# A Multivariate Genome-Wide Association Study of Wing Shape in *Drosophila melanogaster*

William Pitchers,*,1 Jessica Nye,†,2 Eladio J. Márquez,†,3 Alycia Kowalski,* Ian Dworkin,*,4,5
and David Houle†,5

*Department of Integrative Biology, Program in Ecology, Evolutionary Biology and Behavior, Michigan State University, East Lansing, Michigan and †Department of Biological Science, Florida State University, Tallahassee, Florida

ORCID IDs: 0000-0003-0385-5939 (W.P.); 0000-0001-6957-0931 (J.N.); 0000-0002-2874-287X (I.D.); 0000-0002-8095-3156 (D.H.)

**ABSTRACT** Due to the complexity of genotype–phenotype relationships, simultaneous analyses of genomic associations with multiple traits will be more powerful and informative than a series of univariate analyses. However, in most cases, studies of genotype–phenotype relationships have been analyzed only one trait at a time. Here, we report the results of a fully integrated multivariate genome-wide association analysis of the shape of the *Drosophila melanogaster* wing in the *Drosophila* Genetic Reference Panel. Genotypic effects on wing shape were highly correlated between two different laboratories. We found 2396 significant SNPs using a 5% false discovery rate cutoff in the multivariate analyses, but just four significant SNPs in univariate analyses of scores on the first 20 principal component axes. One quarter of these initially significant SNPs retain their effects in regularized models that take into account population structure and linkage disequilibrium. A key advantage of multivariate analysis is that the direction of the estimated phenotypic effect is much more informative than a univariate one. We exploit this fact to show that the effects of knockdowns of genes implicated in the initial screen were on average more similar than expected under a null model. A subset of SNP effects were replicable in an unrelated panel of inbred lines. Association studies that take a phenomic approach, considering many traits simultaneously, are an important complement to the power of genomics.

KEYWORDS multivariate GWAS; genome-wide association analysis; developmental genetics; phenomics; GP map; *Drosophila* wing

UNDERSTANDING the inheritance and evolution of complex traits is an important challenge for geneticists and evolutionary biologists alike. A detailed understanding of how genetic variation affects complex traits is important for the treatment of disease, for our attempts to control the evolution of useful or dangerous organisms, and for understanding and predicting evolution over long timescales. Here, we describe the results of a genome-wide association study (GWAS) of the *Drosophila melanogaster* wing, a model complex trait. We undertook this study as part of our attempt to understand the evolution of the fly wing.

The quantitative genetics of the wing is relatively well studied (Mezey and Houle 2005; Houle and Fierst 2013), yet many aspects of the evolution of wings over short and long timescales are not consistent with the abundant variation we observe (Houle *et al.* 2003, 2017; Carter and Houle 2011; Pitchers *et al.* 2013; Bolstad *et al.* 2015). This suggests that we need a more detailed understanding of the genotype–phenotype map (Lewontin 1974), the relationship between genetic variation and the phenotype, to understand the inheritance and evolution of the wing. Fortunately, the fly wing is also a model system for the study of developmental genetics (*e.g.*, de Celis and Diaz-Benjumea 2003; Blair 2007; Wartlick *et al.* 2011; Matamoro-Vidal *et al.* 2015), suggesting that the genetic variation influencing the wing can be fitted into a causal framework, directing our attention to the critical aspects of development that enable and shape evolution of

the wing. We seek to generate a more precise characterization of the natural genetic variation influencing wing shape than has been possible with previous mapping studies, which utilized far fewer markers than current methods allow (Weber *et al.* 1999, 2001; Zimmerman *et al.* 2000; Palsson *et al.* 2004; Dworkin *et al.* 2005; Mezey *et al.* 2005; Dworkin and Gibson 2006).

The evolutionary patterns that we seek to explain concern the relationship of different parts of the wing, rather than the presence or magnitude of single traits. We can measure many aspects of wing shape (Houle *et al.* 2003), but these are interrelated due to the connections among the cells that make up the wing during development and in the adult. Any developmental change that affects one aspect of the wing, such as the length of a particular vein, must also affect adjacent areas of the wing; any one wing measurement incompletely captures wing variation (Mezey and Houle 2005). This reality of wing morphology is a challenge for association analyses because forward genetic analyses are generally built on analyses of single traits.

Despite growing enthusiasm for a comprehensive phenomic approach (Houle 2010; Houle *et al.* 2010), the majority of GWASs that include more than one trait use multiple, univariate analyses for each site, rather than a single multivariate analysis (*e.g.,* Teslovich *et al.* 2010; Battle *et al.* 2014). Statisticians have long appreciated the value of genuinely multivariate approaches to association studies (Lange *et al.* 2003; Shriner 2012), leading to a recent proliferation of multivariate methods and software (O'Reilly *et al.* 2012; Stephens 2013; van der Sluis *et al.* 2013; Scutari *et al.* 2014; Zhou and Stephens 2014; Schaid *et al.* 2016; Porter and O'Reilly 2017). While these methods are diverse, a consistent result is that multivariate analyses increase power to detect associations and the biological usefulness of the results (Porter and O'Reilly 2017). Given these advantages, it is unfortunate that just a few genuinely multivariate empirical association studies have been published (*e.g.,* Anderson *et al.* 2011; Topp *et al.* 2013). The majority of published multivariate analyses are examples in the method development papers, instead of fully realized studies.

In this paper, we apply a fully integrated multivariate analysis of the fly wing, drawing on genotypes in the *Drosophila* Genome Reference Panel (DGRP; Mackay *et al.* 2012). We analyze the genetic architecture of segregating variation for a 58-dimensional representation of fly wing shape (Figure 1A) plus size. The goal of a GWAS is to determine which SNPs are most likely to have a causal effect on the phenotypes under study, which we refer to as causal SNPs. The challenges are that we have only studied a modest number of genotypes relative to the number of variable sites in the genome and the dimension of the phenotype we measured. Both of these factors make it important to account for the presence of population structure and SNPs in linkage disequilibrium (LD) with potentially causal SNPs before making predictions about the causes of phenotypic variation.

The ideal analysis of such data would be a single mixed-model analysis that simultaneously accounts for the effects of all SNPs and population structure. Unfortunately, limited data preclude fitting such a model genome-wide. Even fitting mixed models incorporating a small number of SNPs proved to be impractical for reasons that we will return to in the *Discussion*. Given this situation, we took a two-stage approach. First, we used an approximation of a mixed-model analysis to find those SNPs with the greatest ability to predict wing phenotypes. This SNP-by-SNP approach will include many false positives due to LD and population structure, and will overestimate effect sizes due to the Beavis effect (Beavis 1994, 1998; Xu 2003). In the second stage, we fit multivariate LASSO (Least Absolute Shrinkage and Selection Operator) regressions (Tibshirani 1996; Hastie and Qian 2016) incorporating SNPs in LD and measures of population structure. The LASSO results in regularization or shrinkage of effect sizes, which is expected to help moderate overestimation of effect sizes, including the Beavis effect. It also results in algorithmic variable selection, that is, only a subset of the predictors in the model may be estimated to have nonzero effects.

We confirm that the inheritance of wing shape is highly polygenic and implicate many pathways known to be involved in wing development, as well as novel candidate genes. We then experimentally validate associations using RNA interference (RNAi)-mediated gene knockdowns to examine the degree of replicability for the direction of phenotypic effects. In addition, we replicate some associations using an independent panel of inbred lines.

## Materials and Methods

To assist in following the interlocking set of analyses that we performed, Figure 2 gives an overview to the flow of information from preexisting stocks and data through to the results presented.

### Drosophila strains

For the GWAS, we used the DGRP, a set of inbred lines established from iso-female lines collected at a farmer's market in Raleigh, NC (Mackay *et al.* 2012). We obtained phenotypic data from 184 lines scored in Freeze 2 of the DGRP genotyping (Huang *et al.* 2014).

For the replication analysis, female *D. melanogaster* were collected in the summer of 2004 at a peach orchard in West End, North Carolina (NC2) by I. Dworkin and in a blueberry field in Cherryfield Maine (ME) by Marty Kreitman (Goering *et al.* 2009; Reed *et al.* 2010). All lines were full-sibling inbred for 15–20 generations. In total, 190 lines were used (~50% from each population).

### Rearing, handling of flies, and imaging of wings

Wings of DGRP flies were phenotyped independently in both the Houle laboratory in Florida and the Dworkin laboratory in Michigan. To allow us to evaluate the robustness
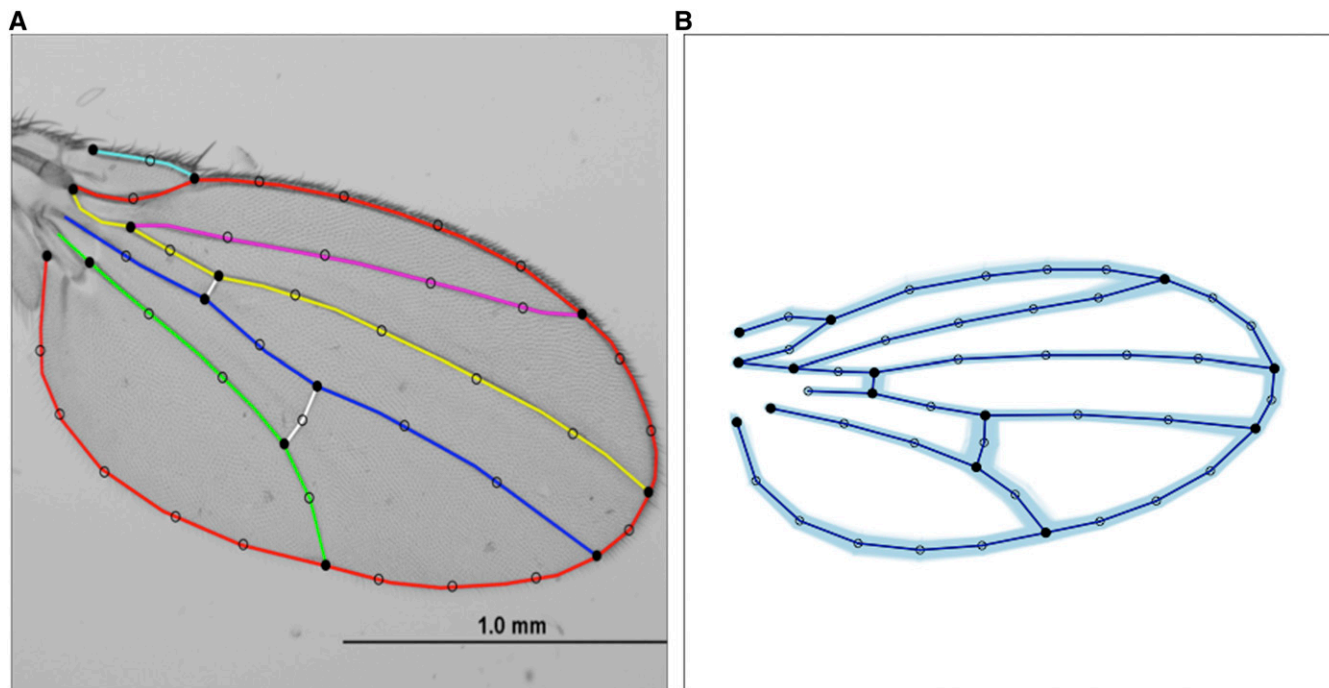
**Figure 1** Capturing wing shape with a spline model. Closed circles are landmarks formed by the intersection of splined veins and open circles are semilandmarks used to represent the locations of veins. (A) Splines fitted to a typical *D. melanogaster* wing. Colored lines are the splines. (B) Blue overlay represents the range of shape variation among *Drosophila* Genome Reference Panel lines.

of the wing phenotypes to experimental conditions, we did not standardize rearing conditions. Each laboratory reared flies and imaged wings according to its own standard. These conditions differed in the following key respects.

In the Houle laboratory, flies were reared at 25° and 55% relative humidity in six-dram vials on a corn meal–sucrose medium preserved with propionic acid, no live yeast added. Wings of live flies were imaged using the "Wingmachine" system (Houle *et al.* 2003). In the Dworkin laboratory, flies were reared at 24° in bottles on a cornmeal–molasses–yeast-based medium with carrageenan as a gelling agent, and propionic acid and methyl paraben as preservatives to which live yeast was added. Dworkin laboratory flies were preserved in 70% ethanol, the wings were dissected, and mounted prior to imaging and analysis. See Supplemental Material, File S1, Supplemental Materials, Methods and Results (SMR) for additional details on the rearing of flies and imaging of wings.

In total, we obtained phenotypic data from 24,672 wings from 184 DGRP lines, with an average sample size of 134.1 wings/line. One-hundred and sixty-three lines were measured in both laboratories. We obtained < 40 wing images (minimum 15) for only four lines.

The ME-NC2 panel flies were reared in the Dworkin laboratory at 25° in a 12:12 light/dark cycle at constant 50% humidity, similar to previously described experiments (Dworkin and Gibson 2006). We dissected ∼20 wings/replicate/line for a total of 7968 males from 153 lines.

### Morphometric data

To capture landmark and semilandmark data from the recorded images, we followed a modified protocol from Houle *et al.* (2003). Splining and error correction was accomplished in the Java program Wings 3.72 (Van der Linde 2004–2014). Wings fits nine cubic B-spline functions to the veins and margins of wings in the saved TIFF images (Figure 1A), using the locations of the two starting guide points to initiate fitting.

The program CPR (Márquez 2012–2014) was used to extract 14 landmark and 34 semilandmark positions from the fitted splines (as shown in Figure 1A). The combined data from the DGRP, validation, and replication data sets (a total of 66,890 wings) were subjected to generalized Procrustes superimposition (Rohlf and Slice 1990), which scales forms to the same size, translates their centroids to the same location, and rotates them to minimize the squared deviations around each point. This separates the useful size and shape information from the nuisance parameters introduced by the arbitrary location and rotation of the specimens within the images. The resulting data set has 58 independent shape variables, which we summarized using principal component (PC) analysis (Table S2). Centroid size was natural log transformed to put it on the same scale as the shape variables. Additional details are given in the SMR.

