Genome-wide analysis reveals novel regulators of growth in *Drosophila melanogaster*

Sibylle Chantal Vonesch[1], David Lamparter[2], Trudy FC Mackay[3], Sven Bergmann[2], Ernst Hafen[1]

[1] Institute of Molecular Systems Biology, ETH Zürich, CH-8093 Zürich

[2] Department of Medical Genetics, University of Lausanne, CH-1005 Lausanne

[3] Department of Biological Sciences, Program in Genetics and W. M. Keck Center for Behavioral Biology, North Carolina State University, Raleigh, NC 27695-7614

ABSTRACT

Organismal size depends on the interplay between genetic and environmental factors. Genome-wide association (GWA) analyses in humans have implied many genes in the control of height but suffer from the inability of controlling the environment. Genetic analyses in *Drosophila* have identified conserved signaling pathways controlling size; however, how these pathways control phenotypic diversity is unclear. We performed GWA of size traits using the *Drosophila* Genetic Reference Panel of inbred, sequenced lines and find that top variants are predominantly sex-specific; do not map to canonical growth pathway genes, but can be linked to these by epistasis analysis; and are enriched in homologs of genes involved in human height regulation. Performing GWA on well-studied developmental traits under controlled conditions expands our understanding of developmental processes underlying phenotypic diversity.

INTRODUCTION

How animals control and coordinate growth among tissues is a fundamental question in developmental biology. A detailed mechanistic but global understanding of the processes taking place during normal physiological development is furthermore relevant for understanding pathological growth in cancers. Classical genetic studies in *Drosophila* have revealed core molecular mechanisms governing growth control and have shed light on the role of humoral factors and the environment on adult size (1,2,3,4). Two major pathways regulate size, the Insulin/TOR pathway, which couples systemic growth to nutrient availability; and the Hippo tumor suppressor pathway, which controls cell survival and proliferation in developing organs (5,6,7). However, growth control is complex (8,9,10), and the interactions between components of these pathways with each other and with unknown molecules and extrinsic factors remain poorly understood. Studies focusing on single or few genes can only capture individual aspects of the entire system of networks underlying this trait, which is especially problematic when individual alleles have subtle and context-dependent effects (11,12). What is needed for a better understanding of the genetic control of size is therefore a global genome-wide approach. One possibility here is to study multifactorial natural genetic perturbations as they occur in a segregating population.

GWA studies (GWAS) are a popular method for linking variation in quantitative traits to underlying genetic loci (13,14). GWAS have been pioneered (15,16) and widely applied in humans and are now a routinely used tool in model organisms such as *Arabidopsis* (17,18), *Drosophila* (19,20,21) and mouse (22) as well as in various crop (23,24) and domestic animal species (25,26,27,28,29), where they have substantially broadened our understanding of the genetics of complex traits. GWAS of height have revealed that many loci with small effect contribute to size variation in humans (9,10,30,31), which contrasts with a much simpler genetic architecture of size in domestic animals, where as a consequence of breeding few loci have relatively large effect sizes that jointly explain a large proportion of size variation (26,32). Although many loci affecting human height have been identified by GWAS, deducing the underlying molecular mechanisms by which they affect size is challenging. Larger genome regions and not single genes are mapped; uncontrolled environmental variability makes it difficult to identify causal links between genotype and phenotype; and functional validation cannot be performed in humans (12,16,33,34,35).

In contrast to human studies, GWAS in model organisms benefit from the feasibility of functional validation, more stringent environmental control and, when using inbred strains, the possibility of measuring several genetically identical individuals to obtain an accurate

estimate of the phenotype for a given trait and genotype. All three factors can substantially improve the power of a GWAS. However, as slight environmental fluctuations occur even under controlled laboratory conditions and genotype-environment interactions often account for a large part of phenotypic variance (36), identifying potential confounders and addressing them, by elimination, randomization or accounting for the source of variation is crucial.

The establishment of the inbred, sequenced lines of the *Drosophila* Genetic Reference Panel (DGRP) (37,38) has made GWAS in *Drosophila* widely applicable. The DGRP lines harbor the substantial natural genetic variation present in the original wild population and show copious phenotypic variation for all traits assayed to date (19,20,21,37,39,40).

