tailed Student's t-test. (C) *piggyBac*-null is a *piggyBac* insertion in the 5' UTR of *Nnk* that contains a fluorescent reporter (DsRed) driven by an eye specific promoter (3xP3) and flanked by stop codons in all three reading frames which terminate translation through *Nnk* (*Schuldiner et al., 2008*). Excision of this *piggyBac* insertion restores viability. (D) Allelic combinations of CRISPR/*piggyBac*-nulls/genetic deficiencies were tested, in some cases in the presence of the *Nnk*-rescue transgene, to investigate and confirm *Nnk* essentiality (*Figure 3—source data 1*).

The online version of this article includes the following source data for figure 3:

Source data 1. Various *Nnk* disruption alleles can be all rescued significantly with a *Nnk-mel* rescue transgene.

complementation with an intact *Nnk-mel* rescue transgene. This *Nnk-mel* rescue transgene was flanked by regulatory sequences from the endogenous locus (1 kb segments upstream and downstream of endogenous *Nnk*) and recoded via mutations in synonymous sites to make it resistant to RNAi knockdown (*Figure 3B*; *Supplementary file 4*). Endogenous levels of *Nnk* expression are too low for us to validate expression of the transgene. However, our ability to rescue *Nnk* knockdown-mediated inviability strongly imply that the recoded transgene is appropriately expressed (*Figure 3B*).

To rule out any indirect effects arising from RNAi knockdown, we also examined two previously generated mutant alleles of *Nnk*. The first is a *piggyBac* insertion two bp upstream of the start codon in the 5' UTR of *Nnk* (*Schuldiner et al., 2008*; *Figure 3C*). This *piggyBac* insertion, which is marked by a DsRed reporter driven by an eye-specific promoter, is flanked by stop codons in all three reading frames, which prevents translation downstream of the insertion. We refer to this insertion as a *piggyBac*-null allele of *Nnk*. We found that this insertion allele is homozygous lethal but viability can be fully rescued upon mobilization of the *piggyBac* element, which repairs the intact 5' UTR in a 'scarless' fashion to restore *Nnk* function (schematized in *Figure 3C*). Furthermore, *piggyBac*-null flies can also be rescued by the *Nnk*-rescue transgene (*Figure 3D*; *Figure 3—source data 1*). Second, we re-examined the previously-generated *CRISPR*-null *Nnk* allele (*Kondo et al., 2017*; *Figure 3A*). Contrary to previous results, we found that this allele is lethal when paired with a *Nnk*-spanning deficiency (BL9207, *Figure 3D*) and also fails to complement the *piggyBac*-null allele (*Figure 3D*; *Figure 3—source data 1*; *Supplementary file 4*). Moreover, the *Nnk*-rescue transgene can restore the viability of the *CRISPR*-null *Nnk* allele (*Figure 3D*; *Figure 3—source data 1*). Based on all these results, we conclude that *Nnk* is unambiguously an essential gene in *D. melanogaster*.

Nicknack is required for larval development in *D. melanogaster*

Next, we investigated the developmental stage at which *Nnk-null* progeny die. For this, we crossed *Nnk-null* heterozygotes to each other (*Figure 4A*). These flies contain the *piggyBac* insertion upstream of the *Nnk* gene (*Figure 3C*) on one chromosome, along with a balancer chromosome (TM3G) marked with *GFP* and carrying a wildtype *Nnk* allele. Progeny homozygous for the balancer chromosome TM3G die as early embryos. Consequently, all larvae lacking *GFP* expression are *Nnk-null* homozygotes whereas those that are heterozygous express the *GFP* encoded on the balancer. We conducted egg-lay experiments from crosses between heterozygote *Nnk-null* flies and tracked the developmental progression of *GFP*-negative, *Nnk-null* progeny relative to their *GFP*-expressing heterozygote siblings. We found that *Nnk-null* progeny progress through embryogenesis at the same rate as their heterozygote siblings and are morphologically indistinguishable from heterozygotes until the L1 larval stage. The *Nnk-null* larvae are able to move toward and consume yeast paste much like their heterozygote siblings. However, when heterozygote siblings molt into the L2 stage 48 hr after egg laying (AEL), *Nnk-null* larvae do not molt (*Figure 4B*). Instead, 48 hr AEL, *Nnk-null* larvae progressively become unable to move or eat, eventually dying by 60 hr AEL.

The *Nnk-null* larval arrest phenotype is reminiscent of previous findings with the *ZAD-ZNF* genes *Séance*, *Ouija Board* and *Molting Defective*, which encode proteins necessary for expression of the *Spookier* and *Neverland* genes required for ecdysone biosynthesis (*Uryu et al., 2018*; *Neubueser et al., 2005*; *Komura-Kawa et al., 2015*). Defects in any of the *Séance*, *Ouija Board* and *Molting Defective ZAD-ZNF* genes leads to arrest of larval development and death (*Uryu et al., 2018*; *Neubueser et al., 2005*; *Komura-Kawa et al., 2015*). However, this lethality can be rescued by supplementing the diet with ecdysone or by overexpression of ecdysone biosynthetic enzymes, bypassing the requirement for these three *ZAD-ZNF* genes. We therefore, tested whether dietary supplementation with an ecdysone precursor, cholesterol (7DH, or 7-Dehydrocholesterol) or ecdysone (20E, or 20-Hydroxyecdysone) could bypass or significantly delay the death of *Nnk-null* larvae (*Figure 4C*). We found that it could not. In contrast, the same dietary supplementation is able to significantly restore the viability of *Npc1a* larvae, which lack an essential transporter of cholesterol (*Figure 4C*). Based on these results, we conclude that *Nnk* plays an essential role in larval progression in *D. melanogaster* that is distinct from known steps of the ecdysone biosynthesis pathway.

Transcriptional consequences of *Nicknack* knockout in *D. melanogaster* larvae

To further investigate our developmental findings, we also performed RNA-seq analyses, comparing the transcriptomes of *Nnk-null* larvae to their age-matched heterozygote siblings. We compared the two genotypes at each of two important time-points. First, we compared transcriptomes 24 hr AEL, when the two genotypes are morphologically indistinguishable L1 larvae. At this timepoint, we find that 249 genes (2.4% of all expressed genes) are differentially expressed between *Nnk-null* larvae and control larvae, with 116 genes at least two-fold upregulated and 133 genes two-fold downregulated in *Nnk-null* larvae (*Figure 4D*). Among genes downregulated in *Nnk-null* L1 larvae, we find a
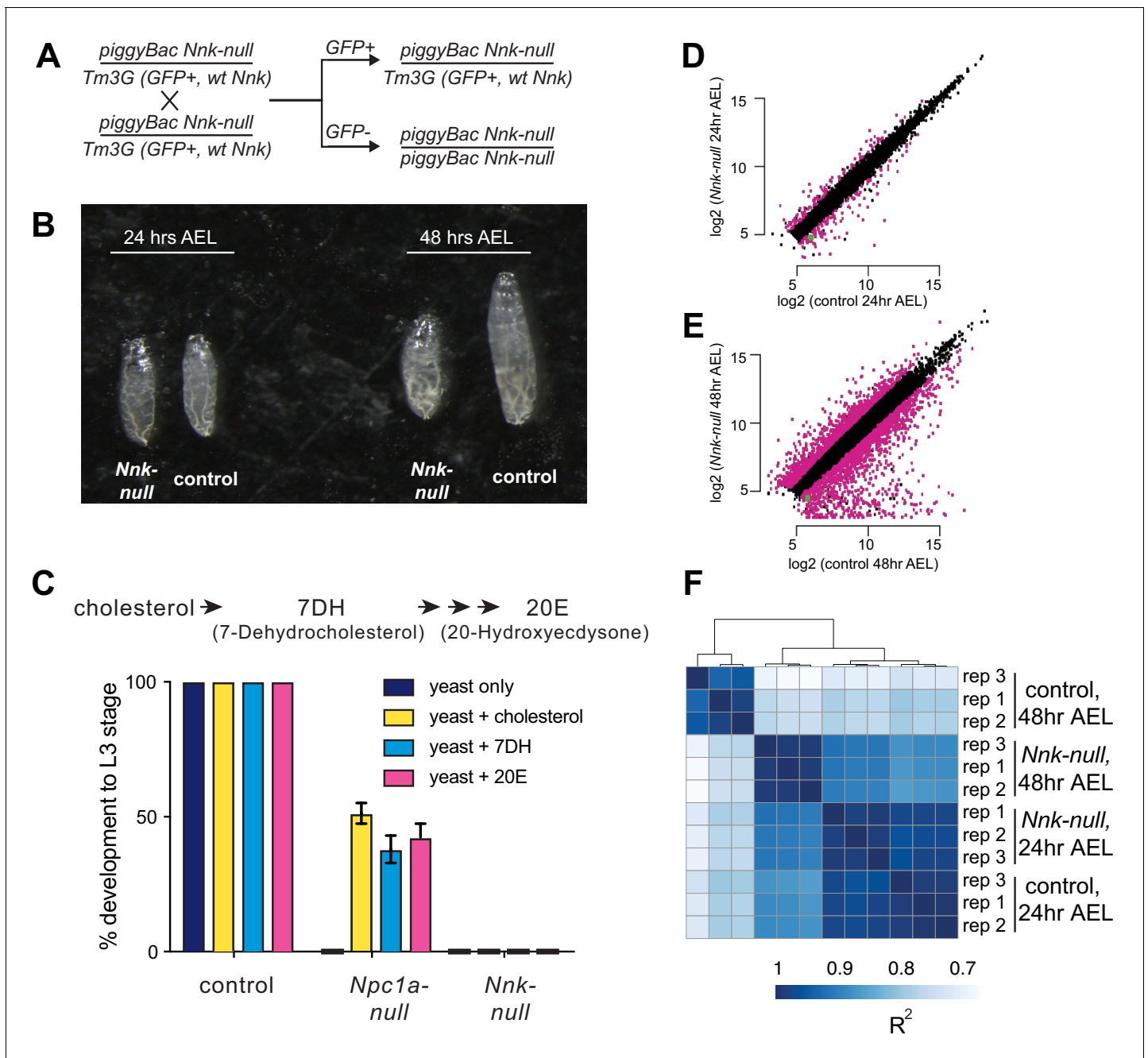
Figure 4. *Nicknack-null* larvae arrest in early larval development. (**A**) Schematic of cross to generate homozygous *piggyBac*-null larvae and age-matched control (heterozygote) siblings. *Tm3G* homozygous progeny are not shown, as they die as early embryos. (**B**) 24 hr after egg laying (AEL), *Nnk-null* larvae are morphologically indistinguishable from control larvae. 48 hr AEL, *Nnk-null* larvae are significantly smaller than the age-matched controls and have failed to undergo the first larval molt. (**C**) Adding sterols (cholesterol, 7DH or 20E) to the food of control (w1118) larvae did not alter their ability to develop to L3 larvae within 60 hr AEL. *Npc1a*-null larvae (*Npc1a57/Npc1a57*) used as a control do not develop on food supplemented with yeast alone but can be partially rescued (>40% molting into L3 larvae) with the addition of cholesterol, 7DH, or 20E. In contrast, *Nnk-null* mutants fail to progress through development even with the addition of dietary sterols. Vertical axis shows the percentage of L1 larvae that progressed to L3 larvae. Graphs show mean of three replicates with error bars showing standard deviation; n > 100 for all genotypes and treatments. (**D, E**) Scatter plots showing RNA-seq results for all expressed genes at 24 hr (**D**) and 48 hr (**E**) AEL. X-axis indicates the expression (normalized abundance) of genes in control larvae whereas the Y-axis shows the expression of genes in *Nnk-null* larvae at the same time point. Magenta dots represent genes significantly over- and under-expressed in *Nnk-null* larvae (green dot represents expression of *Nnk* itself). (**F**) Overall correlations between transcriptome profiles of three replicates each for control and *Nnk-null* larvae reveals that *Nnk-null* transcriptomes 48 hr AEL are more similar to transcriptomes of control and

Figure 4 continued on next page

significant over-representation of functional categories related to proteolysis and sterol transport (both important functions during larval development), as well as dopamine monooxygenase activity (*Supplementary file 3*). In contrast, we find that genes related to lysosome, cytochrome P450s (also important for larval molts), and eye-related functions are upregulated upon *Nnk* loss.

Second, we performed comparisons 48 hr AEL when the *Nnk-null* larvae are significantly smaller and appear developmentally arrested compared to their age-matched controls. At this timepoint, we find that 3027 (28.1% of all expressed genes) genes are differentially expressed, with 1301 genes at least two-fold upregulated and 1726 genes two-fold downregulated in the *Nnk-null* mutants compared to the age-matched controls (*Figure 4E*). Thus, there are significantly more genes affected by *Nnk* loss by 48 hr AEL. Intriguingly, clustering samples by the transcriptional profile of all genes shows that *Nnk-null* larvae at 48 hr AEL are transcriptionally more similar to control larvae at 24 hr AEL of either genotype than they are to age-matched control larvae (*Figure 4F*). The transcriptional status of *Nnk-null* larvae therefore mirrors the phenotypes we observe, displaying a severe developmentally arrested phenotype and transcriptional profile at 48 hr AEL (*Figure 4B*).

