



Structural variants are enriched in deleterious visible phenotypes in *Drosophila*

Alejandra Samano, Matthew Musat, Mihir Junaghere, et al.

Genome Res. published online May 14, 2026

Access the most recent version at doi:[10.1101/gr.281434.125](https://doi.org/10.1101/gr.281434.125)

P<P Published online May 14, 2026 in advance of the print journal.

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).



To subscribe to *Genome Research* go to:
<https://genome.cshlp.org/subscriptions>

Research

Structural variants are enriched in deleterious visible phenotypes in *Drosophila*

Alejandra Samano, Matthew Musat, Mihir Junaghere, Asad Ahmad, Mehlum Ali, Sebastian Alves, Sreeram Pasupuleti, Jelisha Perera, Omar Saada, Brady Sabido, Trevor Smith, Sophie Walz, and Mahul Chakraborty

Department of Biology, Texas A&M University, College Station, Texas 77843, USA

Genome structural variants (SVs) comprise a sizable portion of functionally important genetic variation, yet many evade discovery using short reads. Although long-read sequencing can reveal hidden SVs, their contribution to organismal trait variation remains unclear. To address this gap, we investigate the molecular basis of 50 classical phenotypes in 11 *Drosophila melanogaster* strains using highly contiguous de novo genome assemblies generated with Oxford Nanopore Technologies long reads. These assemblies enable construction of a pangenome graph with nucleotide-resolution maps of SVs, including complex rearrangements such as the interchromosomal inverted duplication $Dp(2;4)ey^D$ and large tandem duplications at the *Bar* locus. We uncover new candidate causal mutations for 15 phenotypes and new molecular alleles for two mutations comprising tandem duplications, transposable element (TE) insertions, and indels. For example, the wing-vein phenotype *plexus* (*px^l*) links to a 1.5 kb partial tandem gene duplication, and the century-old *Curved* (*c^l*) wing phenotype links to a 7.5 kb DM412 retrotransposon disrupting the coding sequence of the muscle protein gene *Strn-Mlck*. We also identify a candidate intergenic enhancer for *Ablp^{eyD}*, supported by CRISPR-Cas9, and uncover eight SV alleles of previously identified causal genes, including uncharacterized SVs underlying the extensively studied white and yellow phenotypes. Overall, 67.4% of genes causing phenotypic changes harbor candidate SVs >100 bp, whereas only 28% are expected based on euchromatic SVs. Together, our results indicate that SVs are strongly enriched among this class of large-effect, deleterious visible phenotypes in *Drosophila*.

[Supplemental material is available for this article.]

Understanding the mutational basis of phenotypic differences between individuals or species is a fundamental puzzle in biology. Mutations that cause large or perceptible changes in phenotypes play an important role in adaptive evolution, agriculture, and medical genetics (Dittmar et al. 2016; Marian 2020). Although genome-wide association and linkage mapping studies have identified thousands of candidate variants, they often focus on small mutations, such as single-nucleotide polymorphisms (SNPs) and short indels in nonrepetitive regions, which are readily detected with short-read sequencing. Genome structural variants (SVs)—including duplications, transpositions, deletions, insertions, and inversions—alter more nucleotides and often exert stronger functional effects yet remain systematically undercharacterized owing to longstanding technical barriers (Alkan et al. 2011).

Population genomics studies across diverse species consistently show that SVs segregate at lower frequencies than SNPs, a pattern interpreted as evidence that SVs are more often deleterious and subject to stronger purifying selection (Chakraborty et al. 2019; Abel et al. 2020; Collins et al. 2020; Hämälä et al. 2021; Collins and Talkowski 2025; Samano et al. 2025). This interpretation is supported by case studies in which SVs are directly implicated in disease and severe phenotypes (Talkowski et al. 2012; Merker et al. 2018; Collins and Talkowski 2025), as well as by evidence that transposable element (TE) insertions disrupt gene function, alter regulation, and reduce fitness (Sorek et al. 2002; Hollister and

Gaut 2009; Lee and Karpen 2017). However, the inference that SVs are generally deleterious largely derives from allele-frequency patterns rather than from direct analyses of phenotypic effects. As a result, despite extensive population genetic evidence suggesting that SVs disproportionately underlie harmful phenotypic changes, this hypothesis has rarely been tested using a defined, phenotypically diverse set of traits.

Visible phenotypic changes provide a powerful model for uncovering genotype–phenotype relationships (Sax 1923; Koornneef et al. 1983; Long et al. 1995; Doebley 2004). Historically, visible phenotypes formed the foundation of classical genetics: Early geneticists used them to construct genetic maps and uncover fundamental principles of inheritance (Bateson et al. 1905; Sturtevant 1913), including landmark discoveries from Mendel’s pea experiments (Mendel 1996), Morgan’s *Drosophila* work (Morgan 1910; Bridges 1922; Muller 1927), and McClintock’s studies in maize (McClintock 1950). In particular, delineating mutations underlying morphological variation has helped illuminate evolution within and between species as well as elucidate genetic mechanisms of pathological conditions (Wild et al. 1997; Jeong et al. 2008; Chan et al. 2010; Imsland et al. 2012; Ghodsinejad Kalahroudi et al. 2014; Van’t Hof et al. 2016). A systematic inquiry into the roles of SVs and SNPs in the molecular basis of morphological changes within a species can help elucidate the molecular properties of mutations underlying these phenotypic changes.

Corresponding author: mahul@tamu.edu

Article published online before print. Article, supplemental material, and publication date are at <https://www.genome.org/cgi/doi/10.1101/gr.281434.125>. Freely available online through the *Genome Research* Open Access option.

© 2026 Samano et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

The model organism *Drosophila melanogaster* has an extensive collection of phenotype markers, many of which are deleterious to the organism. These mutations were collected based on their distinctive, easily scorable phenotypes rather than on prior knowledge of the underlying molecular basis, making them an unbiased set of large-effect variants whose causal mutations remain incompletely resolved. Although TEs are known contributors to several visible phenotypes (Green 1988; Sankaranarayanan 1988), no comprehensive investigation has quantified the contribution of SVs to this foundational genetic resource. Here, we investigate the molecular basis of 50 classical visible phenotypes by generating highly contiguous de novo genome assemblies for 11 *D. melanogaster* strains using Oxford Nanopore Technologies (ONT) long-read sequencing. We employed a pangenomic approach, constructing a graph-based representation of variation across the 11 assemblies relative to the ISO-1 reference. Using this comprehensive map of variation, we identify both known and previously uncharacterized candidate causal alleles, including duplications, TE insertions, complex rearrangements, and regulatory mutations.

Results

De novo genome assembly

We collected deep ONT long-read coverage (average 90×, genome size $G = 140$ Mb) for 11 genomes carrying 50 visible mutations: 45 are spontaneous, four are caused by X-ray irradiation, and one is caused by a chemical mutagen (Supplemental Table S1). Although ONT long reads have an average accuracy of 85%–95%, the high coverage used here (55–141×) yields high consensus accuracy across genomic positions (Supplemental Table S2; Sereika et al. 2022; Kolmogorov et al. 2023). Consistent with this, base-level quality-value (QV) estimates derived from ONT-only error profiling (see Methods) range from 44.6 to 52.3 across assemblies, indicating high consensus accuracy (Table 1). Uniform long-read coverage across assemblies suggests an absence of large-scale assembly errors (Supplemental Fig. S1). All chromosome arms in our assemblies are represented by highly contiguous sequences (median assembly contig $N50 = 23.1$ Mbp), which include all of the euchromatin and a portion of pericentromeric heterochromatin in single contigs (Fig. 1A; Supplemental Fig. S2). However, bal-

ancer chromosomes in two strains (1570 and 6027) are fragmented owing to challenges in assembling heterozygous sequences with long, error-prone reads (Li and Durbin 2024). The number of dipteran Benchmarking Universal Single Copy Orthologs (BUSCOs) in these genomes (complete BUSCO scores 98.6–99.5) and contiguity are comparable to the reference genome ISO-1 (complete BUSCO score 98.8), further underscoring the high quality of the assemblies (Table 1; Supplemental Table S3).

Landscape of genetic variation in 11 genomes

We identified mutations by comparing each genome assembly to the ISO-1 reference genome (Hoskins et al. 2015). To map genetic variation, we constructed a pangenome graph that captures all mutation classes (Hickey et al. 2023). Unlike traditional approaches that compare genomes to a single linear reference, a pangenome graph represents sequences as nodes in a network, enabling identification of both shared and unique variants, including SNPs, small indels, and SVs across all genomes (Eizenga et al. 2020; Sirén et al. 2021). Although we also performed pairwise genome alignments and read mapping to validate genotypes at candidate loci, the pangenome graph provides a comprehensive and accurate view of genetic variation. Focusing on euchromatic regions of the five major chromosome arms (2L, 2R, 3L, 3R, and X), we identified 53,337 small structural variants (SSVs; 11–100 bp) and 11,587 large structural variants (LSVs; >100 bp) across the 11 genomes. Of the LSVs, 7156 were associated with TEs. In total, 1.64 million SNPs were detected, of which 5.2% are located in coding exons.

We examined the minor allele-frequency distribution of LSVs, SSVs, nonsynonymous SNPs (nsSNPs), and synonymous SNPs (sSNPs). LSVs are significantly skewed toward lower frequencies than nsSNPs, a pattern likely driven by TEs (P -value $< 2.2 \times 10^{-16}$, χ^2 test between frequency distributions of LSVs and nsSNPs) (Fig. 1B). Many of the identified LSVs are TE insertions, and their low allele frequencies reflect both invasion dynamics and fitness consequences in *Drosophila*. Newly invading TE families produce large numbers of young insertions that initially segregate at low frequency (Kofler et al. 2015; Pianezza et al. 2025), and because most TE insertions are mildly to strongly deleterious, purifying selection further maintains these variants at low population frequencies (Charlesworth and Langley 1989; Petrov et al. 2011; Cridland et al. 2013). Although SSVs also showed a skew

Table 1. Assembly statistics

Strain name	Contigs			Scaffolds		Diptera complete BUSCOs	QV
	Length (Mb)	No. of contigs	N50 (Mb)	No. of scaffolds	N50 (Mb)		
1282	146.0	56	24.1	36	26.6	99.4	47.7
1349	151.2	49	24.2	28	29.4	99.4	46.5
156	154.3	36	23.8	18	29.3	99.4	49.0
1570	159.2	96	21.8	68	28.8	99.4	47.9
2969	150.1	34	23.8	17	29.7	99.4	44.6
5295	154.1	53	24.3	32	29.9	99.4	49.2
554	150.4	51	24.1	28	28.8	99.4	48.1
576	151.8	44	21.7	23	29.1	99.5	50.8
6027	151.4	69	23.9	49	27.9	98.6	48.6
620	151.9	51	24.3	29	27.6	99.4	52.3
662	153.5	45	18.5	21	29.0	99.4	48.2

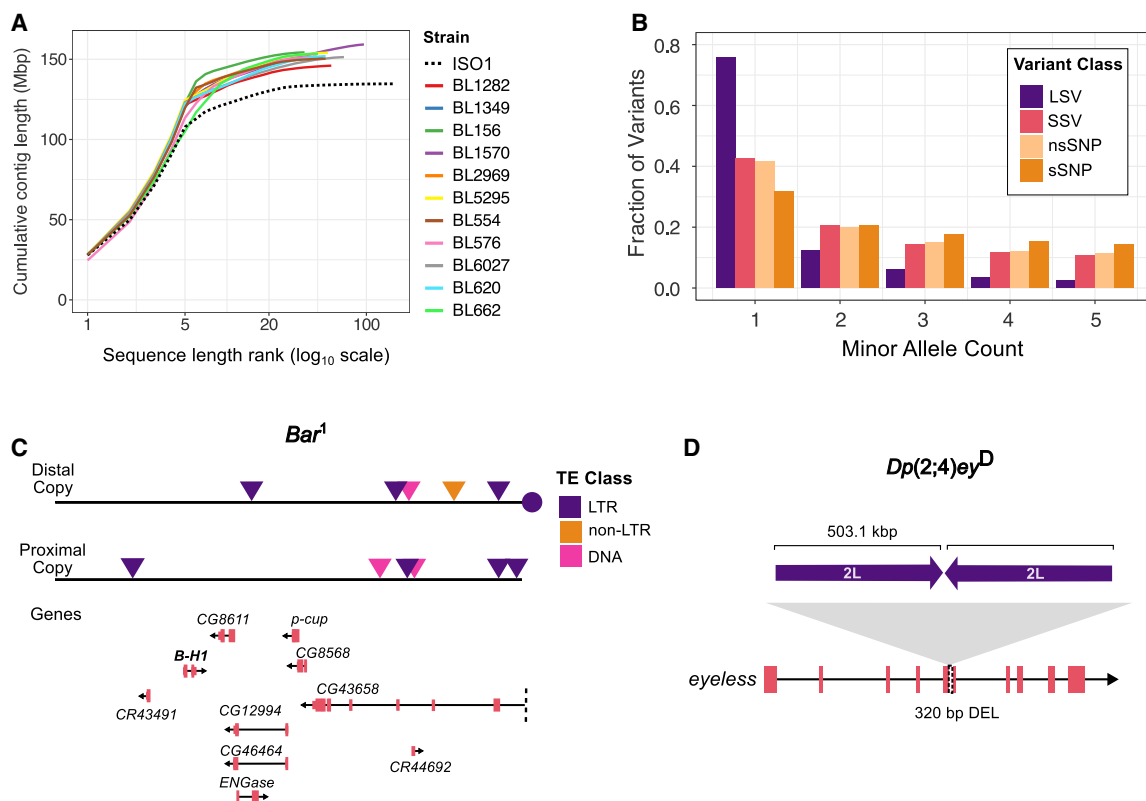


Figure 1. Genome assemblies reveal the structure and frequency of variants underlying visible phenotypes. (A) Contiguity plot comparing the ISO-1 release 6 contig-level reference assembly with the 11 de novo genome assemblies generated in this study. (B) Minor allele counts for large SVs (LSVs; >100 bp), small SVs (SSVs; 11–100 bp), nonsynonymous SNPs (nsSNPs), and synonymous SNPs (sSNPs). (C) Gene and TE content of the duplicated sequences at the *Bar* locus (ISO-1 coordinates X: 17,334,493–17,537,993), which contains nine complete genes and one truncated gene (CG4368). A roo element (circle) separates the two copies, consistent with the hypothesized mechanism of TE-mediated duplication. (D) The *Dp(2;4)ey^D* mutation is a translocation–duplication of a large segment of Chromosome 2L inserted into an exon of the *eyeless* gene on Chromosome 4, replacing 320 bp of coding sequence.

toward lower frequencies, their distribution is more similar to that of nsSNPs. These patterns align with previous population genomics studies, which suggest that SVs are subject to stronger purifying selection, likely owing to their more deleterious effects (Cridland et al. 2013; Chakraborty et al. 2019; Samano et al. 2025).

