

Position Effect Variegation in natural populations not explained by common variation in known modifiers

Keegan J. P. Kelsey and Andrew G. Clark¹

Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853

ABSTRACT Changes in chromatin state may drive changes in gene expression, and it is of growing interest to understand the population genetic forces that drive differences in chromatin state. Here, we use the phenomenon of position effect variegation (PEV), a well-studied proxy for chromatin state, to explore the genetic architecture of natural variation in factors that modify chromatin state. While previous mutation screens have identified over 150 suppressors and enhancers of PEV, it remains unknown to what extent allelic variation in these modifiers mediates inter-individual variation in chromatin state. Is natural variation in PEV mediated by segregating variation in known Su(var) and E(var) genes, or is the trait polygenic, with many variants mapping elsewhere in the genome? We designed a mapping study that directly answers this question and suggests that the bulk of the variance in PEV does not map to genes with prior annotated impact to PEV. Instead, we find enrichment of top *P*-value ranked associations that suggest impact to active promoter and transcription start site proximal regions. This work provides a quantitative view of the role naturally segregating autosomal variants play in modifying PEV, a phenomenon that continues to shape our understanding of chromatin state and epigenetics.

KEYWORDS Position Effect Variegation (PEV); chromatin; natural variation; phenotypic variance; suppressor of variegation, enhancer of variegation

Chromatin states, through their impact on chromatin accessibility and gene expression have clear evolutionary importance (Schulze and Wallrath 2007). The ModEncode project has generated valuable data for use in understanding determinants of chromatin states (modENCODE Consortium *et al.* 2010). This has generated genome-wide assessment of chromatin features and led to identification of broad chromatin features with functional consequences (Filion *et al.* 2016; Kharchenko *et al.* 2011; Ernst and Kellis 2012). Despite rapid progress in chromatin biology, we still know little about naturally segregating variants and involvement in generating differences in chromatin states among individuals. Any natural variant that serves to impact chromatin state may be a target of natural selection, and it is of great interest to understand population genetic forces that drive differences between chromatin features of individuals.

Originally discovered by H. J. Muller, PEV has long been studied in *Drosophila melanogaster* and is widely accepted as a valuable tool in understanding dynamics of chromatin state and gene expression, especially with regards to the boundary be-

tween heterochromatin and euchromatin. The most commonly used form of PEV is the result of a specific X chromosome inversion, *white*^{mottled-4} or simply *w^{m4}* (Muller 1930), that relocated the normally euchromatic *white* gene next to pericentromeric heterochromatin. The molecular impact of this relocation is now understood to be a spreading of chromatin factors that results in altered chromatin structure and gene silencing of *white* without known modification to coding sequence (Wallrath and Elgin 1995). Depending on the genomic background, the phenotypic manifestation of this particular silencing is mosaic across the facets of the eye, resulting in a “mottled” eye phenotype with local clones of cells in each eye showing an apparently stochastic response. Of significance, PEV is triggered by gene-chromosome rearrangements in many organisms, producing modified gene expression as a result of modified chromatin environment (Girton and Johansen 2008; Elgin and Reuter 2013).

Manifestation of PEV in the *Drosophila* eye allows for ease of visible phenotyping and has generated a large body of work leading to the discovery of numerous PEV modifying factors, both genetic and environmental, that also impact chromatin states. These modifiers are typically described as positively influencing gene expression at the *w^{m4}* locus, termed suppressor of variegation, or Su(var), or negatively influencing gene expres-

sion at w^{m4} , termed enhancer of variegation, or E(var). Screens for modifiers have identified over 150 loci that modify PEV (Ebert *et al.* 2004). Two such well characterized modifiers of PEV and chromatin include *Su(var)3-9* and *Su(var)2-5*. *Su(var)3-9*, a histone methyltransferase with action on histone H3 at lysine 9, helps to form heterochromatin (Schotta *et al.* 2002). *Su(var)2-5* or *HP1a* interacts with *Su(var)3-9* to uniquely bind distinct regions of the genome (Greil *et al.* 2003) and is required for the spread of heterochromatin (Hines *et al.* 2009). It is important to understand, however, that the vast majority of these modifiers have been identified through mutation screens and may or may not represent loci harboring natural variation that impact the PEV phenotype. Despite the extensive literature on modifiers of PEV, there is currently little information on natural variation and impact to chromatin state between individuals and populations.

Genome-wide association studies (GWAS) initially gained favor in human studies as a way to perform a relatively unbiased search for common natural variants involved in a phenotype of interest. Using sets of inbred reference lines of *D. melanogaster*, GWAS has proven to be a powerful resource for generating unbiased, data-driven hypotheses, where a genome-wide search of candidate loci may be coupled with the extensive functional annotation and genetic tools already available. The *Drosophila* Genetic Reference Panel (DGRP), a collection of genome-sequenced inbred lines of *D. melanogaster* (Mackay *et al.* 2012; Huang *et al.* 2014), has become a handy resource for initial GWAS screens with *Drosophila*. Several groups have already successfully used the resource to identify novel variants involved in a wide range of traits, including; sleep (Harbison *et al.* 2013), leg development (Grubbs *et al.* 2013), sperm competition (Chow *et al.* 2012), host-microbiota interaction (Dobson *et al.* 2015), fecundity and fitness (Durham *et al.* 2014) and nutritional indices (Unckless *et al.* 2015).

To better understand the underlying genetic architecture of segregating natural variants involved in heterochromatin dynamics, we performed GWAS on F1 progeny of DGRP lines crossed to w^{m4} , a line bearing an X-linked inversion that displays PEV of the white eye phenotype. PEV was quantified by novel digital image analysis of visible images captured with a dissecting microscope. Despite this detailed work, we find little evidence of association for segregating variants within known *Su(var)* and *E(var)* genes. However, we do find variants having association with PEV to be over-represented in regions having a chromatin state indicative of active promoter and transcription start site (TSS)-proximal features. Furthermore, a comprehensive search across binding sites for factors that modify chromatin state link numerous binding sites to PEV-associated variants, emphasizing regions with bi-stable chromatin states. Altogether, the evidence suggests autosomal natural variation interacts with PEV through numerous, small effect loci that are enriched for TF binding and sites of open chromatin, implying influence through gene expression or subtle changes to chromatin balance.

Materials and Methods

Drosophila stocks

Lines from the *Drosophila* Genetic Reference Panel (DGRP) (Mackay *et al.* 2012) were a gift of the Mackay lab. 1712 (Bloomington), which harbors the *white*^{mottled-4}, *In(1)w^{m4}* or simply, w^{m4} , locus on the X chromosome and a second chromosome deletion and balancer, *Df(2L)2802/CyO*, was used to assess variegation across the DGRP population. Canton-S and mutant eye color stocks, 245 (*bw¹*) and 3605 (*w¹¹¹⁸*) (Bloomington) provided biological context in our eye color phenotype assay. All flies were

maintained on a standard cornmeal-molasses-sucrose-yeast media and kept at 25°C on a 12-hour light/dark cycle.

Experimental cross

In each of two replicate vials, ten males from each DGRP line were crossed to three virgin females of the 1712 stock and allowed to mate for one day. Mated 1712 females were then transferred to new food, allowed to lay eggs, and removed from the vial after five days. Variegating male progeny segregated into two phenotypic classes based on the second chromosome, curly winged (*CyO*) and non-curly winged (*Df(2L)2802*), and were aged at least four to eight days before imaging.

Image capture and eye color quantification

The left or right eye was randomly selected from each adult male and imaged using an Olympus SMZ-10 dissecting microscope with an attached Cannon Rebel 6 megapixel digital camera in a windowless room. Images were captured and stored using software from the camera manufacturer. On average, 17.5 eyes (standard deviation = 9.2) were imaged from each line, resulting in a total of 3966 images across all lines and second chromosome combinations. A standard gray card (Kodak, 18% gray) was imaged before and after each set of conditions to normalize against fluctuating light conditions. Ommatidia from images were isolated using a pipeline built in Cell Profiler 2.0 (Lamprecht *et al.* 2007) and visually inspected. Images that failed to process were individually assessed and isolated in Photoshop. Isolated ommatidia and standard gray card images were then processed using custom scripts in R (R Development Core Team 2008). Ommatidia image files were separated into red, green, and blue color channels and values were bounded between 0 and 1. Color channel values from each pixel of every image were normalized against the mean value of individual color channels from matched gray card pairs according to a generalized gamma adjustment using the formula,

$$V^\gamma = V_{out}; \quad \text{where} \quad \gamma = \frac{\log 0.82}{\log \bar{s}}$$

where \bar{s} is the individual mean color channel (red, green or blue) for the gray card imaged before and after samples. 0.82 refers to the idealized 18% gray card value in RGB color space. V is the color channel value for an individual pixel within the image, γ is the normalizing function, and V_{out} is the normalized color value. The final summarized output for each eye and image resulted in three values, including the mean of each normalized red, green and blue color channel. We show that these descriptive values sufficiently separate individuals for our purposes.

Statistical analysis of phenotype

All subsequent analysis was performed in R (R Development Core Team 2008). The mean red, green and blue color values for each image were used input for Principal Component Analysis (PCA). Eye groups were assessed using MANOVA and the formula,

$$Y_{pc1,pc2,pc3} = \mu + S + e$$

where Y is PC1, PC2 and PC3 from each image, and S is the stock of origin (Canton-S, 1712, 245 or 3605). Differences between experimental groups were assessed using a linear model to fit the principal components from each image with the formula,

$$Y = \mu + L + C + V + LxC + e$$

where Y is either PC1, PC2 or PC3, L is the DGRP line of origin, C is the second chromosome background (CyO or $Df(2L)2802$), and V is the replicate vial (A or B). Proportion of variance for each effect is calculated as,

$$eta^2 = \frac{SS_{effect}}{SS_{total}}$$

where SS is the sum of squares. Broad-sense heritability (H^2) was calculated using,

$$H^2 = \frac{V_g}{V_p}$$

and based on the linear model,

$$Y = \mu + L + e$$

where Y is a single principal component and single second chromosome background combination, and L is the DGRP line of origin. The proportion of variance explained by line differences was considered the variance attributed to genetic components (V_g), and the total variance of the sample was considered the variance attributed to phenotype (V_p). For each second chromosome background, H^2 was summed across the three principal components and weighted according to the proportion of variance explained by each PC (Table S1).