Nicknack encodes a heterochromatin-localizing protein in *D. melanogaster*

Most of the *Odj-Nnk* cluster of *ZAD-ZNF* genes are functionally uncharacterized. Since *Odj* encodes a protein that is highly enriched in pericentric heterochromatin (*Swenson et al., 2016*), we speculated that its close paralog, *Nnk,* might also encode a protein with heterochromatic localization in *D. melanogaster* cells. To test this possibility, we used transient transfections to introduce epitope-tagged *Odj* and *Nnk* genes into *D. melanogaster* Schneider 2 (S2) cells and induced their expression with a heat-shock promoter (*Figure 5A*). Upon induction, we confirm that Oddjob has a broad localization pattern within heterochromatin (marked by histone H3 lysine nine methylation, H3K9me3 and outlined with a dashed line, Figure 5A). We found that the *Nnk*-encoded protein also localizes to heterochromatin, but its localization is restricted to discrete foci, unlike Oddjob (*Figure 5A*). Since heat-shock induction can alter chromatin properties in cells, we also employed a complementary transient transfection strategy in which we expressed mCherry epitope-tagged *Odj* and *Nnk* genes under the control of a constitutive pCopia promoter in S2 cells. These analyses also revealed a broad heterochromatic localization of Odj contrasting with discrete foci within heterochromatin for Nnk protein (*Figure 5B*). These foci do not overlap with centromeres (identified by the centromeric histone Cid) or dual-strand piRNA clusters (marked by the piRNA-binding HP1 protein Rhino) (*Figure 5—figure supplement 1*).

Odj has a broad heterochromatic localization in *D. melanogaster* cells, which could result from direct interaction with HP1a (*Swenson et al., 2016*). Indeed, Odj has two potential PxVxL motifs, which are putative interaction sites for HP1a (*Smothers and Henikoff, 2000*), in the linker and ZNF domains (*Figure 5—figure supplement 2A*). Investigating Odj orthologs in other *Drosophila* species revealed that the PxVxL motif in the linker domain is well-conserved, but the one in the ZNF domain is not. We found that mutation of the putative HP1a-interaction site in the linker domain (V164A) converted Odj localization from a broad to a discrete pattern that at least partially overlapped with Nnk (*Figure 5—figure supplement 2B*). In contrast, mutation of the second putative PxVxL site (V321A) did not significantly affect Odj localization. Our results suggest that the putative PxVxL motif in the Odj linker region is a major contributor to Oddjob's broad localization to heterochromatin potentially by mediating a direct interaction with HP1a. In the absence of this interaction, Oddjob and Nnk (which lacks a canonical PxVxL motif) localize similarly to discrete foci within heterochromatin (*Figure 5B*). Our findings suggest that altered protein–protein interactions or DNA-binding specificity via the linker domain may provide a means for functional diversification between closely-related *ZAD-ZNF* paralogs.

To gain deeper insight into the heterochromatic localization of Nnk and Odj, we performed chromatin immunoprecipitation and sequencing (ChIP-seq) using transient transfection of S2 cells with
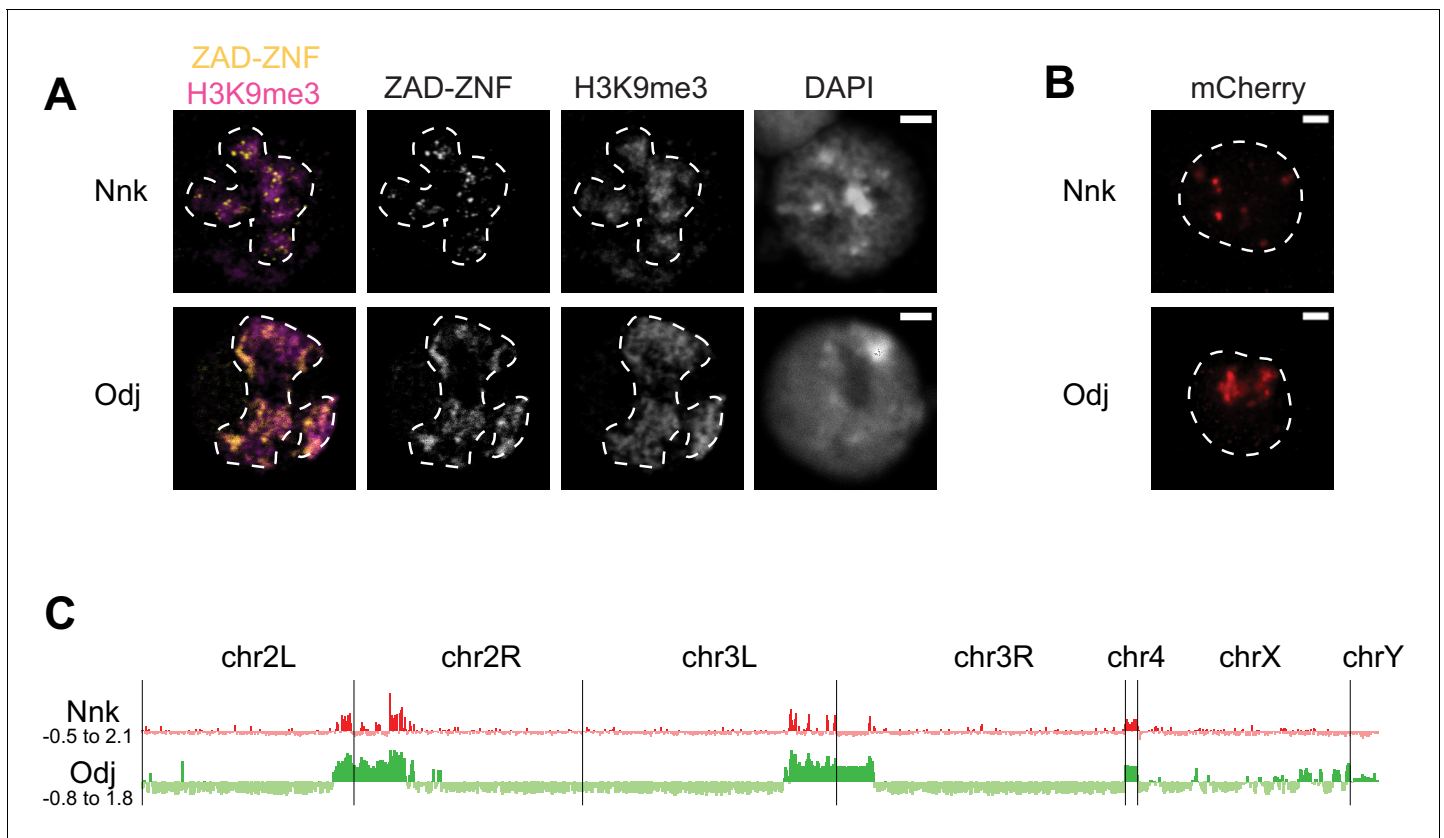
Figure 5. Heterochromatic localization of Odj and Nnk proteins. (**A**) Odj and Nnk proteins localize within the heterochromatic chromocenter, marked by H3K9me3 staining (magenta, outlined with dashed line) of *D. melanogaster* S2 cells. Whereas Oddjob localizes broadly to heterochromatin, Nicknack localizes to discrete foci within heterochromatin. All images are of representative nuclei from S2 cells transfected with a Venus-tagged ZAD-ZNF (yellow) under the control of an inducible heatshock-promoter. DAPI marks DNA in each nucleus. Scale bar = 2 μm. (**B**) Constitutive pCopia-driven expression confirms the broad Odj and discrete Nnk localization within heterochromatin of *D. melanogaster* S2 cells (dashed line outlines nucleus). Scale bar = 1 μm. (**C**) ChIP-seq analyses confirms that Odj binding is highly enriched throughout heterochromatin, whereas Nnk localizes only to some regions within heterochromatin, and to the 4th chromosome; note that S2 cells do not have an intact Y chromosome (*Lee et al., 2014*). The y-axis represents normalized ChIP-versus-input log2 ratios, using IGV to visualize and smooth data. Numbers on the y-axis give the range of ratios displayed for each factor.

The online version of this article includes the following figure supplement(s) for figure 5:

Figure supplement 1. Nnk localization in heterochromatin in *D. melanogaster* S2 cells.

Figure supplement 2. Mutation of the Odj PxVxL motif in the linker disrupts its broad localization within heterochromatin.

Figure supplement 3. Additional analysis of ChIP-seq signals.

Figure supplement 4. Effect of *Nnk* loss on gene and repeat expression.

the epitope-tagged pCopia constructs for *Odj* and *Nnk*. Consistent with its cytological localization (*Figure 5A and B*), we found that Odj was highly enriched throughout the large chromosome regions previously defined as pericentric heterochromatin (*Hoskins et al., 2015*) whereas Nnk-bound regions are more narrowly distributed within heterochromatin (*Figure 5C*). Finer-scale analyses suggested that Nnk and Odj signals are enriched close to transcription start sites classified by modENCODE as residing in TSS-proximal chromatin in S2 cells, both in heterochromatin and euchromatin (*Figure 5—figure supplement 3*; *Ho et al., 2014*). However, many of these regions overlap with false positive 'phantom peaks' previously identified in ChIP-seq experiments in S2 cells (*Jain et al., 2015*; *Figure 5—figure supplement 3*). While some heterochromatic TSSs could still represent true binding, we are more confident that the broader occupancy we observe outside TSSs represents Odj and Nnk's true localization. Moreover, we did not observe a significant effect of *Nnk* loss on heterochromatin-embedded gene expression at the larval L1 stage via RNA-seq analyses (*Figure 4D*, *Figure 5—figure supplement 4*). Instead, we found that some heterochromatin-

embedded repetitive elements are derepressed upon *Nnk* loss (*Figure 5—figure supplement 4*). Based on these results, we hypothesize that rather than acting as a transcription regulator for specific heterochromatin-embedded genes like its paralogs *Séance, Ouija Board* and *Molting Defective* (*Uryu et al., 2018*), *Nnk* may instead repress expression from heterochromatic repeats.

D. simulans Nnk can rescue female, but not male viability of *Nnk-null D. melanogaster*

Our analyses revealed *Nnk* is an evolutionarily young, differentially-retained gene that is essential for larval development in *D. melanogaster*. *Nnk* has also evolved under dramatic positive selection in just the 2.5 million-year divergence between *D. melanogaster* and *D. simulans*, having undergone 52 fixed non-synonymous differences in the 439 aa protein-coding region (*Table 2*). Given the intriguing correlation between essentiality and positive selection that we had observed in the *ZAD-ZNF* genes, we investigated whether positive selection of *Nnk* has affected its function.

We first assayed the subcellular localization of epitope-tagged *Nnk* orthologs from *D. melanogaster* and *D. simulans* in *D. melanogaster* S2 cells. We used transient transfections to introduce these genes into S2 cells and used heat shock to drive their expression. We found that both proteins similarly localize to foci within heterochromatin (*Figure 6A–B*; dashed line marks heterochromatin boundaries). Thus, the rapid evolution of *Nnk* during *D. melanogaster- D. simulans* divergence has not dramatically affected its gross subcellular localization.

Next, we examined the consequences of *Nnk* positive selection on viability in *D. melanogaster*. We created a *D. simulans* Nnk 'rescue' transgene (*Figure 6C*). This rescue transgene is similar to the *D. melanogaster Nnk*-rescue construct (*Figure 3B*), except that it contains the *D. simulans Nnk* coding sequence (codon-optimized to *D. melanogaster*) with 1 kb *Nnk*-flanking sequences from *D. melanogaster* (*Supplementary file 4*). We introduced this *D. simulans Nnk*-rescue transgene into the same *attP* site on *D. melanogaster* X chromosome as the *D. melanogaster Nnk*-rescue transgene via PhiC31-mediated transgenesis (*Figure 6D*). The *attP*-insertion rescue design allowed us to put the transgene in the same genetic location in the *Nnk-mel* and *Nnk-sim* rescue crosses, normalizing for variability in expression of transgenes. This allowed a near-isogenic comparison of the *D. simulans* and *D. melanogaster Nnk* transgenes' ability to rescue inviability of *Nnk-null D. melanogaster* flies, despite their high level of sequence divergence. Unfortunately, low levels of endogenous *Nnk* expression did not allow us to assess whether the expression levels of both *Nnk* transgenes were equivalent.

We crossed heterozygous females either containing one or no copy of the (RNAi-resistant) *Nnk-rescue* transgene and the *Nnk-RNAi* allele to *Act5C-GAL4/CyO-GFP* males. We expected to find that progeny that did not inherit the rescue transgene would be inviable. In contrast, in the resulting progeny from rescue transgene-bearing females, we expect half of the progeny to inherit the *Nnk-rescue* transgene (see Materials and methods). We found that the *Nnk-mel* transgene significantly rescued *Nnk* knockdown compared to no-transgene controls; however, males were recovered at slightly lower levels than females (67 males: 101 females compared to expectation of 1:1 ratio; p=0.08, Fisher's exact test). We similarly found that the *D. simulans Nnk* transgene can significantly rescue the lethality caused by knockdown of endogenous *D. melanogaster Nnk* in females (*Figure 6C*), although at a slightly lower level than *Nnk-mel* rescue. In contrast, rescue of male viability by *D. simulans Nnk* is extremely poor (2 males: 33 females compared to expectation of 1:1 ratio; p<0.0001, Fisher's exact test) resulting in a severe sex-bias. Thus, *Nnk-sim* is much worse than *Nnk-mel* in rescuing male viability (67:101 versus 2:33, p<0.0001, Fisher's exact test). Based on these findings, we infer that the *D. simulans Nnk-rescue* transgene specifically fails to rescue *Nnk*-knockout males, in the presence of the heterochromatin-rich Y chromosome. Our findings suggest that not only is *Nnk* a positively-selected, essential *ZAD-ZNF* gene, but also that its positive selection is required for optimal function in the *D. melanogaster* genome.