Here, we used the DGRP to perform single- and two-locus associations for size-related developmental traits in *Drosophila.* We find pervasive sex-specificity of top variants, validate a substantial number of novel growth regulators, link candidates to existing networks and show candidates are enriched for genes implied in variation for human height.

RESULTS

Quantitative Genetic analysis of size

We cultured 143 DGRP lines under conditions we had previously shown to reduce environmental influences on size (Supplementary Table 1, Fig. 1a) and measured five body and 21 wing traits. We chose two traits as representative measurements for size (wing and eye disc derived) for further analysis: centroid size (CS), and interocular distance (IOD), respectively (Fig. 1b). We observed extensive phenotypic and genetic variation in both traits (Fig. 2a, b, Supplementary Table 1, 2), with substantial broad-sense heritabilities (H^2_{CS}=0.63, H^2_{IOD}=0.69). Both traits showed significant genetic variation in sex dimorphism but similar heritability estimates for males and females and high cross-sex genetic (r_{MF}) and phenotypic correlation (Supplementary Table 1). 15% of phenotypic variance in centroid size could be attributed to raising flies on different foodbatches, which only differed by the day on which they were prepared (according to the same protocol) (Supplementary Table 1). Though nutrition is a well-studied size-determining factor (41), we were surprised that even such small nutritional variation could have substantial phenotypic effects. Although the environmental effect of foodbatch was markedly lower for IOD (3%), we used batch-mean corrected phenotypes in all subsequent analyses to remove this effect.

Single-marker and gene based GWAS identify novel loci associated with size variation

To identify common loci contributing to size variation in *Drosophila*, we performed single marker GWAS (42) for 1,319,937 SNPs for a wing disc derived (CS) and an eye disc derived (IOD) size measure. As the genetic correlation, the extent of genetic influence on two traits that is common to both, between CS and IOD was low (0.46 and 0.51 for females and males, respectively), we expected to map SNPs for the two traits separately with some variants common to both. To find loci that specifically affect variation in wing size unrelated to the overall organismal size variation we constructed an additional phenotype (rCS) that had the effect of IOD on CS removed via regression. Besides the foodbatch, two cosmopolitan inversions, *In(2L)t* and *In(3R)Mo*, were correlated with both CS and IOD, which we addressed by modeling their presence in the homozygous state, yielding the inversion corrected phenotypes CS_{IC} and IOD_{IC}.

Only for one trait (IOD in females) we observed significantly associated SNPs when applying a stringent Bonferroni corrected *p*-value threshold of $3.8x10^{-08}$. The six significant SNPs were all located in a cluster on chromosome *2L*, 12-13kb upstream of the gene encoding the EGFR pathway regulator *kek1* (Fig. 3a,b). As QQ-plots showed a departure from uniformity for *p*-values below 10^{-05} (Supplementary Figs. 1,2), we picked candidate loci using this

nominal significance threshold for further examination and functional validation. This yielded between 31 and 51 SNPs for females and between 17 and 36 SNPs for males, with little overlap in top associations and moderate correlation of overall SNP ranks between sexes (Supplementary Table 3,4; Supplementary Figs. 3-5).

Correcting for inversion presence generally enhanced GWAS power, as was evident by more loci reaching nominal significance. Nevertheless, the majority (65-86%) of SNPs from the GWAS with uncorrected trait values remained candidates in the GWAS with corrected phenotypes. As could be expected based on the relatively low genetic correlation between CS and IOD, no candidate SNPs were shared between these phenotypes (Fig. 3c, Supplementary Table 4). In both sexes about one third of top SNPs was shared between the absolute and relative CS GWAS, suggesting variation in relative versus absolute organ size may be achieved through genetic variation at both shared and private loci.

Nominally associated variants predominantly mapped to intergenic regions, but were nevertheless enriched in gene regions (p<0.001, hypergeometric test) (Fig. 3d, Supplementary Table 5), demonstrating that associations were not randomly distributed across the genome. For gene-level analyses we determined candidates for each phenotype as genes having a nominally significant SNP in or within 1kb of their transcribed region, yielding a total of 107 genes over all phenotypes. Only the candidate gene sets for rCS were enriched for STRING curated interactions (43) and none showed enrichment of specific functional groups (43,44,45), though growth was among the top categories for CSM_{IC} (FDR corrected p=0.08). Surprisingly, we found that only few canonical growth genes contained or were close to nominally associated SNPs. Exceptions included several SNPs near or in the genes coding for Ilp8, TOR and EGFR pathway components and regulators of tissue polarity and patterning. However, some SNPs that narrowly missed the candidate reporting threshold localized to further growth regulatory genes, such as the Hippo pathway components *ex* and *wts*.