Genotypes and association testing

Genotypes and annotation for DGRP lines were downloaded from the website, dgrp.gnets.ncsu.edu, and all variants and findings are reported using build BDGPR5/dm3. As others have observed, the DGRP lines display small amounts of cryptic genetic relatedness (Huang *et al.* 2014; He *et al.* 2014). Here, we used GEMMA (Zhou and Stephens 2012) to both estimate a centered genetic relatedness matrix (GRM), accounting for cryptic relatedness, and implement the univariate mixed linear model (MLM). PC1 from each of the two experimental populations (CyO and $Df(2L)2802$ second chromosome backgrounds) was used as separate input, providing two independent sources of association values. Males were regressed on each variant using single marker association (SMA). Variants were treated as a fixed effect and the GRM was included as a random effect. Effect size was determined using the function, `pes`, within the R package `compute.es` (Del Re 2013). 107 lines were used with a CyO second chromosome background and 109 lines were used with a $Df(2L)2802$ second chromosome background. Testing was performed across 775,689 (CyO) and 928,587 ($Df(2L)2802$) bi-allelic variants (SNPs and indels) with a MAF of 0.05 or greater. Due to the high correlation between phenotype and top associations of the two experimental populations, subsequent analysis was performed using only the $Df(2L)2802$ GWA data. Bootstrap analysis was used to generate expected site class frequencies. Site classes were counted from 1,000 randomly selected common variants, and an expected distribution was achieved through 10,000 iterations of resampling. For variants with more than one annotated class, only one class was selected and priority was ranked as follows; nonsynonymous > ncRNA > synonymous > UTR > intronic > intergenic. The `prop.test` in R was used to assess the proportion of exonic variants that were nonsynonymous between observed counts in the variants with top associations and expected counts through sampling.

Candidate gene analysis

Candidate genes were selected based on known involvement in chromatin modification. Genes were identified in Flybase using the term “Modifier of Variegation.” Location of each gene was extracted from Flybase and SNPs within the gene and +/- 2 kb of the gene were examined using the above mixed linear model for each SNP. Bootstrap analysis was used to generate an expected Site Frequency Spectrum. Proportions of variant frequencies were tallied by selecting 1,000 variants at random and then binning into allele frequency groups of 0.05. An expected distribution was achieved through 10,000 iterations of resampling. Bootstrap analysis was used to generate an expected distribution of segregating sites with sets of genes. Total number of segregating sites (within +/- 2 kb) were counted within sets of 105 randomly selected genes throughout the autosomal genome and normalized by number of base pairs summed across the full set. The expected distribution of proportion of segregating variants was then achieved through 10,000 iterations of resampling.

Genomic analysis

The 9-state genome-wide combinatorial chromatin state annotation is described in Kharchenko *et al.* (2011) and annotation files were sourced from www.modencode.org. Expected 9-state distributions were generated in the experimental population through sampling autosomal variants for state assignment. 1,000 variants were randomly selected and chromatin states were counted. This process was repeated 100 times. ChIP-chip and ChIP-seq files were also downloaded from ModENCODE (www.modencode.org). If replicate samples existed, only one file was randomly selected for analysis and composite files, if they were made available, were used instead of individual samples. Comparison between the expected distribution of variants within binding sites and variants enriched with associations to PEV was performed as described above. 1,000 autosomal variants were selected at random and variants within binding sites were counted. The distribution of counts, as generated through 10,000 iterations, was then compared to observed counts from 1,000 of the top P -value ranked associating variants. A full list of factors with respective ModENCODE IDs and observed and expected counts has been made available (File S3).

Data Availability

Original eye images are available upon request. File S1 contains PEV phenotype values. File S2 contains a full list of variants with respective association P -values. File S3 contains observed and expected counts of top PEV associations within annotated chromatin features.

Results

Image analysis captures multidimensional color space

Pigments in the eye of *D. melanogaster* are synthesized by two, well-characterized metabolic pathways (Summers *et al.* 1982). These pigments are typically quantified through separate extractions based on chemical properties (Ephrussi and Herold 1944). Often, only a single extracted pigment is used to describe eye color, and, although useful for detecting general differences, this method results in considerable loss of information. Even casual inspection of eye color patterns that manifest PEV reveals a far more complex range of differences, including pigment intensity, different hues of pigmentation including yellow, orange, brown and red, and variation in patch size and morphology. To

better capture the multidimensional aspects of eye color and PEV, and improve mapping, we developed an imaging method that retains and fully describes eye color within a single assay (see methods). Eye-color stocks (Figure 1A) were imaged and Principal Component Analysis (PCA) using the values from the individual red, green and blue color channels, of each image, sufficiently described the multivariate data. Non-pigmented (*white*) and pigmented stocks (*bw¹*, *w^{m4}*, and Canton-S) exhibited large differences described by principal component 1 (PC1), and pigmented stocks (*bw¹*, *w^{m4}*, and Canton-S) exhibited differences primarily described across principal component 2 (PC2) (Figure 1B). PC1 accounted for 87.4% (s.d. = 24.4%) of variance in the data and PC2 accounted for 12.5% (s.d. = 9.2%). Component loadings provide detail on how each color channel contributes to the dispersion of the data, where blue and green channels have a similar impact on PC1 (-0.72 and -0.68, respectively) while the red channel has a minor impact (-0.12). This is in contrast to PC2, where the red channel is the primary driver (0.94) and blue and green channels have minor roles (-0.30 and 0.15, respectively). MANOVA using PC1 and PC2 from each individual image highlights the ability of this approach to discriminate across eye groups of the four stocks (P -value < 2.2×10^{-16}).

Natural variation in background genetic effects on *white-mottled-4* expression

To quantify natural variation in PEV, we made use of the *Drosophila* Genetic Reference Panel (DGRP). The PEV phenotype was expressed by crossing males from inbred DGRP lines to virgin females carrying the *white-mottled-4* (*w^{m4}*) allele on the X chromosome. F1 variegating males were identical across a single X, third and fourth chromosome, segregating according to one of two second chromosomes and varying with respect to a full haplotype from each of the DGRP lines assayed (Figure S1). The two second chromosomes differ primarily with respect to an approximate 200 kb deletion in 25F2-25F5 (*Df(2L)2802*) on the non-balancing chromosome and an inversion on the balancing chromosome (*CyO*). Experimental F1 progeny exhibited a wide range of eye pigmentation differences, showing variation that spanned a complete lack of pigmentation to eyes that were heavily pigmented (Figure S2). The quantitative image assay further detailed a broad phenotypic spread in PEV with PC1 and PC2 values spanning between an eye mutant that lacks pigmentation (*white*) and mutants of known pigment deficiencies (Figure 1B). Reapplying PCA to just the mean red, green and blue color values of images from each of the F1 variegation males, indicates that PC1 captures the vast majority of variance in the experimental data, 97.4% (s.d. = 18.6%), while PC2 and PC3 only describe a small proportion of variance, 2.4% (s.d. = 2.9%) and 0.2% (s.d. = 0.9%). Using PCA on color images of PEV individuals, effectively allows the simplification of a multivariate data source to a single describing variable, PC1, with minimal, 2.6%, loss of data and provides a robust univariate phenotype for association mapping.

ANOVA using individual PCs provides an assessment of importance ascribed to each of three experimental variables; among line differences, second chromosome differences and replicate environments (Table 1). Adjusted values show that combined genetic components explain 79.6% of the phenotypic variance, where second chromosome differences separately explained 49.5% of the variance, among-line variation (our source of natural variation) explained 27.4% of the variance and 2.7% of the variance was explained through genetic interactions of

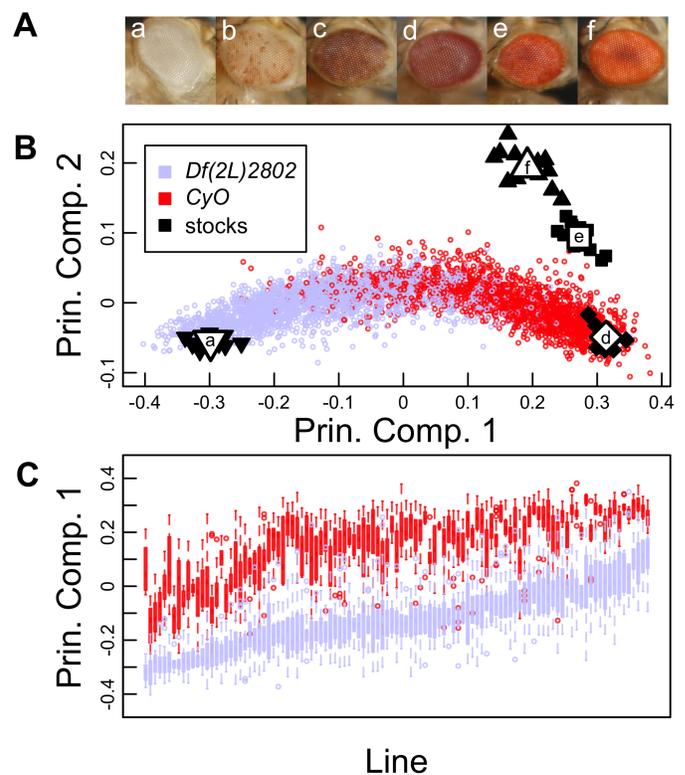


Figure 1 (A) Examples of imaged eyes from various *Drosophila melanogaster* alleles; (a) *w¹¹¹⁸*, (b) *w^{m4}* with the *Df(2L)2802* second chromosome, (c) *w^{m4}* with the *CyO* second chromosome, (d) *bw¹*, (e) *w^{m4}* with *Df(2L)2802/CyO*, and (f) Canton-S. (B) Scatter plot of PC1 and PC2 values for mutant and experimental PEV eyes. Black and white points represent individual images and average values of eye colors from stocks; (a) *w^{m4}*, (d) *bw¹*, (e) *w^{m4}* with *Df(2L)2802/CyO*, (f) Canton-S. Experimental individuals with just the *Df(2L)2802* (periwinkle), in general, show greater variegation and eyes that are closer to the *w¹¹¹⁸* allele (a), an eye that lacks pigmentation. Experimental siblings with the *CyO* allele (red) show less variegation, or more pigmentation. (C) Boxplot of PEV summarized by line (x-axis) and separated by second chromosome, *Df(2L)2802* (periwinkle) vs. *CyO* (red). PEV is represented PC1 (y-axis), where lower values indicate less pigmented eyes and higher values indicate more pigmented eyes. Among-line variance is greater than within-line variance, suggesting natural genetic variation is involved in observed differences in PEV.