Discussion

In this study, we explored the relationship between genetic innovation and essentiality in the *ZAD-ZNF* gene family, which encodes the most abundant class of transcription factors in insects. Due to their lineage-specific amplification, protein structure and expression patterns, *ZAD-ZNF* genes were previously hypothesized to be analogous to the KZNF (KRAB-Zinc Finger) transcription factor-
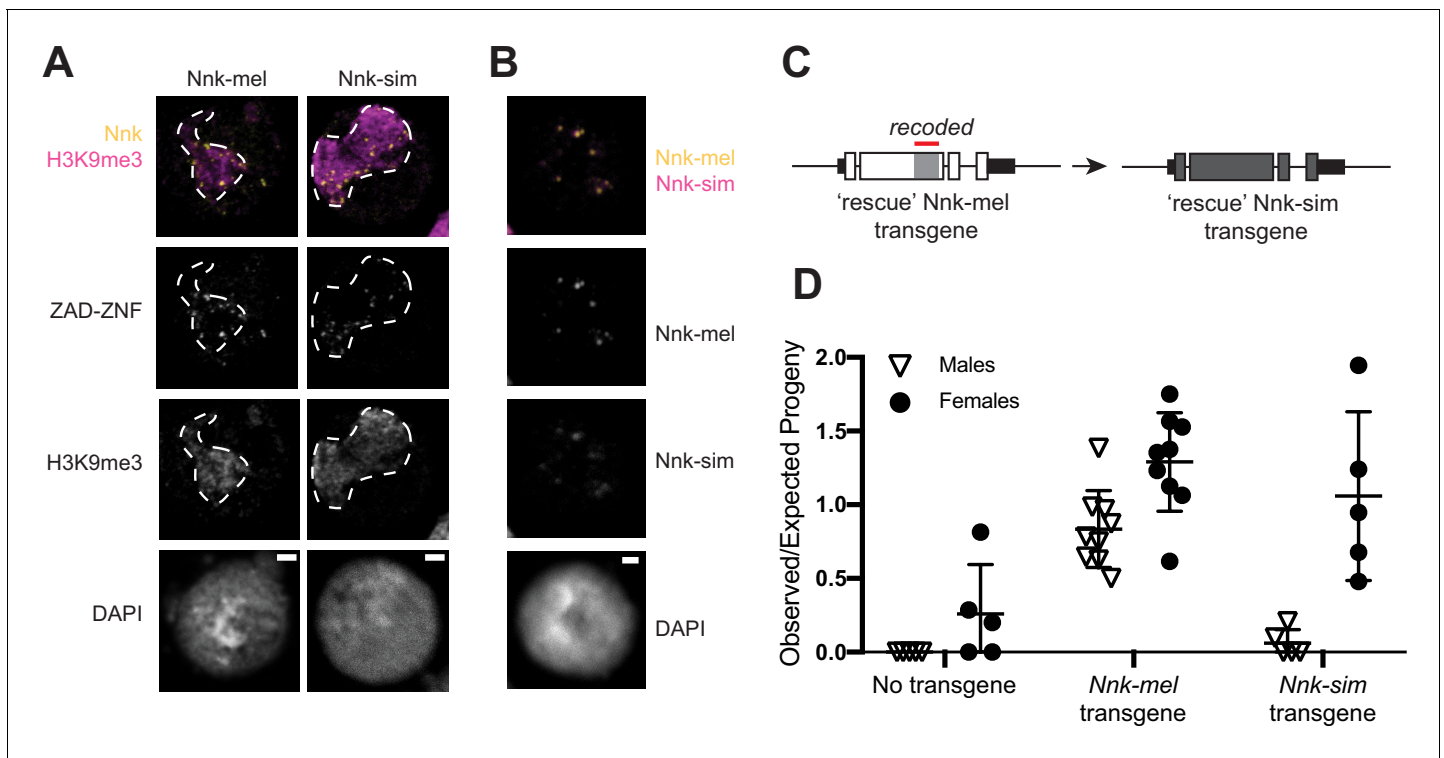
Figure 6. *D. simulans* Nicknack localizes to *D. melanogaster* heterochromatin but can only partially rescue essential function. (**A**) Venus-tagged *D. melanogaster* Nnk (Nnk-mel) or Venus-tagged *D. simulans* Nnk (Nnk-sim) (yellow) localize to foci within the heterochromatic chromocenter marked by H3K9me3 (magenta, dashed line) in *D. melanogaster* S2 cells. Each image is of a representative S2 cell nucleus with DAPI marking DNA. Scale bar = 2 μm. (**B**) Transient co-transfection of Venus-tagged Nnk-sim (magenta) and FLAG-tagged Nnk-mel (yellow) shows that they overlap in their localization to discrete heterochromatic foci in *D. melanogaster* S2 cells. Scale bar = 2 μm. (**C**) Schematic of rescue *Nnk-sim* transgene, which contains the genomic region 1 kb upstream of the *Nnk* start codon and 1 kb downstream of the stop codon, but with the *D. simulans* coding region (dark gray) from start to stop codon, including introns. The region of *Nnk-mel* targeted by the RNAi hairpin is highlighted by the red line; because of its divergence and codon-optimization, the *Nnk-sim* rescue transgene is resistant to RNAi. (**D**) Ability of *Nnk-mel* and *Nnk-sim* transgenes to rescue ubiquitous knockdown of *Nnk* using Act5C-GAL4 drivers. The vertical axis represents the observed/expected progeny in each cross. Without *Nnk* transgene, there are significantly fewer *Nnk* knockdown male and female progeny recovered than expected. *Nnk-mel* transgene is able to rescue *Nnk*-RNAi depletion in both males and females. *Nnk-sim* transgene is able to rescue *Nnk*-RNAi knockdown females significantly better than knockdown males. Filled circles represent number of female adult progeny and open triangles represent number of male adult progeny recovered from single-pair matings. Horizontal bars show the mean and standard deviation.

encoding gene family found in vertebrates (*Chung et al., 2002*; *Chung et al., 2007*), many of which target transposable element sequences (TEs) inserted in the genome. Just like *KZNF* genes in mammals, we find strong evidence for evolutionary dynamism in *Drosophila ZAD-ZNF* genes. For example, we find that only 61 of 91 *ZAD-ZNFs* found in *D. melanogaster* are universally retained in most *Drosophila* species. Furthermore, 12 *ZAD-ZNFs* have evolved under positive selection during *D. melanogaster*- *D. simulans* divergence.

Despite these similarities, however, there are considerable differences between these gene families. First, a direct connection to TEs has only been revealed for one *ZAD-ZNF* gene in *Drosophila*; *CG17801* has a role in regulating *HetA* and *Blood* transposable elements in the female germline (*Czech et al., 2013*). Second, we find that *ZAD-ZNF* genes that evolve under positive selection are more likely to encode essential functions in embryonic axial patterning, larval development, and meiosis (*Chen et al., 2000*; *Harms et al., 2000*; *Page and Orr-Weaver, 1996*; *Uryu et al., 2018*; *Lake et al., 2011*; *Chen et al., 2006*; *Table 2*). In contrast, most *KZNF* genes that have been shown to be essential for sterility or viability in mammals are slowly evolving (*Wolf et al., 2020*; *Imbeault et al., 2017*). Finally, unlike *KZNFs* (*Thomas and Schneider, 2011*), positive selection in *ZAD-ZNFs* is not primarily focused on their DNA-binding C2H2 domains but rather on the poorly-characterized linker domains that connect the ZAD and C2H2 domains (*Table 2*). We speculate that

the linker, which is often comprised of intrinsically disordered domains, may play an important role in chromatin localization of ZAD-ZNF proteins either via direct DNA-binding or via protein–protein interactions (*Brodsky et al., 2020*; *Erijman et al., 2020*). For example, the Oddjob linker domain encodes a PxVxL motif that is important for its broad heterochromatic localization, potentially by mediating direct interaction with HP1a (*Figure 5*, *Figure 5—figure supplement 2*).

Based on these dissimilarities, we do not favor the possibility that ZAD-ZNF innovation is solely driven by arms-races with TEs. Instead, we favor the hypothesis that the recurrent adaptation of a subset of *ZAD-ZNF* stems from their roles in pericentromeric heterochromatin organization or regulation of gene expression. Indeed, many genes that encode crucial heterochromatin functions are often critical for viability or fertility, yet are quite variable even among closely-related species (*Ross et al., 2013*; *Klattenhoff et al., 2009*; *Levine et al., 2016*; *Parhad et al., 2017*; *Vermaak and Malik, 2009*). Heterochromatin is a gene-poor component of most eukaryotic genomes. Yet, its establishment and maintenance is nevertheless essential for many cellular processes including chromosome condensation and segregation, repression of TEs, and genome stability (*Abe et al., 2016*; *Grézy et al., 2016*; *Levine et al., 2015*; *Nambiar and Smith, 2018*; *Okita et al., 2019*; *Vernì and Cenci, 2015*; *Liu et al., 2014*; *Azzaz et al., 2014*; *Inoue et al., 2008*; *Ruiz-Estévez et al., 2014*; *Verschure et al., 2005*; *Brennecke et al., 2007*; *Senti and Brennecke, 2010*; *Goriaux et al., 2014a*; *Goriaux et al., 2014b*). Thus, the rapid evolution of genes encoding heterochromatin functions might reflect lineage-specific mechanisms to package heterochromatic DNA, or silence TEs. In this context, it is important to note the heterochromatic satellite DNA sequences themselves are among the most rapidly evolving component of *Drosophila* genomes (*Wei et al., 2018*; *Chakraborty et al., 2020*; *Jagannathan et al., 2017*).

Several characterized *ZAD-ZNFs* have been found to play key roles at heterochromatin. For example, the Oddjob protein co-immunoprecipitates with HP1a and broadly localizes to heterochromatin (*Figure 5*; *Swenson et al., 2016*). Similarly, Séance and Ouija Board control the expression of heterochromatin-embedded genes necessary for larval development (*Uryu et al., 2018*). Based on these observations, we propose that *ZAD-ZNF* diversification (marked by gene turnover and positive selection) is driven by the high turnover of sequences embedded within heterochromatin. Although the bulk of heterochromatin is made up of highly repetitive elements such as satellite DNAs and TEs, heterochromatin also harbors many genes that are deeply embedded within heterochromatin (*Yasuhara et al., 2005*; *Yasuhara and Wakimoto, 2006*; *Eberl et al., 1993*; *Schulze et al., 2006*; *Schulze et al., 2005*; *Devlin et al., 1990*). These genes, many of which encode essential functions (*Sinclair et al., 2000*), require a heterochromatic environment to ensure their correct expression and regulation (*Wakimoto and Hearn, 1990*). We posit that the constant turnover of flanking and embedded sequence elements such as TEs and satellite DNAs may require constant adaptation of transcription factors required for the proper expression of genes embedded in heterochromatin. *ZAD-ZNF* adaptation might also be necessary to protect against the inappropriate expression of heterochromatin-embedded elements especially at crucial developmental stages. Because of the high rate of evolutionary turnover of heterochromatic sequences, we hypothesize that *ZAD-ZNF* genes that are essential in one species could nevertheless be lost in another, either because the target loci of the *ZAD-ZNF* genes is lost, or because another *ZAD-ZNF* paralog has acquired this essential regulation function.

By testing the causal link between positive selection and essential function, we also find further support for *Nnk*'s essential function being related to heterochromatin biology. We find that *D. simulans Nnk* can significantly rescue *Nnk-null* inviability in females, but not in males. We speculate that this failure to rescue male viability could be due to the heterochromatin-rich Y chromosome in males. For example, if the *D. melanogaster* Nnk protein, but not the *D. simulans* Nnk protein, could appropriately repress the *D. melanogaster* Y chromosome, this might explain the sex-bias seen in the *D. simulans Nnk*-rescue progeny. The Y chromosome is itself not essential for viability or sex determination in *Drosophila*. However, de-repression of the heterochromatin-rich Y chromosome could nevertheless lead to detrimental consequences on larval development. This effect could be direct, leading to inappropriate expression of Y-chromosome-embedded genetic elements that block larval development. Alternatively, this effect could be indirect; de-repression of the Y chromosome could indirectly impact several other chromatin processes genome-wide e.g., due to inappropriate titration of transcription or heterochromatin factors, which could exacerbate an already hypomorphic function of the *D. simulans Nnk* allele (*Branco et al., 1869*; *Francisco and Lemos, 2014*; *Wang et al., 2018*;

Piergentili, 2010; *Brown et al., 2020*). Finally, different functional optimality of male-specific or female-specific *Nnk* functions could also drive its rapid evolution (*VanKuren and Long, 2018*). In any of these scenarios, we hypothesize that rapid co-evolution with heterochromatic sequences might have driven the rapid evolution of Nnk and possibly other heterochromatin-interacting ZAD-ZNF proteins.

Based on our findings, we hypothesize that constant adaptation in *ZAD-ZNF* genes is driven by rapid alterations in heterochromatin across *Drosophila* and other insect species. This co-evolutionary arms-race may provide the explanation for the unexpected correlation we find between gene essentiality and innovation in the largest family of transcription factors in insect genomes.

Materials and methods

Phylogenomic analysis of *ZAD-ZNF* genes in *Drosophila*

We used the Flybase database (http://flybase.org) to identify all ZAD-containing proteins (Pfam motif PF07776) in 12 sequenced and annotated *Drosophila* species (*Drosophila 12 Genomes Consortium, 2007*). Using NCBI's Conserved Domains search, we identified other domains found in the ZAD-containing proteins (*Marchler-Bauer et al., 2017*). To estimate the evolutionary age of these ZAD-containing genes, we used OrthoDB to identify orthologs of the 91 ZAD-containing proteins across *Drosophila* (*Zdobnov et al., 2017*). We used the divergence between the two most distantly-related species that still encoded and ortholog to calculate the minimum age of each gene using timetree.org (*Kumar et al., 2017*). If orthologs were identified in basally-branching but not later-branching species by OrthoDB, we defined this as a loss of the *ZAD-ZNF* gene in the lineage. The repertoire of different *ZAD-ZNF* genes in different *Drosophila* species inferred from our analysis is summarized in *Supplementary file 1*.