### DGRP genotype data

We used the publicly available Freeze 2 genotypes from February 2013 (Huang *et al.* 2014) (ftp://ftp.hgsc.bcm.edu/DGRP/freeze2_Feb_2013/vcf_files/freeze2.vcf.gz).
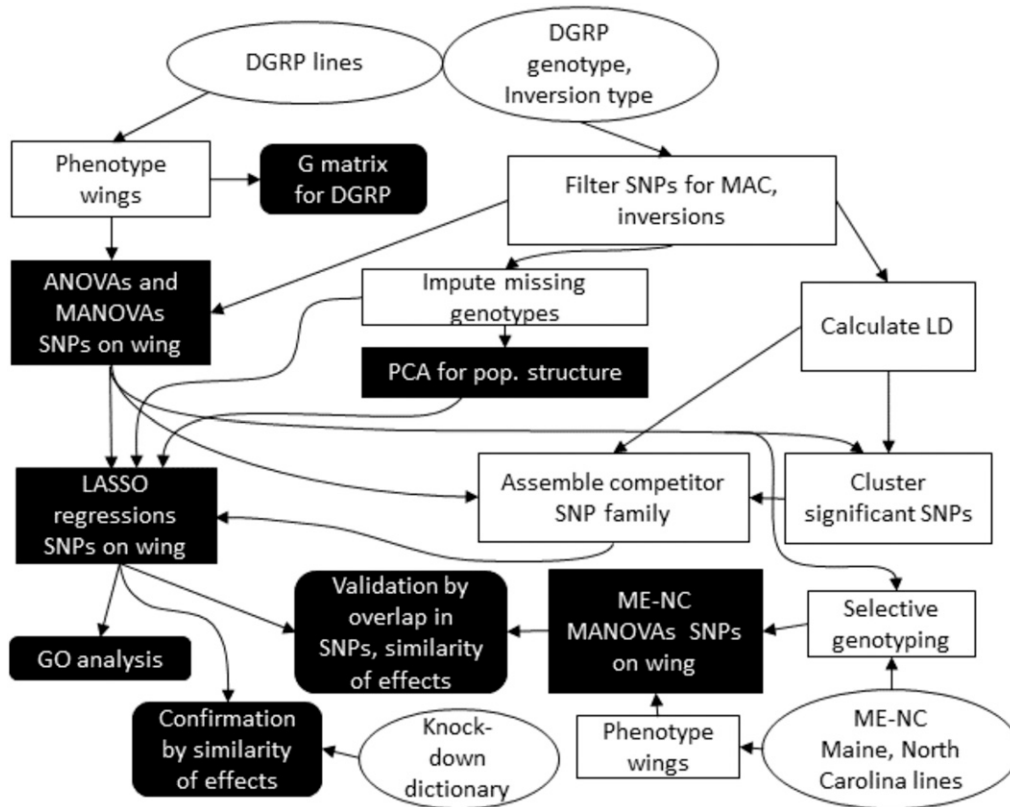
**Figure 2** Relationship between experiments and analyses. Ellipses represents stocks or results that are the input to this work. Boxes with no fill are experiments or preliminary analyses only described in *Materials and Methods*. Filled boxes generate the major results presented. DGRP, *Drosophila* Genome Reference Panel; GO, gene ontology; LASSO, Least Absolute Shrinkage and Selection Operator; LD, linkage disequilibrium; MAC, minor allele count; MANOVA, multivariate analysis of variance analysis; ME-NC, Maine-North Carolina; PCA, principal component analysis.

Our analyses were based primarily on FlyBase GenBank Release 5, but coordinates have been converted to Release 6 for presentation here. We used only calls of homozygous genotypes with genotypic phred scores ≥ 20 and treated others as missing data. All polymorphisms are referred to as SNPs including those involving multiple nucleotides. If two or more variants were found at the same site, we analyzed the one with the highest minor allele count, treating rarer variants as equivalent to the reference.

As noted by Huang *et al.* (2014), some pairs of lines are more closely related than random. We captured population structure using PC analysis with smartpca (Patterson *et al.* 2006). Some of this population structure is due to three common inversions found in the DGRP lines: In(2L)t, In(2R)NS, and In(3R)Mo (Corbett-Detig and Hartl 2012; Langley *et al.* 2012; Huang *et al.* 2014; Houle and Márquez 2015). We used the results of Houle and Márquez (2015) to infer the karyotypes for each of the 184 lines in our analysis.

Initial analyses to choose sites for replication in the ME-NC2 were carried out using Freeze 1 data from August 2010, available at https://www.hgsc.bcm.edu/arthropods/drosophila-genetic-reference-panel.

### Genetic variation for wing shape

We estimated the genetic variance–covariance matrix of the DGRP line effects using restricted maximum likelihood implemented in Wombat (Meyer 2006-2018, 2007). Wombat is limited to analysis of 40 variables, so we analyzed the first 39 PCs of shape, plus the natural log of centroid size.

These 39 PCs capture 88% of the total phenotypic variance. We compared the fits of full (40-d) and reduced-rank model likelihoods (Kirkpatrick and Meyer 2004; Meyer and Kirkpatrick 2005, 2008). We fitted a pooled sex covariance matrix, treating laboratory, sex, and rearing block as fixed effects. To account for relatedness among lines, we calculated a generalized relationship matrix based on all SNPs in the Freeze 2 data set with minor allele counts of five or more using the program smartpca in the EIGENSOFT package (Patterson *et al.* 2006; https://www.hsph.harvard.edu/alkes-price/software/). The line effect was assumed to be distributed proportionally to this relationship matrix with diagonals set to 1, and coefficients < 0.01 set to 0, using the RAN GIN option in Wombat.

### LD and cluster analysis

LD complicates the interpretation of significant associations uncovered in a GWAS. We quantified LD as the squared gametic correlation between sites. To help us interpret our results, we calculated the gametic correlation, $r$, between all sites judged as significant at a false discovery rate (FDR) of 5% (see below) and all other sites throughout the genome using the algorithm described in Houle and Márquez (2015). We retained a list of all those pairs where $r^2 > 0.5$. This cutoff was chosen based on simulations that show that LD of $r^2 < 0.5$ will infrequently generate false positives for a SNP in LD with a SNP that has a phenotype effect size typical of those detected in this study in a similar number of lines (Houle and Márquez 2015).

To help interpret which SNPs are most likely to cause variation in phenotype, we carried out an LD-based cluster analysis on SNPs significantly associated with phenotype in the multivariate analysis of variance (MANOVA) analysis. The algorithm used to determine cluster membership is described in SMR.

### Detecting phenotypic associations

Mean wing forms for each combination of line, laboratory, and sex were used in the association analyses. Our analyses included the 2,517,547 SNPs where the minor allele was homozygous in at least five lines and where $\geq 120$ phenotyped lines had genotypes called.

For our initial MANOVA association analysis, effects of SNPs on morphometric variation were quantified using a multivariate linear model taking into account the effects of laboratory, sex, SNP, and line:

$$\mathbf{Y}_{hijk} = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_k + \boldsymbol{\alpha}\boldsymbol{\beta}_{ij} + \boldsymbol{\alpha}\boldsymbol{\gamma}_{ik} + \boldsymbol{\beta}\boldsymbol{\gamma}_{jk} + \boldsymbol{\alpha}(\mathbf{L})_{hi}$$
$$+ \boldsymbol{\alpha}(\mathbf{L})\boldsymbol{\beta}_{hij} + \boldsymbol{\alpha}(\mathbf{L})\boldsymbol{\gamma}_{hik} + \boldsymbol{\varepsilon}_{hijk} \quad (1)$$

where $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and $\boldsymbol{\gamma}$ represent vectors of fixed effects of the $i^{th}$ SNP, $j^{th}$ sex, and $k^{th}$ laboratory, respectively, $\boldsymbol{\alpha}(\mathbf{L})$ represents the random effect of the $h^{th}$ inbred line nested within SNP, $\boldsymbol{\varepsilon}$ is the residual vector, and higher-order terms represent interactions between these factors. To compare the results of multivariate and univariate analyses, we also calculated univariate tests using the model in Equation 1. Attempts to fit this model as a true mixed model in SAS proved to be infeasible due to computational demands. We approximated the mixed-model tests and estimation using the procedures described in the SMR.

When one of the three common inversions was present in a line, we treated all genotype calls in the inversion and regions that are in high LD with the inversion as missing. For In(2R)NS and In(2L)t, this included sites between the breakpoints inferred by Corbett-Detig et al. (2012) plus 20 kb either side. On chromosome 3R, we masked all sites with coordinates > 16 Mb in lines inferred to carry an In(3R)Mo genotype (Corbett-Detig and Hartl 2012; Houle and Márquez 2015).

Least-squares estimates of SNP effect vectors were obtained from a simpler model neglecting interactions of SNP effects with sex and laboratory,

$$\mathbf{Y}_{hijk} = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_j + \boldsymbol{\gamma}_k + \boldsymbol{\alpha}\boldsymbol{\beta}_{ij} + \boldsymbol{\beta}\boldsymbol{\gamma}_{jk} + \boldsymbol{\alpha}(\mathbf{L})_{hi} + \boldsymbol{\varepsilon}_{hijk} \quad (2)$$

where effect size is the length (2-norm) of this vector of effects, $\|\boldsymbol{\alpha}_i\|$.

To compare the amount of variance explained in different analyses, we projected each original 59-d effect vector into the 40-d space defined by the $\mathbf{G}$ matrix, yielding a new vector $\boldsymbol{\alpha}_{i\cdot40}$. The total amount of variance explained is $p(1-p)\|\boldsymbol{\alpha}_{i\cdot40}\|^2$, where $p$ is minor allele frequency (MAF). To control the FDR, we applied the Storey and Tibshirani (2003) approach as implemented in the R package fdrtools (Strimmer 2008).

### LASSO regressions

SNP associations can be explained if the SNP directly causes phenotypic variation, or if it is in LD with other causal SNPs. To include these confounding elements in our model, we first collapsed our data so that we could use a regression model. To do this, we calculated the vector of least-squares means for the shape and size variables from a linear model with laboratory, sex, and DGRP line as main effects.

We then carried out LASSO multivariate multiple regressions with one significant (focal) SNP, plus a family of competing predictors, comprising three sets of variants. The first group consists of all significant SNPs that were annotated as being within 2 kb of the transcript of a gene that the focal SNP is either in, closest to, or is within 2 kb of the focal SNP. The second group consists of all SNPs anywhere in the genome that have LD $r^2 > 0.5$ with any of the first group of significant SNPs. The third set of predictors are the scores on the 13 significant genotypic PCs with more variation than expected under the null hypothesis of no population structure. See the SMR for details of this analysis.

Regularization was implemented using multivariate LASSO regression calculated in the R package GLMNET (Friedman et al. 2010) with the "mgaussian" option (Hastie and Qian 2016). For each SNP model, the shrinkage parameter was chosen by fivefold cross-validation. Additional description of the LASSO and its implementation is in the SMR.

### Gene ontology analysis

For each SNP significant in the MANOVA analysis at an FDR of 5%, we downloaded the full gene ontology (GO) information for the nearest genes up- and downstream, on both the negative and positive strands. We identified the closest coding region to each of these SNPs for subsequent GO analyses. For the LASSO-significant SNPs, we used WebGestalt (Wang et al. 2017) to test for enrichment of biological process GO categories. We compared nonredundant biological processes with five or more genes assigned against the entire genome. The Benjamini and Hochberg (1995) FDR correction was used to adjust for multiple tests, assuming a hypergeometric distribution.

### Quantitative knockdowns of gene activity for validation

We knocked down expression of genes of interest using the progesterone-inducible Geneswitch Gal4 construct (GS, Roman et al. 2001) engineered to be under the regulation of a ubiquitous *tubulin* driver (Tub-5 GS), generously furnished by Scott Pletcher. GS was used to drive expression of interfering RNA for a gene of interest (UAS-[GOI] RNAi) constructs obtained from the Transgenic RNAi Project (TRiP) (Ni et al. 2008; in a *yv* background), and the Bloomington *Drosophila* Stock Center or the Vienna *Drosophila* RNAi Center (Dietzl et al. 2007; in a *w1118* background). The list of RNAi stocks used is in File S5.

To carry out a knockdown experiment, we crossed Tub-5GS and UAS-[GOI] RNAi stocks, and allowed these flies to lay eggs

on media containing the progesterone analog mifepristone at concentrations of 0, 0.3, 0.9, and 2.7 μM. The parameters of the multivariate regression of phenotype on mifepristone were retained as the effect of the manipulated gene of interest. All knockdown experiments were conducted in the Houle laboratory. Additional details of these analyses are in the SMR.

### Comparing knockdown vectors to SNP vectors

We compared the directions of phenotypic effects using vector correlations, as described in the SMR. The statistical significance of vector correlations between the knockdown vectors and LASSO-significant SNP vectors was determined by comparing the observed correlations to the distribution of correlations under the null hypothesis of no relationship. Even if the SNPs have no real effects, the inferred vectors will tend to fall in the more variable regions of phenotype space, so to ensure that the random vectors were appropriately sampled we took two approaches. First, we assumed that the estimated directions of effects in the overall sample of SNPs were representative of the random distribution of effect directions. Second, we randomly sampled 40-d vectors from a multivariate normal distribution with mean 0 and covariance equal to **G**, and compared these to the observed vectors projected into the corresponding 40-d subspace. These two approaches yielded similar results. We report the results using the random sample of inferred vectors, but both approaches agreed in all the specific cases discussed.

To test for significant SNP-wise correlations, we compared the quantiles of vector correlations between each knockdown vector and 10,000 random vectors. To test for significant experiment-wise correlations, we calculated 10,000 sets of 580 correlations of random SNP effects with each knockdown vector. The quantiles of the maximum of the 580 correlations were calculated and compared to the observed vector correlations. In both cases, the quantiles differ considerably with the direction of the knockdown effect; vectors close to the principal axis of genetic variation (PC1) are much more common in the estimated set of vectors and so have quantiles considerably larger than those in less-common directions. File S5 includes both the quantiles for random sets of 580 vector correlations and the vector correlations of each knockdown vector with the first five PCs of the among-line variance matrix.