second chromosome background and individual lines. Partitioning the sample by presence/absence of the second chromosome deficiency provides two separate measures of broad-sense heritability (H^2) given different genetic backgrounds. Among-line differences explained over half of the phenotypic variance, 59.4% and 57.4%, for each of the populations with *CyO* and *Df(2L)2802* second chromosome backgrounds (Table S1). These data suggest that the bulk of the observed variation in PEV is attributed to segregating genetic variants among the naturally derived DGRP haplotypes.

When comparing second chromosome backgrounds, lines showed strong positive correlation in PEV between PC1 values (Pearson correlation of 0.84), where the *Df(2L)2802* second chromosome acts a clear E(var) with respect to the *CyO* second

Table 1 Partitioning the variance in PEV attributed to genetic and environmental factors

| Variable | Principal Component 1 | | Principal Component 2 | | Adjusted ^a |
|--------------------------------|------------------------|-------------------------|------------------------|-------------------------|------------------------|
| | Proportion of variance | <i>P</i> -value | Proportion of variance | <i>P</i> -value | Proportion of variance |
| Among line | 27.7% | < 1.0x10 ⁻²² | 15.3% | < 1.0x10 ⁻²² | 27.4% |
| Second chromosome | 50.9% | < 1.0x10 ⁻²² | < 0.1% | 0.88 | 49.5% |
| Vial | < 0.1% | 0.33 | < 0.33% | 1.6x10 ⁻⁶ | < 0.1% |
| Among line x second chromosome | 1.9% | < 1.0x10 ⁻²² | 31.2% | < 1.0x10 ⁻²² | 2.7% |
| Within line and Residuals | 19.5% | - | 53.2% | - | 20.4% |

^a Adjusted values are calculated by summing across each specific variable as weighted by the proportion of the variance explained by each PC; 97.4% (PC1), 2.4% (PC2) and 0.2% (PC3). PC3 is not shown above as it explains a negligible amount of the variance in the data.

chromosome balancer; showing greater variegation, or less pigmented eyes, in nearly all lines (Figure S3). Although not the focus of this study, it is important to recognize that individuals differing only by second chromosome accounted for almost half (49.5%) of the phenotypic variance observed, considerably more than explained through natural variation. The consequences of this highlight two potential scenarios. First is the possibility of an unannotated mutation in either a Su(var) or E(var) between the second chromosomes. A second possibility is that the totality of the deletion on *Df(2L)2802*, an approximate 200 kb deletion, acts to enhance variegation through a sponge or sink model, similar to what is proposed to occur with the Y chromosome (Francisco and Lemos 2014); suggesting general amounts of DNA (or chromatin) can act to suppress or enhance the PEV phenotype. We cannot distinguish these possibilities.

Genome-wide Association testing

To assess the contribution of segregating variants to PEV, single marker analysis (SMA) was performed genome-wide using variegating F1 males from crosses between DGRP lines and the *w^{m4}* reporter. Full haplotypes from each of the distinct DGRP lines provided source variation for association mapping. PC1 from each of the two second chromosome populations were used as separate input into a univariate mixed linear model (MLM) accounting for cryptic relatedness (He *et al.* 2014). Testing was performed across common bi-allelic variants (SNPs and indels, MAF ≥ 0.05) using haplotypes extracted from DGRP chromosomes 2, 3 and 4. The X chromosome, carrying the *w^{m4}* reporter, was invariant across experimental populations. Quantile-quantile (Q-Q) plots indicate *P*-values from each experimental population overall conform well to the null distribution (Figure S4). Effect sizes follow a trend where Cohen's *d* increases as MAF decreases (Figure S5). A comparison of the *P*-value rank ordered 1,000 top associations shows 75.3% overlap between the two experimental populations, consistent with a strong correlation in PEV between the two groups. The top-ranked associations are enriched for variants that are located in exons, UTRs and ncRNAs, with reduced representation within intronic and intergenic regions (Figure S6). A comparison of exonic sites further shows no significant difference between proportions of nonsynonymous variants in top associations compared to expected (proportion test, *P*-value = 1), suggesting a broad influence to functional elements of the genome and not strictly toward nonsynonymous variants. In addition, of the 10 variants with the smallest *P*-values in both sets of associations, only two

variants were identified as resulting in a missense mutation and both variants showed weakened association between the independent GWAS samples. Among the smallest *P*-values, only two variants were identified as resulting in a missense mutation and both showed reduced association across independent GWAS samples (Figure S2). Despite many variants identified as having an effect size greater than or equal to 1 and with enrichment proximal to functional regions, our small sample sizes, substantial background effect on phenotype, as noted through second chromosome differences, and abundance of multiple variant classes in top associations reduce confidence in traditional functional follow-up. A full list of variants with association *P*-values has been made available (File S2)

Common variants within known autosomal PEV modifiers fail to fully account for among-line differences

Despite low power to identify individual causal variants with high confidence, an extensive literature on known genic modifiers of PEV affords the opportunity to assess significance of ensembles of variants. To quantify the impact of naturally occurring polymorphism in known modifiers of PEV, we identified variants in Su(var) and E(var) genes. The term, "Modifier of Variegation" was searched within FlyBase, and over 200 genes satisfied this criterion. As our experimental setup resulted in individuals sharing a common X chromosome, the set of modifiers was reduced to 105 autosomal genes (Table S3). Variants within, or extending ±2 kb of identified autosomal modifiers were grouped and results from the above SMA were used. 16,640 total variants (6,153 having MAF ≥ 0.05) were identified in autosomal modifiers of PEV within the 109 lines having the *Df(2L)2802* second chromosome background. Importantly, of common variants (MAF ≥ 0.05) in this reduced set of PEV genes, only 5 were also identified in the top 1,000 genome-wide SMA hits, making up less than 0.5% of top associated variants. These five variants hold overall *P*-value ranks of 379, 553, 772, 820 and 821, indicating that a majority of natural variants with likely impact to phenotypic variance of PEV do not reside in or near genes with known impact to PEV. GCTA was used to assess cumulative statistical explanation of phenotypic variance (Yang *et al.* 2011). Top variants identified through GWA, rank-ordered by *P*-value, explained a far greater proportion of among-line variance than variants within known genic modifiers (Figure 2 and Figure S7). The 1,000 top ranking overall variants statistically explained 93.7% of variance due to genetic differences and the 1,000 top ranking variants in known PEV genes explained considerably

less at 56.0%. These sets are both compared to variants randomly drawn from the autosomal genome which explained, on average, 2.3% (standard deviation = 2.7%) of phenotypic variance.

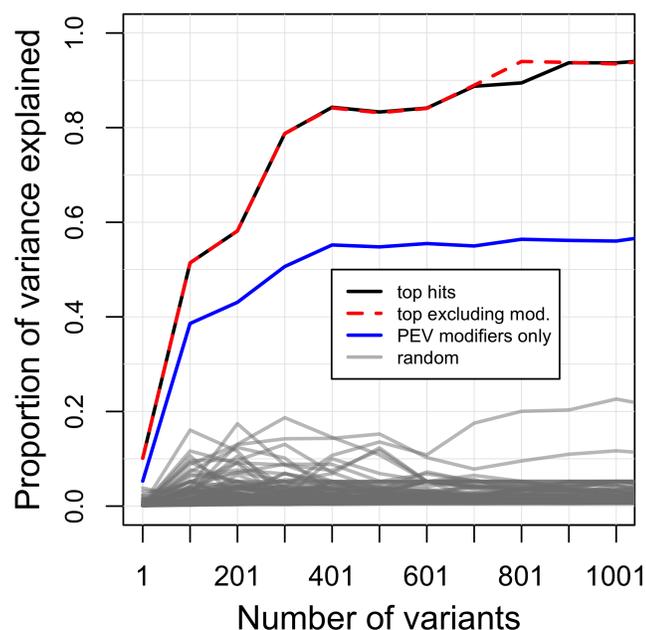


Figure 2 Proportion of between-line phenotypic variance explained within the *Df(2L)2802* second chromosome population using GCTA. Comparisons between SNP groupings include; the most significant GWA variants (black line), top hits excluding variants from known PEV modifiers (red line), variants within known PEV modifiers only (blue line), and randomly selected autosomal variants (gray line).

Although there is strong evidence for involvement of genes from table S3 in PEV, there is little data indicating natural variation within these genes is responsible for differences in PEV among lines. This is not completely surprising however, as many known *Su(var)* and *E(var)* genes show conservation across species (Fodor *et al.* 2010), suggesting little room for variation in coding sequence. Indeed, two additional pieces of data further explain the lack of association with variants in known PEV modifiers and suggest purifying selection within the modifiers. The observed site frequency spectrum (SFS) within the experimental population shows an increase in low MAF variants and a decrease in variants ≥ 0.05 , compared to sets of genes randomly selected from the autosomal genome (Figure 3A). Finally, known PEV modifiers exhibit a paucity of segregating sites compared to an expected distribution, having fewer segregating sites than 99.7% of gene sets randomly selected from autosomes (Figure 3B).