Defining gene essentiality

We used FlyBase gene summaries (FlyBase.org) and published studies when available to define gene essentiality. Our criteria for essentiality were broad: if there was a lethal allele reported in any assay, we counted the gene as essential for viability in *D. melanogaster*. *Supplementary file 1* summarizes the phenotypic data (and its source) that we used to classify the *ZAD-ZNF* genes into either essential or non-essential categories.

Analyses of positive selection

For the McDonald-Kreitman test, we extracted the gene of interest from *D. melanogaster* population genetic datasets available through Popfly (www.popfly.com) and removed low frequency (<0.05) variants from the dataset to minimize the effects of false positives and low-frequency variants that may not have been subject to selection (*Hervas et al., 2017*; *Fay et al., 2002*). We used a manually trimmed alignment of the *D. melanogaster* filtered dataset and the reference *D. simulans* sequence for the McDonald-Kreitman test (http://mkt.uab.es/mkt/; *McDonald and Kreitman, 1991*; *Egea et al., 2008*). We also carried out a polarized McDonald-Kreitman test in which we only analyzed non-synonymous and synonymous sites that were fixed along the lineage leading to *D. melanogaster*, after its divergence from *D. simulans*, inferred using *D. yakuba* as an outgroup species, except for the *Trem* gene where we used *D. erecta* outgroup instead because the *D. yakuba* ortholog aligned poorly to *D. melanogaster* and *D. simulans*.

Viability studies

We used an *Act-GAL4/CyO-GFP* driver for ubiquitous knockdown. The RNAi lines used to specifically target *Oddjob* cluster genes are the VDRC KK or GD lines: *Oddjob* (27971), *CG17803* (38869), *CG17801* (29501), *CG17806* (40106) and *Nicknack* (102311). RNAi controls used for the experiment were *Cid* (43856) and *HP1B* (26097) for the ubiquitous knockdown. Ubiquitous knockdown of *Cid* produced no viable progeny. We crossed five virgin females carrying *Act-GAL4/CyO-GFP* to 3 males of each RNAi line. We allowed the females to lay eggs for 3 days and flipped the flies into fresh vials three times. Each cross was performed in triplicate. Progeny were counted 10–15 days after each cross was set up. A minimum of 20 *CyO-GFP* males and 20 *CyO-GFP* females (*i.e.,* control genotype) were required for us to quantify the crosses. Each cross is represented by a point on the graph and

shown as a ratio of observed (number of knockdown or rescue progeny counted in each cross) over expected (the number of expected knockdown or rescue progeny from Mendelian segregation inferred by inheritance of the balancer chromosome). If the knockdown has no effect on viability or the rescue is 100% effective, the observed/expected = 1. If the knockdown of the essential gene is 100% penetrant or the rescue is ineffective, the observed/expected = 0. Plotting and statistical analyses were conducted using Graphpad Prism 8 software.

Defining orthologs

Since *Oddjob* cluster genes experienced numerous independent segmental duplications, it was not possible to determine orthologs by synteny alone. Instead, we used TBLASTN (*Gertz et al., 2006*) to identify candidate orthologs of *Oddjob* cluster genes, using the genes in *D. melanogaster* as queries. We used a reciprocal blast search strategy to identify potential orthologs and further investigated these candidates by making a maximum likelihood phylogenetic tree (LG substitution model in PhyML with 100 bootstrap replicates) of a manually trimmed protein alignment (constructed using Clustal Omega program [*Thompson et al., 2002*] in the Geneious package, www.geneious.com) (*Supplementary file 2*, *Figure 2—figure supplement 1*). We assigned orthologs based on genes that formed a monophyletic clade with the each of the *D. melanogaster Oddjob* cluster genes. We mapped these *Oddjob* cluster orthologs back to each genome assembly to determine the composition of the *Oddjob* locus across *Drosophila* species. In cases where there were other ZAD-ZNFs present, we blasted them against the *D. melanogaster* genome to examine if there were any orthologs present in *D. melanogaster.* If the top hit was not a member of the *Oddjob* locus, we did not include it in the tree. In the case where there are partial ZAD-ZNFs (containing just the ZAD domain or the zinc-finger domains), we performed a blast search against the *D. melanogaster* genome. If the top hit was a member of the *Oddjob* cluster, we assigned orthology by making a phylogenetic tree with all other *Oddjob* cluster orthologs. Our phylogenomic inferences are dependent on the accuracy and state of completion of genomes from other *Drosophila* species, which can vary.

Nnk-rescue transgene design, construction, and crosses

We designed a *D. melanogaster* recoded transgene comprising a 3.3 kb fragment containing the genomic region of *Nicknack* plus 1 kb upstream and downstream of the start and stop codons, based on the *D. melanogaster* reference assembly. We recoded the sequence targeted by the VDRC RNAi KK line (103211) by making synonymous changes at each codon. The synonymous changes made the gene resistant to RNAi knockdown but did not alter the amino-acid sequence of Nnk. The recoded region spans 270 base pairs that corresponds to the region targeted by the VDRC RNAi KK line (103211). The resulting sequence was synthesized by GENEWIZ Co. Ltd. (Suzhou, China) and cloned into a plasmid we generated that contains 3xP3-DsRed attP, which produces fluorescent red eyes in the adult to mark the presence of the transgene. To generate the *D. simulans* transgenic allele, we codon-optimized the *D. simulans Nnk* coding sequence for the *D. melanogaster* genome using IDT's codon-optimization tool. The resulting sequence was synthesized by GENEWIZ Co. Ltd. (Suzhou, China) and swapped for the *D. melanogaster Nnk* coding sequence in the plasmid described above, using the NEBuilder kit (New England Biolabs). We submitted transgenic constructs to The BestGene Inc (Chino Hills, CA) for injection into the X-chromosome *attP18* line (BL 32107) using *PhiC31* site-specific integration (*Groth et al., 2004*).

For the transgene rescue cross, we crossed five virgin female flies bearing one copy of *Nnk-rescue* transgene (on the X chromosome) and the RNAi allele (on the second chromosome) to 3 *Act5C-GAL4/CyO-GFP* males. We allowed the females to lay eggs for 3 days and flipped the cross three times. We set up each cross in triplicate and progeny were counted after 10–15 days. The description of the observed/expected calculation can be found in 'viability studies' section. Each cross had at least three replicates. All flies were raised at 25˚C.

Characterization of *Nnk-null* mutants

We placed 50–75 flies heterozygous for the *Nnk-null* allele (*Nnk pBac null/TM3G)* or for the *Nnk* CRISPR allele (*Nnk CRISPR/TM3G)* into a small embryo collection cage containing a grape-juice plate with a thin strip of yeast paste and collected embryos for 3 hr at 25˚C. We transferred the larvae to fresh grape-juice plates containing yeast paste daily and scored developmental stage by mouth

hook morphology. We used fluorescence to distinguish between heterozygotes (GFP-positive larvae) and homozygotes (GFP-negative larvae). For the trans-heterozygote evaluation, we crossed 30–40 virgin female *Nnk CRISPR/TM3G* to 10 *Nnk pBac null/TM3G* males. Crosses were done in triplicate and at least 100 progeny were counted per cross.

Larval collection for RNA-sequencing

We placed 50–75 flies heterozygous for the *Nnk-null* allele (*Nnk pBac/TM3G)* into a small embryo collection cage containing a grape-juice plate with a thin strip of yeast paste and collected embryos for 3 hr at 25˚C. The first time point was collected 24 hr after egg laying (AEL) and the second 48 AEL. We transferred the larvae to fresh grape-juice plates containing yeast paste daily.

RNA extractions and library preparation

Whole larvae (~30 animals at 24 hr AEL and ~20 animals at 48 hr AEL for each sample; RNA from each time point and genotype was prepared in triplicate) were ground in a 1.5 mL Eppendorf tube containing 50 µL of TRIzol reagent using a DNase, RNase and DNA free 1.5 mL pestle. 450 µL of TRIzol reagent was added after grinding. Immediately, we added 500 µL of chloroform and the tube was inverted gently 2–3 times. We removed the aqueous phase into a fresh tube containing 1 mL of 200 proof EtOH and mixed by inversion. The mixture was then bound to a Zymo-spin column according to the manufacturer's instructions (Zymo Research). We followed the DNase extraction and purification protocol outlined in the RNA Clean and Concentrator kit (Zymo Research). We eluted the RNA in 15 µL of DNase/RNase-free water and immediately placed the samples at −80˚C. We checked the quality of the samples with a 2200 Tapestation (Agilent Technologies) and selected samples that had an RNA integrity number >9.0 for library preparation. Library construction and Illumina 150 bp paired-end RNA-sequencing were conducted at Novogene Bioinformatics Technology Co., Ltd (Beijing, China).

Transcriptome data analysis

We used Kallisto (*Bray et al., 2016*) to quantify abundances of the *D. melanogaster* reference transcriptome (refMrna.fa for dm6, obtained from UCSC Genome Browser Oct 16th, 2018, which contains 34,114 transcripts). For each transcript, we acquired the gene name using R (org.Dm.eg.db). Kallisto counts were read into R using the tximport package (*Soneson et al., 2015*) and using the summarizeToGene function, we summarized alternative splice-form counts into a single count per gene. We used DESeq2 to identify differentially expressed genes with adjusted p-value<=0.05 and absolute $\log_2$(fold change)>=1 (*Anders and Huber, 2010*), comparing *Nnk-null* larvae to controls for each timepoint separately. Before performing each comparison, we excluded genes with low expression (<100 counts total across all samples): this filtering yields 10,574 'expressed' genes in 24 hr AEL larvae, and 10,758 'expressed' genes in 48 hr AEL larvae. The RNA-seq data has been deposited in the SRA database with the accession number PRJNA643855.

We performed Gene Ontology (GO) enrichment analysis on each of four gene lists: 116 over-expressed genes and 133 under-expressed genes in 24 hr AEL larvae, and 1301 over-expressed genes and 1726 under-expressed genes in 48 hr AEL larvae. We used the Bioconductor GOstats package to perform conditional hypergeometric enrichment tests for each of three ontologies (biological process, molecular function and cellular component) (*Huber et al., 2015*; *Falcon and Gentleman, 2007*). For the 'universe' of all genes examined (*i.e.,* background) we used the corresponding list of 'expressed' genes at each developmental timepoint. We report only over-represented categories with p<0.001, and use annotations found in the org.Dm.eg.db Bioconductor package.

To estimate expression levels aggregated across all instances of each repeat type, we took an approach similar to that of *Day et al., 2010*. Here, rather than mapping reads to the typical reference genome assembly, we constructed a 'repeat assembly', where we used RepeatMasker annotations for dm6 (obtained from UCSC) to extract and concatenate all instances of each repeat type, adding 75 bp flanking sequences (half the length of each read) and inserting 150 N bases between each instance. The repeat assembly therefore consists of a single 'chromosome' for each repeat type. We then used BWA-mem to map sequences as single reads (not paired-ends) to the repeat assembly (*Li, 2013*), and samtools to count reads mapping to each repeat type ('chromosome'). We combined counts for pairs of simple repeats that represent the reverse-complement of one another.

We used DESeq2 to perform statistical analyses on raw counts, using the total number of sequenced fragments as size factors (*Love et al., 2014*). We normalized counts by dividing each by the number of reads sequenced for that sample in millions.

Dietary sterol supplementation

To evaluate the ability of dietary sterols to rescue *Nnk-null* phenotype, we carried out sterol supplementation as previously outlined (*Uryu et al., 2018*). Briefly, we mixed together 20 mg of dry yeast in 38 µL of water. To this yeast paste, we added either 2 µL of EtOH (negative control) or 2 µL of EtOH plus cholesterol (Sigma), 7-dehydrocholestrol (Sigma) or 20-hydroxyecdysone. Stocks used for these experiments were balanced over GFP-balancers (*Npc1a57/CyO-GFP* and *Nnk-null/TM3G*). The control used for these experiments was the *w1118* stock. Eggs were laid on yeasted grape-juice plates for 3 hr at 25°C. 24 hr AEL, GFP-negative larvae were transferred onto grape plates containing fresh yeast paste at 25°C. For 72 hr AEL, larvae were transferred to fresh grape-juice plates containing yeast paste daily and scored for their developmental stage based on the morphology of their mouth hooks. *Npc1a57/CyO-GFP* stocks were a kind gift from Leo Pallanck (*Fluegel et al., 2006*).

Tissue culture and transfection

Oddjob cluster genes were amplified from genomic DNA from *D. melanogaster* and *D. simulans* and directionally cloned into pENTER/D-TOPO (ThermoFisher) according to the manufacturer's instructions. We verified that clones had the appropriate insertions by sequencing. We used LR clonase II (ThermoFisher) to get each *Oddjob* cluster gene into the *Drosophila* Gateway Vector destination vector to express the gene of interest with an N-terminal Venus tag under the control of the *D. melanogaster Hsp70* promoter (pHVW). For Rhino localization in S2 cells, the *Rhino* gene was cloned into the *Drosophila* Gateway Vector destination vector to enable expression of Rhino with an N-terminal 3xFlag tag under the control of the *D. melanogaster Hsp70* promoter (pHFW). *Nnk* and *Odj* were also cloned into pCopia vectors containing N-terminal and C-terminal mCherry epitope tags, respectively, and HP1a was previously tagged with mCerulean at the N-terminus.