Assembly of large, complex SVs

Large and repetitive SVs are often difficult to resolve at the molecular level (Treangen and Salzberg 2011). To assess the capacity of our assemblies to characterize such mutations, we analyzed two visible mutations known to be linked to large duplications. Strain 2969 carries the *Bar¹* allele, an X-linked mutation that causes a slit-eye phenotype in males and homozygous females (Tice 1914). Classical genetic analyses (Sturtevant 1925; Muller 1936), followed by molecular cloning and sequencing (Tsubota et al. 1989; Miller et al. 2016), demonstrated that the *Bar¹* allele arose by unequal crossing over between homologous *roo* elements at cytological positions 16A1 and 16A7 on different chromatids. This results in the tandem duplication of the 16A1–7 sequence with a single *roo* element retained between the duplicates, marking the site of recombination (Supplemental Fig. S3). Consistent with this model, our assembly of strain 2969 captures both copies of the 203.5 kb region between positions 16A1 and 16A7. The breakpoints match prior studies and include a complete *roo* element be-

tween the two copies, supporting the TE-induced duplication model (Supplemental Fig. S4). The duplicated segment contains seven complete protein-coding genes, including *BarH1*, which is associated with the *Bar¹* eye phenotype (Kojima et al. 1993), as well as two long noncoding RNAs (lncRNAs), and one truncated gene (Fig. 1C). Notably, the two copies show substantial sequence divergence, including TE insertions unique to one copy (Fig. 1C). We also analyzed strain 662, which carries the *Dp(2;4)ey^D* mutation, an X-ray-induced translocation–duplication resulting in reduced or absent eyes (Hochman et al. 1964). Consistent with earlier cloning experiments (Kronhamn et al. 2002), we identified a 503.1 kb sequence from Chromosome 2L that was duplicated, with one copy inverted and inserted into the *eyeless* gene on the fourth chromosome (Fig. 1D; Supplemental Fig. S5). This translocation disrupts a coding exon and removes 320 bp of coding sequence from *ey*. In contrast to the *Bar¹* duplicates, the duplicated sequences of the *Dp(2;4)ey^D* mutation show fewer sequence differences.

Discovery of previously uncharacterized mutations

Strains carrying the *Bar¹* and *Dp(2;4)ey^D* mutations were included in this study owing to their known association with large genomic rearrangements. In contrast, the other strains were selected without prior consideration of the molecular basis of their visible

phenotypes. We first examined the molecular nature of mutations in genes previously linked to phenotypes, such as the *white* eye color mutation, as hidden SVs can mislead inferences about the causal mutations underlying phenotypic changes (Chakraborty et al. 2019; Ebert et al. 2021; Fadaie et al. 2021). For each phenotype, we identified candidate mutations as variants predicted to disrupt coding regions or located in putative regulatory elements (e.g., promoters, enhancers, UTRs) that could affect gene expression and were unique to strains with the phenotype. If the previously documented mutation was absent, we checked for another disruptive mutation in the same gene. Among the 50 phenotypes examined, we confirmed the previously characterized causal mutation for 33. We identified 18 uncharacterized mutations in total, including two that co-occurred with the previously reported causal mutation and one that contradicted the documented allele for that genotype reported on FlyBase (Öztürk-Çolak et al. 2024). For 15 phenotypes with no prior molecular characterization, we used our pangenome graph to identify candidate mutations.

The mutation *px*¹ is associated with increased wing veins at the wing margins and tips. However, the mutation responsible for this gene is yet to be characterized. We found a 1.5 kb tandem duplication that copies the 5' region of the first coding exon of *plexus* (*px*), with the duplicated segments separated by a 7.4 kb *DM412* retrotransposon (Fig. 2A). Long-read direct RNA-seq data reveal that this insertion produces a truncated, chimeric *px* transcript that includes 462 bp of transposon-derived sequence and terminates prematurely (Fig. 2B; Supplemental Fig. S6). This truncation removes the cysteine-rich and Q/QREKE motifs thought to be essential for Px function (Matakatsu et al. 1999). Loss of functional *plexus* is therefore likely to impair repression of wing-vein development, providing a mechanistic explanation for the ectopic wing-vein phenotype observed in *px*¹ mutant wings.

For two of the 15 uncharacterized phenotypes, the underlying genes had not yet been mapped in the *D. melanogaster* genome, so we combined our comprehensive variant map with additional genetic mapping experiments to identify candidate mutations

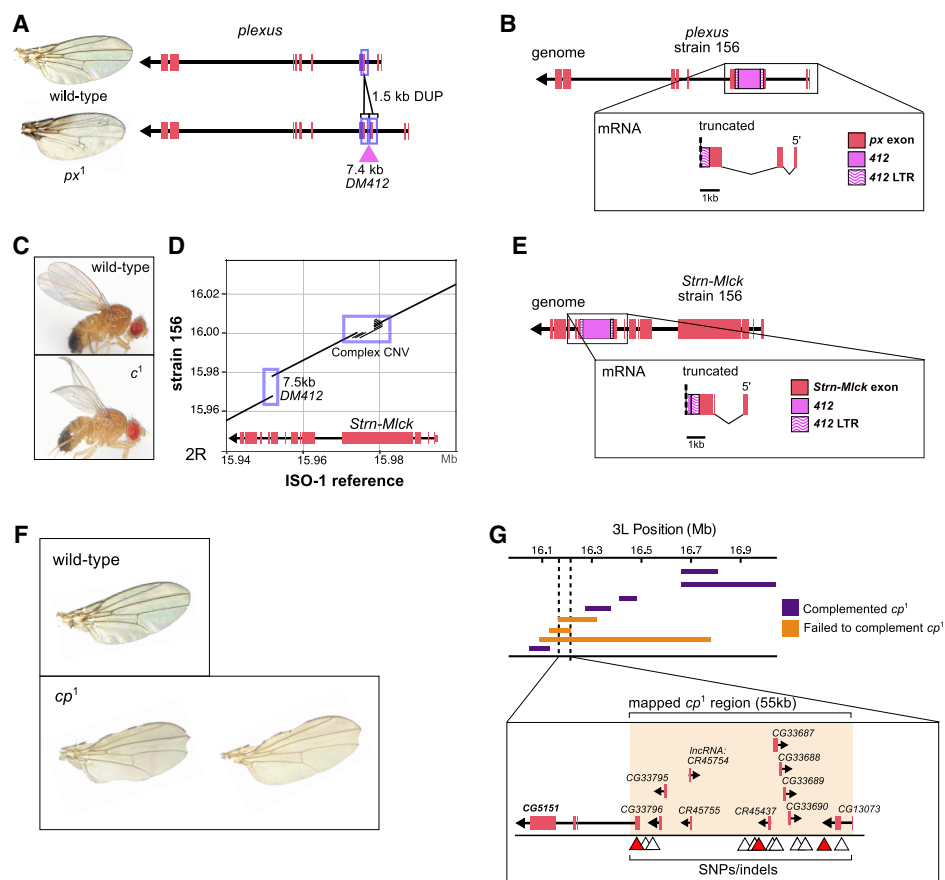


Figure 2. Structural variants underlying the *plexus* and *curved* visible phenotypes and their transcriptional consequences. (A) Wing phenotypes and genomic organization of the *plexus* (*px*¹) locus. Compared to the wild-type, *px*¹ flies show excess wing-vein patterning. The *px*¹ allele is associated with a 1.5 kb tandem duplication of a partial exon and a 7.4 kb insertion of the LTR retrotransposon, *DM412*. (B) Genomic structure and transcript model of *plexus* in strain 156, supported by ONT direct mRNA sequencing. The *DM412* insertion disrupts the gene, producing a truncated, chimeric transcript that includes transposon-derived sequence and lacks downstream exons. (C) Wing phenotypes of wild-type and *c*¹ flies. (D) Genome alignment at the *Strn-Mlck* locus underlying the *curved* (*c*¹) mutation. Relative to the ISO-1 reference, strain 156 carries a 7.5 kb *DM412* insertion and a complex copy-number variant in *Strn-Mlck*. (E) Genomic and transcript models of *Strn-Mlck* in strain 156, supported by direct mRNA sequencing. The *DM412* insertion truncates the transcript, resulting in loss of distal coding exons and incorporation of transposon sequence. (F) Wing phenotypes of wild-type and *cp*¹ flies. *cp*¹ mutants show variable wing notching patterns. (G) Complementation mapping with overlapping deletions narrowed the *cp*¹ candidate region to a 55 kb interval on Chromosome 3L (3L: 16,162,336–16,217,328; ISO-1 r6) containing *CG5151*, a gene implicated in wing development. Within this region, 15 SNPs and small indels unique to *cp*¹ strains were identified (triangles), with three candidates highlighted (red) based on their location within *CG5151* or overlap with wing disc ATAC-seq peaks.

for the phenotypes. The *curved* (c^1) mutation, which produces a characteristic curved wing phenotype (Fig. 2C), has not been mapped to the genome sequence (Supplemental Fig. S7). Previous studies based on deficiency maps narrowed the candidate region to 10 genes, with *Strn-Mlck* proposed as the most likely target because mutations in this gene produce wing phenotypes similar to those of *curved* (Rodriguez 2004; Kahsai and Cook 2018). In strain 156, we identified a 7.5 kb *DM412* LTR retrotransposon insertion within an exon of *Strn-Mlck*, together with nearby copy-number variation (CNV) in a region enriched for PEVK and SAIDE repeats in the protein (Fig. 2D). Long-read direct RNA-seq data demonstrate that the *DM412* insertion generates a truncated, chimeric *Strn-Mlck* transcript containing 673 bp of transposon-derived sequence before premature termination (Fig. 2E; Supplemental Fig. S8). Notably, this *DM412* insertion is absent from strains lacking the c^1 phenotype but is present in six additional c^1 strains (Supplemental Table S4), further supporting a causal link between the TE insertion and the curved wing phenotype. No other genes within the mapped interval harbor obvious disruptive mutations, implicating *Strn-Mlck* as a likely candidate underlying the c^1 phenotype. Definitive confirmation would require additional functional tests, such as targeted CRISPR-based disruption or allele replacement.

A clipped wing phenotype, *clipped* (cp^1) (Fig. 2F), was mapped to a 3-Mbp region on Chromosome 3L. We further narrowed the region by crossing flies carrying the cp^1 mutation with deficiency lines (see Methods), reducing it to a 55 kb window overlapping eight protein-coding genes and three lncRNAs (Fig. 2G; Supplemental Fig. S9; Supplemental Table S5). One of these genes, *CG5151*, has been shown to play an important role in wing development. Knockdown of the gene in the posterior imaginal wing disc results in wing notching similar to that observed in cp^1 mutants (Bageritz et al. 2019). To identify candidate mutations underlying the *clipped* phenotype, we used our pangenome graph to detect variants shared among strain 620 and six additional cp^1 mutant strains but absent from non-*clipped* strains. We identified 15 shared SNPs and small indels within the 55 kb interval harboring the cp^1 locus (Fig. 2G; Supplemental Table S6). Notably, one 7 bp indel disrupts a predicted transcription factor binding site (TFBS) in *CG5151*, whereas an additional indel and one SNP fall within wing disc ATAC-seq peaks, indicating localization to regions of open chromatin. Although any of these variants could potentially contribute to the *clipped* phenotype, additional functional experiments will be required to determine causality. Importantly, we detected no shared SVs within the mapped interval across cp^1 strains. We therefore classify *clipped* as being associated with small variants rather than SVs. Thus, six phenotypes among the 15 without a candidate mutation were associated with SVs, and we infer the rest are caused by SNPs or indels involving 10 or fewer nucleotides.

In addition to the 50 mutations analyzed in this study, we investigated a candidate mutation underlying the unmapped *Abnormal leg pattern* (*Ablp^{eyD}*) allele. This phenotype is associated with the X-ray-induced *Dp(2;4)ey^D* translocation duplication, which was specifically included to evaluate our ability to assemble and analyze large, repetitive mutations; therefore, it was not counted among the set of 50 phenotypes. *Ablp^{eyD}* was previously mapped to a 90 kb region on Chromosome 2L corresponding to the source of the sequence translocated onto Chromosome 4 in the *Dp(2;4)ey^D* mutation. Using the pangenome graph, we examined this 90 kb interval for candidate mutations shared between strain 662 and eight additional strains carrying the *Ablp^{eyD}* allele (see Methods). However, no SVs, SNPs, or indels within this region

were uniquely shared across the nine *Ablp^{eyD}* strains, suggesting that the phenotype may not arise from the same causal mutation in each background. Given this absence of a shared candidate variant, we examined nearby regulatory features within the mapped interval. Within this interval, strain 662 carrying *Ablp^{eyD}* has an 8 kb *roo* element insertion within a predicted TFBS for *paired* (*prd*) (Supplemental Fig. S10; MacArthur et al. 2009), a key regulator of segmental patterning in *Drosophila* development (Kilchherr et al. 1986). This TFBS is located between *drumstick* (*drm*) and *sister of odd and bowl* (*sob*), both members of the odd-skipped gene family involved in leg joint formation (Hao et al. 2003). However, deletion of the TFBS ($\Delta 886$) using CRISPR-Cas9 produced no detectable leg joint phenotype (Supplemental Figs. S10, S11). In contrast, an independent 11 bp deletion ($\Delta 11$) resulted in abnormal development of the first tarsal joint, the same segment affected in the *Ablp^{eyD}* phenotype (Supplemental Figs. S12, S13). Examination of a publicly available ATAC-seq data (Huynh et al. 2023) suggests that the $\Delta 11$ deletion overlaps an accessible chromatin peak in the wing and eye discs but not in the brain or ovary, suggesting a stage- and tissue-specific regulatory role (Supplemental Fig. S10). Notably, the *roo* element itself is absent from the eight other stocks carrying the *Ablp^{eyD}* allele (Supplemental Table S4). Although the *roo* element could epigenetically silence the regulatory site (Lee and Karpen 2017), thereby causing the tarsal phenotype, a different mutation at this regulatory site may underlie the *Ablp^{eyD}* phenotype.