Chromatin and feature enrichment of top associations

As we see little evidence for association within known *Su(var)* and *E(var)*s, we then ask if other genomic features show an over-representation of association with PEV. Chromatin features correlate with functional elements of the genome (modENCODE Consortium *et al.* 2010), and are a natural link to PEV. We next asked if particular chromatin states were over or underrepresented in the top *P*-value ranked associations. Using a previously built combinatorial 9-state (c1-c9) chromatin assignment

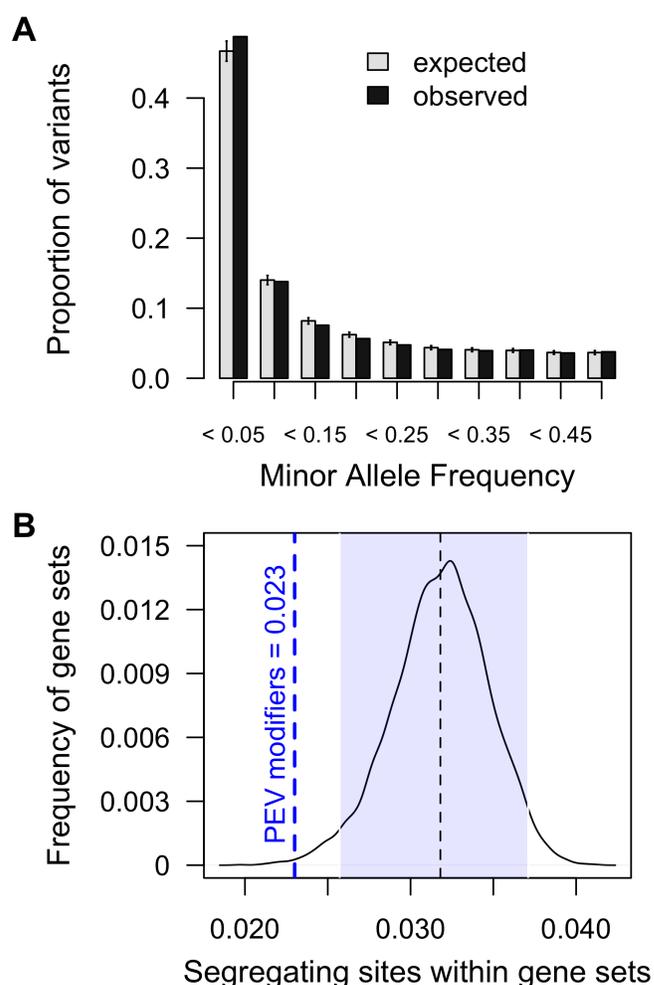


Figure 3 Variants within known autosomal PEV modifiers compared to expected distributions. (A) Site Frequency Spectrum (SFS) of variants within known autosomal modifiers of PEV (black) and variants within sets of randomly selected autosomal genes (gray). Error bars reflect the standard deviation. (B) Proportion of known autosomal PEV modifiers that contain segregating sites compared to sets genes drawn randomly 10,000 times. The shaded area highlights 95% of the expected distribution. 99.7% of randomly selected gene sets contain a greater proportion of segregating sites than 105 known genic PEV modifiers.

from S2 and BG3 cells (Kharchenko *et al.* 2011), we queried all chromatin states for an over-representation of associations to PEV. First, we generated an expected proportion of chromatin states using all autosomal variants (Figure S8). We then selected the top 1,000 SMA variants, as rank-ordered by *P*-value, and compared the represented chromatin state proportions to the genome-wide distribution. We found an over-representation of chromatin state c1 using data from both BG3 and S2 cells (> 1 standard deviation) and lowered representation of chromatin state c4, c6 and c7 (< 1 standard deviation). Chromatin state c1 is described as representing active promoter and transcriptional start site (TSS)-proximal regions and consists of regions having H3K4me2, H3K4me3 and H3K9ac histone marks.

Given that chromatin states are correlated with factors that influence function across the genome, we next performed an

unbiased query of individual ChIP-chip and ChIP-seq data also made publicly available through the modENCODE project (modENCODE Consortium *et al.* 2010; Kharchenko *et al.* 2011). This dataset is comprised of hundreds of factors sampled across a full range of developmental stages, with a bulk of the data coming from S2 and BG3 cells. Again, for each chromatin factor with prior data, we generated the null distribution of variants expected to exist within each site if randomly drawn from the all autosomes (Figure S9). Variants from the top ranked associations were then compared to expected and assessed for enrichment. We note that several features displayed a strong enrichment with PEV associated variants, including histone modifications, transcription factors and non-transcription factors. We repeated these tests for an over-representation of chromatin state, c1, where sites from all three histone marks, H3K4me2, H3K4me3 and H3K9ac are enriched with top associations (Figure 4). We also observed enrichment for binding sites of known PEV factors, such as JIL-1 (Lerach *et al.* 2006), LSD1 (Di Stefano *et al.* 2007), BEAF-32 (Gilbert *et al.* 2006), among others. Further, we note enrichment of top ranked associations in sites that suggest natural variation has a particular impact on TSS regions that show a "balanced" or "bistable" chromatin state. Bistable chromatin sites are sites that may be influenced to either exhibit active or repressed gene expression. We note statistically significant differences in nearly all bound factors that strongly characterize bistable sites; ASH1, H3K4me1, H3K4me2, H3K4me3, RNA Pol II, including depletion of H3K27me3 (Kharchenko *et al.* 2011). The over-representation of associations in sites bound by known chromatin modifiers suggests that natural autosomal variation interacts with chromatin dynamics and PEV through influencing genome-wide expression rates or chromatin state occupancy and balance, not through altering protein function of individual genic modifiers.

Remarkably, these findings have precedent in the PEV system and fit extremely well with prior findings indicating a key driver behind mosaic features of variegation is a bistable equilibrium between TF binding and heterochromatin content (Ahmad and Henikoff 2001). Here, it was found that by simply varying levels of a GAL4 transcriptional activator to a heterochromatin-embedded promoter, heterochromatin state could be disrupted. In the model proposed, termed The Site Exposure Model of Variegated Silencing (Widom 1999; Ahmad and Henikoff 2001), kinetics of DNA-histone contact dictate ability of a TF to bind a promoter or enhancer feature, and thus influence gene expression. Features that increase contact between TF binding and activator, include changes to TF abundance or changes to TF binding efficiency such as through mutated underlying binding sequence, changes to nucleosome occupancy (observed through histone and chromatin marks), or changes to abundance of TF guide molecules. Applied to our study, this suggests that each individual shows differences in PEV due to a large number of sites that impact expression rates of factors that then impact binding efficiency of TFs at the w^{m4} locus. This model also predicts that changes to chromatin content, i.e. an increase or decrease in heterochromatin, can in turn impact sensitive loci throughout the genome and influence gene expression. Indeed, this fits with observations that differing natural Y chromosomes, a giant source of heterochromatin, impact gene expression in autosomes (Lemos *et al.* 2008, 2010).

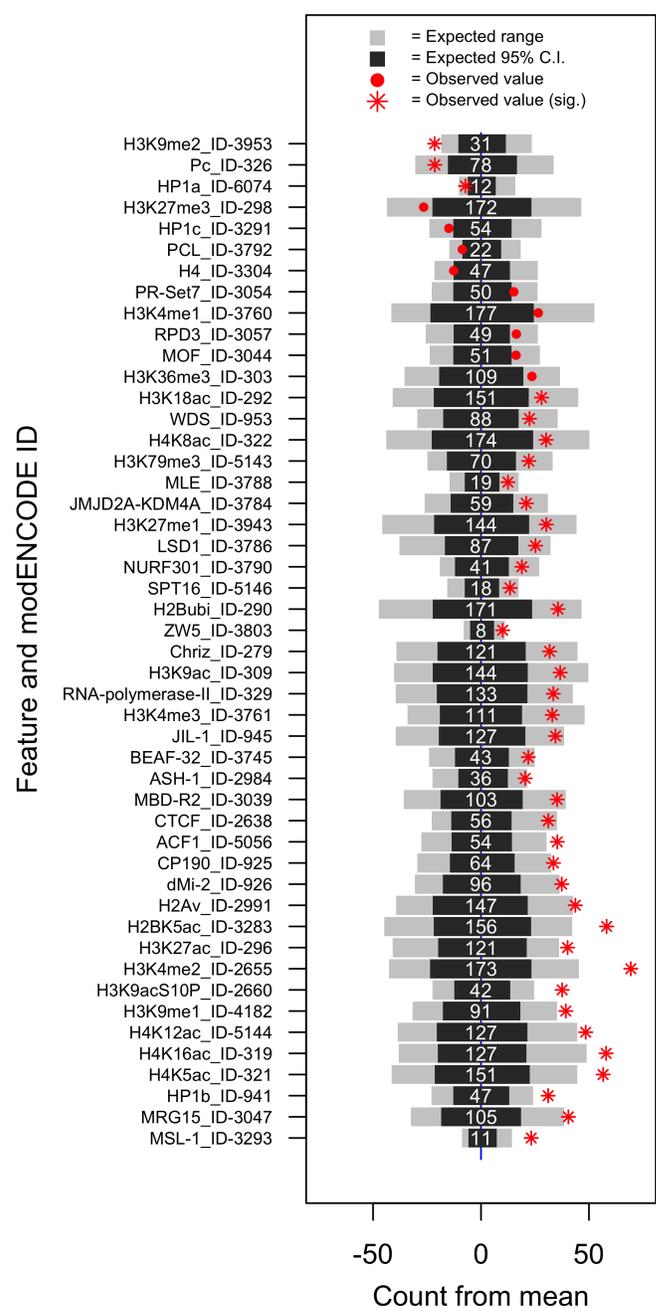


Figure 4 Observed enrichment of variants with association to PEV and chromatin features from S2 cells only. Black bars represent 95% of the expected distribution and gray bars represent the left and right 2.5% tails of the distribution. Numeric values are the average number of variants expected to fall within each chromatin feature. Observed counts from variants within the 1,000 top P -value ranked associations are represented in red. Only features with observed values in the 5% tails of the expected distribution are shown. False Discovery Rate (FDR) was used to assess statistical significance across the multiple tests, and significant values are noted.