### Replication tests with the ME-NC2 populations

Based on a preliminary analysis of Freeze 1 genotypic data, we chose 389 SNPs for reanalysis in an independent sample of flies. We compiled a list of SNPs with the smallest associated *P*-values, removing those where the SNP's effect was unstable (LogRatio of *P* within 1.5 SD of 0) and ranking the remainder by effect size. We then excluded SNPs whose minor allele was present in more than nine DGRP lines that were highly correlated with other SNPs and those far from any gene transcript. Of these 389 SNPs, 342 were included in the Freeze 2 data that we analyzed for associations.

*Genotyping for replication cohort:* The genotyping for our replication SNP set was carried out by KBiosciences (now LGC Genomics) using "Kompetitive Allele-Specific PCR" assays. This is a fluorescence-based genotyping technology that uses allele-specific primers, making it generally more accurate for smaller jobs than high-throughput methods. We designed primers based on 100 bp of the *D. melanogaster* reference genome from FlyBase (version 5.41) on either side of each SNP. Samples of genomic DNA extracted from 15 flies from each of the ME-NC2 lines were submitted to KBiosciences. Several duplicated control samples (same genotype but independently labeled) were included to assess any technical variation in genotyping. We obtained genotype data for 300 SNPs with minor allele counts sufficient for analyses in both the DGRP and ME-NC2 data sets.

*Analysis:* To test for significant associations with the ME-NC2 panel, we used the same pipeline and analysis framework as described above for the DGRP. The model included population (ME or NC2) and SNP as fixed effects, with lines nested within SNPs as a random effect. Laboratory and sex were not included as this study was carried out solely in the Dworkin laboratory, and only males were phenotyped. To provide a null distribution for testing whether the average vector correlation of the 27 matching LASSO-significant DGRP SNP effects with the ME-NC2 effects was greater than expected, we computed 10,000 means of groups of 27 vector correlations between random subsets of DGRP effects and the ME-NC2 effects.

### Data availability

Data and programs on which this study is based have been archived in Figshare https://doi.org/10.25386/genetics.6790526.

## Results

### High repeatability for wing shape across laboratories

Of the 184 DGRP lines phenotyped, 163 were measured in both the Dworkin and Houle laboratories. The means and SDs of the variables by sex and laboratory are in File S2. Wing shape has considerable variation among lines (Figure 1B and Figure 3A). As described in the *Materials and Methods*, each laboratory used different rearing conditions, and imaging hardware. Despite these environmental differences, line effects on wing shape have a high degree of interlaboratory repeatability with respect to both effect sizes (Figure 3A) and directions (Figure S1A). However, wing size was weakly correlated across laboratories (Figure 3B). This is likely to be due to genotype–environment interactions with laboratory rearing practices, rather than measurement error, as repeatability of wing size within a laboratory is high (Figure S1B). A MANOVA on line–sex–laboratory means shows that the effects of laboratory [Wilk's $\lambda = 0.0026$, $F = 3014.6$, d.f. $= (59,460)$, sex (Wilk's $\lambda = 0.027$ and $F = 275.9$] and
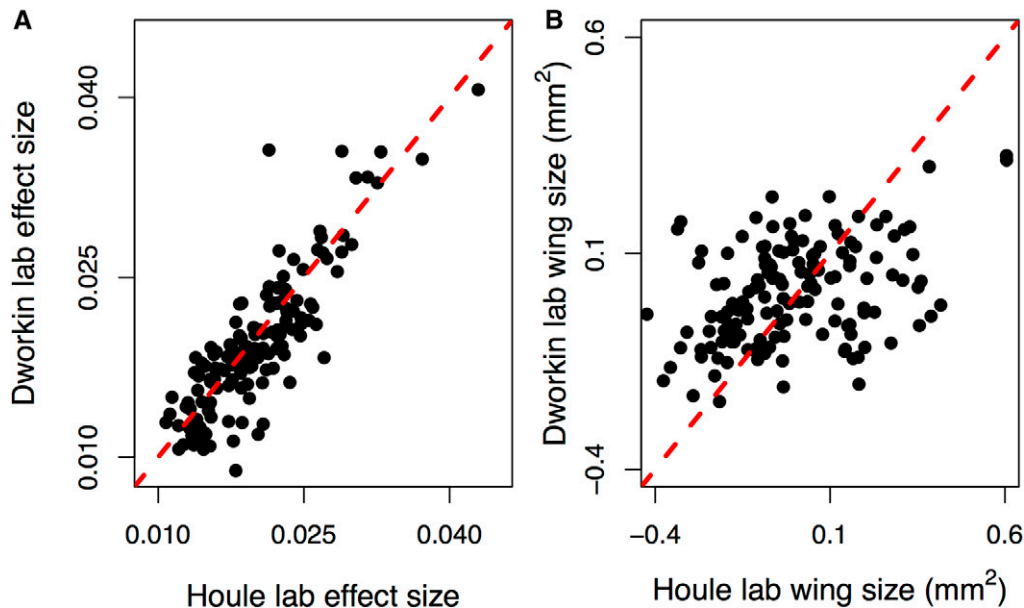
**Figure 3** Interlaboratory repeatability. (A) High repeatability of line effect sizes for shape across laboratories. (B) Low repeatability of size across laboratories, despite high intralaboratory repeatability (Figure S1B).

laboratory by sex interactions (Wilk's $\lambda = 0.35$ and $F = 14.2$) are all highly significant ($P < 0.0001$), reflecting subtle differences in means across laboratories.

### Genetic variance in wing shape in the DGRP

To verify the presence of genetic variance in wing size and shape, we estimated the variance–covariance matrices in wing size and shape, assuming that it is proportional to the genomic relatedness matrix. Due to software limitations, we could test for genetic variance in only 40 dimensions (out of 59 possible), so we chose to fit the first 39 PC scores for wing shape, plus ln centroid size. A model with genetic variance in all 40 possible dimensions fitted better than models with 39 dimensions (Akaike Information Criterion Corrected (AICC), $\Delta$AICC $= 1466$). This is strong evidence that at least 40 independent aspects of wing shape are affected by genotypic variation in the DGRP sample. Estimates of the **G** matrix in both 40-d space and the original x–y coordinate space are given in File S2.

### Chromosomal inversions influence wing shape, but Wolbachia does not

Three inversion karyotypes—(In(2L)t, In(2R)NS, and In(3R)Mo)—were found in more than seven of the DGRP lines that we phenotyped (Huang *et al.* 2014; Houle and Márquez 2015). Approximately half of the DGRP lines carried the intracellular parasite *Wolbachia* (Huang *et al.* 2014). We conducted MANOVAs on the effects of inversion genotypes and *Wolbachia* status, with the results shown in Table 1. Each of the three inversions has a highly significant effect on wing shape–size, but *Wolbachia* infection status was not significant.

### Basic GWAS analysis

We carried out individual MANOVAs of the effect of genotype on wing shape for each of the 2,517,547 polymorphisms with a minor allele count $\geq 5$. To pick SNPs for additional analyses, we used the FDR algorithm of Storey and Tibshirani (2003). A total of 2396 sites had significant effects using a 5% FDR cutoff ($q$-value $< 0.05$). The Storey and Tibshirani algorithm estimates that the $P$-values can be explained by mixture of $\eta_0 = 71.5\%$ SNPs with no phenotypic effect, with the remainder having some effect. Figure 4A shows a Manhattan plot of the multivariate results. A list of the significant sites, test statistics, effect sizes, and variance explained, plus information about genes implicated, are given in File S3.

We calculated the genetic variance in shape–size explained by each of the significant SNPs in the 40-d subspace for which we can estimate **G** as a proportion of the trace of **G**. Estimated effect sizes are modest and no single SNP is estimated to explain $> 3.6\%$ of the variance. In addition, the estimated effect sizes are clearly too large on average, as the median percentage of variance explained is 1.3% (mean is 1.4%). There are two known causes for the upward bias in effect size. First, sampling variation causes effects for SNPs judged to be significant to be overestimated (Beavis 1994, 1998; Xu 2003). Second, these analyses do not compensate for the effects of LD and relatedness among lines, which we return to below. These results are consistent whether considering the shape-only data, or shape and size simultaneously, which are almost perfectly correlated (0.99).

A quantile–quantile plot of the $P$-values is shown in Figure 5. For sites with MAF $< 0.15$, the distribution shows clear evidence of substantial deviation from the expected uniform distribution throughout the range of $P$-values. We interpret this as largely due to spreading the signal of true effects to the large number of sites in LD with rare alleles (see below). The $P$-values are much closer to the null distribution at sites with MAF $> 0.15$. This distribution is also consistent with a very large number of sites each having small phenotypic effects. We return to these issues below.

**Table 1 Summary statistics for MANOVAs of the effects of the three common inversions and *Wolbachia* infection status on wing shape–size**

| Predictor | Numerator d.f. | Denominator d.f. | Wilk's λ | *P* |
|---|---|---|---|---|
| In(2L)t | 128 | 94.7 | 0.364 | $2.08 \times 10^{-7}$ |
| In(2R)NS | 128 | 93.4 | 0.475 | $2.70 \times 10^{-3}$ |
| In(3R)Mo | 128 | 93.8 | 0.327 | $4.53 \times 10^{-9}$ |
| *Wolbachia* | 59 | 124.7 | 0.660 | 0.34 |

### Comparing multivariate and univariate analyses

To understand the relative power of the multivariate analysis, we carried out univariate analyses of each SNP on the scores on PC1 through PC20. When we applied the Storey and Tibshirani (2003) FDR algorithm independently to each of the 20 sets of *P*-values, we observed four significant SNPs on PC1 (shown as green dots in Figure 4) and none on the other 19 axes at an FDR of 5%. Just one of these sites is also significant at the FDR 5% level in the multivariate analysis (3L:17987278, in an intron of *Eip75B*).

To further compare the multivariate and univariate results, we also applied the same critical *P*-value identified as the FDR 5% cutoff in the multivariate analysis ($P = 0.00007$) to all of the univariate analyses. A total of 6990 SNPs were identified as significant at $P < 0.00007$ in at least one (of the 20) univariate analyses. Only 139 of these were also significant in the multivariate analyses. Figure 4B shows the genomic locations of the 565 sites significant at $P = 0.00007$ on PC1.

The discrepancy between the identities of SNPs implicated in the univariate and multivariate analyses was unexpected, and we investigate the possible reasons in Figure 6. The two reference lines show the mean multivariate effect sizes for all SNPs and for the MANOVA-significant SNPs only. Unsurprisingly, significant SNPs have larger effect sizes. The red squares show the effect sizes on each of the first 20 PCs for univariate-significant SNPs (at $P < 0.00007$) while the gray boxes show the total multivariate effect sizes for those same SNPs. Effects of univariate-significant SNPs are unusually concentrated in the direction of the vector on which they are significant. This is particularly apparent for high-ranked PCs (PC1, PC2, *etc*.) where the univariate effect is a very large proportion of the total multivariate effect of that SNP, as demonstrated by how close the red squares are to the gray. For SNPs significant on low-ranked PCs (PC20, PC19, *etc*.), the multivariate vector lengths are close to the average multivariate vector length of all SNPs. A very different pattern is apparent in the univariate effects on each PC of MANOVA-significant SNPs (green diamonds) and of all SNPs (blue circles). The effects of MANOVA-significant SNPs are modestly higher than average across all 20 PCs. These comparisons suggest that the univariate analyses identify SNPs whose effects are unusually concentrated on just one PC vector, but are otherwise unremarkable in the full space. This is reinforced by the fact that only 24 sites were identified as signif-

icant in two different univariate analyses. In contrast, the average score of a SNP that is significant in the multivariate analysis (green diamonds) is modestly higher than average across the full range of PCs.

As noted in the *Introduction*, and treated at greater length in the *Discussion*, the PCs taken as traits in these analyses are biologically arbitrary. Furthermore, this would be true of any set of univariate vectors in wing size–shape space; our expectation is that pleiotropy is the proper null hypothesis of genetic effects on the wing. We have simulated multivariate analyses of shape data under the additional assumption that genetics effects are possible in any direction in phenotype space. These simulations show that MANOVA analyses have the expected Type I errors. Given this, the lack of correspondence between the multivariate and univariate results can be explained by the fact each multivariate analysis finds the direction in the phenotypic space that best distinguishes SNPs, while the univariate analyses sample only a limited set of directions. We believe that false positives are overrepresented in the SNPs implicated in the univariate relative to the multivariate analyses.

### Correcting for population structure and LD

While the DGRP lines were sampled from a large natural population, some pairs of lines are more related than expected at random (Huang *et al*. 2014). Our own analysis of just the 184 phenotyped lines found 13 genotypic eigenvectors that explained more than the expected genetic variance (see the SMR) and that some, but not all, of this relatedness is due to inversion polymorphism. There is also substantial LD among the 184 lines (see the SMR; Huang *et al*. 2014; Houle and Márquez 2015; Pool 2015), some of which is due to departures from this nonrandom coancestry and some due to "rarity disequilibrium" (Houle and Márquez 2015).

These relatedness and LD results suggest that the genes implicated by the MANOVA results may not play a causal role in the genetic effects detected. Furthermore, methods for adjusting the FDR assume that the tests are themselves independent, which is violated for correlated genotypes. To further increase the chance of identifying causal SNPs, we implemented two additional analyses to help judge the likelihood that the MANOVA-significant SNPs had a causal effect or were likely to be closely linked to a causal SNP.

First, we performed a cluster analysis to group the MANOVA-significant SNPs according to their LD. We identified a total of 862 clusters of SNPs uncorrelated (at $r^2 \geq 0.5$) with any other significant SNPs. There were 659 clusters containing just one SNP. At the other extreme, two large clusters contained 236 and 644 SNPs, including correlations over both short and long distances ($> 100$ kb). The clusters and the number of SNPs they contain are shown in Table S3, as are several other indicators of the extent of LD and clustering (see the SMR for details).