Schneider 2 cells were obtained from the *Drosophila* Genomics Resource Center (Bloomington, IN, USA) and grown at 25°C in M3+BPYE+10%FCS. For the transfections, one million cells were seeded, and one day later 2 micrograms of plasmid DNA was transfected into cells using Xtreme-gene HP transfection reagent (Roche) according to the manufacturer's specifications. For the Rhino localization experiment, cells were co-transfected with 1 µg of GFP-tagged ZAD-ZNF vector and 1 µg of Flag-tagged Rhino vector. Cells were allowed to recover for 24 hr post transfection, heat shocked for 1 hr at 37°C and recovered for 3 hr at 25°C prior to fixation. For the pCopia vectors, transient transfections were conducted on S2 cells using the TransIT-2020 reagent (Mirus), and live imaging was performed 72 hr later, using an Applied Precision Deltavision microscope and analyzed using SoftworX software.

For the heatshock vector transfections, cells were transferred to coverslips for 30–45 min prior to starting the immunohistochemistry protocol. 0.5% sodium citrate hypotonic solution was added to the coverslip for 10 min to swell cells, which was then spun at 1900 rpm for 1 min in a Cytospin III to remove the cytoplasm from the cells. The sodium citrate was immediately removed and cells were subsequently fixed. For fixation, 4% PFA + PBST (PBS + 0.1% Triton) was added to the cells for 10 min. Coverslips were then washed in PBST and blocked for 30 min in PBST + 3% BSA. Cells were incubated with primary antibody overnight at 4°C in a humid chamber. For immunolocalization, the following dilutions of primary antibodies were used: GFP (Abcam AB13970) 1:1000, H3K9me3 (Abcam AB8898) 1:500 M2 FLAG (Sigma-Aldrich F4042) 1:1000. After washing with PBST three times for 10 min per wash, the following fluorescent secondary antibodies were used at 1:1000 dilution: goat anti-chicken (Invitrogen Alexa Fluor 488, A-11039), goat anti-rabbit (Invitrogen Alexa Fluor 568, A-11011) and goat anti-mouse (Invitrogen Alexa Fluor 568, A-11031). The cells were incubated with 1x DAPI in the final wash and mounted in SlowFade Gold Mounting Medium (ThermoFisher). We imaged cells on a Leica TCS SP5 II confocal microscope with LASAF software and images were processed using ImageJ and were representative of the cell population.

ChIP-seq analyses

S2 cells transfected with either mCherry-Nnk or Odj-mCherry for 72 hr were fixed with 1% paraformaldehyde for 10 min and sheared with Bioruptor sonicator (Diagenode) to obtain chromatin. Chromatin fragments were confirmed to contain DNA in the 200–500 bp size range using a Bioanalyzer (Agilent). Each immunoprecipitation was performed on chromatin from 2×10^7 S2 cells by overnight incubation with Protein-G Dynabeads and 5 µg of anti-mCherry antibody (Novus). Library construction from immunoprecipitated DNA was conducted using TruSeq sample preparation kits (Illumina). 150 bp paired-end sequences were generated by the Vincent J. Coates Genomics Sequencing Laboratory at UC Berkeley.

We used BWA-mem to map paired reads to version 6 of the *D. melanogaster* genome assembly (dm6) from which we had removed unplaced scaffolds (*Li, 2013*). We used deepTools' bamCompare (with the '–binsize 1 `-extendReads`' options) to obtain log2 ratios of fragment coverage for matched ChIP and input samples, and visualized those ratios in IGV (*Robinson et al., 2011*). We further visualized ChIP-seq signal around TSSs using deepTools' computeMatrix and plotHeatmap tools, using TSS annotations obtained from the TxDb.Dmelanogaster.UCSC.dm6.ensGene BioConductor package, taking the most upstream TSS for genes with alternative start sites. We split TSS annotations according to whether they were within 'TSS-proximal' (active) chromatin according to modENCODE's nine state annotation for S2 cells, obtained from http://intermine.modencode.org and converted from dm3 to dm6 coordinates using UCSC's liftOver tool (*Hinrichs et al., 2006*). We further split the 'active' TSSs according to whether they are within cytogenetic heterochromatin, using coordinates from *Supplementary file 2* of *Hoskins et al., 2015* as well as the whole of chromosomes four and Y. The ChIP-seq data has been deposited in the SRA database with the accession number PRJNA644950.

Additional information

Author contributions

Bhavatharini Kasinathan, Harmit S Malik, Conceptualization, Data curation, Formal analysis, Supervision, Funding acquisition, Validation, Investigation, Visualization, Methodology, Writing - original draft, Project administration, Writing - review and editing; Serafin U Colmenares III, Data curation, Formal analysis, Validation, Investigation, Visualization, Methodology, Writing - review and editing; Hannah McConnell, Validation, Investigation, Methodology, Writing - review and editing; Janet M Young, Resources, Data curation, Software, Supervision, Investigation, Methodology, Writing - original draft, Writing - review and editing; Gary H Karpen, Supervision, Funding acquisition, Validation, Writing - original draft, Project administration, Writing - review and editing

Author ORCIDs

Bhavatharini Kasinathan ⓘ https://orcid.org/0000-0001-7053-5826
Serafin U Colmenares III ⓘ http://orcid.org/0000-0002-4094-4220
Janet M Young ⓘ https://orcid.org/0000-0001-8220-8427
Gary H Karpen ⓘ http://orcid.org/0000-0003-1534-0385
Harmit S Malik ⓘ https://orcid.org/0000-0001-6005-0016

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.63368.sa1
Author response https://doi.org/10.7554/eLife.63368.sa2

Additional files

Supplementary files

• Supplementary file 1. Differential retention in *Drosophila* species and phenotypes conferred by the 91 *ZAD-ZNF* genes in *D. melanogaster*. ZAD-ZNF gene repertoires in different *Drosophila* species were inferred from OrthoDB analyses (Materials and methods). Phenotypic consequences of knockdown or knockout phenotype of ZAD-ZNFs in *D. melanogaster* were retrieved from flybase.org. Key for citations: Neely et al. Cell 2010 (*Neely et al., 2010*) Mummery-Widmer et al. Nature 2009 (*Mummery-Widmer et al., 2009*) Schnorrer et al. Nature 2010 (*Schnorrer et al., 2010*) Chang et al. Dev. Cell 2010 (*Chang et al., 2010*) Burguete et al. Dev. Biol. 2019 (*Burguete et al., 2019*) Gibert et al. Dev. Biol. 2005 (*Gibert et al., 2005*) Dai et al. Dev. Cell 2003 (*Dai et al., 2003*) Komura-Kawa et al. PLOS Genet. 2015 (*Komura-Kawa et al., 2015*) Weiler Genetics 2007 (*Weiler, 2007*) Giulian et al. PLOS ONE 2013 (*Giuliani et al., 2013*) Chen et al. Science 2010 (*Chen et al., 2010*) Schalet Mutat. Res. 1986 (*Schalet, 1986*) Shannon et al. Genetics 1972 (*Shannon et al., 1972*) Judd et al. Genetics 1972 (*Judd et al., 1972*) Liu et al. Genetics 1975 (*Liu and Lim, 1975*) Lim et al. Genet. Res. 1974 (*Lim and Snyder, 1974*) Robbins Genetics 1983 (*Robbins, 1983*) Gaszner et al. Genes Dev. 1999 (*Gaszner et al., 1999*) Schertel et al. Genome Res. 2015 (*Schertel et al., 2015*) Toba et al. Genetics 1999 (*Toba et al., 1999*) Ruppert et al. Cell Rep. 2017 (*Ruppert et al., 2017*) Schupbach et al. Genetics 1989 (*Schüpbach and Wieschaus, 1989*) Chen et al. Development 2000 (*Chen et al., 2000*) Lieberfarb et al. Development 1996 (*Lieberfarb et al., 1996*) Page et al. J. Cell Sci. 1996 (*Page and Orr-Weaver, 1996*) Whitfield et al. PLOS Biol. 2013 (*Whitfield et al., 2013*) Holden et al. Genetics 1973 (*Holden and Suzuki, 1973*) Ishimoto et al. Genetics 2010 (*Ishimoto and Kitamoto, 2010*) Ishimoto et al. Proc. Natl. Acad. Sci. USA 2009 (*Ishimoto et al., 2009*) Walker et al. J. Insect Physiol. 1987 (*Walker et al., 1987*) Ono et al. Dev. Biol. 2006 (*Ono et al., 2006*) Neubueser et al. Dev. Biol. 2005 (*Neubueser et al., 2005*) Laundrie et al. Genetics 2003 (*Laundrie et al., 2003*) Spradling et al. Genetics 1999 (*Spradling et al., 1999*) Page et al. EMBO J. 2005 (*Page et al., 2005*) Luschnig et al. Genetics 2004 (*Luschnig et al., 2004*) Ashburner et al. Genetics 1990 (*Ashburner et al., 1990*) Chen et al. Curr. Biol. 2006 (*Chen et al., 2006*) Kahsai et al. G3 2018 (*Kahsai and Cook, 2018*) Lake et al. PLoS Genet. 2011 (*Lake et al., 2011*) Zhang et al. Genetics 2007 (*Zhang et al., 2007*) Maksimenko et al. Biochim Biophys Acta 2020 (*Maksimenko et al., 1863*) Zolotarev et al. BioTechniques 2019 (*Zolotarev et al., 2019*) Neumüller et al. Cell Stem Cell 2011 (*Neumüller et al., 2011*) Nazario-Yepiz et al. Mech. Dev. 2017 (*Nazario-Yepiz and Riesgo-Escovar, 2017*) Lewandowski et al. Dev. Biol. 2010 (*Lewandowski et al., 2010*) Ellis et al. PLOS ONE 2015

(*Ellis et al., 2015*). (KD: knockdown; VDRC GD; Vienna *Drosophila* Resource Center GG RNAi line; VDRC KK; Vienna *Drosophila* Resource Center KK RNAi line; TRiP: Transgenic RNAi project, Perrimon lab).

• Supplementary file 2. Multiple alignment of the ZNF domain encoded by all genes found within the *Odj-Nnk ZAD-ZNF* cluster across 20 *Drosophila* species.

• Supplementary file 3. Gene ontology analysis of transcriptional dysregulation in *Nnk*-null L1 larvae in *D. melanogaster* reveals categories of genes that are significantly over- or under-expressed upon loss of *Nnk*.

• Supplementary file 4. Nucleotide sequences of the *Nnk-mel* and *Nnk-sim* rescue transgenes. Nnk-mel rescue transgene was recoded by synonymous substitutions to make it impervious to RNAi-mediated knockdown, and the *Nnk-sim* rescue transgene was codon-optimized to match *D. melanogaster* codon preferences.

• Transparent reporting form

Data availability

Our raw data that have been deposited to the SRA database under PRJNA644950 and PRJNA643855.

The following datasets were generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Kasinathan B, Colmenares SU, McConnell H, Young JM, Karpen GH, Malik HS	2020	ChIP-seq of Oddjob and Nicknack	http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA644950	NCBI BioProject, PRJNA644950
Kasinathan B, Colmenares SU, McConnell H, Young JM, Karpen GH, Malik HS	2020	RNA-seq of Nnk null flies	http://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA643855	NCBI BioProject, PRJNA643855

The following previously published dataset was used:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Lack JB, Cardeno CM, Crepeau MW, Taylor W, Corbett-Dettig RB, Stevens KA, Langley CH, Pool JE	2015	The *Drosophila* genome nexus: a population genomic resource of 623 *Drosophila melanogaster* genomes, including 197 from a single ancestral range population	https://www.ncbi.nlm.nih.gov/bioproject/PRJNA268677/	NCBI BioProject, PRJNA268677

References

Abe Y, Sako K, Takagaki K, Hirayama Y, Uchida KS, Herman JA, DeLuca JG, Hirota T. 2016. HP1-Assisted aurora B kinase activity prevents chromosome segregation errors. *Developmental Cell* **36**:487–497. DOI: https://doi.org/10.1016/j.devcel.2016.02.008, PMID: 26954544

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biology* **10**:R106. DOI: https://doi.org/10.1186/gb-2010-11-10-r106

Ashburner M, Thompson P, Roote J, Lasko PF, Grau Y, el Messal M. 1990. The genetics of a small autosomal region of *Drosophila melanogaster* containing the structural gene for alcohol dehydrogenase VII Characterization of the region around the snail and cactus loci. *Genetics* **94**:679–694.