Discussion

We designed an assay to identify autosomal non-recessive variants involved in differences in PEV between naturally derived

lines of *Drosophila melanogaster*. Despite large PEV-induced pigmentation differences between phenotyped lines, we find little evidence for involvement of polymorphic sites within known Su(var) and E(var) genes that contribute to these differences in PEV. Our top SMA associations further indicate that natural differences in response to PEV are not the primary result of changes in protein function among lines. We instead find that regions of enriched association to PEV are over-represented for promoter and TSS-proximal regions with an additional emphasis on sites that display bi-stable chromatin features. This suggests that autosomal interactions with differences in PEV, i.e. differences in heterochromatin dynamics, are the combined result of many small effect loci that accumulate differences and are linked to modified rates of transcription, either through small changes to specific TF binding sites or broad changes in chromatin state and chromatin mark distributions. Furthermore, our data fit the Site Exposure Model of Variegated Silencing (Ahmad and Henikoff 2001), where a variegated state is the result of bistable features between TF binding and chromatin features. However, as the effect sizes for all associated variants are small, the combined set of variants likely do not fully explain differences in PEV across our sample set; indicating there is yet unobserved genetic variation that accounts for differences among lines. It is important to note that this query of natural variation was not exhaustive and only considered autosomal variants with a particular focus on common variants. We reasoned that a heterozygous screen would be informative, because most Su(var) and E(var) allelic effects are dominant, but we note there is every reason to believe that recessive genetic modifiers of PEV exist also and were missed in our screen. Importantly, we did not query GxG interactions or Y-linked variants, known contributors to PEV and gene expression differences in natural populations (Lemos *et al.* 2008, 2010). Finally, it is important to consider that the autosomal loci identified here only show correlation with differences in PEV; it is not known at this time if these loci represent drivers or are instead the consequence of differences in heterochromatin dynamics.

Acknowledgments

We thank Dan Barbash and Jason Mezey for thoughtful critique and comments throughout all stages of the work. Several members of the Clark lab provided support through general discussion and expertise, including; Jen Grenier, Clement Chow, Rob Unckless, Julien Ayroles, Angela Early and Grace Chi. This work was supported by R01 GM119125.

Literature Cited

- Ahmad, K. and S. Henikoff, 2001 Modulation of a transcription factor counteracts heterochromatic gene silencing in *Drosophila*. *Cell* **104**: 839–847.
- Chow, C. Y., M. F. Wolfner, and A. G. Clark, 2012 A large neurological component to genetic differences underlying biased sperm use in *Drosophila*. *Genetics* **193**: 177–185.
- Del Re, A. C., 2013 compute.es: Compute effect sizes. R Package <http://cran.r-project.org/web/packages/compute.es>.
- Di Stefano, L., J.-Y. Ji, N.-S. Moon, A. Herr, and N. Dyson, 2007 Mutation of *Drosophila* Lsd1 disrupts H3-K4 methylation, resulting in tissue-specific defects during development. *Current Biology* **17**: 808–812.
- Dobson, A. J., J. M. Chaston, P. D. Newell, L. Donahue, S. L. Hermann, D. R. Sannino, S. Westmiller, A. C. Wong, A. G. Clark, B. P. Lazzaro, and A. E. Douglas, 2015 Host genetic determinants of microbiota-dependent nutrition revealed by genome-wide analysis of *Drosophila melanogaster*. *Nat. Commun.* **6**: 6312.
- Durham, M. F., M. M. Magwire, E. A. Stone, and J. Leips, 2014 Genome-wide analysis in *Drosophila* reveals age-specific effects of snps on fitness traits. *Nat. Commun.* **5**: 4338.
- Ebert, A., G. Schotta, S. Lein, S. Kubicek, V. Krauss, T. Jenuwein, and G. Reuter, 2004 Su(var) genes regulate the balance between euchromatin and heterochromatin in *Drosophila*. *Genes Dev.* **18**: 2973–2983.
- Elgin, S. C. and G. Reuter, 2013 Position-effect variegation, heterochromatin formation, and gene silencing in *Drosophila*. *Cold Spring Harb. Perspect. Biol.* **5**: a017780.
- Ephrussi, B. and J. L. Herold, 1944 Studies of eye pigments of *Drosophila*. I. Methods of extraction and quantitative estimation of the pigment components. *Genetics* **29**: 148.
- Ernst, J. and M. Kellis, 2012 ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**: 215–216.
- Filion, G. J., J. G. van Bommel, U. Braunschweig, W. Talhout, J. Kind, L. D. Ward, W. Brugman, I. J. de Castro, R. M. Kerkhoven, H. J. Bussemaker, and B. vanSteensel, 2016 Systematic protein location mapping reveals five principal chromatin types in *Drosophila* cells. *Cell* **143**: 212–224.
- Fodor, B. D., N. Shukeir, G. Reuter, and T. Jenuwein, 2010 Mammalian Su(var) genes in chromatin control. *Annu. Rev. Cell Dev. Biol.* **26**: 471–501.
- Francisco, F. O. and B. Lemos, 2014 How do Y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. *J. Genomics* **1**: 94–103.
- Gilbert, M. K., Y. Y. Tan, and C. M. Hart, 2006 The *Drosophila* boundary element-associated factors BEAF-32A and BEAF-32B affect chromatin structure. *Genetics* **173**: 1365–1375.
- Girton, J. R. and K. M. Johansen, 2008 Chromatin structure and the regulation of gene expression: the lessons of PEV in *Drosophila*. *Adv. Genet.* **61**: 1–43.
- Greil, F., I. van der Kraan, J. Delrow, J. F. Smothers, E. de Wit, H. J. Bussemaker, R. van Driel, S. Henikoff, and B. van Steensel, 2003 Distinct HP1 and Su(var)3-9 complexes bind to sets of developmentally coexpressed genes depending on chromosomal location. *Genes & Development* **17**: 2825–2838.
- Grubbs, N., M. Leach, X. Su, T. Petrisko, J. B. Rosario, and J. W. Mahaffey, 2013 New components of *Drosophila* leg development identified through genome wide association studies. *PLoS one* **8**: e60261.
- Harbison, S. T., L. J. McCoy, and T. F. Mackay, 2013 Genome-wide association study of sleep in *Drosophila melanogaster*. *BMC Genomics* **14**: 281.
- He, B. Z., M. Z. Ludwig, D. A. Dickerson, L. Barse, B. Arun, B. J. Vilhjálmsson, P. Jiang, S.-Y. Park, N. A. Tamarina, S. B. Selleck, P. J. Wittkopp, and M. Kreitman, 2014 Effect of genetic variation in a *Drosophila* model of diabetes-associated misfolded human proinsulin. *Genetics* **196**: 557–567.
- Hines, K. A., D. E. Cryderman, K. M. Flannery, H. Yang, M. W. Vitalini, T. Hazelrigg, C. A. Mizzen, and L. L. Wallrath, 2009 Domains of heterochromatin protein 1 required for *Drosophila melanogaster* heterochromatin spreading. *Genetics* **182**: 967–977.
- Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Ràmia, A. M. Tarone, L. Turlapati, T. Zichner, D. Zhu, R. F. Lyman, *et al.*, 2014 Natural variation in genome architecture among 205 *Drosophila melanogaster* genetic reference panel lines. *Genome*

- Research **24**: 1193–1208.
- Kharchenko, P. V., A. A. Alekseyenko, Y. B. Schwartz, A. Minoda, N. C. Riddle, J. Ernst, P. J. Sabo, E. Larschan, A. A. Gorchakov, T. Gu, *et al.*, 2011 Comprehensive analysis of the chromatin landscape in *Drosophila melanogaster*. *Nature* **471**: 480–485.
- Lamprecht, M. R., D. M. Sabatini, A. E. Carpenter, *et al.*, 2007 Cell-profiler™: free, versatile software for automated biological image analysis. *Biotechniques* **42**: 71.
- Lemos, B., L. O. Araripe, and D. L. Hartl, 2008 Polymorphic Y chromosomes harbor cryptic variation with manifold functional consequences. *Science* **319**: 91–93.
- Lemos, B., A. T. Branco, and D. L. Hartl, 2010 Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proc. Nat. Acad. Sci. USA* **107**: 15826–15831.
- Lerach, S., W. Zhang, X. Bao, H. Deng, J. Girton, J. Johansen, and K. M. Johansen, 2006 Loss-of-function alleles of the JIL-1 kinase are strong suppressors of position effect variegation of the *w^{m4}* allele in *Drosophila*. *Genetics* **173**: 2403–2406.
- Mackay, T. F., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, S. Casillas, Y. Han, M. M. Magwire, J. M. Cridland, *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. *Nature* **482**: 173–178.
- modENCODE Consortium, T., S. Roy, J. Ernst, P. V. Kharchenko, P. Kheradpour, *et al.*, 2010 Identification of functional elements and regulatory circuits by *Drosophila* modENCODE. *Science* **330**: 1787–1797.
- Muller, H. J., 1930 Types of visible variations induced by x-rays in *Drosophila*. *Journal of Genetics* **22**: 299–334.
- R Development Core Team, 2008 *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0.
- Schotta, G., A. Ebert, V. Krauss, A. Fischer, J. Hoffmann, S. Rea, T. Jenuwein, R. Dorn, and G. Reuter, 2002 Central role of *Drosophila* Su(var)3-9 in histone H3-K9 methylation and heterochromatic gene silencing. *The EMBO journal* **21**: 1121–1131.
- Schulze, S. R. and L. L. Wallrath, 2007 Gene regulation by chromatin structure: Paradigms established in *Drosophila melanogaster*. *Annu. Rev. Entomol.* **52**: 171–192, PMID: 16881818.
- Summers, K. M., A. J. Howells, and N. A. Pyliotis, 1982 Biology of eye pigmentation in insects. *Advances in Insect Physiology* **16**: 119–166.
- Unckless, R. L., S. M. Rottschaefer, and B. P. Lazzaro, 2015 A genome-wide association study for nutritional indices in *Drosophila*. *G3* **5**: 417–25.
- Wallrath, L. L. and S. C. Elgin, 1995 Position effect variegation in *Drosophila* is associated with an altered chromatin structure. *Genes & Development* **9**: 1263–1277.
- Widom, J., 1999 Equilibrium and dynamic nucleosome stability. *Methods Mol. Biol.* **119**: 61–77.
- Yang, J., S. H. Lee, M. E. Goddard, and P. M. Visscher, 2011 GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics* **88**: 76–82.
- Zhou, X. and M. Stephens, 2012 Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: 821–824.