Second, we performed a series of LASSO multivariate multiple regressions (Tibshirani 1996; Friedman *et al*. 2010; Hastie and Qian 2016) to examine the influence of
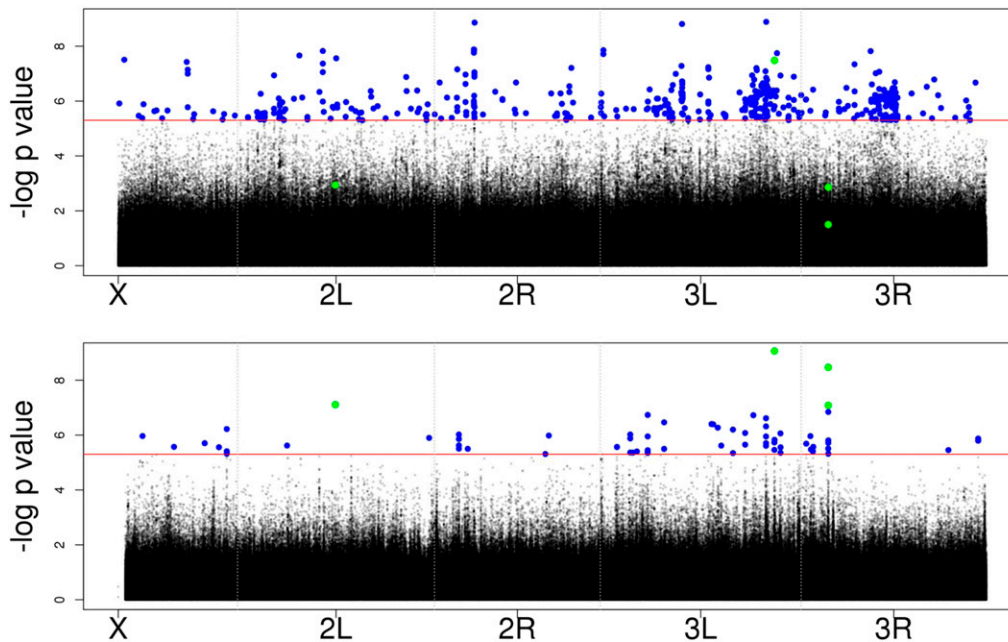
**Figure 4** Manhattan plots of the $\log_{10}$ inverse *P*-values from multivariate analysis (upper panel) and univariate analysis of PC1 (principal component 1) (lower panel). Solid red line is *P* = 0.00007, the cutoff for a 5% false discovery rate (FDR) using the Storey and Tibshirani analysis of the multivariate data. The same cutoff is also applied in to the PC1 analysis. Green points are the four SNPs that reach the 5% FDR cutoff from analysis of just the PC1 *P*-values.

population structure and correlated SNPs on the signal from each of the 2396 significant SNPs from the MANOVA analysis. In each of these multivariate multiple regression models, we used both the significant (focal) SNP plus a family of competitor predictors assembled for that focal SNP. The competitor family consisted of all other MANOVA-significant SNPs closest to the same gene as the focal SNP, all SNPs correlated (at $r^2 \geq 0.5$) with any of those significant SNPS, and scores on the 13 significant population structure eigenvectors.

The focal SNP retained a nonzero phenotypic effect in 580 of the LASSO regressions. We refer to these as LASSO-significant SNPs. These have nonzero estimates of effect size (variable glmeffsz) in File S3. When we fit LASSO models of the focal SNP and the 13 population structure PC scores, just 120 of the models resulted in a 0 focal effect-size estimate. Sites in LD with the focal SNP are responsible for the majority of the reduction from 2396 MANOVA-significant SNPs to 580 LASSO-significant SNPs.

In addition to variable selection, the LASSO also shrinks the estimates of effect size, moderating the Beavis effect. The distribution of MANOVA-significant effect sizes is approximately unimodal, with a mean effect size of 1.094 shape units ($\pm$ 0.429 SD and median 1.06) and a minimum effect size of 0.174 shape units. For the same SNPs, the distribution of LASSO-estimated effect sizes had a very strong mode at 0 due to the many zero estimates; the distribution of nonzero effect sizes was also approximately exponential with a strong mode close to 0. The average effect size after the LASSO analysis was just 0.072 ($\pm$ 0.177 SD and median 0.0), for an average reduction in length of 93%. Taking just the 580 SNPs with nonzero effects in the LASSO analyses, the MANOVA estimates averaged 0.945 shape units ($\pm$ 0.401 SD and median 0.893), while the LASSO estimates had a mean

of 0.296 ($\pm$ 0.252 SD) and a median of 0.225 for an average reduction in length of 70%. The median proportion of the 40-d **G** explained by these 580 SNPs was 0.1% (mean 0.2 $\pm$ 0.3%). The maximum variance explained by one LASSO-significant SNP was 2.21%.

For SNPs retained in the LASSO analysis, the estimated direction of SNP effects in multivariate space was usually very similar to the MANOVA results. The median vector correlation was 0.97, although correlations as low as 0.18 were obtained. Ninety percent of the vector correlations were > 0.85.

Significant SNPS in both the MANOVA and LASSO analyses were enriched for rare SNPs. The overall median MAF is 0.14, while it is 0.04 for the MANOVA-significant SNPs and 0.06 for the LASSO SNPs. One potential explanation for this is that mutation–selection balance on wing shape keeps alleles with phenotypic effects at low frequencies. The fact that rare alleles tended to be dropped in the LASSO analysis is consistent with the higher level of LD for rare SNPs. The LASSO-significant SNPs were in smaller LD clusters of significant SNPs (median cluster size 8 for MANOVA *vs.* 1 for the LASSO), and had on average much smaller families of correlated SNPs entered into the multiple multivariate regression models (median 66 *vs.* 8).

### Known wing development processes implicated by GO analysis

We performed biological process GO analysis for the 479 genes implicated by the LASSO analyses and recognized by WebGestalt. Of these, 336 had biological annotations in *D. melanogaster*. We observed significant enrichment of 61 nonredundant biological processes overall (File S4). Biological processes relevant to wing development are prominent in these categories, starting with wing development (GO 0035220, 42 genes and FDR $1.2 \times 10^{-5}$),
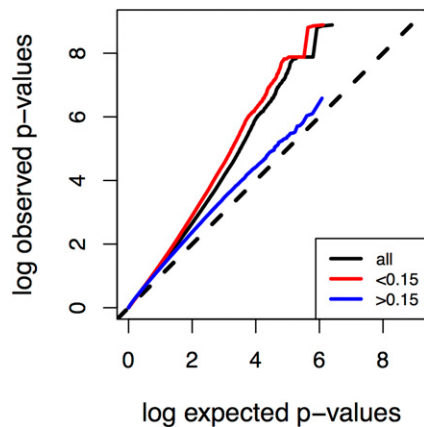
**Figure 5** Quantile–quantile plot of observed *vs.* expected *P*-values genome-wide in the multivariate analysis. Black: all SNPs; red: SNPs with minor allele frequency (MAF) < 0.15; blue: SNPs with MAF > 0.15.

and including appendage development and growth. Overall, 46 of the enriched categories involve aspects of development. Comparison to the Kyoto Encyclopedia of Genes and Genomes database revealed two significantly enriched pathways. The Hippo pathway, which has a well-known role in regulating wing growth and integrating inputs from major wing morphogen pathways, was enriched at FDR = 0.02. Eight hippo genes were implicated (*Actin 5C, dachsous, fat, friend of echinoid, dally, kibra, stardust,* and *Grunge*) against an expected number of 1.74. Apoptosis genes were also enriched (FDR = 0.0007).

### Validation of SNP effects by phenotypic effects of expression knockdowns

As one validation of putative causal SNPs, we utilized quantitative knockdowns of gene expression at LASSO-significant genes using RNAi with a Geneswitch (mifepristone-dependent) tubulin-GAL4 line (see Supplemental Methods; genes listed in File S5). As an example, Figure 7A shows the effects of knockdowns at *Egfr* on wing shape at four different levels of mifepristone. To summarize these results, we performed a multivariate regression of size and shape on mifepristone levels to obtain a single vector. The *Egfr* regression vector is shown in the left wing in Figure 7B. We call the set of phenotypic alterations observed on knockdown a "dictionary" of genetic effects. We note that dictionary knockdowns reduce gene expression throughout the body during the entire duration of wing development. Thus, the effects of the knockdowns may be different from those of SNPs, even if the regions implicated in our analyses have phenotypic effects mediated by changes in gene expression.

We compared the directions of effects in the phenotype space as the absolute value of the vector correlations between SNPs annotated as closest to a gene with the dictionary effect of that gene (see SMR). Table 2 gives the results of tests for greater than expected vector correlations, as well as additional information about each SNP. Four SNPs of the 26 tested were significantly correlated at $P < 0.05$; these suggested

similar effects for genes *Egfr*, *RhoGEF64c*, *knirps*, and *MRP* (*Multidrug-Resistance like Protein 1*). SNP and dictionary effects at *Egfr* and *kni* are shown in Figure 7, B and C. All four of these genes with significant dictionary correlations are highly expressed in the wing disc during larval and pupal development (O'Keefe *et al.* 2012).

*Egfr*, *RhoGEF64c*, and *knirps* are each well-known developmental genes, with known or plausible roles in wing development (Gramates *et al.* 2017). They are each members of multiple significantly enriched developmental GO categories enriched for LASSO-significant genes. *Egfr* is a key receptor of the MAPK pathway that is involved in the specification of the primordial wing disc, notum *vs.* wing determination, and wing vein development (Guichard *et al.* 1999; Wang *et al.* 2000; Paul *et al.* 2013). *Egfr* has previously been implicated in natural variation in wing shape (Palsson and Gibson 2000, 2004; Zimmerman *et al.* 2000; Palsson *et al.* 2004; Dworkin *et al.* 2005). The gene *knirps* is an embryonic gap gene that plays a key role in the development of wing vein L2 (Lunde *et al.* 1998, 2003). As shown in Figure 7C, the intersection of vein L2 with the wing margin shows the largest change for both the knockdown and the correlated SNP. *RhoGEF64c* is a regulatory protein of the Rho GTPase subfamily that regulates intracellular actin dynamics and consequently cell shape, adhesion, and motility (Gramates *et al.* 2017). *RhoGEF64c* is important in leg morphogenesis (Greenberg and Hatini 2011), making a connection to wing development plausible. *MRP* is an active transmembrane transporter that has no known connection to development (Gramates *et al.* 2017). *MRP* is a member of highly enriched LASSO-significant biological GO categories concerning response to chemicals and membrane transport, but nothing closely related to development. It is extremely highly expressed in the wing disc (O'Keefe *et al.* 2012), as well as in most fly tissues (Gramates *et al.* 2017).

In addition to the excess of significant vector correlations, the overall results show that vector correlations are biased toward high correlations, as 19 of the 26 SNPs had dictionary correlations above the median expected under the null hypothesis. To check this, we compared the entire distribution of observed correlations with random sets of correlations. The mean of the observed vector correlations was 0.40; the average random correlation was 0.197, and all of the 1000 sets of 26 random correlations had lower mean values than the observed.

### Vector correlations between gene knockdowns and other SNP effects

We examined the vector correlations between each of the dictionary (gene knockdown) vectors and the 580 LASSO-significant DGRP SNPs. Given the large number of nonindependent tests, significance of the vector correlations was evaluated using two different approaches (see the SMR). Vector correlations $r > 0.5$ are listed in File S6.

The more conservative SNP-set method of determining significance yielded just five significant vector correlations,
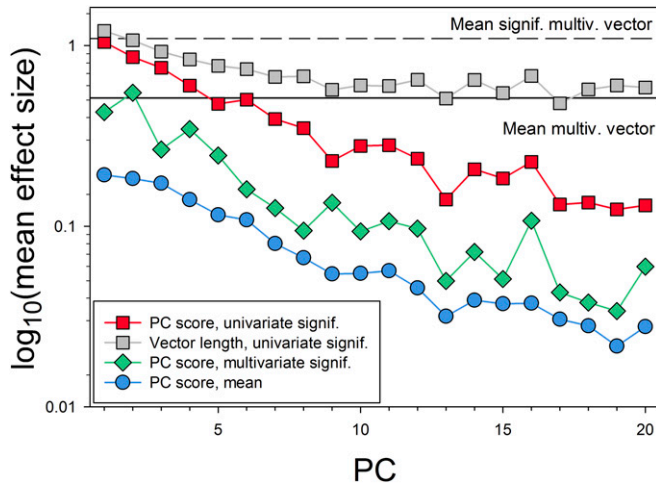
**Figure 6** Mean measures of multivariate and univariate effect size for SNPs categorized by significance of the univariate test on each PC using $P = 0.00007$ as a cutoff. Gray squares: total multivariate effect size for SNPs significant in the corresponding univariate analysis; red squares: univariate effect size for SNPs significant in the corresponding univariate analysis (also shown in Figure 5); green diamonds: univariate effect size score for SNPs significant in the multivariate analysis; and blue circles: univariate effect size score for all SNPs. Horizontal reference lines show the mean multivariate effect size for all SNPs (solid line) and for all SNPs significant in the multivariate analysis of variance analysis (dashed line). multiv., multivariate; PC, principal component; signif., significant.

none of which suggest functional relationships between the genes involved. The more liberal SNP-by-SNP correlations did suggest some intriguing connections that may be worth exploring in future work.