The small number of detected canonical growth pathway genes might be explained by the lack of SNPs with large effects in these genes, which is plausible considering the essential role of many growth regulators. We therefore wanted to test whether the combined signal of SNPs with small effects (each too small to reach significance on its own) across known growth genes might be significant. To this end we determined gene-based statistics using the sum of chi-squares VEGAS method (46), which computes a p-value for each gene considering all SNPs within a gene while correcting for gene length and linkage disequilibrium between SNPs. None of the genes reached genome-wide significance ($p < 3.75\text{x}10^{-06}$) (Supplementary Table 6). The overlap between the 20 top-scoring genes from this analysis with our GWAS candidate genes was small for each individual phenotype and

even when combining the VEGAS analyses from all phenotypes only 11 of our 97 VEGAS top-scoring genes contained a SNP that reached significance on its own in one of our GWAS. We did not find GO or interaction enrichment (43,44,45) and as in the individual GWAS, top candidates were largely novel for a putative role in growth control.

Functional validation of candidate genes reveals novel regulators of size

We selected a subset (41% to 69%) of candidates identified by each of our wing size GWAS for functional validation by tissue-specific RNAi. A total of 64% to 79% of tested genes had significant effects on wing area (p<0.001, Wilcoxon rank sum test, Supplementary Table 7, Fig. 4a, Supplementary Fig. 6). We achieved similar validation rates for gene-based candidates. In contrast, only 42% of a set of 24 randomly selected genes had significant effects on wing size (Supplementary Table 7). The overall proportion of validated candidates versus random genes was significantly different (p=0.02, Fisher's exact test) and Wilcoxon test p-values showed different distributions between candidate and random knockdowns (p=0.02, Wilcoxon test, Supplementary Fig. 7). This combined evidence suggests an advantage in power for identifying growth regulators by GWAS over randomly testing genes. The validated candidates constitute 33 functionally diverse novel growth regulators (Supplementary Fig. 8).

Two-locus association reveals novel interactions

To place novel genes within the network of known growth pathways, we next performed tests for two-locus associations (47) to CS_{IC}, IOD_{IC} and rCS in both sexes with SNPs in 306 growth genes as focal SNPs (Supplementary Table 8). Overall, 15 interactions reached Bonferroni-corrected significance (p<7.9×10^{-13}), but we observed none of our GWAS candidates among the significant epistasis partners. Generally, more interactions reached genome wide significance in males than in females. The most significant interaction (CSM_{IC}, p=5.79×10^{-15}) occurred between *mask*, a positive regulator of JAK/STAT signaling (48) and *tutl*, a JAK/STAT target gene during optic lobe development (49)(Fig. 5). Furthermore, among the top five interactions we found one between *nkd,* a downstream target of Dpp (50), and the tyrosine phosphatase *Ptp99A* (CSF_{IC}, p=8.79×10^{-14}), which has been shown to interact with InR and the Ras signaling pathway (51,52). Due to their already known growth-related functions we consider the interactions between these genes as prime candidates for functional validation.

As Bonferroni correction is very conservative, we analyzed interactions passing a nominal significance threshold of p<10^{-09} for the presence of our GWAS candidates or previously known genes, counting only those interactions where the interacting SNP lay in or within 1kb

of a gene. All interactor gene lists (Supplementary Table 8) were enriched for development, morphogenesis and signaling categories (Bonferroni corrected $p<0.001$)(43). Notably, the rCSM list (Fig. 5) was additionally enriched for genes involved in regulation of metabolic processes. The 1353 epistasis partners (collected over all three phenotypes and both sexes) included 73 of the 306 known growth genes and 35 of our 107 overall GWAS candidates. However, these overlaps did not reach significance (*p*-value of 0.46 and 0.22, respectively). We next asked whether the candidate gene sets identified by normal GWAS and the epistasis approach were nevertheless biologically related to each other. To this end we used the STRING database (43), which revealed that the number of observed curated interactions between the two gene sets was much larger than expected by chance ($p<<0.001$). Analyzing pairwise interactions may thus help to place genes into pre-established networks.