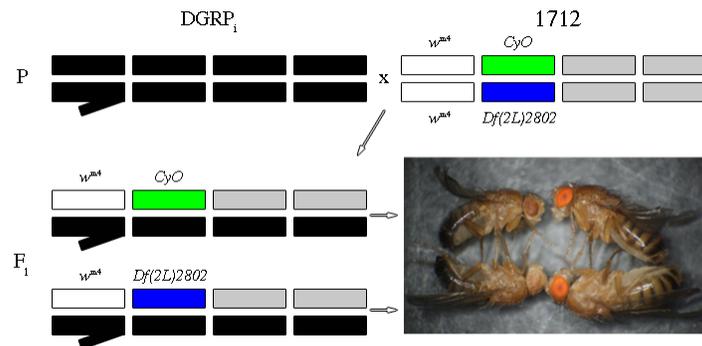


Figure S 1 Experimental cross to generate males with unique DGRP haplotypes, *i*, that also carry *w^{m4}* and exhibit PEV. F1 males segregate into two variegating classes, those with a *CyO* second chromosome (visible curly wing) and those with *Df(2L)2802* second chromosome (visible straight wing). F1 males were imaged. Females do not display variegation in this cross and were not studied.

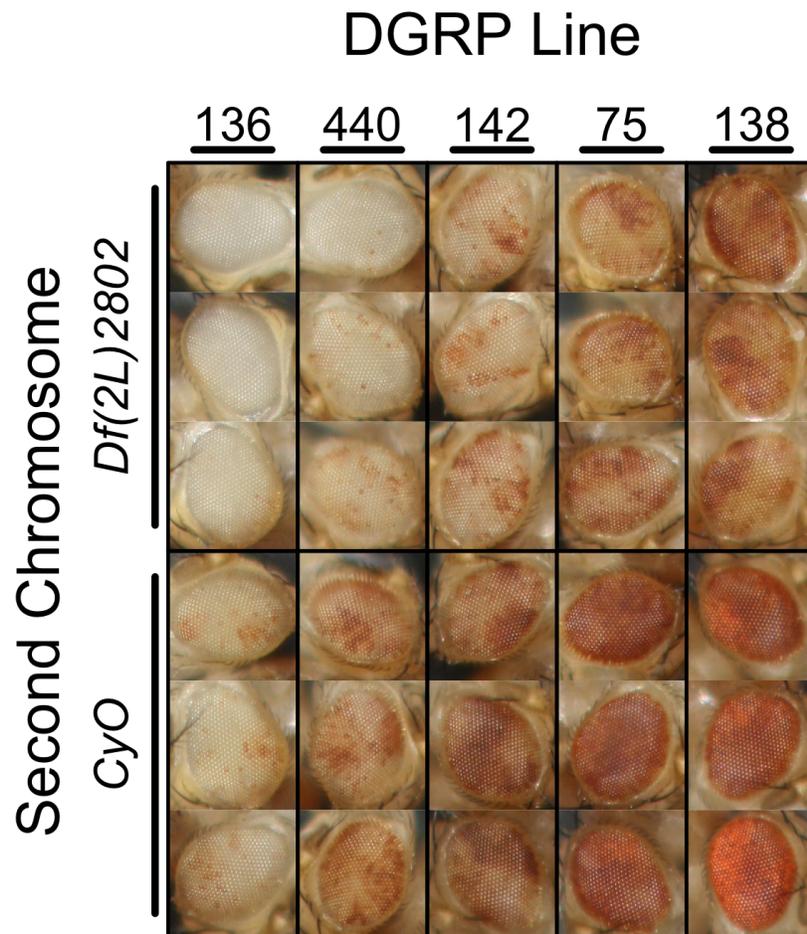


Figure S 2 F1 experimental males show a wide range of visible differences in eye pigmentation, spanning a range that lacks pigmentation (DGRP line 136 with a *Df(2L)2802* second chromosome background) to full pigmentation and near-wild-type eyes (DGRP line 138 with a *CyO* second chromosome background). To demonstrate the range of pigmentation in F1 progeny, three replicates are shown from a subset of DGRP line by second chromosome combinations. Even at a gross level, the PEV phenotype shows greater similarities in pigmentation between replicate individuals than between DGRP line or second chromosome background.

Table S 1 Heritability of PEV

| Variable | CyO background | | | | Df(2L)2802 background | | | |
|-----------|----------------|-------|--------|-----------------------|-----------------------|-------|-------|-----------------------|
| | PC1 | PC2 | PC3 | Adjusted ^a | PC1 | PC2 | PC3 | Adjusted ^b |
| Line | 59.8% | 45.3% | 44.0% | 59.4% | 57.7% | 47.8% | 55.8% | 57.4% |
| Vial | < 0.1% | 0.5% | < 0.1% | < 0.1% | < 0.1% | 0.2% | 0.2% | < 0.1% |
| Residuals | 40.2% | 54.3% | 56.0% | 40.6% | 42.3% | 52.0% | 44.0% | 42.6% |

^a Adjusted values are calculated by summing across each specific variable as weighted by the proportion of data explained by each PC; 97.4% (PC1), 2.4% (PC2) and 0.2% (PC3). PC3 is not shown above as it explains a negligible amount of data.

^b See footnote ^a

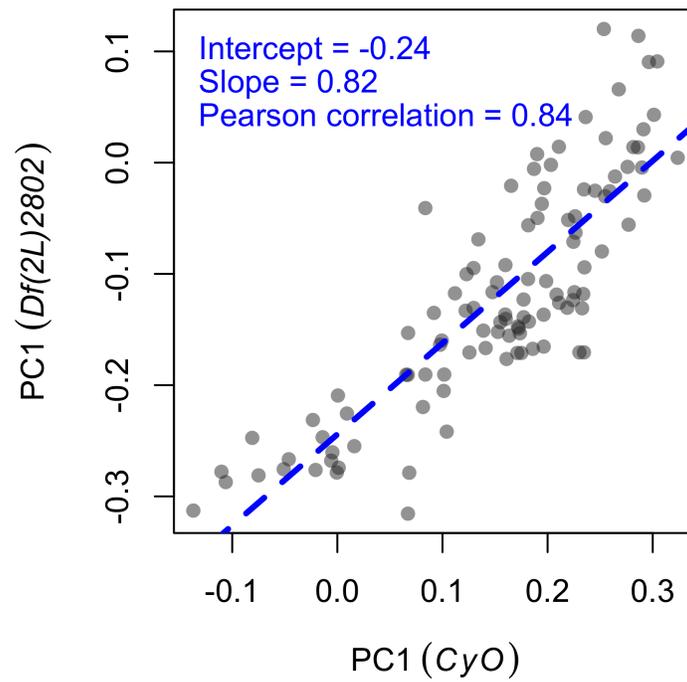


Figure S 3 Comparison between PC1 values of PEV populations with *CyO* and *Df(2L)2802* second chromosome backgrounds. Although there is strong correlation in PEV between the two backgrounds, the *CyO* second chromosome shows less variegation (more red eyes and higher PC1 values) with respect to the *Df(2L)2802* second chromosome background.

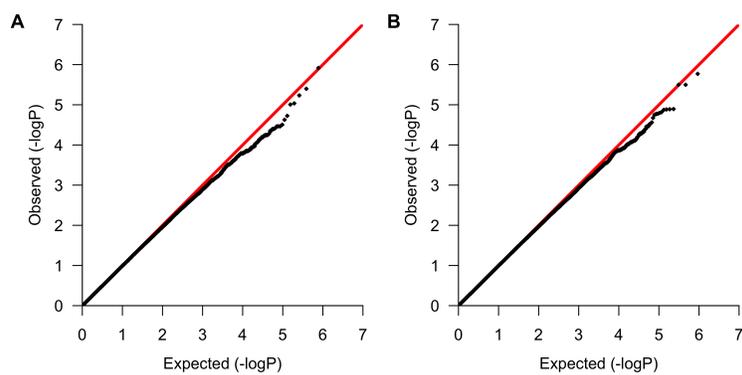


Figure S 4 Q-Q plots of expected (red line) and observed (black points) P -values from GWAS. **(A)** Experimental population with *CyO* second background. **(B)** Experimental population with *Df(2L)2802* second chromosome background.