We highlight a cluster of correlations involving genes with known interactions between genes in the *Notch* (*N*), *hippo* (*hpo*), *wingless* (*wg*), and *decapentaplegic* (*dpp*) pathways. Vector correlations among the relevant dictionary and SNP effects are summarized in Table S1. Representative phenotypic effects are shown in Figure S4. Knockdowns of *scalloped* (*sd*), *Serrate* (*Ser*), and *dachsous* (*ds*) were all reasonably highly correlated. The transcription factor *sd* interacts physically with transcription factor *vestigial* (*vg*), whose expression is a key outcome of the wg, hpo, and dpp pathways (Matamoro-Vidal *et al.* 2015), and interacts genetically with genes in the N, hpo, wg, and dpp pathways (Shyamala and Chopra 1999; Djiane *et al.* 2013; Gramates *et al.* 2017). We also knocked down *Ser*, which is a ligand of the Notch protein (Rebay *et al.* 1991), and interacts genetically with the vg- and hpo-pathway genes (Gramates *et al.* 2017). SNPs in or near the genes *kuzbanian* (*kuz*), *mastermind* (*mam*), and *Actin5C* (*Act5C*) show high and significant vector correlations, with both *sd* and *Ser* knockdowns. Proteins coded for by *kuz* and *mam* physically and genetically interact with the N pathway (Qi *et al.* 1999; Petcherski and Kimble 2000; Dornier *et al.* 2012; Yuan *et al.* 2016; Gramates *et al.* 2017). Actin interacts physically with the *achaete-scute* complex proteins (Hsiao *et al.* 2014), which then signal neighboring cells through the N pathway (Heitzler *et al.* 1996). While a SNP in the

intron of *N* was significant in our initial GWAS, it did not survive in the LASSO analysis (its effect shrank down to zero). We note SNP-wise significant correlations, ranging from $r = 0.53$ to $0.89$, between the effect of a SNP near the gene *tout-velu* (*ttv*) and knockdowns of several hpo-pathway genes (*dachs*, *dachsous*, *four-jointed*, *fat*, and *mob as tumor suppressor*). These are potentially interesting because Ttv encodes a glycosyltransferase involved with heparin sulfate synthesis, which is known to be involved with diffusion of the major wing morphogens Hedgehog, wg and dpp (Bornemann *et al.* 2004; Han *et al.* 2004; Takei *et al.* 2004).

### Replication in the ME and NC2 populations

Of the 342 Freeze 2 SNPs selected for replication testing in the ME-NC2 panel, there were 45 significant tests at $P < 0.05$, higher than the expected 17 tests. Fifty-five of the 342 were significant in the MANOVA analyses and 27 were also estimated to have nonzero effects in the LASSO analysis (Table 3). Of these 27, there were 4 nominally significant results at $P < 0.05$, higher than the expected 1.35 cases, and 9 had $P < 0.1$. In addition, 4 of 27 vector correlations between the LASSO and ME-NC2 phenotypic effects were significant at $P < 0.05$.

SNP 2L:19,596,734 was significant in both the LASSO and the ME-NC2 tests. As shown in Figure 8, the vector correlation is a strikingly high 0.84, higher than any of the 10,000 random vector correlations used to judge significance. This SNP is in an intron of the gene *Lar* (*Leukocyte-antigen-related-like*). *Lar* is a promising candidate because it regulates signaling through Abl tyrosine kinase, a key component of a signaling complex that regulates cell adhesion, motility (Srinivasan *et al.* 2012; Barlan *et al.* 2017), and nervous system morphogenesis (Krueger *et al.* 1996). *Lar* is strongly expressed in the wing disc (O'Keefe *et al.* 2012).

The second SNP with a high vector correlation between the two data sets is in an intron of *RhoGEF64c*, which was also implicated by the dictionary vector analyses presented earlier. A SNP in the intron of the well-known regulator of wing-vein proliferation *plexus* (Matakatsu *et al.* 1999) is nearly significant in the ME-NC2 analysis. A ME-NC2-significant SNP is in an intron of *Dgk*, which is conjectured to have a role in nervous system development (Gramates *et al.* 2017). The other SNPs implicated by either significance or high vector correlations have no known functional connection to processes of wing development.

The mean vector correlation between ME-NC2 and LASSO-significant estimates was 0.23, while the mean random correlation was 0.21 with an SD of 0.15 over sets of 27 correlations.

### Discussion

Our results have implications for the usefulness of multivariate association analyses, for the genetic architecture of quantitative traits, in particular the inheritance and evolution of the *D. melanogaster* wing, the study of pleiotropy, and for the
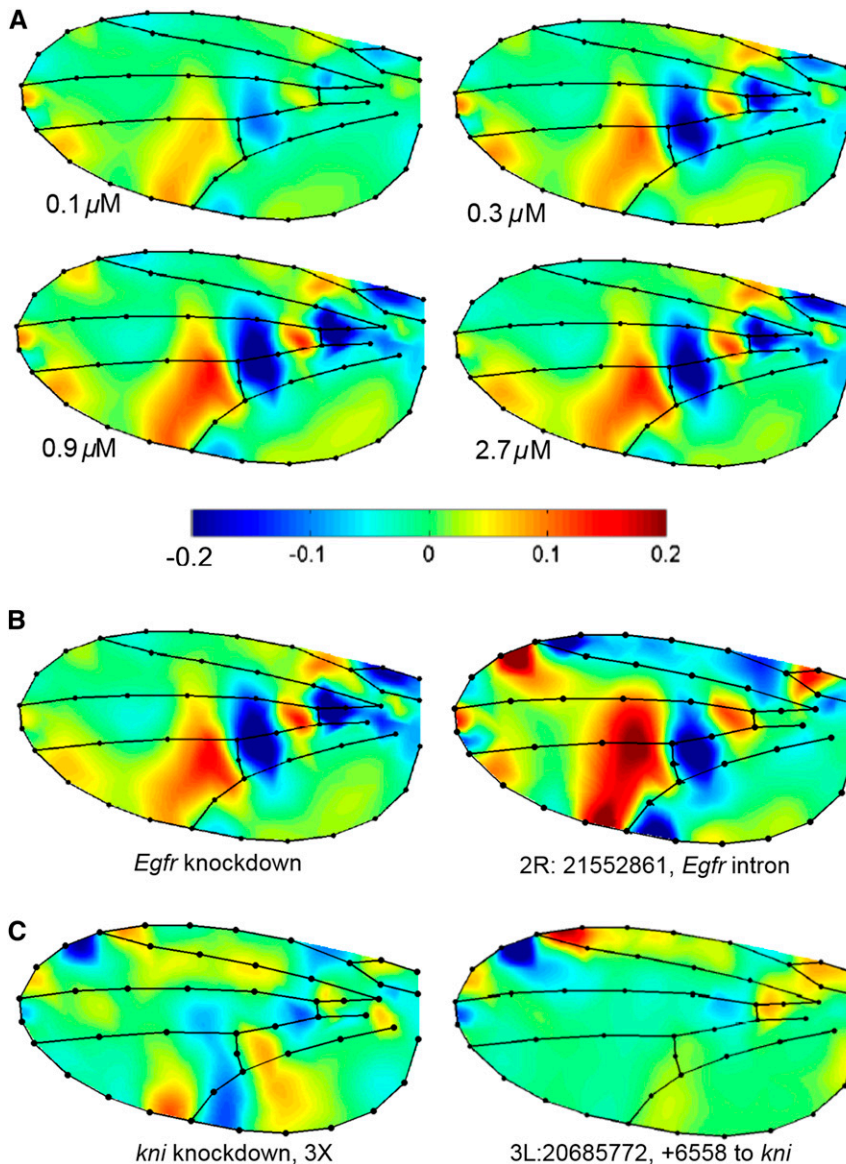
**Figure 7** Wing shape deformations inferred for gene knockdowns and SNP effects implicating corresponding genes. All deformations shown magnified 2× except for knirps knockdown. Deformation scale is log2, so +0.2 represents a 15% increase in the corresponding area of the wing. (A) Effects of different levels of *Egfr* knockdown on wing shape. (B) Comparison of knockdown (left) and LASSO SNP vector (right) for 2R:17440366, which is in an intron of *Egfr*. The *Egfr* knockdown is the regression of the shape changes shown in (A) on the level of mifepristone applied. The correlation between these vectors is 0.68. (C) Comparison of knockdown of knirps (left) and LASSO SNP vector for 3L:20685772, 6558-bp downstream from knirps. The correlation between these vectors is 0.54. LASSO, Least Absolute Shrinkage and Selection Operator.

study of the genotype–phenotype map. We discuss each of these in turn.

### Multivariate association analyses

Multivariate analyses increase both the power of association studies and the interpretability of the results obtained over a series of univariate analyses. For our data, the gain in power in the multivariate analysis was dramatic. At an FDR of 5%, 2396 SNPs were identified as potentially significant in the fully multivariate analysis. In comparison, the univariate analyses of PC scores identified just four significant variants on PC1, and none on the next 19 PCs when using the same FDR algorithm on each axis. When we utilized the *P*-value estimated from the FDR from the multivariate analysis ($P = 0.00007$), almost 7000 SNPs were nominally significant, but just 24 of those had significant effects on two or more PC axes. It is particularly notable that just 139 SNPs had significant effects in both the multivariate and at least one univariate analysis. As discussed above, we interpret this to mean that the FDR of the more liberal ($P = 0.00007$) univariate criterion is quite high.

The increased power of multivariate association studies has repeatedly been demonstrated in simulation studies using a variety of statistical methodologies (O'Reilly *et al.* 2012; Stephens 2013; van der Sluis *et al.* 2013; Zhou and Stephens 2014; Márquez and Houle 2015; Porter and O'Reilly 2017). This, in conjunction with our results, suggests that the expectation of increased power is general. Except in special cases, any multivariate analysis will be more powerful than the corresponding set of univariate analyses.

As noted in the *Introduction*, full multivariate mixed-model analyses might be preferable to the approximate and regression-based analyses that we actually used. We did attempt to fit mixed models using R, SAS, GEMMA (Zhou and Stephens 2014), and Wombat (Meyer and Tier 2012), but for various reasons this proved impossible or impractical due to long run

**Table 2 Comparison of knockdown (dictionary) effects on genes with LASSO-significant SNPs close to the coding region**

| Gene | Dictionary vector length | SNP effect | SNP | Location | Minor allele frequency | LD cluster | Vector correlation | P | Median | 95% |
|------|------|------|------|------|------|------|------|------|------|------|
| bru1 | 0.469 | 0.509 | 2L:12286157 | Intron | 0.04 | 214 | 0.358 | > | 0.283 | 0.635 |
| Bx | 1.200 | 0.659 | X:18441698 | Intron | 0.03 | 99 | 0.292 | | 0.303 | 0.703 |
| CG12934 | 1.916 | 0.038 | 2R:6539390 | +2055 | 0.04 | 326 | 0.338 | > | 0.191 | 0.481 |
| | | 0.191 | 2R:6539191 | +1856 | 0.06 | 327 | 0.451 | + | 0.188 | 0.505 |
| | | 0.448 | 2R:6539555 | +2220 | 0.08 | 327 | 0.191 | > | 0.191 | 0.459 |
| CG8654 | 0.414 | 0.551 | 2R:15872087 | 3′ UTR | 0.04 | 405 | 0.298 | > | 0.244 | 0.561 |
| dally | 0.896 | 0.478 | 3L:8864892 | Intron | 0.06 | 589 | 0.538 | > | 0.321 | 0.699 |
| ds | 5.766 | 0.072 | 2L:655894 | Intron | 0.45 | 114 | 0.696 | + | 0.346 | 0.759 |
| | | 0.159 | 2L:702560 | Intron | 0.06 | 115 | 0.528 | > | 0.377 | 0.750 |
| | | 0.100 | 2L:702798 | Intron | 0.09 | 115 | 0.263 | | 0.333 | 0.764 |
| | | 0.224 | 2L:718623 | −3640 | 0.03 | 116 | 0.472 | > | 0.360 | 0.744 |
| | | 0.110 | 2L:718627 | −3644 | 0.03 | 116 | 0.537 | > | 0.367 | 0.745 |
| Dys | 0.321 | 0.043 | 3R:15365550 | Intron | 0.09 | 855 | 0.237 | > | 0.150 | 0.410 |
| Egfr | 0.939 | 0.955 | 2R:17440366 | Intron | 0.05 | 445 | 0.683 | * | 0.309 | 0.646 |
| foxo | 0.266 | 0.051 | 3R:9898374 | Intron | 0.07 | 524 | 0.221 | | 0.244 | 0.549 |
| | | 0.176 | 3R:9904321 | Intron | 0.05 | 524 | 0.479 | > | 0.242 | 0.538 |
| ft | 4.302 | 0.050 | 2L:4220302 | Intron | 0.03 | 149 | 0.481 | > | 0.323 | 0.713 |
| RhoGEF64c | 1.664 | 0.515 | 3L:4740113 | Intron | 0.06 | 535 | 0.054 | | 0.181 | 0.472 |
| | | 0.709 | 3L:4744803 | Intron | 0.03 | 536 | 0.488 | * | 0.189 | 0.478 |
| kay | 0.453 | 0.030 | 3R:25594742 | Intron | 0.03 | 920 | 0.587 | + | 0.262 | 0.601 |
| | | 0.082 | 3R:25594743 | Intron | 0.03 | 920 | 0.545 | > | 0.271 | 0.623 |
| kni | 0.359 | 0.484 | 3L:20678872 | +6558 | 0.06 | 727 | 0.541 | * | 0.220 | 0.53 |
| | | 0.076 | 3L:20679438 | +5992 | 0.03 | 728 | 0.195 | | 0.225 | 0.527 |
| luna | 0.369 | 0.004 | 2R:6881826 | Intron | 0.03 | 332 | 0.213 | | 0.300 | 0.686 |
| MRP | 0.442 | 0.451 | 2L:12747188 | +2381 | 0.06 | 217 | 0.546 | * | 0.191 | 0.481 |
| tx | 0.968 | 0.140 | 3R:22260272 | −6228 | 0.03 | 549 | 0.251 | | 0.332 | 0.709 |

Quantiles calculated for vector correlations of dictionary vectors with 10,000 random vectors estimated in the multivariate analysis of variance analysis. LD, linkage disequilibrium; >, $P < 0.5$; +, + $P < 0.1$; *, $P < 0.05$.

times. Wombat readily estimates SNP effects but does not directly estimate P-values for multivariate fixed effects.

A second important advantage of multivariate analyses is that multivariate effects are far more informative than the set of yes-or-no decisions about which traits are affected by each SNP that result from standard univariate testing. Multivariate effect vectors enable us to assess the overall degree of similarity of effects using the correlation between vectors. We performed two sets of validation experiments using this approach and their results provided confirmatory evidence for the effects of some segregating sites. The phenotypic effects of knockdowns of genes implicated in the initial GWAS provided good evidence that SNP effects are overall more similar to these than expected under the null hypothesis of no similarity. In several cases, the effects of particular SNP-knockdown pairs are individually more similar than random vectors (Figure 7 and Table 2). Replication of SNPs in a second panel of lines from natural populations from ME-NC2 again suggested that some SNPs had very similar effects in both mapping populations.