Azzaz AM, Vitalini MW, Thomas AS, Price JP, Blacketer MJ, Cryderman DE, Zirbel LN, Woodcock CL, Elcock AH, Wallrath LL, Shogren-Knaak MA. 2014. Human heterochromatin protein 1α promotes nucleosome associations that drive chromatin condensation. *Journal of Biological Chemistry* **289**:6850–6861. DOI: https://doi.org/10.1074/jbc.M113.512137, PMID: 24415761

Baumann DG, Dai M-S, Lu H, Gilmour DS. 2017. GFZF, a Glutathione *S* -Transferase Protein Implicated in Cell Cycle Regulation and Hybrid Inviability, Is a Transcriptional Coactivator . *Molecular and Cellular Biology* **38**: e00476-17. DOI: https://doi.org/10.1128/MCB.00476-17

Branco AT, Brito RM, Lemos B. 1869. Sex-specific adaptation and genomic responses to Y chromosome presence in female reproductive and neural tissues. *Proceedings. Biological Sciences* **284**:20172062. DOI: https://doi.org/10.1098/rspb.2017.2062

Bray NL, Pimentel H, Melsted P, Pachter L. 2016. Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology* **34**:525–527. DOI: https://doi.org/10.1038/nbt.3519, PMID: 27043002

Brennecke J, Aravin AA, Stark A, Dus M, Kellis M, Sachidanandam R, Hannon GJ. 2007. Discrete small RNA-generating loci as master regulators of transposon activity in *Drosophila*. *Cell* **128**:1089–1103. DOI: https://doi.org/10.1016/j.cell.2007.01.043, PMID: 17346786

Brodsky S, Jana T, Mittelman K, Chapal M, Kumar DK, Carmi M, Barkai N. 2020. Intrinsically disordered regions direct transcription factor in Vivo Binding Specificity. *Molecular Cell* **79**:459–471. DOI: https://doi.org/10.1016/j.molcel.2020.05.032, PMID: 32553192

Brown EJ, Nguyen AH, Bachtrog D. 2020. The *Drosophila* Y chromosome affects heterochromatin integrity Genome-Wide. *Molecular Biology and Evolution* **20**:msaa082. DOI: https://doi.org/10.1093/molbev/msaa082

Burguete AS, Francis D, Rosa J, Ghabrial A. 2019. The regulation of cell size and branch complexity in the terminal cells of the *Drosophila* tracheal system. *Developmental Biology* **85**:79–85. DOI: https://doi.org/10.1016/j.ydbio.2019.02.001

Chakraborty M, Chang C-H, Khost DE, Vedanayagam J, Adrion JR, Liao Y, Montooth K, Meiklejohn CD, Larracuente AM, Emerson JJ. 2020. Evolution of genome structure in the *Drosophila* simulans species complex. *bioRxiv*. DOI: https://doi.org/10.1101/2020.02.27.968743

Chang KC, Garcia-Alvarez G, Somers G, Sousa-Nunes R, Rossi F, Lee YY, Soon SB, Gonzalez C, Chia W, Wang H. 2010. Interplay between the transcription factor zif and aPKC regulates neuroblast polarity and self-renewal. *Developmental Cell* **19**:778–785. DOI: https://doi.org/10.1016/j.devcel.2010.10.007, PMID: 21074726

Chen B, Harms E, Chu T, Henrion G, Strickland S. 2000. Completion of meiosis in *Drosophila* oocytes requires transcriptional control by grauzone, a new zinc finger protein. *Development* **127**:1243–1251. PMID: 10683177

Chen LY, Wang JC, Hyvert Y, Lin HP, Perrimon N, Imler JL, Hsu JC. 2006. Weckle is a zinc finger adaptor of the toll pathway in dorsoventral patterning of the *Drosophila* embryo. *Current Biology* **16**:1183–1193. DOI: https://doi.org/10.1016/j.cub.2006.05.050, PMID: 16782008

Chen S, Zhang YE, Long M. 2010. New genes in *Drosophila* quickly become essential. *Science* **330**:1682–1685. DOI: https://doi.org/10.1126/science.1196380, PMID: 21164016

Chen Z-X, Sturgill D, Qu J, Jiang H, Park S, Boley N, Suzuki AM, Fletcher AR, Plachetzki DC, FitzGerald PC, Artieri CG, Atallah J, Barmina O, Brown JB, Blankenburg KP, Clough E, Dasgupta A, Gubbala S, Han Y, Jayaseelan JC, et al. 2014. Comparative validation of the D. Melanogaster modENCODE transcriptome annotation. *Genome Research* **24**:1209–1223. DOI: https://doi.org/10.1101/gr.159384.113

Chung HR, Schäfer U, Jäckle H, Böhm S. 2002. Genomic expansion and clustering of ZAD-containing C2H2 zinc-finger genes in *Drosophila*. *EMBO Reports* **3**:1158–1162. DOI: https://doi.org/10.1093/embo-reports/kvf243, PMID: 12446571

Chung HR, Löhr U, Jäckle H. 2007. Lineage-specific expansion of the zinc finger associated domain ZAD. *Molecular Biology and Evolution* **24**:1934–1943. DOI: https://doi.org/10.1093/molbev/msm121, PMID: 1756 9752

Czech B, Preall JB, McGinn J, Hannon GJ. 2013. A transcriptome-wide RNAi screen in the *Drosophila* ovary reveals factors of the germline piRNA pathway. *Molecular Cell* **50**:749–761. DOI: https://doi.org/10.1016/j.molcel.2013.04.007, PMID: 23665227

Dai P, Akimaru H, Ishii S. 2003. A hedgehog-responsive region in the *Drosophila* wing disc is defined by debra-mediated ubiquitination and lysosomal degradation of ci. *Developmental Cell* **4**:917–928. DOI: https://doi.org/10.1016/S1534-5807(03)00158-8, PMID: 12791275

Day DS, Luquette LJ, Park PJ, Kharchenko PV. 2010. Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biology* **11**:R69. DOI: https://doi.org/10.1186/gb-2010-11-6-r69, PMID: 20584328

Devlin RH, Bingham B, Wakimoto BT. 1990. The organization and expression of the light gene, a heterochromatic gene of *Drosophila melanogaster*. *Genetics* **40**:129–140.

Drinnenberg IA, Henikoff S, Malik HS. 2016. Evolutionary turnover of kinetochore proteins: a ship of theseus? *Trends in Cell Biology* **26**:498–510. DOI: https://doi.org/10.1016/j.tcb.2016.01.005, PMID: 26877204

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **12**:203–218. DOI: https://doi.org/10.1038/nature06341

Eberl DF, Duyf BJ, Hilliker AJ. 1993. The role of heterochromatin in the expression of a heterochromatic gene, the rolled locus of *Drosophila melanogaster*. *Genetics* **134**:277–292. PMID: 8514136

Egea R, Casillas S, Barbadilla A. 2008. Standard and generalized McDonald-Kreitman test: a website to detect selection by comparing different classes of DNA sites. *Nucleic Acids Research* **36**:W157–W162. DOI: https://doi.org/10.1093/nar/gkn337, PMID: 18515345

Ellis K, Friedman C, Yedvobnick B. 2015. *Drosophila* domino exhibits genetic interactions with a wide spectrum of chromatin Protein-Encoding loci. *PLOS ONE* **10**:e0142635. DOI: https://doi.org/10.1371/journal.pone.0142635, PMID: 26555684

Emerson RO, Thomas JH. 2009. Adaptive evolution in zinc finger transcription factors. *PLOS Genetics* **5**: e1000325. DOI: https://doi.org/10.1371/journal.pgen.1000325, PMID: 19119423

Erijman A, Kozlowski L, Sohrabi-Jahromi S, Fishburn J, Warfield L, Schreiber J, Noble WS, Söding J, Hahn S. 2020. A High-Throughput screen for transcription activation domains reveals their sequence features and

permits prediction by deep learning. *Molecular Cell* **78**:890–902. DOI: https://doi.org/10.1016/j.molcel.2020.04.020

Falcon S, Gentleman R. 2007. Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**:257–258. DOI: https://doi.org/10.1093/bioinformatics/btl567, PMID: 17098774

Fay JC, Wyckoff GJ, Wu CI. 2002. Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature* **415**:1024–1026. DOI: https://doi.org/10.1038/4151024a, PMID: 11875569

Fluegel ML, Parker TJ, Pallanck LJ. 2006. Mutations of a *Drosophila* NPC1 gene confer sterol and ecdysone metabolic defects. *Genetics* **172**:185–196. DOI: https://doi.org/10.1534/genetics.105.046565, PMID: 16079224

Francisco FO, Lemos B. 2014. How do y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. *Journal of Genomics* **2**:94–103. DOI: https://doi.org/10.7150/jgen.8043, PMID: 25057325

Gaszner M, Vazquez J, Schedl P. 1999. The Zw5 protein, a component of the scs chromatin domain boundary, is able to block enhancer-promoter interaction. *Genes & Development* **13**:2098–2107. DOI: https://doi.org/10.1101/gad.13.16.2098, PMID: 10465787

Gertz EM, Yu YK, Agarwala R, Schäffer AA, Altschul SF. 2006. Composition-based statistics and translated nucleotide searches: improving the TBLASTN module of BLAST. *BMC Biology* **4**:41. DOI: https://doi.org/10.1186/1741-7007-4-41, PMID: 17156431

Gibert JM, Marcellini S, David JR, Schlötterer C, Simpson P. 2005. A major bristle QTL from a selected population of *Drosophila* uncovers the zinc-finger transcription factor poils-au-dos, a repressor of achaete-scute. *Developmental Biology* **288**:194–205. DOI: https://doi.org/10.1016/j.ydbio.2005.09.032, PMID: 16216235

Giuliani F, Giuliani G, Bauer R. 2013. Innexin 3, a new gene required for dorsal closure in *Drosophila* embryo. *PLOS ONE* **7**:e69212. DOI: https://doi.org/10.1371/journal.pone.0069212

Goriaux C, Desset S, Renaud Y, Vaury C, Brasset E. 2014a. Transcriptional properties and splicing of the flamenco piRNA cluster. *EMBO Reports* **15**:411–418. DOI: https://doi.org/10.1002/embr.201337898, PMID: 24562610

Goriaux C, Théron E, Brasset E, Vaury C. 2014b. History of the discovery of a master locus producing piRNAs: the flamenco/COM locus in *Drosophila melanogaster*. *Frontiers in Genetics* **5**:257. DOI: https://doi.org/10.3389/fgene.2014.00257, PMID: 25136352

Green EW, Fedele G, Giorgini F, Kyriacou CP. 2014. A *Drosophila* RNAi collection is subject to dominant phenotypic effects. *Nature Methods* **11**:222–223. DOI: https://doi.org/10.1038/nmeth.2856

Grézy A, Chevillard-Briet M, Trouche D, Escaffit F. 2016. Control of genetic stability by a new heterochromatin compaction pathway involving the Tip60 histone acetyltransferase. *Molecular Biology of the Cell* **27**:599–607. DOI: https://doi.org/10.1091/mbc.E15-05-0316, PMID: 26700317

Groth AC, Fish M, Nusse R, Calos MP. 2004. Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. *Genetics* **166**:1775–1782. DOI: https://doi.org/10.1534/genetics.166.4.1775, PMID: 15126397

Harms E, Chu T, Henrion G, Strickland S. 2000. The only function of grauzone required for *Drosophila* oocyte meiosis is transcriptional activation of the cortex gene. *Genetics* **155**:1831–1839. PMID: 10924478

Hervas S, Sanz E, Casillas S, Pool JE, Barbadilla A. 2017. PopFly: the *Drosophila* population genomics browser. *Bioinformatics* **33**:2779–2780. DOI: https://doi.org/10.1093/bioinformatics/btx301, PMID: 28472360

Hinrichs AS, Karolchik D, Baertsch R, Barber GP, Bejerano G, Clawson H, Diekhans M, Furey TS, Harte RA, Hsu F, Hillman-Jackson J, Kuhn RM, Pedersen JS, Pohl A, Raney BJ, Rosenbloom KR, Siepel A, Smith KE, Sugnet CW, Sultan-Qurraie A, et al. 2006. The UCSC genome browser database: update 2006. *Nucleic Acids Research* **34**:D590–D598. DOI: https://doi.org/10.1093/nar/gkj144, PMID: 16381938

Ho JW, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn KA, Minoda A, Tolstorukov MY, Appert A, Parker SC, Gu T, Kundaje A, Riddle NC, Bishop E, Egelhofer TA, Hu SS, Alekseyenko AA, Rechtsteiner A, Asker D, et al. 2014. Comparative analysis of metazoan chromatin organization. *Nature* **512**:449–452. DOI: https://doi.org/10.1038/nature13415, PMID: 25164756

Holden JJ, Suzuki DT. 1973. Temperature-sensitive mutations in *Drosophila melanogaster* XII the genetic and developmental effects of dominant lethals on chromosome 3. *Genetics* **73**:445–458. PMID: 4349421

Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, Krzywinski M, Schein J, Accardo MC, Damia E, Messina G, Méndez-Lago M, de Pablos B, Demakova OV, Andreyeva EN, Boldyreva LV, et al. 2015. The release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Research* **25**:445. DOI: https://doi.org/10.1101/gr.185579.114, PMID: 25589440

Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, et al. 2015. Orchestrating high-throughput genomic analysis with bioconductor. *Nature Methods* **12**:115–121. DOI: https://doi.org/10.1038/nmeth.3252, PMID: 25633503

Imbeault M, Helleboid PY, Trono D. 2017. KRAB zinc-finger proteins contribute to the evolution of gene regulatory networks. *Nature* **543**:550–554. DOI: https://doi.org/10.1038/nature21683, PMID: 28273063

Inoue A, Hyle J, Lechner MS, Lahti JM. 2008. Perturbation of HP1 localization and chromatin binding ability causes defects in sister-chromatid cohesion. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis* **657**:48–55. DOI: https://doi.org/10.1016/j.mrgentox.2008.08.010, PMID: 18790078

Ishida T, Kinoshita K. 2007. PrDOS: prediction of disordered protein regions from amino acid sequence. *Nucleic Acids Research* **35**:W460–W464. DOI: https://doi.org/10.1093/nar/gkm363, PMID: 17567614

Ishimoto H, Sakai T, Kitamoto T. 2009. Ecdysone signaling regulates the formation of long-term courtship memory in adult *Drosophila melanogaster*. *PNAS* **6**:6381–6386. DOI: https://doi.org/10.1073/pnas.0810213106

Ishimoto H, Kitamoto T. 2010. The steroid molting hormone ecdysone regulates sleep in adult *Drosophila melanogaster*. *Genetics* **185**:269–281. DOI: https://doi.org/10.1534/genetics.110.114587, PMID: 20215472

Jagannathan M, Warsinger-Pepe N, Watase GJ, Yamashita YM. 2017. Comparative analysis of satellite DNA in the *Drosophila melanogaster* Species Complex. *G3: Genes, Genomes, Genetics* **7**:693–704. DOI: https://doi.org/10.1534/g3.116.035352, PMID: 28007840