Prevalence of SVs in deleterious phenotypes

We found that 66% (33/50) of markers are associated with LSVs and 6% (3/50) with SSVs (Fig. 3A; Table 2). The remaining phenotypes are associated with SNPs or small indels that disrupt protein-coding sequences or regulatory sequences. Of the phenotypes in our data set caused by spontaneous or natural mutations, 71% (32/45) are associated with SVs. Similar to previous observations that TEs cause many visible phenotypic changes in *D. melanogaster* (Sankaranarayanan 1988), 46% (23/50) of mutations are associated with TEs. Notably, 21 of the 23 TE insertions associated with visible phenotypes in our data set are LTR retrotransposons, consistent with prior genomic analyses showing that LTR elements represent a major and functionally impactful class of TEs in *D. melanogaster* (Kaminker et al. 2002; Bergman et al. 2006). We also find duplication CNVs, indels, and inversions underlying the mutations (Table 2; Supplemental Table S7).

To determine whether SVs are associated with visible phenotypic changes disproportionately compared with the average SV burden in euchromatic genes, we compared the prevalence of SVs in the 43 genes linked to the 50 phenotypic markers with their genome-wide distribution across the 10 genome assemblies in which these mutations were identified. We excluded strain 2969, which was included in this study to assess assembly performance on the *Bar* tandem duplication and carries no additional marker mutations. We performed a Monte Carlo simulation, randomly drawing 100,000 gene sets matched for gene length to our marker gene set to generate a null distribution for the expected number of SVs (Methods). Based on this null model, we would expect SVs in 12 genes by chance; however, we observed SVs in 29 of the 43 marker genes, representing an enrichment of 141.67% (P -value = 9.99×10^{-6}) (Fig. 3B). Because the analyzed genomes were generated by crossing multiple strains to combine marker mutations, their SV content may differ from the spectra of mutations in chromosomes segregating in natural populations. To address this, we

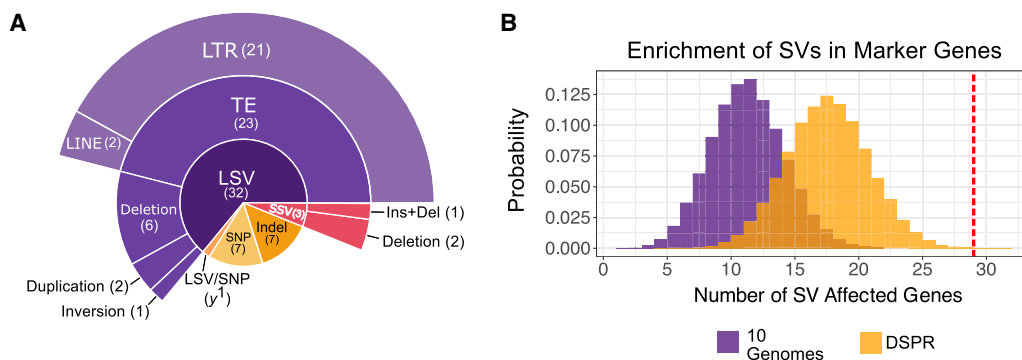


Figure 3. Enrichment of SVs among genes underlying visible phenotypes. (A) Distribution of candidate mutation classes identified across the 50 visible phenotypes analyzed in this study. Mutation classes include large structural variants (LSVs; >100 bp), small structural variants (SSVs; 11–100 bp), and transposable element (TE) insertions, with TEs further classified as long terminal repeat (LTR) or long interspersed nuclear element (LINE) retrotransposons. (B) Monte Carlo null distributions of the expected number of genes harboring at least one SV, generated by randomly sampling 43 genes matched for gene length from the genome-wide SV map across 10 assemblies (purple) and from the DSPR SV map (orange). The observed number of marker genes containing a candidate SV (29 of 43) is indicated by the red vertical line.

repeated the analysis using SV calls from the genome assemblies of 14 inbred strains collected from various geographical locations worldwide (Chakraborty et al. 2019). Given the abundance of SVs in this population sample, we expect to observe SVs in 18 genes by chance. The observation of 29 genes with SVs in the marker set thus represents a 61.1% enrichment (P -value = 7.90×10^{-4}) (Fig. 3B). To test whether this enrichment reflects a generally elevated mutation rate in these genes, we performed analogous length-matched enrichment analyses for nsSNPs and small indels. In contrast to SVs, marker genes show no enrichment for either mutation type, indicating that the excess of SVs is not attributable to broadly hypermutable or fragile genomic regions (Supplemental Fig. S14).

All 50 phenotypes examined in this study are associated with visible phenotypic changes, and 44 among these have deleterious effects on health and behavior (Supplemental Table S8). For instance, the vermilion eye color mutation (v^1) causes slow and irregular heart rates (Beasley and Dowse 2016); *white* eye color mutations (w^{1118} and w^1) are linked to defects affecting mobility, life span, and courtship behavior (Krstic et al. 2013; Xiao et al. 2017; Arimoto et al. 2020); Bristle mutants (*forked*) exhibit a reduced response to courtship sounds (Cosetti et al. 2008); and yellow body color mutant (*yellow*) males have lower mating success owing to reduced melanization of their sex combs (Massey et al. 2019). Among the 44 markers associated with deleterious fitness effects, 75% (33/44) are associated with an SV.

Allelic diversity underlying phenotypic changes

Similar phenotypes can often result from distinct mutations in the same gene (Schmidt et al. 2010; King et al. 2014; Chakraborty et al. 2019; GTEX Consortium 2020), although the prevalence of multiple alleles at loci underlying variation in deleterious organismal phenotypes remains unclear. To examine the diversity of molecular alleles underlying the deleterious phenotypes in our data set, we inspected the genes underlying the 12 phenotypes present in more than one strain. We identified multiple SV alleles linked to four phenotypes shared between strains, including three previously unknown.

Notably, two phenotypes are classic, well-characterized mutations: *yellow*, which results in a yellow body color, and *forked*, which results in forked bristles. Our analysis revealed previously

uncharacterized molecular diversity underlying these phenotypes. For example, strains 1570, 5295, and 6027 possess the γ^1 mutation, although only strains 1570 and 5295 carry the previously characterized start-codon-loss mutation caused by an SNP in the initiation codon of the γ gene (Geyer et al. 1990). The strain 6027 instead has a 622 bp insertion in a protein-coding exon of γ (Fig. 4A,B). This insertion disrupts the ORF and introduces a premature stop codon, likely leading to a truncated γ protein. Similarly, f^1 is thought to be caused by an intronic *gypsy* TE insertion (Hoover et al. 1993). Although strains 1570, 5295, and 6027 show the forked mutant phenotype, only strains 5295 and 6027 have the TE. The *f* gene in 1570 harbors a tandem duplication that copies exonic and intronic sequences, likely disrupting the ORF and producing a mutant phenotype (Fig. 4C,D).

We also identified a new molecular basis for mutations documented as alleles of the same gene. For instance, w^1 and w^{1118} are alleles of the *white* gene, both of which are associated with loss of eye pigmentation. The w^1 allele was previously linked to a TE insertion near the transcription start site, and we confirm this by finding a 4.7 kb *Doc* insertion in strains 1570 and 6027 (Fig. 4E,F). Although the strain 5295 was originally labeled as w^{1118} , our analysis revealed that it lacks the canonical deletion previously associated with this allele (Hazelrigg et al. 1984). Instead, comparison to additional w^{1118} assemblies indicates that the allele in strain 5295 more closely resembles w^1 , with an additional nested 3.5 kb *cop* element inserted within the *Doc* element (Fig. 4E). These insertions occur at the same genomic site, suggesting that recurrent TE insertions at this locus may give rise to the *white* eye phenotype. We also find evidence that distinct mutation types can have similar phenotypic effects. g^1 and g^2 are alleles of the eye color gene *garnet* (*g*). We show that g^1 is associated with an intronic 7.4 kb *blood* insertion, whereas g^2 is linked to an in-frame 3 bp deletion and an nsSNP in the last exon (Fig. 4G,H).

Additionally, two sets of mutations listed as distinct alleles in the stock genotypes may share the same molecular basis. ct^1 and ct^6 are both alleles of the *cut* gene and are associated with wing notches. The molecular basis of ct^6 has been characterized and is caused by a *gypsy* TE inserted between the *cut* promoter and a distant wing-margin enhancer (Dorsett 1993; Cai and Levine 1997). We found that strains 6027 and 1570, which are genotyped as ct^1 and ct^6 , respectively, both carry this same *gypsy* TE insertion at the same position, with no other unique disruptive mutations

Table 2. Fifty visible phenotypes and candidate mutations

Mutation	Strain name(s)	Mutation type	Origin	Reference	Description
<i>ac</i> ¹	6027	LSV/deletion	Spontaneous	Campuzano et al. 1985	2.8 kb deletion, located 5.8 kb upstream of <i>ac</i>
<i>AdC^{b-1}</i>	156	Indel	Spontaneous	Phillips et al. 2005	4 bp replaced by an 8 bp insertion in coding sequence; nonsynonymous SNP
<i>al</i> ¹	156	Indel	Spontaneous	Uncharacterized	10 bp deletion of coding sequence
<i>ast</i> ¹	1349	LSV/TE insertion	Spontaneous	Uncharacterized, but mutation type predicted by Higson et al. (1993)	Duplication of 6.4 kb region containing complete asteroid gene, copies separated by 33.8 kb insertion of <i>HMS-Beagle</i> , <i>gypsy</i> , and <i>roo</i> elements
<i>c</i> ¹	156	LSV/TE insertion	Spontaneous	Uncharacterized	7.5 kb <i>DM412</i> element inserted in coding sequence
<i>ca</i> ¹	576	SSV/deletion	Spontaneous	Ma et al. 2004	14 bp deletion of coding sequence
<i>car</i> ¹	6027	SNP	X-ray	Sevrioukov et al. 1999	Nonsynonymous SNP
<i>d</i> ¹	1349	LSV/TE insertion	Spontaneous	Giordano et al. 2003	7.4 kb <i>blood</i> element inserted in coding sequence
<i>cm</i> ¹	1282, 1570, 6027	LSV/TE insertion	Spontaneous	Mullins et al. 1999	9.2 kb <i>roo</i> element inserted in coding sequence
<i>cp</i> ¹	620	SNP/indel	Spontaneous	Uncharacterized	Potentially a 7 bp indel near a regulatory site in <i>CG5151</i>
<i>ct</i> ¹	6027	LSV/TE insertion	Spontaneous	Uncharacterized	7.3 kb <i>gypsy</i> element inserted in wing enhancer
<i>ct</i> ⁶	1570	LSV/TE insertion	Spontaneous	Tchurikov et al. 1989; Dorsett 1993	7.3 kb <i>gypsy</i> element inserted in wing enhancer
<i>cu</i> ¹	576	Indel	Spontaneous	Grönke et al. 2009	Substitution of 2 bp at splice acceptor site
<i>Diap1^{th-1}</i>	576, 1349, 620	SNP/indel	Spontaneous	Uncharacterized	576 and 620 do not share any unique mutations in or near the gene; both have indels/SNPs in close proximity located in an intron, a putative enhancer region
<i>dpp^{d+ho}</i>	1349	LSV/deletion	Spontaneous	Masucci et al. 1990	3066 bp deletion, 22.7 kbp from 3' end of <i>dpp</i>
<i>dpy^{ov1}</i>	156, 1349	LSV/TE insertion	Spontaneous	Carmon et al. 2010	7.4 kb <i>blood</i> element intronic insertion
<i>dy</i> ¹	1570	LSV/deletion	spontaneous	DiBartolomeis et al. 2002	358 bp deletion, 1.8 kb from start of <i>dy</i> ; 464 bp deletion, 2.7 kb from start of <i>dy</i>
<i>e</i> ⁵	554, 576	SNP	Spontaneous	Uncharacterized	Nonsynonymous SNP
<i>ed</i> ¹	1349	LSV/TE insertion	Spontaneous	Uncharacterized	7.8 kb intronic insertion of <i>gypsy</i> and <i>Stalker</i> elements

(continued)

Table 2. Continued

Mutation	Strain name(s)	Mutation type	Origin	Reference	Description
<i>f</i> ¹	1570	LSV/duplication	Spontaneous	New allele	2.7 kb duplication of coding exon and intronic sequence
	5295, 6027	LSV/TE insertion		Hoover et al. 1993	7.4 kb <i>gypsy</i> element intronic insertion
<i>g</i> ¹	6027	LSV/TE insertion	Spontaneous	Uncharacterized, but mutation type predicted by Lloyd et al. (1999)	7.4 kb <i>blood</i> element intronic insertion
<i>g</i> ²	1570, 5295	SNP/indel	Spontaneous	Uncharacterized, but mutation type predicted by Lloyd et al. (1999)	3 bp in-frame deletion, nonsynonymous SNP
<i>hry</i> ¹	576	LSV/TE insertion	Spontaneous	Hooper et al. 1989	7.4 kb <i>gypsy</i> element insertion, 4.8 kb from start of <i>hry</i>
<i>in</i> ¹	620	SNP	Spontaneous	Uncharacterized	Nonsynonymous SNP
<i>kni</i> ^{rt-1}	620	LSV/deletion	Spontaneous	Lunde et al. 2003	253 bp deletion of regulatory sequence, 11.7 kb from start of <i>kni</i>
<i>m</i> ¹	6027	SSV/indel	Spontaneous	Roch et al. 2003	11 bp deletion of coding sequence
<i>m</i> ^{74f}	1282	Indel	Ethyl methanesulfonate	Uncharacterized	1 bp deletion, 3 bp intronic insertion
<i>oc</i> ¹	1570	LSV/inversion	X-ray	Spradling et al. 1979; Spradling and Mahowald 1981	15.4 kb inversion, including <i>oc</i> gene
<i>p</i> ^p	620	SNP	Spontaneous	Falcón-Pérez et al. 2007; Syrzycka et al. 2007	Nonsynonymous SNP
<i>pn</i> ¹	1570, 6027	LSV/TE insertion	Spontaneous	Frolov et al. 1994; Timmons and Shearn 1996	7.5 kb <i>DM412</i> element inserted in coding exon
<i>pr</i> ¹	156	LSV/TE insertion	Spontaneous	Kim et al. 1996	7.5 kb <i>DM412</i> element intronic insertion
<i>px</i> ¹	156	LSV/Duplication	Spontaneous	Uncharacterized	Partial duplication of coding exon with copies separated by 7.4 kb <i>DM412</i> element
<i>ras</i> ²	1570	LSV/TE insertion	Spontaneous	Location and size approximated by Nash et al. (1994)	5 kb <i>blastopia</i> intronic insertion
<i>ras</i> ⁴	6027	LSV/TE insertion	Spontaneous	Uncharacterized, might be misgenotyped, same mutation as <i>ras</i> ²	5 kb <i>blastopia</i> intronic insertion
<i>rb</i> ¹	6027	LSV/TE insertion	Spontaneous	Uncharacterized	7.8 kb insertion of <i>gypsy</i> and <i>Stalker</i> elements into coding exon
<i>ru</i> ¹	576	SNP	Spontaneous	Yu et al. 2010	Premature stop codon
<i>sc</i> ¹	6027	LSV/TE insertion	Spontaneous	Campuzano et al. 1985	1.6 kb <i>gypsy</i> , 7 kb from start of <i>sc</i>
<i>sd</i> ¹	1282	SSV/deletion	X-ray	Uncharacterized	22 bp deletion of coding sequence
<i>sn</i> ³	5295, 6027	LSV/deletion	Spontaneous	Paterson and O'Hare 1991	292 bp deletion from 5' UTR
<i>speck</i> ¹	156	LSV/TE insertion	Spontaneous	Chakraborty et al. 2019; Spana et al. 2020	7.5 kb <i>DM412</i> element inserted into intron, present in ISO-1 reference
<i>sr</i> ¹	554, 576	SNP	Spontaneous	Uncharacterized	576 and 554 share four SNPs in 5' UTR and three intronic SNPs
<i>st</i> ¹	576, 620	LSV/TE insertion	Spontaneous	ten Have et al. 1995	7.4 kb <i>DM412</i> element inserted in coding exon