### Inheritance of Drosophila wing shape

Our study implicated a large number of sites as potentially affecting wing shape and size, consistent with previous studies suggesting that the inheritance of wing shape is highly polygenic. Previous association studies on aspects of wing shape in D. melanogaster have also detected relatively large numbers

of QTL, given the number of markers employed. Weber et al. (1999, 2001) generated recombinant inbred lines (RILs) between populations selected for high and low values of a univariate wing shape index, and found ≥ 20 sites over the two largest chromosomes with uniformly small effects. Zimmerman et al. (2000) found evidence for a dozen QTL for several aspects of wing shape in each of two small mapping populations, each founded by crossing two inbred lines. Mezey et al. (2005) mapped at least 21 QTL for the first seven PCs of wing shape in a set of RILs derived from the cross of a single pair of wild-collected flies. This consistent finding of many significant effects is reminiscent of the genetic architecture of human height, where a large number of sites with individually small effects are responsible for the standing variation (Lango Allen et al. 2010; Yang et al. 2010; Wood et al. 2014).

The identity of the genes, as well as the validation studies with gene knockdowns (the dictionary) and the replication in independent mapping populations from ME-NC2 populations, give us confidence that we have identified at least some of the SNPs causing shape variation in the DGRP population. We found significant phenotypic similarities between some SNPs mapping close to a particular gene and the phenotypic effects of knockdowns of those genes. The GO analysis shows that our LASSO-significant SNPs tend to map in or close to genes involved in biological processes known to be involved in wing development (Matamoro-Vidal et al. 2015). More

**Table 3 LASSO-significant SNPs scored in the ME and NC2 populations**

| SNP | Nearest gene | Distance[b] | Loc. | ME-NC2 $P^a$ | Effect size ME-NC2 | Effect size LASSO | Vector correlation $r$ | Vector correlation $P^c$ |
|---|---|---|---|---|---|---|---|---|
| X:7053042 | CG12541 | 0 | Intron | 0.67 | 0.280 | 0.257 | 0.28 | < |
| X:8105108 | Gclc | 0 | Syn. | 0.51 | 0.214 | 0.254 | 0.14 | |
| X:9749045 | Sp1 | 0 | Intron | 0.35 | 0.656 | 0.403 | 0.26 | < |
| X:19619597 | CG14221 | 0 | Syn. | 0.12 | 0.277 | 0.026 | 0.03 | |
| 2L:5147351 | Msp300 | 0 | Intron | 0.52 | 0.278 | 0.137 | 0.07 | |
| 2L:11348992 | CG14926 | −5729 | Inter. | 0.09 | 0.635 | 0.003 | 0.00 | |
| 2L:12747188 | MRP | 2381 | Inter. | 0.61 | 0.570 | 0.451 | 0.36 | < |
| 2L:19219145 | Gamma Tub37 | −33 k | Inter. | 0.07 | 0.531 | 0.147 | 0.02 | |
| 2L:19596734 | Lar | 0 | Intron | 0.03 | 0.438 | 0.468 | 0.84 | *** |
| 2R:7599329 | CG1358 | 0 | Intron | 0.11 | 0.580 | 0.240 | 0.54 | < |
| 2R:7844508 | Dgk | 0 | Intron | 0.03 | 0.425 | 0.210 | 0.10 | |
| 2R:10652050 | CG12934 | 2220 | Inter. | 0.30 | 0.405 | 0.449 | 0.13 | |
| 2R:12571322 | stil | 0 | Intron | 0.55 | 0.645 | 0.064 | 0.03 | |
| 2R:14822284 | Pgm2b | −2297 | Inter. | 0.04 | 0.274 | 0.150 | 0.06 | |
| 2R:15922844 | bdg | 196 | Inter. | 0.35 | 0.239 | 0.052 | 0.08 | |
| 2R:20505363 | Obp57a | 0 | Syn. | 0.06 | 0.390 | 0.187 | 0.29 | < |
| 2R:22511711 | px | 0 | Intron | 0.05 | 0.351 | 0.173 | 0.11 | < |
| 3L:3632969 | dar1 | 0 | Intron | 0.32 | 0.578 | 0.314 | 0.36 | < |
| 3L:4740113 | RhoGEF64C | 0 | Intron | 0.35 | 0.605 | 0.515 | 0.49 | *** |
| 3L:4962783 | Con | 0 | Intron | 0.20 | 0.576 | 0.433 | 0.34 | < |
| 3L:12202436 | Adk1 | 0 | Intron | 0.15 | 0.309 | 0.299 | 0.54 | * |
| 3L:15043133 | ind | 0 | Syn. | 0.58 | 0.252 | 0.224 | 0.04 | |
| 3L:19310451 | pip | 0 | Syn. | 0.99 | 0.696 | 0.490 | 0.47 | + |
| 3L:20560300 | Rcd2 | −20 k | Inter. | 0.01 | 1.232 | 0.354 | 0.04 | |
| 3R:8491058 | CG43462 | −13 k | Inter. | 0.11 | 0.517 | 0.361 | 0.12 | |
| 3R:16194246 | sra | 0 | Intron | 0.45 | 0.567 | 0.339 | 0.51 | * |
| 3R:19539828 | Dys | 0 | Intron | 0.08 | 1.003 | 0.043 | 0.05 | |

ME-NC2, Maine-North Carolina 2; Loc., location; LASSO, Least Absolute Shrinkage and Selection Operator; Syn., Coding region SNP with synonymous substitution; Inter., Integenic SNP.

[a] $P$-value from the ME and NC2 multivariate analysis of variance analysis.

[b] Distance to transcript in base pairs, $k = 1000$.

[c] <, $P < 0.5$; +, $P < 0.1$; *, $P < 0.05$; and ***, $P < 0.001$.

speculatively, we identified significant similarities between the effects of SNPs near genes with known roles in the Notch, wingless, decapentaplegic, and hippo pathways, and the effects of knockdowns of genes potentially involved in the integration of signaling of those same pathways.

The regularized effect sizes estimated by the LASSO analysis were on average quite small, explaining, on average, just 0.1% of the genetic variance in wing shape and size. Nevertheless, 17 SNPs were estimated to explain > 1% of the variance. The LASSO effects were 70% smaller than the corresponding initial estimates obtained from the MANOVA analyses, suggesting a substantial Beavis effect for the initial analyses. Simulations of data sets similar to ours (Márquez and Houle 2015) that show the power of our experiments is quite modest for sites that explain just 1% of the variation, perhaps just 20%. Such low power means that our first-round, SNP-by-SNP analyses would only have detected a minority of all the variants with an effect on the phenotype.

The LASSO is a useful tool for variable selection, but does have limitations that are important to bear in mind when interpreting our results. When the number of predictors, $p$, is larger than the d.f. in the data, $N$, the LASSO algorithm will return at most $N - 1$ nonzero effects (Zou and Hastie 2005).

The LASSO can select the causal SNPs from a set of correlated SNPs (that is, it will be statistically consistent) when $p$ is large, as long as the SNPs are not too correlated with each other (Knight and Fu 2000; Zhao and Yu 2006; Bunea et al. 2007; Wainwright 2009; reviewed in section 3 of Meinshausen and Bühlmann 2010) and the number of predictors with causal effects is less than the d.f. Unfortunately, these conditions are not met for the genome as a whole, given the evidence that the inheritance of wing shape is polygenic. More specifically, for many of our SNP-by-SNP LASSO analyses, some of the predictor SNPs are in substantial LD (Houle and Márquez 2015). In such cases, the LASSO will select one of each set of highly correlated SNPs for inclusion in the model, but it will not necessarily be the actual causal SNP.

Consequently, we also track the degree of disequilibrium of each SNP in our models. Table S2 gives summary diagnostics about this degree of correlation, such as the number and map distance to perfectly correlated SNPs, the total number of competitor SNPs, and how many of the potentially significant SNPs are correlated with the focal SNP. In most cases, the 580 SNPs retained in our final LASSO models had relatively small numbers of highly correlated competitor predictors. This suggests that our filtered list of 580 SNPs is relatively conservative.
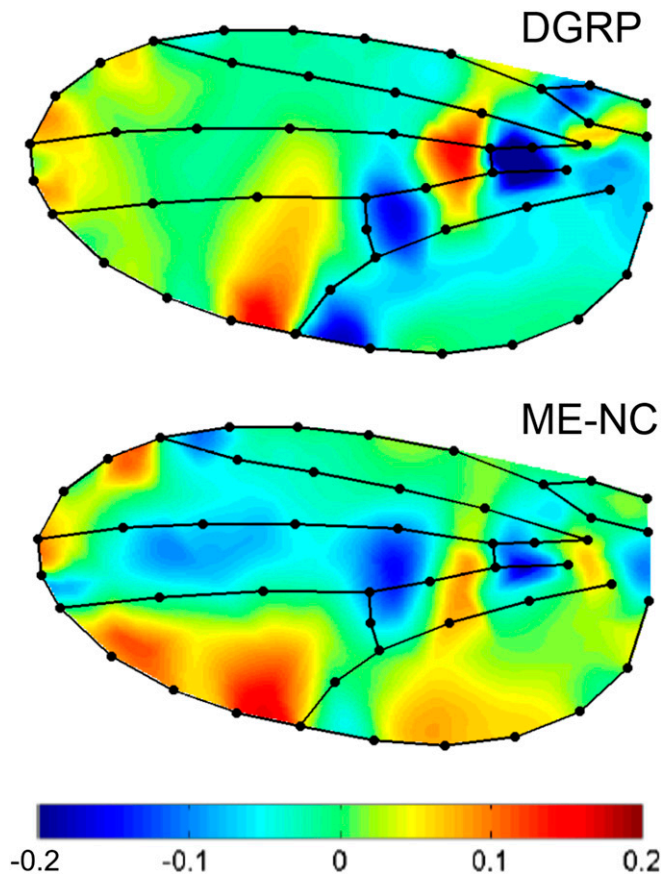
**Figure 8** Wing-shape deformations inferred for a SNP in the intron of Lar. Above: effect of SNP 2L:19596734 in the DGRP. Below: effect of SNP 2L:19596734 in the ME-NC2 replication population. Vector correlation between the vectors of effects from the DGRP and the ME-NC2 was 0.84. Both effects shown at 3×. DGRP, *Drosophila* Genome Reference Panel; ME-NC, Maine-North Carolina.

We also carried out a third round of analyses with all of the LASSO-significant SNPs as predictors, as suggested by Meinshausen (2007). This further reduces the effect sizes of SNPs, suggesting that our SNP models are still overfitted. These models fall afoul of the sample size limitation mentioned above, where the number of selected variables must be less than the number of samples (Zou and Hastie 2005). We simply do not have enough data to fit a single overarching model of genetic effects.

### Pleiotropy in the Drosophila wing

The *Drosophila* wing is a single structure, consisting of veins connected by wing blade tissue. The integration enforced by the physical connection between each part of the wing and the continuity of these structures throughout development makes it a natural subject for a multivariate genetic analysis. Any change during development that affects one aspect of the wing, such as the length of a particular vein, must also affect adjacent areas of the wing. The processes most likely to affect wing shape and size are the pattern of growth of the wing tissue, the differentiation of veins from nonvein tissue, and

the rearrangement and planar polarization of cells during pupariation (Matamoro-Vidal *et al.* 2015). The known candidate genetic pathways that affect these key developmental events have effects across broad regions of the wing, rather than being confined to one small area. These considerations suggest that all effects on wing shape will be pleiotropic.

A second layer of dependency among measurements of the wing is introduced by the geometric morphometric analysis that we used (Zelditch *et al.* 2004). There is no one reference structure on a complex integrated morphological structure like a wing that can be used as a standard for comparison with the locations of other structures. One can only interpret the relative locations of all measured structures to each other. Even if we could imagine a developmental change that just moved the location of a single landmark (Figure 1A) on the wing in just one dimension, say the location of the distal end of vein L5 in the X-direction, then geometric morphometrics would recover an effect vector with not only a large change in that single X-dimension, but also compensatory changes in the locations of all other landmarks. With geometric morphometrics it is not possible to define a set of shape traits that can be measured independently of all other shape traits.

While it is useful to ask whether there is pleiotropy between autonomous morphological structures, such as the wing and the eye of a fly, or between morphology and other types of traits, such as life history or behavior, both geometry and biology lead to a strong expectation of within-structure pleiotropy.

The MANOVA that we used in this study can be thought of as consisting of two steps. The first step defines a trait: the direction in phenotype space that maximizes the distinctness of the means of the two genotypes. The second step estimates the statistical significance of a difference of that magnitude across the whole-phenotype space. In the more typical series of univariate analyses that are applied to multivariate phenotypes, for example analyses of trait scores on PC axes, only a finite set of traits are chosen. This finite set cannot sample all possible types of effects and statistical tests generally ignore the fact that such tests implicitly are sampled from a larger set of directions. This is why multivariate analyses can be both more powerful than a set of univariate analyses and univariate tests may have higher false positive rates, as suggested by the mismatch between our univariate and multivariate results.

In addition, this aspect of MANOVA corresponds to our intuition about what variant genotypes that actually affect the integrated wing phenotype should do: in principle, every site affecting wing development could do so in a slightly different way, and each of those changes will have pleiotropic effects that extend across the wing. The plots of wing shape change in Figure 7, Figure 8, and Figure S5 represent estimates of those directions of some of our significant SNPs.

A critical justification for transitioning from univariate to multivariate association studies is to enable the study of the genotype–phenotype map, *i.e.,* how genomic variation is translated into phenotypic variation (Houle 2010; Houle *et al.* 2010). Every phenotypic effect will have a molecular

origin, for example in gene expression, which then ramifies outward to cells, tissues, and finally to the outward aspects of organismal form and function, such as morphology and behavior. Each such molecular change may have effects on many whole-organism phenotypes. For example, the study of even the simplest monogenic human genetic diseases, such as sickle cell anemia, often reveals a host of disorders tracing back to the single genetic cause. Decisions about how to treat genetic disease, the value of a genetic variant in plant or animal breeding, or whether an endangered population is likely to adapt to a changing environment will be improved when we have information about all of the pleiotropic effects of genetic variation, and not just the few that happen to have been studied.