Jain D, Baldi S, Zabel A, Straub T, Becker PB. 2015. Active promoters give rise to false positive 'Phantom Peaks' in ChIP-seq experiments. *Nucleic Acids Research* **43**:6959–6968. DOI: https://doi.org/10.1093/nar/gkv637

Jauch R, Bourenkov GP, Chung HR, Urlaub H, Reidt U, Jäckle H, Wahl MC. 2003. The zinc finger-associated domain of the *Drosophila* transcription factor grauzone is a novel zinc-coordinating protein-protein interaction module. *Structure* **11**:1393–1402. DOI: https://doi.org/10.1016/j.str.2003.09.015, PMID: 14604529

Judd BH, Shen MW, Kaufman TC. 1972. The anatomy and function of a segment of the X chromosome of *Drosophila melanogaster*. *Genetics* **71**:139–156. PMID: 4624778

Kahsai L, Cook KR. 2018. Mapping second chromosome mutations to defined genomic regions in *Drosophila melanogaster*. *G3: Genes, Genomes, Genetics* **8**:9–16. DOI: https://doi.org/10.1534/g3.117.300289, PMID: 29066472

Klattenhoff C, Xi H, Li C, Lee S, Xu J, Khurana JS, Zhang F, Schultz N, Koppetsch BS, Nowosielska A, Seitz H, Zamore PD, Weng Z, Theurkauf WE. 2009. The *Drosophila* HP1 homolog rhino is required for transposon silencing and piRNA production by dual-strand clusters. *Cell* **138**:1137–1149. DOI: https://doi.org/10.1016/j.cell.2009.07.014, PMID: 19732946

Komura-Kawa T, Hirota K, Shimada-Niwa Y, Yamauchi R, Shimell M, Shinoda T, Fukamizu A, O'Connor MB, Niwa R. 2015. The *Drosophila* zinc finger transcription factor ouija board controls ecdysteroid biosynthesis through specific regulation of spookier. *PLOS Genetics* **11**:e1005712. DOI: https://doi.org/10.1371/journal.pgen.1005712, PMID: 26658797

Kondo S, Vedanayagam J, Mohammed J, Eizadshenass S, Kan L, Pang N, Aradhya R, Siepel A, Steinhauer J, Lai EC. 2017. New genes often acquire male-specific functions but rarely become essential in *Drosophila*. *Genes & Development* **31**:1841–1846. DOI: https://doi.org/10.1101/gad.303131.117, PMID: 29051389

Krylov DM, Wolf YI, Rogozin IB, Koonin EV. 2003. Gene loss, protein sequence divergence, gene dispensability, expression level, and interactivity are correlated in eukaryotic evolution. *Genome Research* **13**:2229–2235. DOI: https://doi.org/10.1101/gr.1589103, PMID: 14525925

Krystel J, Ayyanathan K. 2013. Global analysis of target genes of 21 members of the ZAD transcription factor family in *Drosophila melanogaster*. *Gene* **512**:373–382. DOI: https://doi.org/10.1016/j.gene.2012.09.114, PMID: 23085320

Kumar S, Stecher G, Suleski M, Hedges SB. 2017. TimeTree: a resource for timelines, timetrees, and divergence times. *Molecular Biology and Evolution* **34**:1812–1819. DOI: https://doi.org/10.1093/molbev/msx116, PMID: 28387841

Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A thousand fly genomes: an expanded *Drosophila* genome nexus. *Molecular Biology and Evolution* **3313**:3308. DOI: https://doi.org/10.1093/molbev/msw195

Lake CM, Nielsen RJ, Hawley RS. 2011. The *Drosophila* zinc finger protein trade embargo is required for double strand break formation in meiosis. *PLOS Genetics* **7**:e1002005. DOI: https://doi.org/10.1371/journal.pgen.1002005, PMID: 21383963

Langley CH, Stevens K, Cardeno C, Lee YC, Schrider DR, Pool JE, Langley SA, Suarez C, Corbett-Detig RB, Kolaczkowski B, Fang S, Nista PM, Holloway AK, Kern AD, Dewey CN, Song YS, Hahn MW, Begun DJ. 2012. Genomic variation in natural populations of *Drosophila melanogaster*. *Genetics* **192**:533–598. DOI: https://doi.org/10.1534/genetics.112.142018, PMID: 22673804

Laundrie B, Peterson JS, Baum JS, Chang JC, Fileppo D, Thompson SR, McCall K. 2003. Germline cell death is inhibited by P-element insertions disrupting the dcp-1/pita nested gene pair in *Drosophila*. *Genetics* **165**:1881–1888. PMID: 14704173

Lee H, McManus CJ, Cho DY, Eaton M, Renda F, Somma MP. 2014. DNA copy number evolution in *Drosophila* cell lines. *Genome Biology* **8**:R70. DOI: https://doi.org/10.1186/gb-2014-15-8-r70

Lee YCG, Yang Q, Chi W, Turkson SA, Du WA, Kemkemer C, Zeng ZB, Long M, Zhuang X. 2017. Genetic architecture of natural variation underlying adult foraging behavior that is essential for survival of *Drosophila melanogaster*. *Genome Biology and Evolution* **9**:1357–1369. DOI: https://doi.org/10.1093/gbe/evx089, PMID: 28472322

Lee YCG, Ventura IM, Rice GR, Chen DY, Colmenares SU, Long M. 2019. Rapid evolution of gained essential developmental functions of a young gene via interactions with other essential genes. *Molecular Biology and Evolution* **36**:2212–2226. DOI: https://doi.org/10.1093/molbev/msz137, PMID: 31187122

Lespinet O, Wolf YI, Koonin EV, Aravind L. 2002. The role of lineage-specific gene family expansion in the evolution of eukaryotes. *Genome Research* **12**:1048–1059. DOI: https://doi.org/10.1101/gr.174302, PMID: 12097341

Levine MT, Vander Wende HM, Malik HS. 2015. Mitotic fidelity requires transgenerational action of a testis-restricted HP1. *eLife* **4**:e07378. DOI: https://doi.org/10.7554/eLife.07378, PMID: 26151671

Levine MT, Vander Wende HM, Hsieh E, Baker EP, Malik HS. 2016. Recurrent gene duplication diversifies genome defense repertoire in *Drosophila*. *Molecular Biology and Evolution* **33**:1641–1653. DOI: https://doi.org/10.1093/molbev/msw053, PMID: 26979388

Lewandowski JP, Sheehan KB, Bennett PE, Boswell RE. 2010. Mago nashi, tsunagi/Y14, and ranshi form a complex that influences oocyte differentiation in *Drosophila melanogaster*. *Developmental Biology* **339**:307–319. DOI: https://doi.org/10.1016/j.ydbio.2009.12.035

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. https://arxiv.org/abs/1303.3997.

Li J, Gilmour DS. 2013. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *The EMBO Journal* **32**:1829–1841. DOI: https://doi.org/10.1038/emboj.2013.111, PMID: 23708796

Lieberfarb ME, Chu T, Wreden C, Theurkauf W, Gergen JP, Strickland S. 1996. Mutations that perturb poly(A)-dependent maternal mRNA activation block the initiation of development. *Development* **122**:579–588. PMID: 8625809

Lim JK, Snyder LA. 1974. Cytogenetic and complementation analyses of recessive lethal mutations induced in the X chromosome of *Drosophila* by three alkylating agents. *Genetical Research* **24**:1–10. DOI: https://doi.org/10.1017/S0016672300015020, PMID: 4373306

Liu X, Song Z, Huo Y, Zhang J, Zhu T, Wang J, Zhao X, Aikhionbare F, Zhang J, Duan H, Wu J, Dou Z, Shi Y, Yao X. 2014. Chromatin protein HP1 interacts with the mitotic regulator borealin protein and specifies the centromere localization of the chromosomal passenger complex. *Journal of Biological Chemistry* **289**:20638–20649. DOI: https://doi.org/10.1074/jbc.M114.572842, PMID: 24917673

Liu CP, Lim JK. 1975. Complementation analysis of methyl methane-sulfonate-induced recessive lethal mutations in the zeste-white region of the X chromosome of *Drosophila melanogaster*. *Genetics* **79**:601–611. PMID: 166020

Long M, VanKuren NW, Chen S, Vibranovski MD. 2013. New gene evolution: little did we know. *Annual Review of Genetics* **47**:307–333. DOI: https://doi.org/10.1146/annurev-genet-111212-133301, PMID: 24050177

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**:8. DOI: https://doi.org/10.1186/s13059-014-0550-8

Luschnig S, Moussian B, Krauss J, Desjeux I, Perkovic J, Nüsslein-Volhard C. 2004. An F1 genetic screen for maternal-effect mutations affecting embryonic pattern formation in *Drosophila melanogaster*. *Genetics* **167**:325–342. DOI: https://doi.org/10.1534/genetics.167.1.325, PMID: 15166158

Maksimenko O, Kyrchanova O, Klimenko N, Zolotarev N, Elizarova A, Bonchuk A. 1863. Small *Drosophila* zinc finger C2H2 protein with an N-terminal zinc finger-associated domain demonstrates the architecture functions. *Biochimica Et Biophysica Acta. Gene Regulatory Mechanisms* **2020**:194446. DOI: https://doi.org/10.1016/j.bbagrm.2019.194446

Malik HS, Henikoff S. 2001. Adaptive evolution of cid a centromere-specific histone in *Drosophila*. *Genetics* **157**:1293–1298. PMID: 11238413

Marchler-Bauer A, Bo Y, Han L, He J, Lanczycki CJ, Lu S, Chitsaz F, Derbyshire MK, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lu F, Marchler GH, Song JS, Thanki N, Wang Z, Yamashita RA, Zhang D, Zheng C, et al. 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. *Nucleic Acids Research* **45**:D200–D203. DOI: https://doi.org/10.1093/nar/gkw1129, PMID: 27899674

McDonald JH, Kreitman M. 1991. Adaptive protein evolution at the adh locus in *Drosophila*. *Nature* **351**:652–654. DOI: https://doi.org/10.1038/351652a0, PMID: 1904993

Miklos GL, Rubin GM. 1996. The role of the genome project in determining gene function: insights from model organisms. *Cell* **86**:521–529. DOI: https://doi.org/10.1016/S0092-8674(00)80126-9, PMID: 8752207

Mummery-Widmer JL, Yamazaki M, Stoeger T, Novatchkova M, Bhalerao S, Chen D, Dietzl G, Dickson BJ, Knoblich JA. 2009. Genome-wide analysis of notch signalling in *Drosophila* by transgenic RNAi. *Nature* **458**:987–992. DOI: https://doi.org/10.1038/nature07936, PMID: 19363474

Nambiar M, Smith GR. 2018. Pericentromere-Specific cohesin complex prevents meiotic pericentric DNA Double-Strand breaks and lethal crossovers. *Molecular Cell* **71**:540–553. DOI: https://doi.org/10.1016/j.molcel.2018.06.035, PMID: 30078721

Nazario-Yepiz NO, Riesgo-Escovar JR. 2017. Piragua encodes a zinc finger protein required for development in *Drosophila*. *Mechanisms of Development* **144**:171–181. DOI: https://doi.org/10.1016/j.mod.2016.12.003, PMID: 28011160

Neely GG, Hess A, Costigan M, Keene AC, Goulas S, Langeslag M, Griffin RS, Belfer I, Dai F, Smith SB, Diatchenko L, Gupta V, Xia CP, Amann S, Kreitz S, Heindl-Erdmann C, Wolz S, Ly CV, Arora S, Sarangi R, et al. 2010. A genome-wide *Drosophila* screen for heat nociception identifies α2δ3 as an evolutionarily conserved pain gene. *Cell* **143**:628–638. DOI: https://doi.org/10.1016/j.cell.2010.09.047, PMID: 21074052

Neubueser D, Warren JT, Gilbert LI, Cohen SM. 2005. Molting defective is required for ecdysone biosynthesis. *Developmental Biology* **280**:362–372. DOI: https://doi.org/10.1016/j.ydbio.2005.01.023, PMID: 15882578

Neumüller RA, Richter C, Fischer A, Novatchkova M, Neumüller KG, Knoblich JA. 2011. Genome-wide analysis of self-renewal in *Drosophila* neural stem cells by transgenic RNAi. *Cell Stem Cell* **8**:580–593. DOI: https://doi.org/10.1016/j.stem.2011.02.022, PMID: 21549331

Okita AK, Zafar F, Su J, Weerasekara D, Kajitani T, Takahashi TS, Kimura H, Murakami Y, Masukata H, Nakagawa T. 2019. Heterochromatin suppresses gross chromosomal rearrangements at centromeres by repressing Tfs1/TFIIS-dependent transcription. *Communications Biology* **2**:17. DOI: https://doi.org/10.1038/s42003-018-0251-z, PMID: 30652128

Ono H, Rewitz KF, Shinoda T, Itoyama K, Petryk A, Rybczynski R, Jarcho M, Warren JT, Marqués G, Shimell MJ, Gilbert LI, O'Connor MB. 2006. Spook and spookier code for stage-specific components of the ecdysone

biosynthetic pathway in diptera. *Developmental Biology* **298**:555–570. DOI: https://doi.org/10.1016/j.ydbio.2006.07.023, PMID: 16949568

Page AR, Kovacs A, Deak P, Torok T, Kiss I, Dario P, Bastos C, Batista P, Gomes R, Ohkura H, Russell S, Glover DM. 2005. Spotted-dick, a zinc-finger protein of *Drosophila* required for expression of Orc4 and S phase. *The EMBO Journal* **24**:4304–4315. DOI: https://doi.org/10.1038/sj.emboj.7600890, PMID: 16369566