(continued)

Table 2. Continued

Mutation	Strain name(s)	Mutation type	Origin	Reference	Description
<i>sv^{de}</i>	662	LSV/TE insertion	Spontaneous	Fu et al. 1998	5 kb Tirant 1.6 Kb from 5' end of gene
<i>Ubx^{bxc-1}</i>	554	LSV/TE insertion	Spontaneous	Bender et al. 1983; Castelli-Gair and García-Bellido 1990	7.5 kb <i>DM412</i> element insertion into intronic regulatory element
<i>Df(1)os-o(upd1^{os-o}, upd3^{os-o})</i>	1570	LSV/deletion	X-ray	Wang et al. 2014	3.1 kb deletion upstream of <i>upd</i> ³
<i>Ab(1)os-s(upd1^{os-s}, upd3^{os-s})</i>	1282	SNP/Indel	Spontaneous	Uncharacterized	Nonsynonymous SNP and 3 bp in-frame deletion of coding sequence in <i>upd</i> ³
<i>v¹</i>	1570, 5295, 6027	LSV/TE insertion	Spontaneous	Searles et al. 1990	7.5 kb <i>DM412</i> element inserted in 5' UTR
<i>w¹</i>	1570, 6027	LSV/TE insertion	Spontaneous	Driver et al. 1989	4.7 kb <i>Doc</i> element insertion, 17 bp from start of <i>w</i>
<i>w¹¹¹⁸</i>	5295	LSV/TE insertion	Spontaneous	Hazelrigg et al. 1984	4.7 kb <i>Doc</i> element insertion, 4 kb <i> copia</i> within <i>Doc</i> element, 17 bp from start of <i>w</i>
<i>γ¹</i>	6027	LSV/insertion	Spontaneous	New allele	622 bp insertion in coding exon, introduces premature stop codon
	1570, 5295	SNP		Geyer et al. 1990	Start codon loss caused by SNP

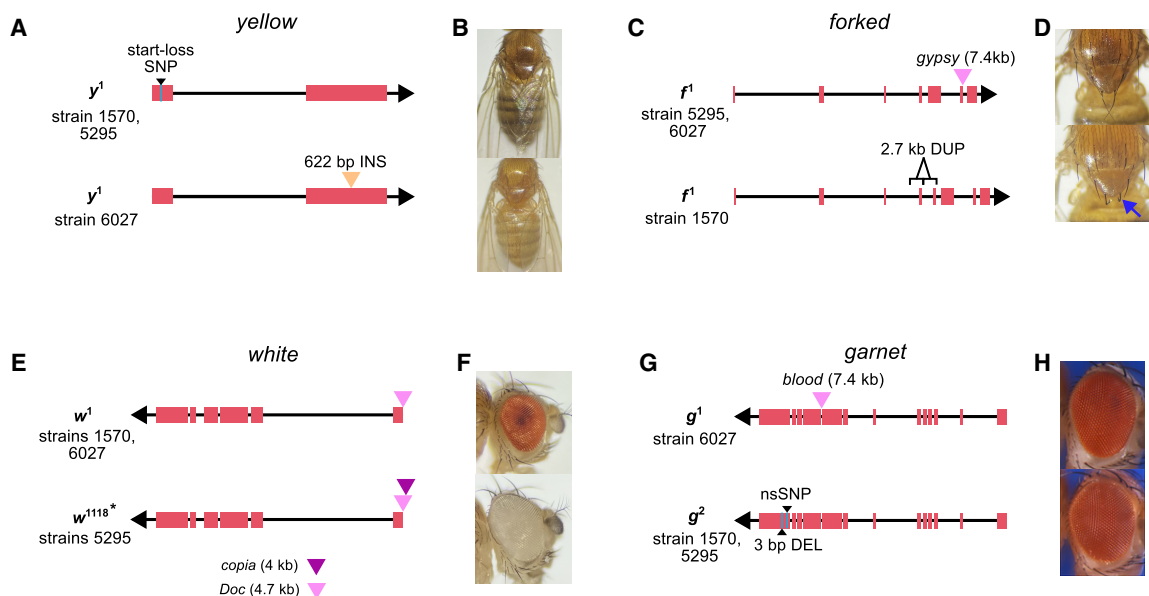


Figure 4. Molecular diversity underlying classical visible phenotypes. (A) Two strains exhibiting the yellow body color phenotype carry the previously characterized SNP that disrupts the start codon. In contrast, strain 6027 harbors a 622 bp insertion in the second exon, resulting in a frameshift and a premature stop codon. (B) Phenotype images of wild-type (top) and *y*¹ mutant (bottom) body color. (C) The *forked* (*f*¹) allele was previously associated with a TE insertion; however, strain 1570 lacks this insertion and instead carries a 2.7 kb tandem duplication encompassing a complete exon. (D) Phenotype images of wild-type (top) and *f*¹ mutant forked bristles (bottom, blue arrow). (E) Two alleles of the *white* gene are associated with TE insertions near the transcription start site. A 4.7 kb *Doc* element insertion is present in strains 1570 and 6027, and a nested *copia* element within the *Doc* insertion is observed in strain 5295, consistent with recurrent insertions at the same genomic site. (F) Phenotype images of wild-type (top) and *w*^{1118*} mutant (bottom) eye color. (G) Two alleles of *garnet* involve either an intronic *blood* element insertion (*g*¹) or a nonsynonymous SNP with an in-frame deletion in the final coding exon (*g*²). (H) Phenotype images of wild-type (top) and *g*¹ mutant eye color, obtained from FlyBase (FBrf0220532).

in or near the *cut* gene. Likewise, *ras*² and *ras*⁴ are alleles of the eye color gene *raspberry*. *ras*² is linked to a 5 kb *blastopia* insertion, and both strains 1570 and 6027 carry this same TE insertion at the same site, again with no other clear disruptions in the gene. These findings suggest that recorded genotypes for these stocks may be incorrect or that the alleles may share the same underlying mutation. The presence of phenotypic differences despite identical disruptions could also mean that additional genetic modifiers, located outside of the previously mapped gene regions, may influence the observed phenotypes. Variation in piRNA-mediated silencing across genetic backgrounds may also modulate TE activity or nearby gene expression, providing another potential source of phenotypic divergence (Kelleher et al. 2012; Sentmanat and Elgin 2012).

Discussion

SVs underlie both deleterious and adaptive phenotypic changes and have been hypothesized to account for a portion of the missing causal variants in complex trait variation (Manolio et al. 2009; Eichler et al. 2010). However, their overall impact on phenotypic variation, including fitness-related traits, remains poorly defined. In *D. melanogaster*, early studies showed that many spontaneous visible mutations involve TEs (Green 1988; Sankaranarayanan 1988), but these analyses focused on phenotypes already known to be associated with TEs. As a result, the broader contribution of SVs—both TE and non-TE—has remained unclear. By constructing a comprehensive variant map across 11 strains carrying 50 classical visible mutations, we show that SVs are associated with a disproportionately large fraction of phenotypic changes compared with SNPs, including eight previously undetected SVs. These find-

ings mirror growing evidence from model organisms (Chakraborty et al. 2019), agriculture (Chia et al. 2012; Alonge et al. 2020), livestock (Li et al. 2024; Yang et al. 2024), and humans (Abel et al. 2020), indicating that hidden SVs likely underlie a wide range of traits, from physiology to behavior.

Although the variants underlying these visible phenotypes were not sampled from natural populations, their nature—often single, large-effect mutations—is similar to phenotypic differences observed in diverse biological contexts, including morphological and behavioral divergence across taxa and domestication traits in crops and livestock (Naisbit et al. 2003; Shapiro et al. 2004; Hoekstra et al. 2006; Andersson 2013; Wills et al. 2013). Our data set consists mostly of spontaneous mutations identified in laboratory stocks; however, these mutations serve as useful models for understanding how large mutations contribute to phenotypic variation. Because purifying selection is weakened in laboratory conditions, owing to both controlled environments and small effective population sizes, deleterious mutations can persist longer in laboratory stocks than in natural populations. This context helps explain why these large-effect alleles remain observable in laboratory strains despite their likely negative fitness effects in wild populations.

Consistent with population genetic analyses showing that SVs, particularly TEs and duplications, segregate at low frequencies (Chakraborty et al. 2019; Collins and Talkowski 2025), we infer that 44 of the phenotypes examined in this study are deleterious, with the majority (75%) associated with SVs. Although our conclusion of enrichment refers specifically to this historically unbiased set of classical visible phenotypes, and not to all trait categories genome-wide, deleterious recessive mutations, such as these marker phenotypes, often segregate at low frequencies in natural

populations. Because the shape of the allele-frequency distribution of SVs in our strains closely resembles that observed in natural populations (Chakraborty et al. 2019), the enrichment of SVs among deleterious phenotypes could provide a biological basis for the population genetic inferences that SVs often exert stronger harmful effects than nonsynonymous SNPs. Additionally, our data show that similar deleterious phenotypic changes can result from different SV alleles, with the proportion of phenotypes (3/12) exhibiting allelic heterogeneity comparable to that of *D. melanogaster* genes (3/12). The prominent role of hidden and multiallelic SVs in deleterious phenotypes highlights their potential contribution to disease genetics and unexplained heritability in complex traits (Manolio et al. 2009; Eichler et al. 2010), as well as to inbreeding depression and extinction risk in small populations experiencing weak natural selection (Rogers and Slatkin 2017).

SVs can influence gene structure and function through diverse mechanisms. Although duplications of complete genes have long been recognized as drivers of phenotypic variation and adaptation (Hughes 1994; Chakraborty and Fry 2015; Cardoso-Moreira et al. 2016), genomic rearrangements involving partial genes can also have profound effects. Classic examples include the origin of the *jingwei* gene in *Drosophila* (Long and Langley 1993) and disease-causing exon duplications in humans (White et al. 2006). Consistent with these patterns, we show that partial gene duplications and noncoding SVs—including variants affecting UTRs, introns, and intergenic regions—account for more than one-half (27/50) of the candidate mutations identified here, reinforcing the importance of noncoding variation in shaping phenotypes (Maurano et al. 2012; Alsheikh et al. 2022; Schipper and Posthuma 2022).

Although reference-based methods are widely used, they often fail to detect complex or large SVs (Chakraborty et al. 2018). In contrast, de novo assemblies and pangenome approaches can capture the full size and structure of complex variants, as illustrated by tandem duplications and TE-associated rearrangements such as *plexus* and *asteroid*. Together, these examples show that assembly- and pangenome-based approaches can reveal functionally important SVs, including novel molecular alleles in highly studied genes, that would be hidden with read mapping alone and thereby lead to misleading inferences about the mutational basis of phenotypic variation.

Beyond variant discovery, long-read assemblies enable improved genome annotation. *D. melanogaster*, despite being one of the best-annotated metazoan genomes, still contains genes that have not been mapped to the genome sequence (Dean et al. 2022). We examined three such cases—*curved*, *clipped*, and *Ablp*—and identified genes and a putative regulatory element likely responsible for the associated mutant phenotypes. These results highlight how modern genomic approaches can close persistent gaps in the genotype–phenotype map, even in *D. melanogaster*. Inspired by A.H. Sturtevant’s pioneering undergraduate work in genetic mapping, we integrated this model in a resource called genomics and long reads education (GALORE) that embodies the same spirit of discovery in the modern classroom, while also supporting more advanced training and research applications, such as recombination landscape inference (see Data access).

The discovery of the *Bar*¹ mutation in *Drosophila* provided the earliest evidence supporting the role of genome structural changes in phenotypic variation (Hurles et al. 2008). Since then, SVs have been implicated in several Mendelian and complex diseases as well as adaptations (Merker et al. 2018; Quan et al. 2021; Collins and

Talkowski 2025). However, the extent to which SVs contribute to phenotypic variation remains unclear, partly because of the challenges of detecting comprehensive SVs. Comparative genomics using highly contiguous genome assemblies have largely solved that problem, although our understanding of the contribution of SVs in phenotypic variation remains incomplete. Consistent with previous findings of enrichment of SVs in candidate genes in QTL mapping experiments, our results suggest a disproportionate role of SVs in spontaneous, deleterious changes in phenotypes with both Mendelian and complex genetic bases (Chakraborty et al. 2019). Thus, our results further show that SVs can act as rare, large-effect alleles and may account for undetected causal mutations underlying variation in Mendelian and complex traits, particularly those affecting organismal fitness.