### Dimensionality as a blessing rather than a curse

With all the advantages of multivariate association analyses detailed above, why are they still rare? In some cases, there are substantial statistical barriers to a fully multivariate analysis. For example, it is challenging to combine binomial and normal variates in the same analysis, although solutions have been proposed (*e.g.*, O'Reilly *et al.* 2012). Many multivariate data sets have incomplete phenotypic data, and restricting the analysis to just those individuals with complete data may reduce sample size too much for reasonable inference. Multivariate methods are unfamiliar to many researchers, posing a relatively simple hurdle to their adoption.

A final factor interfering with the widespread adoption of multivariate methods is the "curse of dimensionality." This phrase was originally coined by Bellman (1957) and has since become a meme useful for causing unease about multivariate analyses, even when the nature of the curse remains implicit. It generally denotes the notion that the hypervolume of sample space increases rapidly with the number of dimensions measured, while the sample size remains fixed, resulting in data that is ever sparser as dimensionality increases (Hastie *et al.* 2009). Zimek *et al.* (2012) identified eight separate challenges that increase with dimensionality of the data just in the realm of distance-based analyses (such as detecting neighbors, hubs, outliers, *etc.*). They also noted that many of these are problematic only in the limiting case where all variables are independently and identically distributed. However, real biological data are always correlated and often clustered. In particular, we know that genetic effects must cause the clustering of individuals with similar genotypes in phenotype space. Our argument that the relationship between vectors of effects is more informative in a high-dimensional data space is essentially the other side of the standard sparsity argument. Effects become more informative because a finite set of real effects must be sparser in a larger space, and therefore both similarities and differences become more informative.

Another challenge frequently posited is that a large proportion of the measurements in a high-dimensional data set may be irrelevant. Indeed, our simulations show that the power of an association study declines when traits without

any genetic basis are measured (Márquez and Houle 2015). We are confident that the number of traits can usually be increased without reaching this limit. With the exception of eQTL studies, the current standard approach to GWAS includes just a few traits. Until other forms of high-throughput and automated phenotyping become available, biological measurements will usually be expensive and time-consuming to make, ensuring that considerable thought is often expended on what to measure. Furthermore, the appropriate dimension for analysis can be estimated from data on related individuals (Kirkpatrick and Meyer 2004; Meyer and Kirkpatrick 2005, 2008). In general, PC analysis can reveal how much new information is added when another trait is measured and a cutoff that seems likely to capture most genetic variation chosen.

The best answer to the concern that dimensionality can be a curse are the many independent simulation studies that consistently show that the power of multivariate analyses is higher than univariate analyses. Most of these studies also analyze real data sets and invariably find more associations in multivariate than univariate analyses (O'Reilly *et al.* 2012; Stephens 2013; Scutari *et al.* 2014; Zhou and Stephens 2014; Porter and O'Reilly 2017). These simulations make different assumptions, and apply a wide variety of well-established or experimental multivariate analyses.

We believe that researchers should invoke the blessings of dimensionality rather than its potential to be a curse. Multivariate analyses will generally be more powerful. The ability to estimate the direction of effects becomes more salient with the dimension of the space studied. The phenomenon of pleiotropy simply cannot be studied unless multiple traits are studied together. Those interested in the inheritance of complex traits and the genotype–phenotype map should adopt multivariate approaches whenever it is feasible to do so.

## Literature Cited

Anderson, J. T., C.-R. Lee, and T. Mitchell-Olds, 2011   Life-history QTLs and natural selection on flowering time in *Boechera stricta*, a perennial relative of *Arabidopsis*. Evolution 65: 771–787. https://doi.org/10.1111/j.1558-5646.2010.01175.x

Barlan, K., M. Cetera, and S. Horne-Badovinac, 2017   Fat2 and Lar define a basally localized planar signaling system controlling collective cell migration. Dev. Cell 40: 467–477.e5. https://doi.org/10.1016/j.devcel.2017.02.003

Battle, A., S. Mostafavi, X. Zhu, J. B. Potash, M. M. Weissman *et al.*, 2014   Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. Genome Res. 24: 14–24. https://doi.org/10.1101/gr.155192.113

Beavis, W. D., 1994   The power and deceit of QTL experiments: lessons from comparative QTL studies, pp. 250–266 in *Proceedings of the Forty-Ninth Annual Corn & Sorghum Industry Research Conference*. American Seed Trade Association, Washington, DC.

Beavis, W. D., 1998   QTL analyses: power, precision, and accuracy, pp. 145–162 in *Molecular Dissection of Complex Traits*. CRC Press, Boca Raton, FL.

Bellman, R., 1957   *Dynamic Programming*. Princeton University, Princeton.

Benjamini, Y., and Y. Hochberg, 1995   Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. B 57: 289–300.

Blair, S. S., 2007   Wing vein patterning in *Drosophila* and the analysis of intercellular signaling. Annu. Rev. Cell Dev. Biol. 23: 293–319. https://doi.org/10.1146/annurev.cellbio.23.090506.123606

Bolstad, G. H., J. A. Cassara, E. Márquez, T. F. Hansen, K. Van Der Linde *et al.*, 2015   Complex constraints on allometry revealed by artificial selection on the wing of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 112: 13284–13289. https://doi.org/10.1073/pnas.1505357112

Bornemann, D. J., J. E. Duncan, W. Staatz, S. Selleck, and R. Warrior, 2004   Abrogation of heparan sulfate synthesis in *Drosophila* disrupts the Wingless, Hedgehog and Decapentaplegic signaling pathways. Development 131: 1927–1938. https://doi.org/10.1242/dev.01061

Bunea, F., A. Tsybakov, and M. Wegkamp, 2007   Sparsity oracle inequalities for the Lasso. Electron. J. Stat. 1: 169–194. https://doi.org/10.1214/07-EJS008

Carter, A. J. R., and D. Houle, 2011   Artificial selection reveals heritable variation for developmental instability. Evolution 65: 3558–3564. https://doi.org/10.1111/j.1558-5646.2011.01393.x

Corbett-Detig, R. B., and D. L. Hartl, 2012   Population genomics of inversion polymorphisms in *Drosophila melanogaster*. PLoS Genet. 8: e1003056 [corrigenda: PLoS Genet. 9 (2013)]. https://doi.org/10.1371/journal.pgen.1003056

Corbett-Detig, R. B., C. Cardeno, and C. H. Langley, 2012   Sequence-based detection and breakpoint assembly of polymorphic inversions. Genetics 192: 131–137. https://doi.org/10.1534/genetics.112.141622

de Celis, J. F., and F. J. Diaz-Benjumea, 2003   Developmental basis for vein pattern variations in insect wings. Int. J. Dev. Biol. 47: 653–663.

Dietzl, G., D. Chen, F. Schnorrer, K. C. Su, Y. Barinova *et al.*, 2007   A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. Nature 448: 151–156. https://doi.org/10.1038/nature05954

Djiane, A., A. Krejci, F. Bernard, S. Fexova, K. Millen *et al.*, 2013   Dissecting the mechanisms of Notch induced hyperplasia. EMBO J. 32: 60–71. https://doi.org/10.1038/emboj.2012.326

Dornier, E., F. Coumailleau, J. F. Ottavi, J. Moretti, C. Boucheix *et al.*, 2012   TspanC8 tetraspanins regulate ADAM10/Kuzbanian trafficking and promote Notch activation in flies and mammals. J. Cell Biol. 199: 481–496 [corrigenda: J. Cell Biol. 213: 495–496 (2016)]. https://doi.org/10.1083/jcb.201201133

Dworkin, I., and G. Gibson, 2006   Epidermal growth factor receptor and transforming growth factor-β signaling contributes to variation for wing shape in *Drosophila melanogaster*. Genetics 173: 1417–1431. https://doi.org/10.1534/genetics.105.053868

Dworkin, I., A. Palsson, and G. Gibson, 2005   Replication of an *Egfr*-wing shape association in a wild-caught cohort of *Drosophila melanogaster*. Genetics 169: 2115–2125. https://doi.org/10.1534/genetics.104.035766

Friedman, J., T. Hastie, and R. Tibshirani, 2010   Regularization paths for generalized linear models via coordinate descent. J. Stat. Softw. 33: 1. https://doi.org/10.18637/jss.v033.i01

Goering, L. M., P. K. Hunt, C. Heighington, C. Busick, P. S. Pennings *et al.*, 2009   Association of *orthodenticle* with natural variation for early embryonic patterning in *Drosophila melanogaster*. J. Exp. Zoolog. B Mol. Dev. Evol. 312B: 841–854. https://doi.org/10.1002/jez.b.21299

Gramates, L. S., S. J. Marygold, G. D. Santos, J. M. Urbano, G. Antonazzo *et al.*, 2017   FlyBase at 25: looking to the future. Nucleic Acids Res. 45: D663–D671. https://doi.org/10.1093/nar/gkw1016

Greenberg, L., and V. Hatini, 2011   Systematic expression and loss-of-function analysis defines spatially restricted requirements for *Drosophila* RhoGEFs and RhoGAPs in leg morphogenesis. Mech. Dev. 128: 5–17. https://doi.org/10.1016/j.mod.2010.09.001

Guichard, A., B. Biehs, M. A. Sturtevant, L. Wickline, J. Chacko *et al.*, 1999   rhomboid and Star interact synergistically to promote EGFR/MAPK signaling during *Drosophila* wing vein development. Development 126: 2663–2676.

Han, C., T. Y. Belenkaya, M. Khodoun, M. Tauchi, X. D. Lin *et al.*, 2004   Distinct and collaborative roles of *Drosophila* EXT family proteins in morphogen signalling and gradient formation. Development 131: 1563–1575. https://doi.org/10.1242/dev.01051

Hastie, T., and J. Qian, 2016   Glmnet Vignette. Available at: https://web.stanford.edu/~hastie/Papers/Glmnet_Vignette.pdf. Accessed: March 11, 2018

Hastie, T., R. Tibshirani, and J. Friedman, 2009   *The Elements of Statistical Learning: Data Mining, Inference and Predition*. Springer, New York. https://doi.org/10.1007/978-0-387-84858-7

Heitzler, P., M. Bourouis, L. Ruel, C. Carteret, and P. Simpson, 1996   Genes of the Enhancer of split and achaete-scute complexes are required for a regulatory loop between Notch and Delta during lateral signalling in *Drosophila*. Development 122: 161–171.

Houle, D., 2010   Colloquium papers: numbering the hairs on our heads: the shared challenge and promise of phenomics. Proc. Natl. Acad. Sci. USA 107: 1793–1799. https://doi.org/10.1073/pnas.0906195106

Houle, D., and J. Fierst, 2013   Properties of spontaneous mutational variance and covariance for wing size and shape in *Drosophila melanogaster*. Evolution 67: 1116–1130. https://doi.org/10.1111/j.1558-5646.2012.01838.x

Houle, D., and E. Márquez, 2015   Linkage disequilibrium and inversion-typing of the *Drosophila* melanogaster genome reference panel. G3 (Bethesda) 5: 1695–1701. https://doi.org/10.1534/g3.115.019554

Houle, D., J. Mezey, P. Galpern, and A. Carter, 2003  Automated measurement of *Drosophila* wings. BMC Evol. Biol. 3: 25. https://doi.org/10.1186/1471-2148-3-25

Houle, D., D. R. Govindaraju, and S. W. Omholt, 2010  Phenomics: the next challenge. Nat. Rev. Genet. 11: 855–866. https://doi.org/10.1038/nrg2897

Houle, D., G. H. Bolstad, K. Van Der Linde, and T. F. Hansen, 2017  Mutation predicts 40 million years of fly wing evolution. Nature 548: 447–450. https://doi.org/10.1038/nature23473

Hsiao, Y.-L., Y.-J. Chen, Y.-J. Chang, H.-F. Yeh, Y.-C. Huang *et al.*, 2014  Proneural proteins Achaete and Scute associate with nuclear actin to promote formation of external sensory organs. J. Cell Sci. 127: 182–190. https://doi.org/10.1242/jcs.134718

Huang, W., A. Massouras, Y. Inoue, J. Peiffer, M. Rámia *et al.*, 2014  Natural variation in genome architecture among 205 - *Drosophila melanogaster* Genetic Reference Panel lines. Genome Res. 24: 1193–1208. https://doi.org/10.1101/gr.171546.113

Kirkpatrick, M., and K. Meyer, 2004  Direct estimation of genetic principal components: simplified analysis of complex phenotypes. Genetics 168: 2295–2306. https://doi.org/10.1534/genetics.104.029181

Knight, K., and W. Fu, 2000  Asymptotics for lasso-type estimators. Ann. Stat. 28: 1356–1378. https://doi.org/10.1214/aos/1015957397

Krueger, N. X., D. Van Vactor, H. I. Wan, W. M. Gelbart, C. S. Goodman *et al.*, 1996  The transmembrane tyrosine phosphatase DLAR controls motor axon guidance in *Drosophila*. Cell 84: 611–622. https://doi.org/10.1016/S0092-8674(00)81036-3

Lange, C., S. K. Silverman, X. Xu, S. T. Weiss, and N. M. Laird, 2003  A multivariate family-based association test using generalized estimating equations: FBAT-GEE. Biostatistics 4: 195–206. https://doi.org/10.1093/biostatistics/4.2.195

Langley, C. H., K. Stevens, C. Cardeno, Y. C. G. Lee, D. R. Schrider *et al.*, 2012  Genomic variation in natural populations of *Drosophila melanogaster*. Genetics 192: 533–598. https://doi.org/10.1534/genetics.112.142018

Lango Allen, H., K. Estrada, G. Lettre, S. I. Berndt, M. N. Weedon *et al.*, 2010  Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467: 832–838. https://doi.org/10.1038/nature09410

Lewontin, R. C., 1974  *The Genetic Basis of Evolutionary Change*. Columbia University Press, New York.

Lunde, K., B. Biehs, U. Nauber, and E. Bier, 1998  The knirps and knirps-related genes organize development of the second wing vein in *Drosophila*. Development 125: 4145–4154.