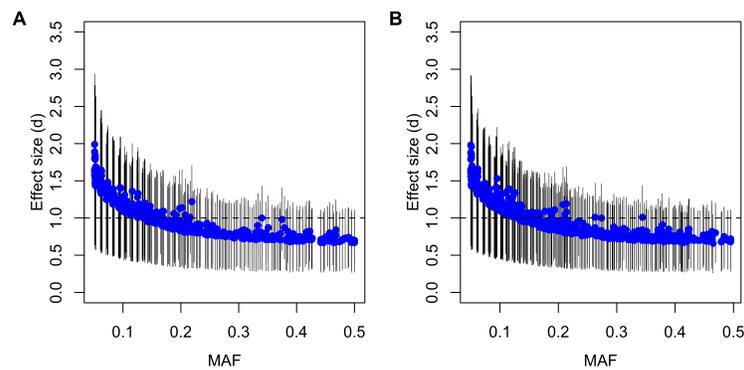


Figure S 5 Effect size (Cohen's d) with upper and lower confidence limits of variants having association P -values ≤ 0.001 . **(A)** Effect sizes of 584 variants from the experimental population with *CyO* second chromosome background. **(B)** Effect sizes of 770 variants from the experimental population with *Df(2L)2802* second chromosome background. 310 and 369 variants from each respective background have effect sizes greater than or equal to 1.

Table S 2 GWAS top ranked associations by *P*-value

| Position | Alleles | Gene | Class | CyO background | | | <i>Df(2L)2802</i> background | | |
|-------------|---------|----------------|------------|----------------|----------------------|------|------------------------------|----------------------|------|
| | | | | AF | <i>P</i> -value | Rank | AF | <i>P</i> -value | Rank |
| 2L 12252172 | T/C | <i>Aret</i> | Intron | NA | NA | NA | 0.095 | 1.3x10 ⁻⁵ | 5 |
| 2R 11802982 | C/T | <i>Sli</i> | Intron | NA | NA | NA | 0.274 | 1.6x10 ⁻⁵ | 9 |
| 2R 12408574 | T/C | <i>Sema-2a</i> | Intron | 0.126 | 1.8x10 ⁻⁵ | 6 | 0.124 | 1.7x10 ⁻⁵ | 10 |
| 2R 18531570 | T/C | NA | Intergenic | 0.051 | 3.1x10 ⁻⁵ | 8 | 0.05 | 3.5x10 ⁻⁵ | 19 |
| 3L 1775145 | A/G | <i>CG13933</i> | 5'UTR | 0.34 | 5.8x10 ⁻⁶ | 3 | 0.344 | 3.2x10 ⁻⁶ | 3 |
| 3L 2081047 | A/G | <i>sls</i> | Synonymous | 0.116 | 2.4x10 ⁻⁵ | 7 | 0.113 | 2.1x10 ⁻⁵ | 13 |
| 3L 2288054 | G/T | <i>DmsR-2</i> | Intron | 0.198 | 9.8x10 ⁻⁶ | 5 | 0.194 | 1.3x10 ⁻⁵ | 6 |
| 3L 2568884 | T/C | <i>msn</i> | Intron | 0.375 | 4.0x10 ⁻⁶ | 2 | 0.378 | 1.3x10 ⁻⁵ | 7 |
| 3L 3595434 | T/G | <i>Eip63E</i> | Intron | 0.219 | 1.2x10 ⁻⁶ | 1 | 0.214 | 1.7x10 ⁻⁶ | 1 |
| 3L 5909455 | T/C | <i>CG13288</i> | Intron | 0.202 | 9.2x10 ⁻⁶ | 4 | 0.208 | 3.2x10 ⁻⁶ | 2 |
| 3L 5909456 | G/C | <i>CG13288</i> | Intron | 0.208 | 3.5x10 ⁻⁵ | 11 | 0.214 | 1.3x10 ⁻⁵ | 4 |
| 3L 12852675 | C/G | <i>CG32107</i> | Nonsyn. | 0.333 | 3.4x10 ⁻⁵ | 9 | 0.337 | 1.2x10 ⁻⁴ | 78 |
| 3L 12852676 | T/C | <i>CG32107</i> | Nonsyn. | 0.333 | 3.4x10 ⁻⁵ | 10 | 0.337 | 1.2x10 ⁻⁴ | 79 |
| 3R 4014584 | C/T | NA | Intergenic | NA | NA | NA | 0.116 | 1.5x10 ⁻⁵ | 8 |

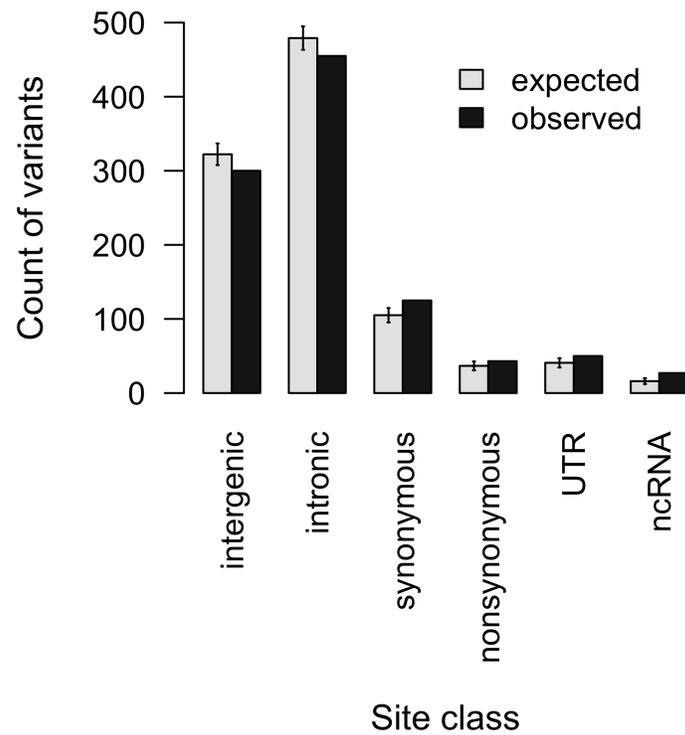


Figure S 6 Distribution of site classes between expected counts (average) and observed counts from the 1000 top associated variants. All observed counts show a significant difference from expected (two-sided, one sample *t*-test, Bonferroni corrected *P*-values all $\leq 2.2 \times 10^{-16}$). Error bars represent the standard deviation across 10,000 randomly drawn samples.

Table S 3 Known autosomal genic modifiers of PEV

| Chromosome | | | | |
|-------------------|---------------------|-------------------|--------------------|--|
| 2L | 2R | 3L | 3R | |
| <i>abo</i> | <i>Asx</i> | <i>AGO2</i> | <i>Ace</i> | |
| <i>aub</i> | <i>Atf-2</i> | <i>asf1</i> | <i>Acf1</i> | |
| <i>barr</i> | <i>BEAF-32</i> | <i>ash1</i> | <i>Bin1</i> | |
| CG8677 | <i>bw</i> | <i>brm</i> | <i>bon</i> | |
| <i>Chd1</i> | <i>Cap-G</i> | <i>E(z)</i> | <i>cav</i> | |
| <i>chm</i> | <i>Dcr-2</i> | <i>FRG1</i> | <i>D1</i> | |
| <i>dp</i> | <i>dom</i> | <i>HP4</i> | <i>Dcr-1</i> | |
| <i>Etl1</i> | <i>Dp1</i> | <i>Jarid2</i> | <i>E(var)3-9</i> | |
| <i>glu</i> | <i>E(Pc)</i> | <i>JIL-1</i> | <i>E2f</i> | |
| <i>Hel25E</i> | <i>egg</i> | <i>msl-3</i> | <i>Fmr1</i> | |
| <i>kis</i> | <i>LamC</i> | <i>not</i> | <i>gpp</i> | |
| <i>lid</i> | <i>lat</i> | <i>Pc</i> | <i>His2Av</i> | |
| <i>lt</i> | <i>Nap1</i> | <i>pzg</i> | <i>Invadolysin</i> | |
| <i>lwr</i> | <i>Pcl</i> | <i>rept</i> | <i>jumu</i> | |
| <i>Mcm10</i> | <i>PCNA</i> | <i>Rpd3</i> | <i>MBD-like</i> | |
| <i>Mt2</i> | <i>Psc</i> | <i>RpLP0</i> | <i>Mekk1</i> | |
| <i>piwi</i> | <i>psq</i> | <i>Sgf11</i> | <i>mod</i> | |
| <i>Psf2</i> | <i>SMC2</i> | <i>sti</i> | <i>mod(mdg4)</i> | |
| <i>r2d2</i> | <i>Su(var)2-10</i> | <i>stwl</i> | <i>mor</i> | |
| <i>Sam-S</i> | <i>Su(var)2-HP2</i> | <i>Su(var)3-3</i> | <i>MRG15</i> | |
| <i>Sir2</i> | <i>Su(z)2</i> | <i>Su(z)12</i> | <i>Orc2</i> | |
| <i>Sos</i> | | <i>SuUR</i> | <i>osa</i> | |
| <i>Su(var)205</i> | | <i>Trl</i> | <i>polybromo</i> | |
| <i>vig</i> | | <i>Ubp64E</i> | <i>Pp1-87B</i> | |
| | | | <i>pr-set7</i> | |
| | | | <i>puc</i> | |
| | | | <i>Rga</i> | |
| | | | <i>Sce</i> | |
| | | | <i>Scm</i> | |
| | | | <i>Snr1</i> | |
| | | | <i>spn-E</i> | |
| | | | <i>Su(var)3-7</i> | |
| | | | <i>Su(var)3-9</i> | |
| | | | <i>trx</i> | |
| | | | <i>vig2</i> | |
| | | | <i>XNP</i> | |

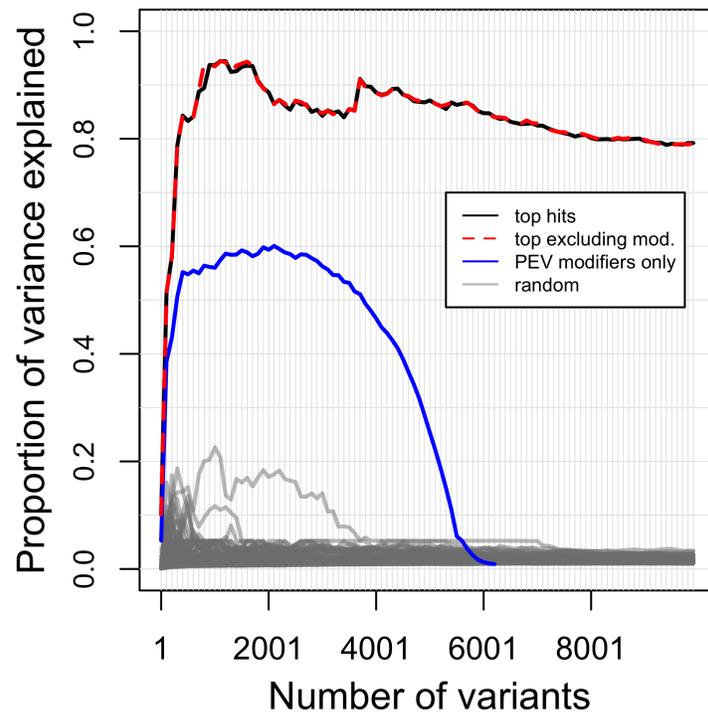


Figure S 7 Proportion of between-line phenotypic variance explained within the *Df(2L)2802* second chromosome population using GCTA. Comparison of SNP groupings include; the most significant GWA variants (black line), top hits excluding variants from known PEV genic modifiers (red line), variants within know PEV genic modifiers only (blue line), and randomly selected autosomal variants (gray line).