Methods

Fly stocks and DNA extraction

We obtained the *D. melanogaster* stocks from the Bloomington Stock Center (ordered on December 10, 2023; received on December 18, 2023). The presence of visual markers listed on the stock center website was verified for each strain. We collected 150 females from each stock and extracted high-molecular-weight DNA using the method described by Chakraborty et al. (2016). Briefly, we flash-froze the flies in liquid nitrogen and ground them into fine powder using a mortar and pestle. We extracted DNA from the fly powder using the Qiagen Blood and Cell Culture DNA Midi Kit and then spooled the DNA at the final stage using a glass hook.

Library preparation and sequencing

We prepared the ONT library for each strain following the manufacturer’s ligation kit protocol. The initial concentration and total volume of DNA for each sample are provided in Supplemental Table S2. For high-molecular-weight DNA from strains 2969, 5295, 576, and 1349, we used the Pacific Biosciences (PacBio) short-read eliminator XL kit to remove DNA fragments <40 kb in length. DNA was end-repaired using the NEBNext companion module for ONT ligation sequencing kit (New England Biolabs), followed by adaptor ligation with the ONT duplex-enabled ligation sequencing kit v14. Libraries were sequenced on R10.4.1 flow cells using a MinION Mk1B for 72 h.

Base calling and assembly

We performed base calling using a Dorado duplex base calling model on a laptop with 64 GB of memory and 2 TB of SSD storage. Although each run produced a small proportion of duplex reads, we did not assemble them separately because of their low coverage (Supplemental Table S3). Raw ONT reads were filtered using Chopper (De Coster and Rademakers 2023), keeping only reads with an average Phred quality score greater than 10 and lengths greater than 10 kbp. We generated draft assemblies for each genome using Hifiasm v0.25.0 (Cheng et al. 2021). Microbial contigs in the draft assembly were identified using BlobTools v1.1 (Laetsch and Blaxter 2017). For input to the BlobTools analysis, we generated a taxonomic annotation file by aligning contigs against the NCBI Nucleotide database (<https://www.ncbi.nlm.nih.gov/nucleotide/>) (downloaded on April 23, 2024) using BLASTN. The read-mapping input was generated by mapping the quality-filtered ONT reads to the draft assembly using minimap2 v2.26 (Li 2017). Only contigs classified as “Arthropoda” or “no-hit” were retained. Cleaned, draft assemblies were polished using Medaka

(<https://github.com/nanoporetech/medaka>) Polished contigs were scaffolded using *mscaffolder* (Chakraborty et al. 2018) with release 6.49 of the ISO-1 genome assembly (Hoskins et al. 2015) as the reference.

Strain 662 carries the *ey^D* mutation, which involves a translocation duplication of a segment from Chromosome 2L onto the fourth chromosome. Hifiasm was unable to resolve this complex SV, so we used Flye v2.9.3 to generate and inspect the repeat graph (Kolmogorov et al. 2019). The repeat graph was visualized using Bandage (Wick et al. 2015), which indicated that the duplicated sequence may have been collapsed into a single copy (Supplemental Fig. S5). To recover the entire structure, we manually expanded the collapsed region by exporting the path as a FASTA sequence. We confirmed the new breakpoints by mapping reads back to the duplicated sequence (Supplemental Fig. S5).

Assembly quality assessment

To evaluate assembly quality, we first identified potential large-scale misassemblies by remapping long reads to the contig assemblies using minimap2 v2.26 (Li 2017) and visually inspecting regions with abnormal coverage profiles. Assembly contiguity was quantified using QAST v5.0.2 (Gurevich et al. 2013), based on standard summary statistics (contig N50, contig number, total assembly size). Completeness was assessed using BUSCO v5.7.1 (Simão et al. 2015) with the Diptera ortholog database downloaded on July 8, 2024 (https://busco-data.ezlab.org/v5/data/lineages/diptera_odb10.2024-01-08.tar.gz).

Assembly consensus accuracy (QV) was estimated by mapping raw ONT reads back to each assembly and calling variants with PEPPER-Margin-DeepVariant using the model trained on ONT R10 sequencing data (Shafin et al. 2021). We treated homozygous nonreference (1/1) variants with filter=pass and genotype quality > 10 as high-confidence discrepancies between the reads and the assembly. The number of such sites was divided by the assembly size (in base pairs) to obtain an estimated error rate, which was converted to a Phred-scaled quality value:

$$QV = 10 \log_{10} (\text{number of errors/assembly size}).$$

This ONT-only, read-mapping-based approach is consistent with prior methods used to assess consensus accuracy in long-read genome assemblies when Illumina data are unavailable (Koren et al. 2017; Solares et al. 2018).

Repeat annotation and SV calling

We annotated repeats in each assembly using RepeatMasker v4.1.2 (Smit et al. 2013). To identify SVs, we combined whole-genome alignment and read-mapping approaches. We utilized the Minigraph-Cactus pipeline (Hickey et al. 2023) to construct a pan-genome graph, encompassing all forms of genetic variation across the 11 genomes. From the ISO1-based VCF file, we identified mutations located within annotated genes that were unique to strains carrying marker mutations. Additionally, each assembly was aligned to the ISO-1 reference genome, specifically the major chromosome arms and the dot chromosome (X, 2L, 2R, 3L, 3R, 4), using MUMmer v4 (Marçais et al. 2018). Structural differences between assemblies were then classified as insertions, deletions, duplications, or inversions using SVMU (Chakraborty et al. 2019). Finally, we mapped the ONT reads to the ISO-1 reference using minimap2 v2.26 (Li 2017) and inspected them in the Integrative Genomics Viewer (IGV) (Thorvaldsdóttir et al. 2013) to confirm that the read data supported SVs detected by whole-genome alignment-based methods. Strains 1570 and 6027 carry an X Chromosome balancer, so the assembled X Chromosomes are

highly fragmented. To identify mutations linked to X-Chromosome marker genes in these strains, we used Sniffles v2.6.1 (Smolka et al. 2024) to detect SVs from reads mapped to the ISO-1 reference. Candidate SVs were similarly verified in IGV.

SNP and indel calling

SNPs and indels were called using PEPPER-Margin-DeepVariant (Shafin et al. 2021), which uses models trained on ONT R10 data and provides higher base-level accuracy than that of graph-derived calls from Minigraph-Cactus. A comparison of small variant calls (Supplemental Fig. S15) demonstrated that although both tools share a high concordance of variants, PMDV captures a large number of high-quality SNPs and small indels, particularly heterozygous sites, that are absent or collapsed in the pangenome graph. Only PMDV variant calls with a “pass” filter were included in downstream analyses. Individual strain VCF files were normalized and merged with BCFtools v1.19 (Danecek et al. 2021), and variant effects were predicted using SnpEff v5.2 (Cingolani et al. 2012).

CRISPR genome editing

To examine the functional significance of the predicted regulatory region for the unmapped *Ablp* gene, we selected two guide RNAs (gRNA1: 5'-TGATTGCGAAGAAACCTCTG-3'; gRNA2: 5'-TTGACAGGCAACTGGCGATC-3') flanking the predicted enhancer site using the CRISPOR web tool (Concordet and Haussler 2018). gRNA2 is positioned 84 bp downstream from the *roo* insertion site and near the 3' end of the predicted *paired* binding site; gRNA1 targets the opposite flank, 806 bp from the *roo* insertion site, to enable full-site deletion. As the gRNA2 sequence is located very close to the 3' end of the predicted *Paired* binding site, deletions resulting from CRISPR-Cas9-mediated DNA breaks may overlap and disrupt the regulatory site. The synthesized gRNAs (Synthego) were incubated with the Cas9 enzyme (Synthego) to form RNPs and then injected into embryos of the BDSC strain 54591 (Port et al. 2014). The presence of the gRNA sequence was verified by aligning the putative regulatory sequence from the ISO-1 reference sequence to the genome sequence of strain 54591 available at GitHub (<https://github.com/chakrabortymlab/DLPD>). Embryo injections were performed by GenetiVision. Surviving females were individually crossed to single males and allowed to mate for 5 days. Following the observation of larval activity, females (G_0) were genotyped by PCR and amplicon sequencing to assess CRISPR-mediated deletions.

To genotype individuals for CRISPR-mediated deletion alleles, we isolated genomic DNA from individual flies using the Monarch genomic DNA purification kit (New England Biolabs) according to the manufacturer's genomic DNA extraction from insects protocol. The genomic DNA was amplified using Q5 polymerase with the primers 5'-TCAGCGAGTACAACCTCAGCA-3' and 5'-TTGTTGTCGCTGGAGATTCTGA-3'. The PCR amplicons were purified using the Monarch PCR & DNA cleanup kit and sequenced with Oxford Nanopore (Plasmidsaurus). The sequencing reads were mapped to the 54591 assembly using minimap2 v2.26 (Li 2017). The read alignments in BAM format were viewed in IGV to examine the CRISPR-induced deletions. We recovered two deletion alleles: a full deletion between the two gRNAs and an 11 bp deletion at the site of the second gRNA. The presence of only reads carrying a single deletion allele was considered as evidence for homozygosity of the deletion, whereas the presence of different edited alleles or a mix of edited and unedited alleles was considered as evidence of heterozygous mutations.

F_1 males and females from G_0 females homozygous for the deletion were crossed and genotyped. Because of the unknown

genotype of the G_0 male, not all F_1 males and females were homozygous for the deletion. Thus, we isolated F_2 males and females homozygous for the deletion and inspected their leg phenotypes.

Deletion mapping of *clipped* gene

Strain 620 carries a wing mutation, cp^1 , a mutant allele of the *clipped* gene, whose precise genomic location has not yet been determined. According to unpublished data on FlyBase (<https://flybase.org/reports/FBfrf0198635>), *clipped* was predicted to reside within a 3.12 Mb region between the genes *asf1* and *st*. To refine the location of the *clipped* gene, we performed deletion mapping using 14 lines, each carrying a chromosomal deletion spanning different portions of the predicted interval (Supplemental Table S5). These deletion stocks were crossed to strain 620 as well as to strain 466, which also carries the cp^1 allele, and the F_1 s were examined for the wing notching phenotype associated with cp^1 .

Eleven deletions complemented the cp^1 phenotype, indicating that the mutation lies outside the regions deleted in those lines. Three deletions failed to complement, suggesting that these *overlapping* deletions uncover the cp^1 mutation. From the overlap of the three noncomplementing deletions, we defined a minimal candidate interval of 54,992 bp on Chromosome 3L (Supplemental Fig. S7).

Enrichment of SVs in mapped genes

To determine whether the number of marker genes with candidate SVs was significantly higher than the genome-wide distribution of SVs, we generated a null distribution from our SV map across 10 genomes. We excluded SVs from strain 2969 from this analysis, as we did not include the *Bar*¹ mutation among our 50 markers because it is known to be an SV. We considered SVs >100 bp and identified euchromatic protein-coding genes located within 1000 bp of any SV. To construct a null distribution, we randomly sampled 43 genes (the number of genes linked to the 50 phenotypes) from the genome, controlling for gene length by matching the distribution of gene lengths in the marker gene set to those in the sampled sets using a kernel density estimation (KDE) and rejection sampling approach. This sampling procedure was repeated 100,000 times. The P -value was calculated using the formula $P = (r + 1)/(n + 1)$ (North et al. 2002), where n is the total number of replicates, and r is the number of replicates in which the number of genes affected by SVs is equal to or larger than the observed number of marker genes with candidate SVs. This analysis was repeated with a map of euchromatic SVs in the *Drosophila* Synthetic Population Resource (DSPR), a panel of 14 isogenic lines derived from globally diverse populations (Chakraborty et al. 2019). We also performed an analogous enrichment analysis for nsSNPs and small indels (2–10 bp).

Additional sequencing and analysis of newly characterized mutations

To further investigate candidate causal mutations underlying the *curved*, *clipped*, and *Ablp*^{eyD} phenotypes, we obtained independent strains carrying these mutations from the Bloomington *Drosophila* Stock Center (eight *Ablp*^{eyD}, six c^1 , and six cp^1 strains). We verified that all strains carried their associated phenotypes. High-molecular-weight DNA was extracted from all the strains and end-repaired as described above. Barcoded sequencing libraries were prepared using the Oxford Nanopore native barcoding kit 24 (v14) and sequenced on a PromethION R10.4.1 flow cell for 72 h.

SVs were genotyped by mapping reads from each strain to the pangenome graph using the vg toolkit, v1.65.0 (Garrison et al.

2018). Long reads were aligned with vg giraffe using an index optimized for Oxford Nanopore R10 data, producing GAF-format alignments. These alignments were converted into a packed coverage representation using vg pack. Variants were then genotyped directly from the pangenome graph using vg call, generating VCF files for each strain. SNPs and indels were genotyped using PEPPER-Margin-DeepVariant (Shafin et al. 2021). Individual strain VCF files were normalized and merged with BCFtools v1.19 (Danecek et al. 2021). A custom Python script, *merged_vcf_strain_filter.py*, was used to filter the merged VCF file to retain variants present in a specified group of strains and absent from all other samples. Candidate genotypes were further validated by mapping reads to the ISO-1 reference genome using minimap2 v2.26 (Li 2017), followed by visual inspection in IGV.

ATAC-seq peak identification

To define regions of accessible chromatin, we used publicly available ATAC-seq data from *D. melanogaster* wing disc, eye disc, brain, and ovary (Huynh et al. 2023). Reads were trimmed with Trimmomatic v0.39 (Bolger et al. 2014) and aligned to the ISO-1 r6 reference genome using BWA-mem2 v2.2.1 (Vasimuddin et al. 2019). PCR duplicates were removed with Picard MarkDuplicates (<http://broadinstitute.github.io/picard>), and chromatin accessibility peaks were called using MACS2 v2.2.9.1 (Zhang et al. 2008). To identify candidate regulatory mutations underlying the *clipped* phenotype, we intersected variants unique to the seven cp^1 strains within the mapped 55 kb interval with ATAC-seq peaks and visually inspected peak support in IGV (Thorvaldsdóttir et al. 2013). Similarly, the functional relevance of CRISPR-induced deletions at the *Ablp*^{eyD} locus was evaluated based on their overlap with tissue-specific accessibility peaks.

Direct RNA sequencing

Total RNA was extracted from 7-day-old male flies of strain 156 using TRIzol. RNA was reverse-transcribed to produce the DNA–RNA hybrids using the Induro reverse transcriptase kit, and sequencing libraries were prepared with the Oxford Nanopore direct RNA sequencing kit. Libraries were sequenced on a PromethION RNA flow cell for 72 h. Reads were mapped to both the ISO-1 reference genome and our de novo assembly for strain 156 using minimap2 v2.26 with splice-aware alignment parameters optimized for Oxford Nanopore RNA sequencing (-ax splice -uf -k14) (Li 2017). Gene annotations for *Strn-Mlck* and *plexus* were lifted over to the 156 assembly using LiftOff v1.6.3 (Shumate and Salzberg 2021).