Intergenic SNPs are preferentially located in regions with enhancer signatures and overlap lincRNA loci

Intergenic SNPs may be functional by changing the sequence of more distant regulatory elements or noncoding RNAs. We therefore tested whether intergenic GWAS candidate SNPs located to putative functional regions. We found enrichment ($p<0.01$, hypergeometric test) of SNPs lying in regions with H3K4Me1 or H3K27Ac, epigenetic signatures of active enhancers (Supplementary Table 5) (53), and in lincRNA loci (54), which have been implied in developmental regulation and are often enriched for trait-associated loci (55). Though only loci associated with IOD in females were enriched for SNPs localizing to lincRNA loci, we found one SNP lying in a lincRNA among the top variants for rCSF and $IODF_{IC}$ (Supplementary Table 5).

A SNP 2kb upstream (position *2L*: 429144) of the Hippo pathway regulator *ex* narrowly missed the candidate threshold ($p=1.7\times10^{-05}$, CSM). However, its genomic location suggests this variant could affect a novel regulatory region for this gene. The region surrounding it was annotated with the enhancer methylation signatures H3K4Me1 and H3K27Ac and had assigned state 4 of the 9 state chromatin model suggestive of a strong enhancer (53,56). Further annotations included H3K9Ac, a mark of transcriptional start sites, histone deacetylase binding sites and an origin of replication. To further assess functionality, we investigated whether the sequence around this SNP was conserved across taxa by performing multiple sequence alignment using BLAST (57) (Supplementary Table 9). Indeed, the region immediately upstream of the *D. melanogaster ex* gene showed high similarity to ~3kb regions slightly more upstream of *expanded* orthologs in the genomes of *D. sechellia, D. yakuba* and *D. erecta* (Supplementary Table 9, Fig. 4b). This combined evidence suggests a putative functional region immediately upstream of the *D. melanogaster ex* gene,

but additional experiments are required to corroborate functionality and to establish a mechanism for influencing size.

Human orthologs of candidate genes are associated with height, obesity and a variety of other traits

To investigate conservation to humans and further elucidate putative functions of candidate genes, we searched for orthologous proteins in humans. We found human orthologs (58) for 62 of our 107 GWA candidate genes, seven of which had a good confidence ortholog (score ≥ 3) associated with height, pubertal anthropometrics or growth defects (Supplementary Table 10). Importantly, the GWAS candidates were enriched for human orthologs associated with height in a recent meta-analysis (p= 0.001) (10). The evidence for an involvement in growth control from GWAS in both organisms and experimental support from validation in *Drosophila* corroborates a biological function of these genes in the determination of body size.

DISCUSSION

We applied several GWAS methods to developmental traits that have been extensively studied by single gene analyses in *Drosophila* as a complementary approach for identifying loci underlying size variation. Our single-marker GWAS revealed only one SNP cluster close to the known growth gene *kek1* to be significantly associated with body size when using a conservative Bonferroni correction. Yet, in contrast to human GWAS, which require independent replication, we exploited the fact that our model organism is amenable to direct validation strategies and tested candidates corresponding to a much lower significance threshold of 10^{-05} for an involvement in size determination. Using tissue-specific RNAi, we validated 33 novel genes affecting *Drosophila* wing size. Nominally significant intergenic associations were preferentially located in regions with an enhancer signature and overlapped lincRNA loci. A SNP upstream of the *expanded* locus was in an evolutionarily conserved region, indicating the presence of a putatively functional element. A two-locus epistasis screen identified several genome-wide significant interactions between known growth genes and novel loci, showing that targeted epistasis analysis can be used to extend existing networks. Finally, we showed that our GWAS top candidate set was enriched for homologs of genes associated with height variation in humans. Our study shows that despite limited statistical power, insights into the genetic basis of trait variation can be gained from analyzing nominal associations through functional and enrichment analyses and performing targeted locus-locus interaction studies.