Page AW, Orr-Weaver TL. 1996. The *Drosophila* genes grauzone and cortex are necessary for proper female meiosis. *Journal of Cell Science* **109**:1707–1715. PMID: 8832393

Parhad SS, Tu S, Weng Z, Theurkauf WE. 2017. Adaptive evolution leads to Cross-Species incompatibility in the piRNA transposon silencing machinery. *Developmental Cell* **43**:60–70. DOI: https://doi.org/10.1016/j.devcel.2017.08.012, PMID: 28919205

Piergentili R. 2010. Multiple roles of the Y chromosome in the biology of *Drosophila melanogaster*. *The Scientific World JOURNAL* **10**:1749–1767. DOI: https://doi.org/10.1100/tsw.2010.168, PMID: 20842320

Robbins LG. 1983. Maternal-Zygotic lethal interactions in *Drosophila melanogaster*: zeste-white region Single-Cistron mutations. *Genetics* **103**:633–648. PMID: 17246119

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nature Biotechnology* **29**:24–26. DOI: https://doi.org/10.1038/nbt.1754, PMID: 21221095

Ross BD, Rosin L, Thomae AW, Hiatt MA, Vermaak D, de la Cruz AF, Imhof A, Mellone BG, Malik HS. 2013. Stepwise evolution of essential centromere function in a *Drosophila* Neogene. *Science* **340**:1211–1214. DOI: https://doi.org/10.1126/science.1234393, PMID: 23744945

Rowe HM, Jakobsson J, Mesnard D, Rougemont J, Reynard S, Aktas T, Maillard PV, Layard-Liesching H, Verp S, Marquis J, Spitz F, Constam DB, Trono D. 2010. KAP1 controls endogenous retroviruses in embryonic stem cells. *Nature* **463**:237–240. DOI: https://doi.org/10.1038/nature08674, PMID: 20075919

Ruiz-Estévez M, Bakkali M, Cabrero J, Camacho JP, López-León MD. 2014. HP1 knockdown is associated with abnormal condensation of almost all chromatin types in a grasshopper (Eyprepocnemis plorans). *Chromosome Research* **22**:253–266. DOI: https://doi.org/10.1007/s10577-013-9399-z, PMID: 24398928

Ruppert M, Franz M, Saratsis A, Velo Escarcena L, Hendrich O, Gooi LM, Schwenkert I, Klebes A, Scholz H. 2017. Hangover links nuclear RNA signaling to cAMP regulation via the phosphodiesterase 4d ortholog dunce. *Cell Reports* **18**:533–544. DOI: https://doi.org/10.1016/j.celrep.2016.12.048, PMID: 28076795

Schalet AP. 1986. The distribution of and complementation relationships between spontaneous X-linked recessive lethal mutations recovered from crossing long-term laboratory stocks of *Drosophila melanogaster*. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* **163**:115–144. DOI: https://doi.org/10.1016/0027-5107(86)90042-4

Schertel C, Albarca M, Rockel-Bauer C, Kelley NW, Bischof J, Hens K, van Nimwegen E, Basler K, Deplancke B. 2015. A large-scale, in vivo transcription factor screen defines bivalent chromatin as a key property of regulatory factors mediating *Drosophila* wing development. *Genome Research* **25**:514–523. DOI: https://doi.org/10.1101/gr.181305.114, PMID: 25568052

Schnorrer F, Schönbauer C, Langer CC, Dietzl G, Novatchkova M, Schernhuber K, Fellner M, Azaryan A, Radolf M, Stark A, Keleman K, Dickson BJ. 2010. Systematic genetic analysis of muscle morphogenesis and function in *Drosophila*. *Nature* **464**:287–291. DOI: https://doi.org/10.1038/nature08799, PMID: 20220848

Schuldiner O, Berdnik D, Levy JM, Wu JS, Luginbuhl D, Gontang AC, Luo L. 2008. piggyBac-based mosaic screen identifies a postmitotic function for cohesin in regulating developmental axon pruning. *Developmental Cell* **14**:227–238. DOI: https://doi.org/10.1016/j.devcel.2007.11.001, PMID: 18267091

Schulze SR, Sinclair DA, Fitzpatrick KA, Honda BM. 2005. A genetic and molecular characterization of two proximal heterochromatic genes on chromosome 3 of *Drosophila melanogaster*. *Genetics* **169**:2165–2177. DOI: https://doi.org/10.1534/genetics.103.023341, PMID: 15687284

Schulze SR, McAllister BF, Sinclair DA, Fitzpatrick KA, Marchetti M, Pimpinelli S, Honda BM. 2006. Heterochromatic genes in *Drosophila*: a comparative analysis of two genes. *Genetics* **173**:1433–1445. DOI: https://doi.org/10.1534/genetics.106.056069, PMID: 16648646

Schüpbach T, Wieschaus E. 1989. Female sterile mutations on the second chromosome of *Drosophila melanogaster* I. Maternal effect mutations. *Genetics* **121**:101–117. PMID: 2492966

Senti KA, Brennecke J. 2010. The piRNA pathway: a fly's perspective on the guardian of the genome. *Trends in Genetics* **26**:499–509. DOI: https://doi.org/10.1016/j.tig.2010.08.007, PMID: 20934772

Shannon MP, Kaufman TC, Shen MW, Judd BH. 1972. Lethality patterns and morphology of selected lethal and semi-lethal mutations in the zeste-white region of *Drosophila melanogaster*. *Genetics* **72**:615–638. PMID: 4631596

Sinclair DA, Schulze S, Silva E, Fitzpatrick KA, Honda BM. 2000. Essential genes in autosomal heterochromatin of *Drosophila melanogaster*. *Genetica* **109**:9–18. DOI: https://doi.org/10.1023/a:1026500620158, PMID: 11293800

Smothers JF, Henikoff S. 2000. The HP1 chromo shadow domain binds a consensus peptide pentamer. *Current Biology* **10**:27–30. DOI: https://doi.org/10.1016/S0960-9822(99)00260-2, PMID: 10660299

Soneson C, Love MI, Robinson MD. 2015. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* **4**:1521. DOI: https://doi.org/10.12688/f1000research.7563.1, PMID: 26925227

Spradling AC, Stern D, Beaton A, Rhem EJ, Laverty T, Mozden N, Misra S, Rubin GM. 1999. The Berkeley *Drosophila* genome project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes. *Genetics* **153**:135–177. PMID: 10471706

Swenson JM, Colmenares SU, Strom AR, Costes SV, Karpen GH. 2016. The composition and organization of *Drosophila* heterochromatin are heterogeneous and dynamic. *eLife* **5**:e16096. DOI: https://doi.org/10.7554/eLife.16096, PMID: 27514026

Talbert PB, Bryson TD, Henikoff S. 2004. Adaptive evolution of centromere proteins in plants and animals. *Journal of Biology* **3**:18. DOI: https://doi.org/10.1186/jbiol11, PMID: 15345035

Thomas JH, Schneider S. 2011. Coevolution of retroelements and tandem zinc finger genes. *Genome Research* **21**:1800–1812. DOI: https://doi.org/10.1101/gr.121749.111

Thompson JD, Gibson TJ, Higgins DG. 2002. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics* **2**:3. DOI: https://doi.org/10.1002/0471250953.bi0203s00, PMID: 18792934

Toba G, Ohsako T, Miyata N, Ohtsuka T, Seong KH, Aigaki T. 1999. The gene search system A method for efficient detection and rapid molecular identification of genes in *Drosophila melanogaster*. *Genetics* **151**:725–737. PMID: 9927464

Uryu O, Ou Q, Komura-Kawa T, Kamiyama T, Iga M, Syrzycka M, Hirota K, Kataoka H, Honda BM, King-Jones K, Niwa R. 2018. Cooperative control of ecdysone biosynthesis in *Drosophila* by transcription factors séance, ouija board, and molting defective. *Genetics* **208**:605–622. DOI: https://doi.org/10.1534/genetics.117.300268, PMID: 29187506

VanKuren NW, Long M. 2018. Gene duplicates resolving sexual conflict rapidly evolved essential gametogenesis functions. *Nature Ecology & Evolution* **2**:705–712. DOI: https://doi.org/10.1038/s41559-018-0471-0, PMID: 29459709

Vermaak D, Malik HS. 2009. Multiple roles for heterochromatin protein 1 genes in *Drosophila*. *Annual Review of Genetics* **43**:467–492. DOI: https://doi.org/10.1146/annurev-genet-102108-134802, PMID: 19919324

Vernì F, Cenci G. 2015. The *Drosophila* histone variant H2A.V works in concert with HP1 to promote kinetochore-driven microtubule formation. *Cell Cycle* **14**:577–588. DOI: https://doi.org/10.4161/15384101.2014.991176, PMID: 25591068

Verschure PJ, van der Kraan I, de Leeuw W, van der Vlag J, Carpenter AE, Belmont AS, van Driel R. 2005. In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Molecular and Cellular Biology* **25**:4552–4564. DOI: https://doi.org/10.1128/MCB.25.11.4552-4564.2005, PMID: 15899859

Vissers JH, Manning SA, Kulkarni A, Harvey KF. 2016. A *Drosophila* RNAi library modulates hippo pathway-dependent tissue growth. *Nature Communications* **7**:10368. DOI: https://doi.org/10.1038/ncomms10368, PMID: 26758424

Wakimoto BT, Hearn MG. 1990. The effects of chromosome rearrangements on the expression of heterochromatic genes in chromosome 2L of *Drosophila melanogaster*. *Genetics* **125**:141–154. PMID: 2111264

Walker VK, Watson KL, Holden JJA, Steel CGH. 1987. Vitellogenesis and fertility in *Drosophila* females with low ecdysteroid titres; the L(3)3DTS mutation. *Journal of Insect Physiology* **33**:137–142. DOI: https://doi.org/10.1016/0022-1910(87)90139-9

Wang M, Branco AT, Lemos B. 2018. The Y chromosome modulates splicing and Sex-Biased intron retention rates in *Drosophila*. *Genetics* **208**:1057–1067. DOI: https://doi.org/10.1534/genetics.117.300637, PMID: 29263027

Wei KH, Lower SE, Caldas IV, Sless TJS, Barbash DA, Clark AG. 2018. Variable rates of simple satellite gains across the *Drosophila* phylogeny. *Molecular Biology and Evolution* **35**:925–941. DOI: https://doi.org/10.1093/molbev/msy005, PMID: 29361128

Weiler KS. 2007. E(var)3-9 of *Drosophila melanogaster* encodes a zinc finger protein. *Genetics* **177**:167–178. DOI: https://doi.org/10.1534/genetics.107.076521, PMID: 17660546

Whitfield ZJ, Chisholm J, Hawley RS, Orr-Weaver TL. 2013. A meiosis-specific form of the APC/C promotes the oocyte-to-embryo transition by decreasing levels of the polo kinase inhibitor matrimony. *PLOS Biology* **11**:e1001648. DOI: https://doi.org/10.1371/journal.pbio.1001648, PMID: 24019759

Wolf G, de Iaco A, Sun MA, Bruno M, Tinkham M, Hoang D, Mitra A, Ralls S, Trono D, Macfarlan TS. 2020. KRAB-zinc finger protein gene expansion in response to active retrotransposons in the murine lineage. *eLife* **9**:e56337. DOI: https://doi.org/10.7554/eLife.56337, PMID: 32479262

Wolf D, Goff SP. 2009. Embryonic stem cells use ZFP809 to silence retroviral DNAs. *Nature* **458**:1201–1204. DOI: https://doi.org/10.1038/nature07844, PMID: 19270682

Yasuhara JC, DeCrease CH, Wakimoto BT. 2005. Evolution of heterochromatic genes of *Drosophila*. *PNAS* **102**:10958–10963. DOI: https://doi.org/10.1073/pnas.0503424102, PMID: 16033869

Yasuhara JC, Wakimoto BT. 2006. Oxymoron no more: the expanding world of heterochromatic genes. *Trends in Genetics* **22**:330–338. DOI: https://doi.org/10.1016/j.tig.2006.04.008, PMID: 16690158

Zdobnov EM, Tegenfeldt F, Kuznetsov D, Waterhouse RM, Simão FA, Ioannidis P, Seppey M, Loetscher A, Kriventseva EV. 2017. OrthoDB v9.1: cataloging evolutionary and functional annotations for animal, fungal, plant, archaeal, bacterial and viral orthologs. *Nucleic Acids Research* **45**:D744–D749. DOI: https://doi.org/10.1093/nar/gkw1119, PMID: 27899580

Zhang J, Schulze KL, Hiesinger PR, Suyama K, Wang S, Fish M, Acar M, Hoskins RA, Bellen HJ, Scott MP. 2007. Thirty-one flavors of *Drosophila* rab proteins. *Genetics* **176**:1307–1322. DOI: https://doi.org/10.1534/genetics.106.066761, PMID: 17409086

Zolotarev N, Fedotova A, Kyrchanova O, Bonchuk A, Penin AA, Lando AS, Eliseeva IA, Kulakovskiy IV, Maksimenko O, Georgiev P. 2016. Architectural proteins pita, Zw5,and ZIPIC contain homodimerization domain and support specific long-range interactions in *Drosophila*. *Nucleic Acids Research* **44**:7228. DOI: https://doi.org/10.1093/nar/gkw371, PMID: 27137890

Zolotarev N, Georgiev P, Maksimenko O. 2019. Removal of extra sequences with I-SceI in combination with CRISPR/Cas9 technique for precise gene editing in *Drosophila*. *BioTechniques* **66**:198–201. DOI: https://doi.org/10.2144/btn-2018-0147, PMID: 30987444