Data access

All sequencing reads and genome assemblies generated in this study have been submitted to the NCBI BioProject database (<https://www.ncbi.nlm.nih.gov/bioproject/>) under accession number PRJNA1214913. All scripts for genome assembly and analysis are available as Supplemental Code and at GitHub (https://github.com/chakrabortymlab/visible_phenotype). Additional undergraduate-level, step-by-step instructions to carry out the genomic analyses reported here are available at GitHub (<https://github.com/chakrabortymlab/GALORE>).

Competing interest statement

The authors declare no competing interests.

Acknowledgments

We thank all students of the undergraduate genomics course (BIOL 450) in the Department of Biology at Texas A&M University who took the course in the spring 2024 semester for their assistance with data collection and analysis. We also thank Anthony Long and Trevor Millar for helpful suggestions and feedback on the manuscript. We thank Kevin Cook for providing an initial list of BDSC stocks with important visual markers. We also thank High-Performance Research Computing at Texas A&M University for providing the computational resources used in this study. This work was supported by funding from the Department of Biology at Texas A&M University, Texas A&M University startup grant, and a National Institutes of Health grant (R00GM129411) to M.C.

Author contributions: M.C. conceived the study, and M.C. and A.S. designed the experiments. A.S. generated genomic data and performed all genomic analyses. M.M., A.A., M.A., S.A., S.P., J.P., O.S., B.S., T.S., and S.W. assisted with SV analysis and validation. A.S. and M.J. performed CRISPR experiments, and A.S. conducted all other molecular biology experiments. M.C. and A.S. wrote the manuscript draft, and all authors edited it.

References

- Abel HJ, Larson DE, Regier AA, Chiang C, Das I, Kanchi KL, Layer RM, Neale BM, Salerno WJ, Reeves C, et al. 2020. Mapping and characterization of structural variation in 17,795 human genomes. *Nature* **583**: 83–89. doi:10.1038/s41586-020-2371-0
- Alkan C, Coe BP, Eichler EE. 2011. Genome structural variation discovery and genotyping. *Nat Rev Genet* **12**: 363–376. doi:10.1038/nrg2958
- Alonge M, Wang X, Benoit M, Soyk S, Pereira L, Zhang L, Suresh H, Ramakrishnan S, Maumus F, Ciren D, et al. 2020. Major impacts of widespread structural variation on gene expression and crop improvement in tomato. *Cell* **182**: 145–161.e23. doi:10.1016/j.cell.2020.05.021
- Alsheikh AJ, Wollenhaupt S, King EA, Reeb J, Ghosh S, Stolzenburg LR, Tamim S, Lazar J, Davis JW, Jacob HJ. 2022. The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. *BMC Med Genomics* **15**: 74. doi:10.1186/s12920-022-01216-w
- Andersson L. 2013. Molecular consequences of animal breeding. *Curr Opin Genet Dev* **23**: 295–301. doi:10.1016/j.gde.2013.02.014
- Arimoto E, Kawashima Y, Choi T, Unagami M, Akiyama S, Tomizawa M, Yano H, Suzuki E, Sone M. 2020. Analysis of a cellular structure observed in the compound eyes of *Drosophila white*; *yata* mutants and *white* mutants. *Biol Open* **9**: bio047043. doi:10.1242/bio.047043
- Bageritz J, Willnow P, Valentini E, Leible S, Boutros M, Teleman AA. 2019. Gene expression atlas of a developing tissue by single cell expression correlation analysis. *Nat Methods* **16**: 750–756. doi:10.1038/s41592-019-0492-x
- Bateson W, Saunders ER, Punnett RC. 1905. *Reports to the evolution committee of the Royal Society. Report II: experimental studies in the physiology of heredity*. Electronic Scholarly Publishing Project, Bellingham, WA. <http://www.esp.org/foundations/genetics/classical/holdings/b/wb-04a.pdf>.
- Beasley V, Dowse H. 2016. Suppression of tryptophan 2,3-dioxygenase produces a slow heartbeat phenotype in *Drosophila melanogaster*. *J Exp Zool A Ecol Genet Physiol* **325**: 651–664. doi:10.1002/jez.2057
- Bender W, Akam M, Karch F, Beachy PA, Peifer M, Spierer P, Lewis EB, Hogness DS. 1983. Molecular genetics of the bithorax complex in *Drosophila melanogaster*. *Science* **221**: 23–29. doi:10.1126/science.221.4605.23
- Bergman CM, Quesneville H, Anxolabéhère D, Ashburner M. 2006. Recurrent insertion and duplication generate networks of transposable element sequences in the *Drosophila melanogaster* genome. *Genome Biol* **7**: R112. doi:10.1186/gb-2006-7-11-r112
- Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114–2120. doi:10.1093/bioinformatics/btu170
- Bridges CB. 1922. The origin of variations in sexual and sex-limited characters. *Am Nat* **56**: 51–63. doi:10.1086/279847
- Cai HN, Levine M. 1997. The gypsy insulator can function as a promoter-specific silencer in the *Drosophila* embryo. *EMBO J* **16**: 1732–1741. doi:10.1093/emboj/16.7.1732
- Campuzano S, Carramolino L, Cabrera CV, Ruiz-Gómez M, Villares R, Boronat A, Modolell J. 1985. Molecular genetics of the achaete-scute gene complex of *D. melanogaster*. *Cell* **40**: 327–338. doi:10.1016/0092-8674(85)90147-3
- Cardoso-Moreira M, Roman Arguello J, Gottipati S, Harshman LG, Grenier JK, Clark AG. 2016. Evidence for the fixation of gene duplications by positive selection in *Drosophila*. *Genome Res* **26**: 787–798. doi:10.1101/gr.199323.115
- Carmon A, Guertin MJ, Grushko O, Marshall B, MacIntyre R. 2010. A molecular analysis of mutations at the complex dumpy locus in *Drosophila melanogaster*. *PLoS One* **5**: e12319. doi:10.1371/journal.pone.0012319
- Castelli-Gair JE, García-Bellido A. 1990. Interactions of Polycomb and trithorax with cis regulatory regions of Ultrabithorax during the development of *Drosophila melanogaster*. *EMBO J* **9**: 4267–4275. doi:10.1002/j.1460-2075.1990.tb07875.x
- Chakraborty M, Fry JD. 2015. Parallel functional changes in independent testis-specific duplicates of aldehyde dehydrogenase in *Drosophila*. *Mol Biol Evol* **32**: 1029–1038. doi:10.1093/molbev/msu407
- Chakraborty M, Baldwin-Brown JG, Long AD, Emerson JJ. 2016. Contiguous and accurate de novo assembly of metazoan genomes with modest long read coverage. *Nucleic Acids Res* **44**: e147. doi:10.1093/nar/gkw654
- Chakraborty M, VanKuren NW, Zhao R, Zhang X, Kalsow S, Emerson JJ. 2018. Hidden genetic variation shapes the structure of functional elements in *Drosophila*. *Nat Genet* **50**: 20–25. doi:10.1038/s41588-017-0010-y
- Chakraborty M, Emerson JJ, Macdonald SJ, Long AD. 2019. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat Commun* **10**: 4872. doi:10.1038/s41467-019-12884-1
- Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a *Pitx1* enhancer. *Science* **327**: 302–305. doi:10.1126/science.1182213
- Charlesworth B, Langley CH. 1989. The population genetics of *Drosophila* transposable elements. *Annu Rev Genet* **23**: 251–287. doi:10.1146/annurev.ge.23.120189.001343
- Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nat Methods* **18**: 170–175. doi:10.1038/s41592-020-01056-5
- Chia J-M, Song C, Bradbury PJ, Costich D, de Leon N, Doebley J, Elshire RJ, Gaut B, Geller L, Glaubitz JC, et al. 2012. Maize HapMap2 identifies extant variation from a genome in flux. *Nat Genet* **44**: 803–807. doi:10.1038/ng.2313
- Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. 2012. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain *w¹¹¹⁸*, Iso-2; Iso-3. *Fly (Austin)* **6**: 80–92. doi:10.4161/fly.19695
- Collins RL, Talkowski ME. 2025. Diversity and consequences of structural variation in the human genome. *Nat Rev Genet* **26**: 443–462. doi:10.1038/s41576-024-00808-9
- Collins RL, Brand H, Karczewski KJ, Zhao X, Alféöldi J, Francioli LC, Khera AV, Lowther C, Gauthier LD, Wang H, et al. 2020. A structural variation reference for medical and population genetics. *Nature* **581**: 444–451. doi:10.1038/s41586-020-2287-8
- Concordet J-P, Haeussler M. 2018. CRISPOR: intuitive guide selection for CRISPR/Cas9 genome editing experiments and screens. *Nucleic Acids Res* **46**: W242–W245. doi:10.1093/nar/gky354
- Cosetti M, Culang D, Kotla S, O'Brien P, Eberl DF, Hannan F. 2008. Unique transgenic animal model for hereditary hearing loss. *Ann Otol Rhinol Laryngol* **117**: 827–833. doi:10.1177/000348940811701106
- Cridland JM, Macdonald SJ, Long AD, Thornton KR. 2013. Abundance and distribution of transposable elements in two *Drosophila* QTL mapping resources. *Mol Biol Evol* **30**: 2311–2327. doi:10.1093/molbev/mst129
- Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, et al. 2021. Twelve years of SAMtools and BCFtools. *GigaScience* **10**: giab008. doi:10.1093/gigascience/giab008
- Dean DM, Deitcher DL, Paster CO, Xu M, Loehlin DW. 2022. “A fly appeared”: *sable*, a classic *Drosophila* mutation, maps to *Yippee*, a gene affecting body color, wings, and bristles. *G3 (Bethesda)* **12**: jkac058. doi:10.1093/g3journal/jkac058
- De Coster W, Rademakers R. 2023. NanoPack2: population-scale evaluation of long-read sequencing data. *Bioinformatics* **39**: btad311. doi:10.1093/bioinformatics/btad311
- DiBartolomeis SM, Akten B, Genova G, Roberts MA, Jackson FR. 2002. Molecular analysis of the *Drosophila* *miniature-dusky* (*m-dy*) gene complex: *m-dy* mRNAs encode transmembrane proteins with similarity to *C. elegans* cuticulin. *Mol Genet Genomics* **267**: 564–576. doi:10.1007/s00438-002-0700-7