Lunde, K., J. L. Trimble, A. Guichard, K. A. Guss, U. Nauber *et al.*, 2003  Activation of the knirps locus links patterning to morphogenesis of the second wing vein in *Drosophila*. Development 130: 235–248. https://doi.org/10.1242/dev.00207

Mackay, T. F. C., S. Richards, E. A. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012  The *Drosophila melanogaster* Genetic Reference Panel. Nature 482: 173–178. https://doi.org/10.1038/nature10811

Márquez, E. J., 2012–2014  CPR: using *Drosophila* wing shape data. Available at: https://www.bio.fsu.edu/~dhoule/wings.html

Márquez, E. J., and D. Houle, 2015  Dimensionality and the statistical power of multivariate genome-wide association studies. bioRxiv. Available at: https://doi.org/10.1101/016592.

Matakatsu, H., R. Tadokoro, S. Gamo, and S. Hayashi, 1999  Repression of the wing vein development in *Drosophila* by the nuclear matrix protein Plexus. Development 126: 5207–5216.

Matamoro-Vidal, A., I. Salazar-Ciudad, and D. Houle, 2015  Making quantitative morphological variation from basic developmental processes: where are we? The case of the *Drosophila* wing. Dev. Dyn. 244: 1058–1073. https://doi.org/10.1002/dvdy.24255

Meinshausen, N., 2007  Relaxed lasso. Comput. Stat. Data Anal. 52: 374–393. https://doi.org/10.1016/j.csda.2006.12.019

Meinshausen, N., and P. Bühlmann, 2010  Stability selection. J. R. Stat. Soc. Series B Stat. Methodol. 72: 417–473. https://doi.org/10.1111/j.1467-9868.2010.00740.x

Meyer, K., 2006-2018  *Wombat: A Program for Mixed Model Analyses by Restricted Maximum Likelihood*. Animal Genetics and Breeding Unit, Amidale. http://didgeridoo.une.edu.au/km/wombat.php

Meyer, K., 2007  WOMBAT–A tool for mixed model analyses in quantitative genetics by restricted maximum likelihood (REML). J. Zhejiang Univ. Sci. B 8: 815–821. https://doi.org/10.1631/jzus.2007.B0815

Meyer, K., and M. Kirkpatrick, 2005  Restricted maximum likelihood estimation of genetic principal components and smoothed covariance matrices. Genet. Sel. Evol. 37: 1–30. https://doi.org/10.1186/1297-9686-37-1-1

Meyer, K., and M. Kirkpatrick, 2008  Perils of parsimony: properties of reduced-rank estimates of genetic covariance matrices. Genetics 180: 1153–1166. https://doi.org/10.1534/genetics.108.090159

Meyer, K., and B. Tier, 2012  "SNP snappy": a strategy for fast genome-wide association studies fitting a full mixed model. Genetics 190: 275–277. https://doi.org/10.1534/genetics.111.134841

Mezey, J. G., and D. Houle, 2005  The dimensionality of genetic variation for wing shape in *Drosophila melanogaster*. Evolution 59: 1027–1038. https://doi.org/10.1111/j.0014-3820.2005.tb01041.x

Mezey, J. G., D. Houle, and S. V. Nuzhdin, 2005  Naturally segregating quantitative trait loci affecting wing shape of *Drosophila melanogaster*. Genetics 169: 2101–2113. https://doi.org/10.1534/genetics.104.036988

Ni, J. Q., M. Markstein, R. Binari, B. Pfeiffer, L. P. Liu *et al.*, 2008  Vector and parameters for targeted transgenic RNA interference in *Drosophila melanogaster*. Nat. Methods 5: 49–51. https://doi.org/10.1038/nmeth1146

O'Keefe, D. D., S. R. Thomas, K. Bolin, E. Griggs, B. A. Edgar *et al.*, 2012  Combinatorial control of temporal gene expression in the *Drosophila* wing by enhancers and core promoters. BMC Genomics 13: 498. https://doi.org/10.1186/1471-2164-13-498

O'Reilly, P. F., C. J. Hoggart, Y. Pomyen, F. C. F. Calboli, P. Elliott *et al.*, 2012  MultiPhen: joint model of multiple phenotypes can increase discovery in GWAS. PLoS One 7: e34861. https://doi.org/10.1371/journal.pone.0034861

Palsson, A., and G. Gibson, 2000  Quantitative developmental genetic analysis reveals that the ancestral dipteran wing vein prepattern is conserved in *Drosophila melanogaster*. Dev. Genes Evol. 210: 617–622. https://doi.org/10.1007/s004270000107

Palsson, A., and G. Gibson, 2004  Association between nucleotide variation in *Egfr* and wing shape in *Drosophila melanogaster*. Genetics 167: 1187–1198. https://doi.org/10.1534/genetics.103.021766

Palsson, A., A. Rouse, R. Riley-Berger, I. Dworkin, and G. Gibson, 2004  Nucleotide variation in the *Egfr* locus of *Drosophila melanogaster*. Genetics 167: 1199–1212. https://doi.org/10.1534/genetics.104.026252

Patterson, N., A. L. Price, and D. Reich, 2006  Population structure and eigenanalysis. PLoS Genet. 2: e190. https://doi.org/10.1371/journal.pgen.0020190

Paul, L., S.-H. Wang, S. N. Manivannan, L. Bonanno, S. Lewis *et al.*, 2013  Dpp-induced Egfr signaling triggers postembryonic wing development in *Drosophila*. Proc. Natl. Acad. Sci. USA 110: 5058–5063. https://doi.org/10.1073/pnas.1217538110

Petcherski, A. G., and J. Kimble, 2000  Mastermind is a putative activator for Notch. Curr. Biol. 10: R471–R473. https://doi.org/10.1016/S0960-9822(00)00577-7

Pitchers, W., J. E. Pool, and I. Dworkin, 2013  Altitudinal clinal variation in wing size and shape in African *Drosophila melanogaster*: one cline or many? Evolution 67: 438–452. https://doi.org/10.1111/j.1558-5646.2012.01774.x

Pool, J. E., 2015 The mosaic ancestry of the *Drosophila* Genetic Reference Panel and the *D. melanogaster* reference genome reveals a network of epistatic fitness interactions. Mol. Biol. Evol. 32: 3236–3251.

Porter, H. F., and P. F. O'Reilly, 2017 Multivariate simulation framework reveals performance of multi-trait GWAS methods. Sci. Rep. 7: 38837. https://doi.org/10.1038/srep38837

Qi, H. L., M. D. Rand, X. H. Wu, N. Sestan, W. Y. Wang *et al.*, 1999 Processing of the Notch ligand delta by the metalloprotease kuzbanian. Science 283: 91–94. https://doi.org/10.1126/science.283.5398.91

Rebay, I., R. J. Fleming, R. G. Fehon, L. Cherbas, P. Cherbas *et al.*, 1991 Specific EGF repeats of Notch mediate interactions with delta and serrate: implications for notch as a multifunctional receptor. Cell 67: 687–699. https://doi.org/10.1016/0092-8674(91)90064-6

Reed, L. K., S. Williams, M. Springston, J. Brown, K. Freeman *et al.*, 2010 Genotype-by-diet interactions drive metabolic phenotype variation in *Drosophila melanogaster*. Genetics 185: 1009–1019. https://doi.org/10.1534/genetics.109.113571

Rohlf, F. J., and D. Slice, 1990 Extensions of the Procrustes method for the optimal superimposition of landmarks. Syst. Zool. 39: 40–59. https://doi.org/10.2307/2992207

Roman, G., K. Endo, L. Zong, and R. L. Davis, 2001 P[Switch], a system for spatial and temporal control of gene expression in *Drosophila* melanogaster. Proc. Natl. Acad. Sci. USA 98: 12602–12607. https://doi.org/10.1073/pnas.221303998

Schaid, D. J., X. Tong, B. Larrabee, R. B. Kennedy, G. A. Poland *et al.*, 2016 Statistical methods for testing genetic pleiotropy. Genetics 204: 483–497. https://doi.org/10.1534/genetics.116.189308

Scutari, M., P. Howell, D. J. Balding, and I. Mackay, 2014 Multiple quantitative trait analysis using Bayesian networks. Genetics 198: 129–137. https://doi.org/10.1534/genetics.114.165704

Shriner, D., 2012 Moving toward system genetics through multiple trait analysis in genome-wide association studies. Front. Genet. 3: 1–7. https://doi.org/10.3389/fgene.2012.00001

Shyamala, B. V., and A. Chopra, 1999 *Drosophila melanogaster* chemosensory and muscle development: identification and properties of a novel allele of scalloped and of a new locus, SG18.1, in a Gal4 enhancer trap screen. J. Genet. 78: 87–97. https://doi.org/10.1007/BF02924560

Srinivasan, S., A. P. Mahowald, and M. T. Fuller, 2012 The receptor tyrosine phosphatase Lar regulates adhesion between *Drosophila* male germline stem cells and the niche. Development 139: 1381–1390. https://doi.org/10.1242/dev.070052

Stephens, M., 2013 A unified framework for association analysis with multiple related phenotypes. PLoS One 8: e65245. https://doi.org/10.1371/journal.pone.0065245

Storey, J. D., and R. Tibshirani, 2003 Statistical significance for genomewide studies. Proc. Natl. Acad. Sci. USA 100: 9440–9445. https://doi.org/10.1073/pnas.1530509100

Strimmer, K., 2008 fdrtool: a versatile R package for estimating local and tail area-based false discovery rates. Bioinformatics 24: 1461–1462. https://doi.org/10.1093/bioinformatics/btn209

Takei, Y., Y. Ozawa, M. Sato, A. Watanabe, and T. Tabata, 2004 Three *Drosophila* EXT genes shape morphogen gradients through synthesis of heparan sulfate proteoglycans. Development 131: 73–82. https://doi.org/10.1242/dev.00913

Teslovich, T. M., K. Musunuru, A. V. Smith, A. C. Edmondson, I. M. Stylianou *et al.*, 2010 Biological, clinical and population relevance of 95 loci for blood lipids. Nature 466: 707–713. https://doi.org/10.1038/nature09270

Tibshirani, R., 1996 Regression shrinkage and selection via the Lasso. J. R. Stat. Soc. B 58: 267–288.

Topp, C. N., A. S. Iyer-Pascuzzi, J. T. Anderson, C.-R. Lee, P. R. Zurek *et al.*, 2013 3D phenotyping and quantitative trait locus mapping identify core regions of the rice genome controlling root architecture. Proc. Natl. Acad. Sci. USA 110: E1695–E1704. https://doi.org/10.1073/pnas.1304354110

Van der Linde, K., 2004–2014 Wings: automated capture of *Drosophila* wing shape. Tallahassee, FL. Available at: https://www.bio.fsu.edu/~dhoule/wings.html

van der Sluis, S., D. Posthuma, and C. V. Dolan, 2013 TATES: efficient multivariate genotype-phenotype analysis for genome-wide association studies. PLoS Genet. 9: e1003235. https://doi.org/10.1371/journal.pgen.1003235

Wainwright, M. J., 2009 Sharp thresholds for high-dimensional and noisy sparsity recovery using $L_1$-Constrained Quadratic Programming (Lasso). IEEE Trans. Inf. Theory 55: 2183–2202. https://doi.org/10.1109/TIT.2009.2016018

Wang, J., S. Vasaikar, Z. Shi, M. Greer, and B. Zhang, 2017 WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. 45: W130–W137. https://doi.org/10.1093/nar/gkx374

Wang, S. H., A. Simcox, and G. Campbell, 2000 Dual role for *Drosophila* epidermal growth factor receptor signaling in early wing disc development. Genes Dev. 14: 2271–2276. https://doi.org/10.1101/gad.827000

Wartlick, O., P. Mumcu, F. Juelicher, and M. Gonzalez-Gaitan, 2011 Understanding morphogenetic growth control - lessons from flies. Nat. Rev. Mol. Cell Biol. 12: 594–604. https://doi.org/10.1038/nrm3169

Weber, K., R. Eisman, L. Morey, A. Patty, J. Sparks *et al.*, 1999 An analysis of polygenes affecting wing shape on chromosome 3 in *Drosophila melanogaster*. Genetics 153: 773–786.

Weber, K., R. Eisman, S. Higgins, L. Morey, A. Patty *et al.*, 2001 An analysis of polygenes affecting wing shape on chromosome 2 in *Drosophila melanogaster*. Genetics 159: 1045–1057.

Wood, A. R., T. Esko, J. Yang, S. Vedantam, T. H. Pers *et al.*, 2014 Defining the role of common variation in the genomic and biological architecture of adult human height. Nat. Genet. 46: 1173–1186. https://doi.org/10.1038/ng.3097

Xu, S., 2003 Theoretical basis of the Beavis effect. Genetics 165: 2259–2268.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders *et al.*, 2010 Common SNPs explain a large proportion of the heritability for human height. Nat. Genet. 42: 565–569. https://doi.org/10.1038/ng.608

Yuan, Z., H. Praxenthaler, N. Tabaja, R. Torella, A. Preiss *et al.*, 2016 Structure and function of the Su(H)-Hairless repressor complex, the major antagonist of notch signaling in *Drosophila melanogaster*. PLoS Biol. 14: e1002509. https://doi.org/10.1371/journal.pbio.1002509

Zelditch, M. L., D. L. Swiderski, H. D. Sheets, and W. L. Fink, 2004 *Geometric Morphometrics for Biologists: A Primer*. Elsevier, Amsterdam.

Zhao, P., and B. Yu, 2006 On model selection consistency of Lasso. J. Mach. Learn. Res. 7: 2541–2563.

Zhou, X., and M. Stephens, 2014 Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nat. Methods 11: 407–409. https://doi.org/10.1038/nmeth.2848

Zimek, A., E. Schubert, and H.-P. Kriegel, 2012 A survey on unsupervised outlier detection in high-dimensional numerical data. Stat. Anal. Data Min. 5: 363–387. https://doi.org/10.1002/sam.11161

Zimmerman, E., A. Palsson, and G. Gibson, 2000 Quantitative trait loci affecting components of wing shape in *Drosophila melanogaster*. Genetics 155: 671–683.

Zou, H., and T. Hastie, 2005 Regularization and variable selection via the elastic net. J. R. Stat. Soc. Series B Stat. Methodol. 67: 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x

*Communicating editor: S. Chenoweth*