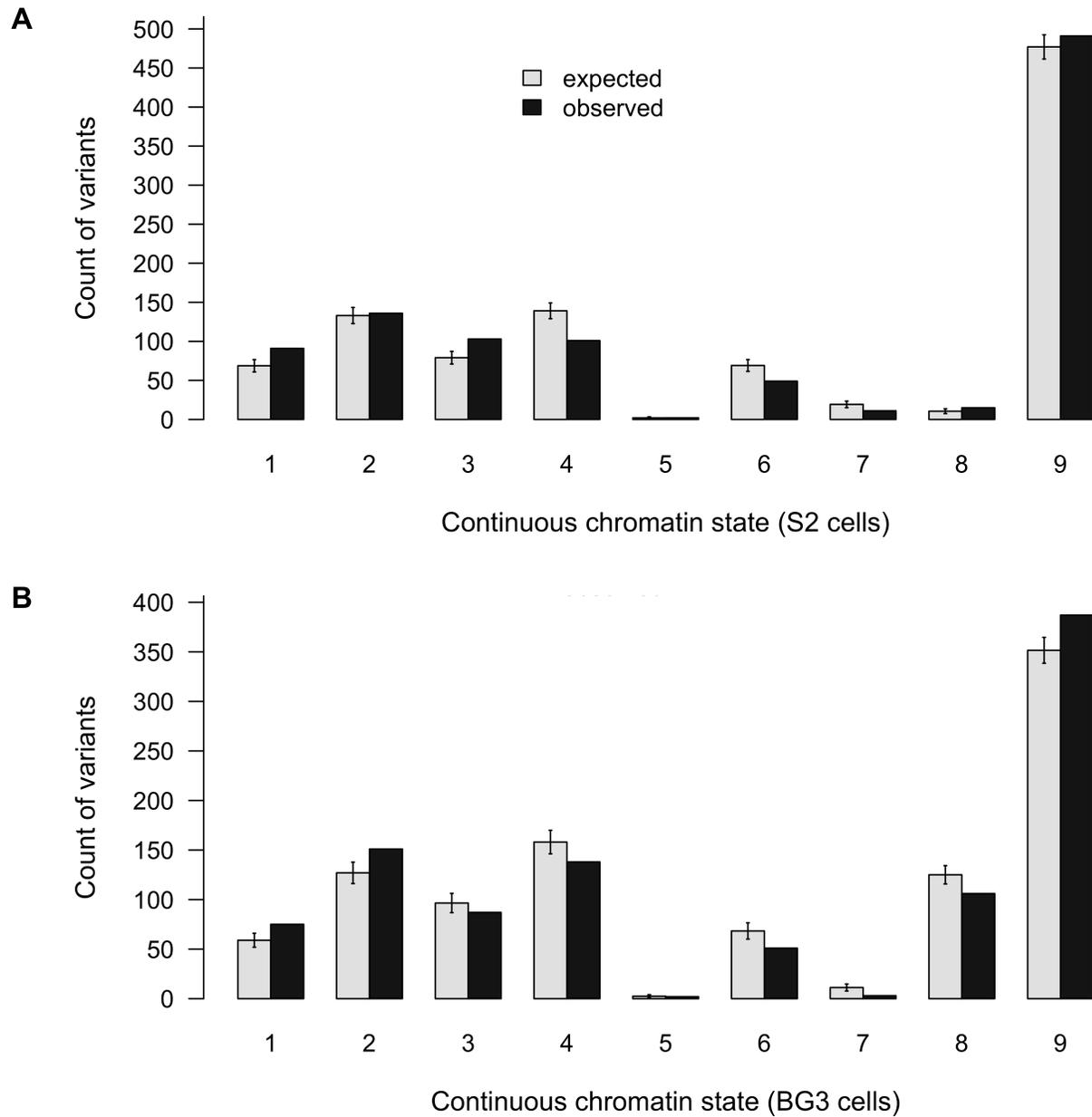


Figure S 8 Counts of continuous 9-state chromatin assignments in randomly select autosomal variants (expected) and the top 1,000 *P*-value ranked SMA variants (observed). **(A)** Counts of chromatin state assignments in S2 cells. **(B)** Counts of chromatin state assignments in BG3 cells.

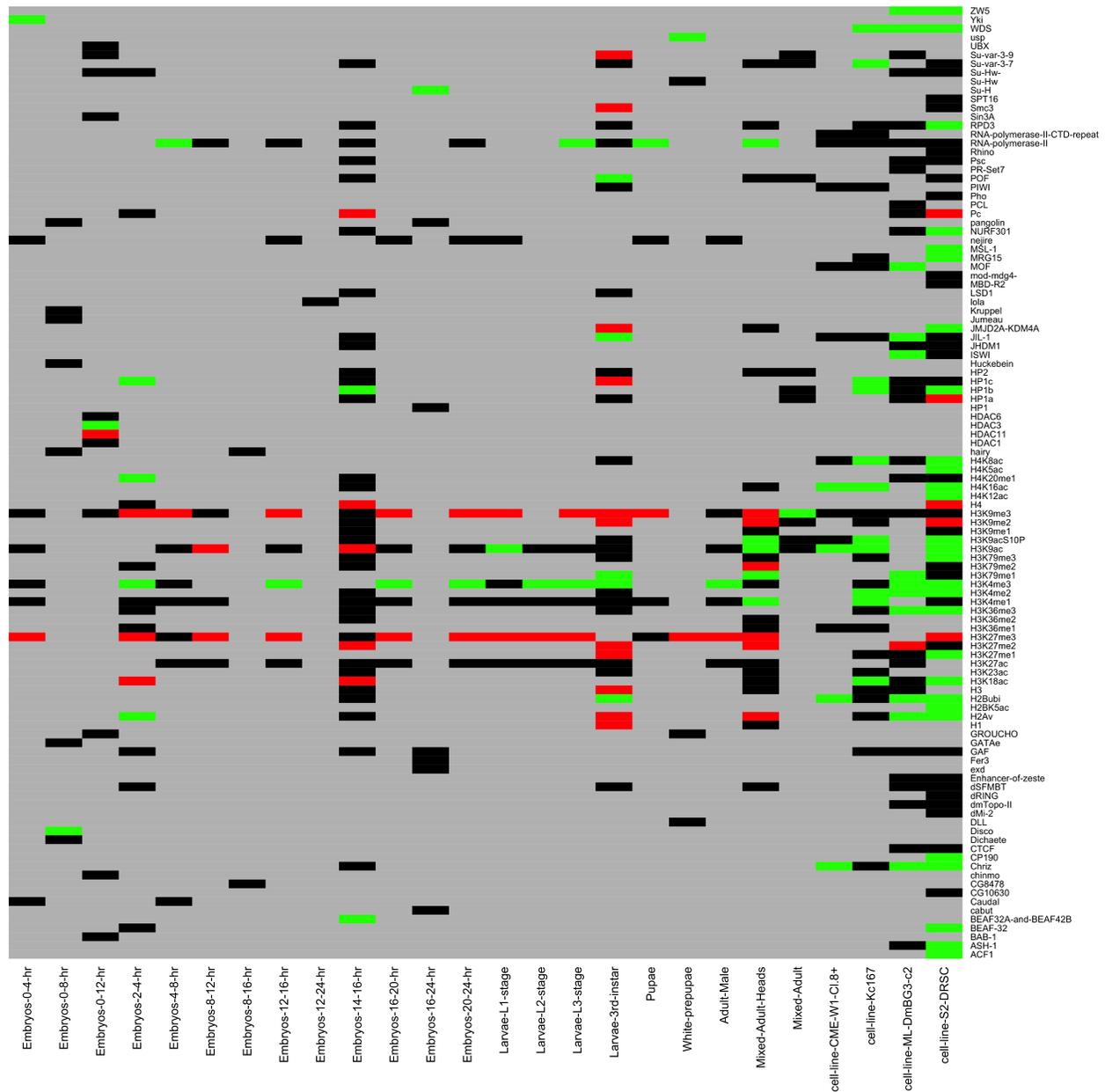


Figure S 9 Expected and observed counts of variants within binding sites of chromatin-associating factors (y-axis) and across various developmental stages (x-axis), as defined through ChIP-chip and ChIP-seq binding assays. For each factor and developmental stage, expected distributions were generated through 10,000 iterations of randomly sampling 1,000 variants from all autosomes, and the number of variants residing within marked binding locations were counted. This expected distribution was then compared to the number of variants residing within marked binding locations from the set of 1,000 top ranked associations. Observed counts with respect the expected distribution are as follows; green represents observed counts that are > 97.5% of the expected distribution, red represents observed counts that are < 2.5% of the expected distribution, black represents observed counts that are within 95% of the expected mass, and gray represents missing sample data.