- Dittmar EL, Oakley CG, Conner JK, Gould BA, Schemske DW. 2016. Factors influencing the effect size distribution of adaptive substitutions. *Proc Biol Sci* **283**: 20153065. doi:10.1098/rspb.2015.306
- Doebley J. 2004. The genetics of maize evolution. *Annu Rev Genet* **38**: 37–59. doi:10.1146/annurev.genet.38.072902.092425
- Dorsett D. 1993. Distance-independent inactivation of an enhancer by the suppressor of hairy-wing DNA-binding protein of *Drosophila*. *Genetics* **134**: 1135–1144. doi:10.1093/genetics/134.4.1135
- Driver A, Lacey SF, Cullingford TE, Mitchelson A, O'Hare K. 1989. Structural analysis of doc transposable elements associated with mutations at the white and suppressor of forked loci of *Drosophila melanogaster*. *Mol Gen Genet* **220**: 49–52. doi:10.1007/BF00260854
- Ebert P, Audano PA, Zhu Q, Rodriguez-Martin B, Porubsky D, Bonder MJ, Sulovari A, Ebler J, Zhou W, Serra Mari R, et al. 2021. Haplotype-resolved diverse human genomes and integrated analysis of structural variation. *Science* **372**: eabf7117. doi:10.1126/science.abf7117
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* **11**: 446–450. doi:10.1038/nrg2809
- Eizenga JM, Novak AM, Sibbesen JA, Heumos S, Ghaffaari A, Hickey G, Chang X, Seaman JD, Rounthwaite R, Ebler J, et al. 2020. Pangenome graphs. *Annu Rev Genomics Hum Genet* **21**: 139–162. doi:10.1146/annurev-genom-120219-080406
- Fadaie Z, Neveling K, Mantere T, Derks R, Haer-Wigman L, den Ouden A, Kwint M, O'Gorman L, Valkenburg D, Hoyng CB, et al. 2021. Long-read technologies identify a hidden inverted duplication in a family with choroideremia. *HGG Adv* **2**: 100046. doi:10.1016/j.xhgg.2021.100046
- Falcón-Pérez JM, Romero-Calderón R, Brooks ES, Krantz DE, Dell'Angelica EC. 2007. The *Drosophila* pigmentation gene *pink* (*p*) encodes a homologue of human Hermansky-Pudlak syndrome 5 (HPS5). *Traffic* **8**: 154–168. doi:10.1111/j.1600-0854.2006.00514.x
- Frolov MV, Zverlov VV, Alatorsev VE. 1994. The mRNA product of the *Drosophila* gene *prune* is spliced and encodes a protein containing a putative transmembrane domain. *Mol Gen Genet* **242**: 478–483. doi:10.1007/BF00281800
- Fu W, Duan H, Frei E, Noll M. 1998. *Shaven* and *sparkling* are mutations in separate enhancers of the *Drosophila Pax2* homolog. *Development* **125**: 2943–2950. doi:10.1242/dev.125.15.2943
- Garrison E, Sirén J, Novak AM, Hickey G, Eizenga JM, Dawson ET, Jones W, Garg S, Markello C, Lin MF, et al. 2018. Variation graph toolkit improves read mapping by representing genetic variation in the reference. *Nat Biotechnol* **36**: 875–879. doi:10.1038/nbt.4227
- Geyer PK, Green MM, Corces VG. 1990. Tissue-specific transcriptional enhancers may act in trans on the gene located in the homologous chromosome: the molecular basis of transvection in *Drosophila*. *EMBO J* **9**: 2247–2256. doi:10.1002/j.1460-2075.1990.tb07395.x
- Ghodsinejad Kalahroudi V, Kamalidehghan B, Kani AA, Aryani O, Tondar M, Ahmadipour F, Chung LY, Houshmand M. 2014. Two novel tyrosinase (TYR) gene mutations with pathogenic impact on oculocutaneous albinism type 1 (OCA1). *PLoS One* **9**: e106656. doi:10.1371/journal.pone.0106656
- Giordano E, Peluso I, Rendina R, Digilio A, Furia M. 2003. The clot gene of *Drosophila melanogaster* encodes a conserved member of the thiodoxin-like protein superfamily. *Mol Genet Genomics* **268**: 692–697. doi:10.1007/s00438-002-0792-0
- Green MM. 1988. Mobile DNA elements and spontaneous gene mutation. In *Banbury Report 30: Eukaryotic transposable elements as mutagenic agents* (ed. Lambert ME, et al.), pp. 41–50. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY. <https://cir.nii.ac.jp/crid/1574231873873419392>.
- Grönke S, Bickmeyer I, Wunderlich R, Jäckle H, Kühnlein RP. 2009. Curled encodes the *Drosophila* homolog of the vertebrate circadian deadenylase nocturnin. *Genetics* **183**: 219–232. doi:10.1534/genetics.109.105601
- GTEX Consortium. 2020. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**: 1318–1330. doi:10.1126/science.aaz1776
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUILT: quality assessment tool for genome assemblies. *Bioinformatics* **29**: 1072–1075. doi:10.1093/bioinformatics/btt086
- Hämälä T, Wafula EK, Guiltinan MJ, Ralph PE, dePamphilis CW, Tiffin P. 2021. Genomic structural variants constrain and facilitate adaptation in natural populations of *Theobroma cacao*, the chocolate tree. *Proc Natl Acad Sci* **118**: e2102914118. doi:10.1073/pnas.2102914118
- Hao I, Green RB, Dunaevsky O, Lengyel JA, Rauskolb C. 2003. The odd-skipped family of zinc finger genes promotes *Drosophila* leg segmentation. *Dev Biol* **263**: 282–295. doi:10.1016/j.ydbio.2003.07.011
- Hazlerigg T, Levis R, Rubin GM. 1984. Transformation of white locus DNA in *Drosophila*: dosage compensation, zeste interaction, and position effects. *Cell* **36**: 469–481. doi:10.1016/0092-8674(84)90240-X
- Hickey G, Monlong J, Ebler J, Novak AM, Eizenga JM, Gao Y, Abel HJ, Antonacci-Fulton LL, Asri M, Baid G, et al. 2023. Pangenome graph construction from genome alignments with Minigraph-Cactus. *Nat Biotechnol* **42**: 663–673. doi:10.1038/s41587-023-01793-w
- Higson TS, Tassiore JE, Bennett SA, Derk RC, Kotarski MA. 1993. The molecular organization of the star/asteroid region, a region necessary for proper eye development in *Drosophila melanogaster*. *Genome* **36**: 356–366. doi:10.1139/g93-049
- Hochman B, Gloor H, Green MM. 1964. Analysis of chromosome 4 in *Drosophila melanogaster*. I. Spontaneous and X-ray-induced lethals. *Genetica* **35**: 109–126. doi:10.1007/BF01804879
- Hoekstra HE, Hirschmann RJ, Bunday RA, Insel PA, Crossland JP. 2006. A single amino acid mutation contributes to adaptive beach mouse color pattern. *Science* **313**: 101–104. doi:10.1126/science.1126121
- Hollister JD, Gaut BS. 2009. Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res* **19**: 1419–1428. doi:10.1101/gr.091678.109
- Hooper KL, Parkhurst SM, Ish-Horowitz D. 1989. Spatial control of hairy protein expression during embryogenesis. *Development* **107**: 489–504. doi:10.1242/dev.107.3.489
- Hoover KK, Chien AJ, Corces VG. 1993. Effects of transposable elements on the expression of the forked gene of *Drosophila melanogaster*. *Genetics* **135**: 507–526. doi:10.1093/genetics/135.2.507
- Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, et al. 2015. The release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res* **25**: 445–458. doi:10.1101/gr.185579.114
- Hughes AL. 1994. The evolution of functionally novel proteins after gene duplication. *Proc Biol Sci* **266**: 119–124. doi:10.1098/rspb.1994.0058
- Hurles ME, Dermitzakis ET, Tyler-Smith C. 2008. The functional impact of structural variation in humans. *Trends Genet* **24**: 238–245. doi:10.1016/j.tig.2008.03.001
- Huynh K, Smith BR, Macdonald SJ, Long AD. 2023. Genetic variation in chromatin state across multiple tissues in *Drosophila melanogaster*. *PLoS Genet* **19**: e1010439. doi:10.1371/journal.pgen.1010439
- Imsland F, Feng C, Boije H, Bed'hom B, Fillon Y, Dorshorst B, Rubin C-J, Liu R, Gao Y, Gu X, et al. 2012. The rose-comb mutation in chickens constitutes a structural rearrangement causing both altered comb morphology and defective sperm motility. *PLoS Genet* **8**: e1002775. doi:10.1371/journal.pgen.1002775
- Jeong S, Rebeiz M, Andolfatto P, Werner T, True J, Carroll SB. 2008. The evolution of gene regulation underlies a morphological difference between two *Drosophila* sister species. *Cell* **132**: 783–793. doi:10.1016/j.cell.2008.01.014
- Kahsai L, Cook KR. 2018. Mapping second chromosome mutations to defined genomic regions in *Drosophila melanogaster*. *G3 (Bethesda)* **8**: 9–16. doi:10.1534/g3.117.300289
- Kaminker JS, Bergman CM, Kronmiller B, Carlson J, Svirskas R, Patel S, Frise E, Wheeler DA, Lewis SE, Rubin GM, et al. 2002. The transposable elements of the *Drosophila melanogaster* euchromatin: a genomics perspective. *Genome Biol* **3**: RESEARCH0084. doi:10.1186/gb-2002-3-12-research0084
- Kelleher ES, Edelman NB, Barbash DA. 2012. *Drosophila* interspecific hybrids phenocopy piRNA-pathway mutants. *PLoS Biol* **10**: e1001428. doi:10.1371/journal.pbio.1001428
- Kilchherr F, Baumgartner S, Bopp D, Frei E, Noll M. 1986. Isolation of the paired gene of *Drosophila* and its spatial expression during early embryogenesis. *Nature* **321**: 493–499. doi:10.1038/321493a0
- Kim N, Kim J, Park D, Rosen C, Dorsett D, Yim J. 1996. Structure and expression of wild-type and suppressible alleles of the *Drosophila* purple gene. *Genetics* **142**: 1157–1168. doi:10.1093/genetics/142.4.1157
- King EG, Sanderson BJ, McNeil CL, Long AD, Macdonald SJ. 2014. Genetic dissection of the *Drosophila melanogaster* female head transcriptome reveals widespread allelic heterogeneity. *PLoS Genet* **10**: e1004322. doi:10.1371/journal.pgen.1004322
- Kofler R, Hill T, Nolte V, Betancourt AJ, Schlötterer C. 2015. The recent invasion of natural *Drosophila* simulans populations by the P-element. *Proc Natl Acad Sci* **112**: 6659–6663. doi:10.1073/pnas.1500758112
- Kojima T, Sone M, Michiue T, Saigo K. 1993. Mechanism of induction of bar-like eye malformation by transient overexpression of bar homeobox genes in *Drosophila melanogaster*. *Genetica* **88**: 85–91. doi:10.1007/BF02424465
- Kolmogorov M, Yuan J, Lin Y, Pevzner PA. 2019. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol* **37**: 540–546. doi:10.1038/s41587-019-0072-8
- Kolmogorov M, Billingsley KJ, Mastoras M, Meredith M, Monlong J, Lorig-Roach R, Asri M, Alvarez Jerez P, Malik L, Dewan R, et al. 2023. Scalable nanopore sequencing of human genomes provides a comprehensive view of haplotype-resolved variation and methylation. *Nat Methods* **20**: 1483–1492. doi:10.1038/s41592-023-01993-x

- Koornneef M, van Eden J, Hanhart CJ, Stam P, Braaksmá FJ, Feenstra WJ. 1983. Linkage map of *Arabidopsis thaliana*. *J Hered* **74**: 265–272. doi:10.1093/oxfordjournals.jhered.a109781
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res* **27**: 722–736. doi:10.1101/gr.215087.116
- Kronhamn J, Frei E, Daube M, Jiao R, Shi Y, Noll M, Rasmuson-Lestander Å. 2002. Headless flies produced by mutations in the paralogous Pax6 genes eyeless and twin of eyeless. *Development* **129**: 1015–1026. doi:10.1242/dev.129.4.1015
- Krstic D, Boll W, Noll M. 2013. Influence of the white locus on the courtship behavior of *Drosophila* males. *PLoS One* **8**: e77904. doi:10.1371/journal.pone.0077904
- Laetsch DR, Blaxter ML. 2017. BlobTools: interrogation of genome assemblies. *F1000Res* **6**: 1287. doi:10.12688/f1000research.12232.1
- Lee YCG, Karpen GH. 2017. Pervasive epigenetic effects of *Drosophila* euchromatic transposable elements impact their evolution. *eLife* **6**: e25762. doi:10.7554/eLife.25762
- Li H. 2017. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**: 3094–3100. doi:10.1093/bioinformatics/bty191
- Li H, Durbin R. 2024. Genome assembly in the telomere-to-telomere era. *Nat Rev Genet* **25**: 658–670. doi:10.1038/s41576-024-00718-w
- Li X, Liu Q, Fu C, Li M, Li C, Li X, Zhao S, Zheng Z. 2024. Characterizing structural variants based on graph-genotyping provides insights into Pig domestication and local adaptation. *J Genet Genomics* **51**: 394–406. doi:10.1016/j.jgg.2023.11.005
- Lloyd VK, Sinclair DA, Wennberg R, Warner TS, Honda BM, Griigli TA. 1999. A genetic and molecular characterization of the garnet gene of *Drosophila melanogaster*. *Genome* **42**: 1183–1193. doi:10.1139/g99-088
- Long M, Langley CH. 1993. Natural selection and the origin of *jingwei*, a chimeric processed functional gene in *Drosophila*. *Science* **260**: 91–95. doi:10.1126/science.7682012
- Long AD, Mullaney SL, Reid LA, Fry JD, Langley CH, Mackay TF. 1995. High resolution mapping of genetic factors affecting abdominal bristle number in *Drosophila melanogaster*. *Genetics* **139**: 1273–1291. doi:10.1093/genetics/139.3.1273
- Lunde K, Trimble JL, Guichard A, Guss KA, Nauber U, Bier E. 2003. Activation of the knirps locus links patterning to morphogenesis of the second wing vein in *Drosophila*. *Development* **130**: 235–248. doi:10.1242/dev.00207
- Ma J, Plesken H, Treisman JE, Edelman-Novemsky I, Ren M. 2004. Lightoid and Claret: a Rab GTPase and its putative guanine nucleotide exchange factor in biogenesis of *Drosophila* eye pigment granules. *Proc Natl Acad Sci* **101**: 11652–11657. doi:10.1073/pnas.0401926101
- MacArthur S, Li X-Y, Li J, Brown JB, Chu HC, Zeng L, Grondona BP, Hechmer A, Simirenko L, Keränen SVE, et al. 2009. Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol* **10**: R80. doi:10.1186/gb-2009-10-7-r80
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, et al. 2009. Finding the missing heritability of complex diseases. *Nature* **461**: 747. doi:10.1038/nature08494
- Marçais G, Delcher AL, Phillippy AM, Coston R, Salzberg SL, Zimin A. 2018. MUMmer4: a fast and versatile genome alignment system. *PLoS Comput Biol* **14**: e1005944. doi:10.1371/journal.pcbi.1005944
- Marian AJ. 2020. Clinical interpretation and management of genetic variants. *JACC Basic Transl Sci* **5**: 1029–1042. doi:10.1016/j.jacbs.2020.05.013
- Massey JH, Chung D, Siwanowicz I, Stern DL, Wittkopp PJ. 2019. The *yellow* gene influences *Drosophila* male mating success through sex comb melanization. *eLife* **8**: e49388. doi:10.7554/eLife.49388
- Masucci JD, Miltenberger RJ, Hoffmann FM. 1990. Pattern-specific expression of the *Drosophila* decapentaplegic gene in imaginal disks is regulated by 3' cis-regulatory elements. *Genes Dev* **4**: 2011–2023. doi:10.1101/gad.4.11.2011
- Matakatsu H, Tadokoro R, Gamo S, Hayashi S. 1999. Repression of the wing vein development in *Drosophila* by the nuclear matrix protein plexus. *Development* **126**: 5207–5216. doi:10.1242/dev.126.23.5207
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**: 1190–1195. doi:10.1126/science.1222794
- McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proc Natl Acad Sci* **36**: 344–355. doi:10.1073/pnas.36.6.344
- Mendel G. 1996. *Experiments in plant hybridization* (trans. Blumberg R). Electronic Scholarly Publishing Project, Bellingham, WA. <http://old.esp.org/foundations/genetics/classical/gm-65-a.pdf>.
- Merker JD, Wenger AM, Sneddon T, Grove M, Zappala Z, Fresard L, Waggott D, Utiramerur S, Hou Y, Smith KS, et al. 2018. Long-read genome sequencing identifies causal structural variation in a Mendelian disease. *Genet Med* **20**: 159–163. doi:10.1038/gim.2017.86
- Miller DE, Cook KR, Kazemi NY, Smith CB, Cockrell AJ, Hawley RS, Bergman CM. 2016. Rare recombination events generate sequence diversity among balancer chromosomes in *Drosophila melanogaster*. *Proc Natl Acad Sci* **113**: E1352–E1361. doi:10.1073/pnas.1601232113
- Morgan TH. 1910. Sex limited inheritance in *Drosophila*. *Science* **32**: 120–122. doi:10.1126/science.32.812.120
- Muller HJ. 1927. Artificial transmutation of the gene. *Science* **66**: 84–87. doi:10.1126/science.66.1699.84
- Muller HJ. 1936. Bar duplication. *Science* **83**: 528–530. doi:10.1126/science.83.2161.528.b
- Mullins C, Hartnell LM, Wassarman DA, Bonifacino JS. 1999. Defective expression of the μ 3 subunit of the AP-3 adaptor complex in the *Drosophila* pigmentation mutant carmine. *Mol Gen Genet* **262**: 401–412. doi:10.1007/s004380051099
- Naisbit RE, Jiggins CD, Mallet J. 2003. Mimicry: developmental genes that contribute to speciation. *Evol Dev* **5**: 269–280. doi:10.1046/j.1525-142X.2003.03034.x
- Nash D, Hu S, Leonard NJ, Tiong SY, Phillips D. 1994. The raspberry locus of *Drosophila melanogaster* includes an inosine monophosphate dehydrogenase like coding sequence. *Genome* **37**: 333–344. doi:10.1139/g94-046
- North BV, Curtis D, Sham PC. 2002. A note on the calculation of empirical *P* values from Monte Carlo procedures. *Am J Hum Genet* **71**: 439–441. doi:10.1086/341527
- Öztürk-Çolak A, Marygold SJ, Antonazzo G, Attrill H, Goutte-Gattat D, Jenkins VK, Matthews BB, Millburn G, dos Santos G, Tabone CJ, et al. 2024. FlyBase: updates to the *Drosophila* genes and genomes database. *Genetics* **227**: iyad211. doi:10.1093/genetics/iyad211
- Paterson J, O'Hare K. 1991. Structure and transcription of the singed locus of *Drosophila melanogaster*. *Genetics* **129**: 1073–1084. doi:10.1093/genetics/129.4.1073
- Petrov DA, Fiston-Lavier A-S, Lipatov M, Lenkov K, González J. 2011. Population genomics of transposable elements in *Drosophila melanogaster*. *Mol Biol Evol* **28**: 1633–1644. doi:10.1093/molbev/msq337
- Phillips AM, Smart R, Strauss R, Brembs B, Kelly LE. 2005. The *Drosophila* black enigma: the molecular and behavioural characterization of the *black¹* mutant allele. *Gene* **351**: 131–142. doi:10.1016/j.gene.2005.03.013
- Pianezza R, Scarpa A, Haider A, Signor S, Kofler R. 2025. Spatiotemporal tracking of three novel transposable element invasions in *Drosophila melanogaster* over the last 30 years. *Mol Biol Evol* **42**: msaf143. doi:10.1093/molbev/msaf143
- Port F, Chen H-M, Lee T, Bullock SL. 2014. Optimized CRISPR/Cas tools for efficient germline and somatic genome engineering in *Drosophila*. *Proc Natl Acad Sci* **111**: E2967–E2976. doi:10.1073/pnas.1405500111
- Quan C, Li Y, Liu X, Wang Y, Ping J, Lu Y, Zhou G. 2021. Characterization of structural variation in Tibetans reveals new evidence of high-altitude adaptation and introgression. *Genome Biol* **22**: 159. doi:10.1186/s13059-021-02382-3
- Roch F, Alonso CR, Akam M. 2003. *Drosophila* miniature and dusky encode ZP proteins required for cytoskeletal reorganization during wing morphogenesis. *J Cell Sci* **116**: 1199–1207. doi:10.1242/jcs.00298
- Rodriguez DM. 2004. "Isolation and characterization of stretchin-myosin light chain kinase mutants in *Drosophila melanogaster*." *PhD thesis*, Ohio State University, Columbus, OH. <https://search.proquest.com/openview/2cb26c4fe5f54d127d445859a10e2c4c/1?pq-origsite=gscholar&cbl=18750&diss=y>.
- Rogers RL, Slatkin M. 2017. Excess of genomic defects in a woolly mammoth on Wrangel Island. *PLoS Genet* **13**: e1006601. doi:10.1371/journal.pgen.1006601
- Samano A, Kumar N, Liao Y, Ishtiaq F, Chakraborty M. 2025. Genome structural variants shape adaptive success of an invasive urban malaria vector *Anopheles Stephensi*. *Mol Biol Evol* **42**: msaf140. doi:10.1093/molbev/msaf140
- Sankaranarayanan K. 1988. Mobile genetic elements, spontaneous mutations and the assessment of genetic radiation hazards in man. In *Banbury Report 30: Eukaryotic transposable elements as mutagenic agents* (ed. Lambert ME, et al.), pp. 319–336. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
- Sax K. 1923. The association of size differences with seed-coat pattern and pigmentation in *PHASEOLUS VULGARIS*. *Genetics* **8**: 552–560. doi:10.1093/genetics/8.6.552
- Schipper M, Posthuma D. 2022. Demystifying non-coding GWAS variants: an overview of computational tools and methods. *Hum Mol Genet* **31**: R73–R83. doi:10.1093/hmg/ddac198
- Schmidt JM, Good RT, Appleton B, Sherrard J, Raymant GC, Bogwitz MR, Martin J, Daborn PJ, Goddard ME, Batterham P, et al. 2010. Copy number variation and transposable elements feature in recent, ongoing

- adaptation at the *cyp6g1* locus. *PLoS Genet* **6**: e1000998. doi:10.1371/journal.pgen.1000998
- Searles LL, Ruth RS, Pret AM, Fridell RA, Ali AJ. 1990. Structure and transcription of the *Drosophila melanogaster* vermilion gene and several mutant alleles. *Mol Cell Biol* **10**: 1423–1431. doi:10.1128/mcb.10.4.1423-1431.1990
- Sentmanat MF, Elgin SCR. 2012. Ectopic assembly of heterochromatin in *Drosophila melanogaster* triggered by transposable elements. *Proc Natl Acad Sci* **109**: 14104–14109. doi:10.1073/pnas.1207036109
- Sereika M, Kirkegaard RH, Karst SM, Michaelsen TY, Sørensen EA, Wollenberg RD, Albertsen M. 2022. Oxford Nanopore R10.4 long-read sequencing enables the generation of near-finished bacterial genomes from pure cultures and metagenomes without short-read or reference polishing. *Nat Methods* **19**: 823–826. doi:10.1038/s41592-022-01539-7
- Sevrioukov EA, He JP, Moghrabi N, Sunio A, Krämer H. 1999. A role for the deep orange and carnation eye color genes in lysosomal delivery in *Drosophila*. *Mol Cell* **4**: 479–486. doi:10.1016/S1097-2765(00)80199-9
- Shafin K, Pesout T, Chang P-C, Nattestad M, Kolesnikov A, Goel S, Baid G, Kolmogorov M, Eizenga JM, Miga KH, et al. 2021. Haplotype-aware variant calling with PEPPER-Margin-DeepVariant enables high accuracy in nanopore long-reads. *Nat Methods* **18**: 1322–1332. doi:10.1038/s41592-021-01299-w
- Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM. 2004. Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature* **428**: 717–723. doi:10.1038/nature02415
- Shumate A, Salzberg SL. 2021. Liftoff: accurate mapping of gene annotations. *Bioinformatics* **37**: 1639–1643. doi:10.1093/bioinformatics/btaa1016
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**: 3210–3212. doi:10.1093/bioinformatics/btv351
- Sirén J, Monlong J, Chang X, Novak AM, Eizenga JM, Markello C, Sibbesen JA, Hickey G, Chang P-C, Carroll A, et al. 2021. Pangenomics enables genotyping of known structural variants in 5202 diverse genomes. *Science* **374**: abg8871. doi:10.1126/science.abg8871
- Smit AFA, Hubley R, Green P. 2013. RepeatMasker Open-4.0, version 4.0. <http://www.repeatmasker.org>.
- Smolka M, Paulin LF, Grochowski CM, Horner DW, Mahmoud M, Behera S, Kalef-Ezra E, Gandhi M, Hong K, Pehlivan D, et al. 2024. Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol* **42**: 1571–1580. doi:10.1038/s41587-023-02024-y
- Solares EA, Chakraborty M, Miller DE, Kalsow S, Hall K, Perera AG, Emerson JJ, Hawley RS. 2018. Rapid low-cost assembly of the *Drosophila melanogaster* reference genome using low-coverage, long-read sequencing. *G3 (Bethesda)* **8**: 3143–3154. doi:10.1534/g3.118.200162
- Sorek R, Ast G, Graur D. 2002. *Alu*-containing exons are alternatively spliced. *Genome Res* **12**: 1060–1067. doi:10.1101/gr.229302
- Spana EP, Abrams AB, Ellis KT, Klein JC, Ruderman BT, Shi AH, Zhu D, Stewart A, May S. 2020. Speck, first identified in *Drosophila melanogaster* in 1910, is encoded by the arylalkaline N-acetyltransferase (AANAT1) gene. *G3 (Bethesda)* **10**: 3387–3398. doi:10.1534/g3.120.401470
- Spradling AC, Mahowald AP. 1981. A chromosome inversion alters the pattern of specific DNA replication in *Drosophila* follicle cells. *Cell* **27**: 203–209. doi:10.1016/0092-8674(81)90374-3
- Spradling AC, Waring GL, Mahowald AP. 1979. *Drosophila* bearing the ocelliless mutation underproduce two major chorion proteins both of which map near this gene. *Cell* **16**: 609–616. doi:10.1016/0092-8674(79)90034-5
- Sturtevant AH. 1913. The linear arrangement of six sex-linked factors in *Drosophila*, as shown by their mode of association. *J Exp Zool* **14**: 43–59. doi:10.1002/jez.1400140104
- Sturtevant AH. 1925. The effects of unequal crossing over at the bar locus in *Drosophila*. *Genetics* **10**: 117–147. doi:10.1093/genetics/10.2.117
- Szyrzycka M, McEachern LA, Kinnear J, Prabhu K, Fitzpatrick K, Schulze S, Rawls JM, Lloyd VK, Sinclair DAR, Honda BM. 2007. The pink gene encodes the *Drosophila* orthologue of the human Hermansky-Pudlak syndrome 5 (HPS5) gene. *Genome* **50**: 548–556. doi:10.1139/G07-032
- Talkowski ME, Rosenfeld JA, Blumenthal I, Pillalamarri V, Chiang C, Heilbut A, Ernst C, Hanscom C, Rossin E, Lindgren AM, et al. 2012. Sequencing chromosomal abnormalities reveals neurodevelopmental loci that confer risk across diagnostic boundaries. *Cell* **149**: 525–537. doi:10.1016/j.cell.2012.03.028
- Tchurikov NA, Gerasimova TI, Johnson TK, Barbakar NI, Keszior AL, Georgiev GP. 1989. Mobile elements and transposition events in the cut locus of *Drosophila melanogaster*. *Mol Genet* **219**: 241–248. doi:10.1007/BF00261183
- ten Have JF, Green MM, Howells AJ. 1995. Molecular characterization of spontaneous mutations at the scarlet locus of *Drosophila melanogaster*. *Mol Genet* **249**: 673–681. doi:10.1007/BF00418037
- Thorvaldsdóttir H, Robinson JT, Mesirov JP. 2013. Integrative genomics viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* **14**: 178–192. doi:10.1093/bib/bbs017
- Tice SC. 1914. A new sex-linked character in *Drosophila*. *Biol Bull* **26**: 221–230. doi:10.2307/1536196
- Timmons L, Shearn A. 1996. Germline transformation using a prune cDNA rescues prune/killer of prune lethality and the prune eye color phenotype in *Drosophila*. *Genetics* **144**: 1589–1600. doi:10.1093/genetics/144.4.1589
- Treangen TJ, Salzberg SL. 2011. Repetitive DNA and next-generation sequencing: computational challenges and solutions. *Nat Rev Genet* **13**: 36–46. doi:10.1038/nrg3117
- Tsubota SI, Rosenberg D, Szostak H, Rubin D, Schedl P. 1989. The cloning of the bar region and the B breakpoint in *Drosophila melanogaster*: evidence for a transposon-induced rearrangement. *Genetics* **122**: 881–890. doi:10.1093/genetics/122.4.881
- Van't Hof AE, Campagne P, Rigden DJ, Yung CJ, Lingley J, Quail MA, Hall N, Darby AC, Saccheri IJ. 2016. The industrial melanism mutation in British peppered moths is a transposable element. *Nature* **534**: 102–105. doi:10.1038/nature17951
- Vasimuddin M, Misra S, Li H, Aluru S. 2019. Efficient architecture-aware acceleration of BWA-MEM for multicore systems. In *IEEE International Parallel and Distributed Processing Symposium (IPDPS)*, Rio de Janeiro, Brazil, pp. 314–324. doi:10.1109/IPDPS.2019.00041
- Wang L, Sexton TR, Venard C, Giedt M, Guo Q, Chen Q, Harrison DA. 2014. Pleiotropy of the *Drosophila* JAK pathway cytokine unpaired 3 in development and aging. *Dev Biol* **395**: 218–231. doi:10.1016/j.ydbio.2014.09.015
- White SJ, Aartsma-Rus A, Flanigan KM, Weiss RB, Kneppers ALJ, Lalic T, Janson AAM, Ginjaar HB, Breuning MH, den Dunnen JT. 2006. Duplications in the *DMD* gene. *Hum Mutat* **27**: 938–945. doi:10.1002/humu.20367
- Wick RR, Schultz MB, Zobel J, Holt KE. 2015. Bandage: interactive visualization of *de novo* genome assemblies. *Bioinformatics* **31**: 3350–3352. doi:10.1093/bioinformatics/btv383
- Wild A, Kalf-Suske M, Vortkamp A, Bornholdt D, König R, Grzeschik KH. 1997. Point mutations in human *GLI3* cause Greig syndrome. *Hum Mol Genet* **6**: 1979–1984. doi:10.1093/hmg/6.11.1979
- Wills DM, Whipple CJ, Takuno S, Kursel LE, Shannon LM, Ross-Ibarra J, Doebley JF. 2013. From many, one: genetic control of prolificacy during maize domestication. *PLoS Genet* **9**: e1003604. doi:10.1371/journal.pgen.1003604
- Xiao C, Qiu S, Robertson RM. 2017. The *white* gene controls copulation success in *Drosophila melanogaster*. *Sci Rep* **7**: 7712. doi:10.1038/s41598-017-08155-y
- Yang J, Wang D-F, Huang J-H, Zhu Q-H, Luo L-Y, Lu R, Xie X-L, Salehian-Dehkordi H, Esmailzadeh A, Liu GE, et al. 2024. Structural variant landscapes reveal convergent signatures of evolution in sheep and goats. *Genome Biol* **25**: 148. doi:10.1186/s13059-024-03288-6
- Yu L, Lee T, Lin N, Wolf MJ. 2010. Affecting rhomboid-3 function causes a dilated heart in adult *Drosophila*. *PLoS Genet* **6**: e1000969. doi:10.1371/journal.pgen.1000969
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based Analysis of ChIP-Seq (MACS). *Genome Biol* **9**: R137. doi:10.1186/gb-2008-9-9-r137

Received September 15, 2025; accepted in revised form April 21, 2026.