Single-marker and two-locus GWA of size: Our study adds 33 novel growth genes and 15 genomic loci that may interact with known growth genes to the extensive number of loci already implied in size regulation from single gene studies. That only a few *bona fide* growth genes were among the nominally significant candidates could be due to selection against functional variation in natural populations and/or during subsequent inbreeding and/or that the effect sizes of SNPs in these genes are too small to be detected in the DGRP. Our validation of a substantial number of novel genes underscores the complementarity of the GWAS approach to classical genetics and highlights the importance of probing natural variants. However, future studies would benefit from utilizing bigger population sizes in order to improve statistical power.

While a cluster of SNPs lying 12-13kb upstream of the EGFR pathway regulator Kekkon-1 was among the top associations in all body size GWA, all the most significant wing size associations mapped to putative novel growth genes: *CG6091*, a de-ubiquitinating enzyme whose human ortholog has a role in innate immunity, *CG34370*, which was recently identified in a GWAS for lifespan and lifetime fecundity in *Drosophila* (59) and, surprisingly, *dsx*, a gene well characterized for its involvement in sex determination, fecundity and courtship behavior. As genes affecting growth also impact on general and reproductive fitness of organisms it is not surprising that most of the candidate variants in or close to genes lie in regulatory regions, potentially modulating splicing, RNA turnover or RNA/protein abundance. Our data support the general notion that intergenic SNPs can impact phenotypes, either by affecting transcript abundance of protein coding genes (e.g. through distal enhancer elements) or via noncoding RNAs, which have been shown to regulate many biological functions including cellular processes underlying growth (60,61,62).

Nonsense and missense SNPs are sparse among our candidates (one and six, respectively), and represent prime contenders for effects on protein function. However, confirming such effects requires testing the SNP in an isogenic background. Knockdown of most candidate genes resulted in a small change in median wing size (-19.4% to 10.1%), indicating a redundant or mildly growth enhancing or suppressing role in this tissue, which may explain why they were not discovered by classical mutagenesis screens. However, larger effects might be observed upon ubiquitous knockout, knockdown or overexpression.

Epistasis analysis revealed 15 loci showing Bonferroni-significant interactions with SNPs in previously known growth genes, demonstrating the usefulness of this approach for extending existing biological networks. Furthermore, we found putative biological interactors for several GWAS candidates among the top interactions that did not reach genome-wide significance e.g. *Lar* with *InR*. Lar can phosphorylate InR (63), so polymorphisms at these two loci could act synergistically to modulate InR activity. The enrichment of annotated interactions

between our GWAS candidates and epistasis partners shows that different analyses yielding different top associations uncover common underlying genetic networks. A similar combinatorial approach has been successful in another DGRP study (21), underscoring that combinatorial approaches can help placing candidates from different analyses into a joint biological network, and provide a basis for further hypothesis driven investigation of the roles and connectivity of novel and known genes.

Biological roles of novel growth genes: Apart from expected processes like signaling, transcription, translation and morphogenesis, we validated genes involved in transmembrane transport, planar cell polarity (PCP), metabolism and immunity. A total of 24 GWAS or epistasis candidate genes were discovered to be enhancers or suppressors of major growth pathways in another study (64, Supplementary Table 11**)** and 15 were associated with nutritional variation in *Drosophila* (65), supporting their role in growth control. Importantly, the enrichment of GWAS candidates for genes associated with human height not only supports their role in growth but also shows that results of GWAS studies in *Drosophila* can be translated to orthologous traits and genes in humans.

The yeast ortholog of Mid1, is a stretch activated Ca^{2+} channel with a role in the polarized growth of mating projections (66,67). As mechanical tension plays a role during growth of imaginal discs, this channel could act in translating such signals to intracellular signaling pathways via the second messenger Ca^{2+}. The human ortholog of another validated candidate, the transmembrane channel Trpm, showed association with anthropometric traits during puberty, indicating a role during the postnatal growth phase. The mucin Muc68Ca showed one of the largest knockdown effects. Mucins form a protective layer around vital organs, and the expression pattern of Muc68Ca in the larval midgut concurs with a putative effect on growth via the control of intestinal integrity (68).

A dual role in PCP, the establishment of cell polarity within a plane in an epithelium, and growth control has been shown for many genes, which regulate these two processes via distinct but coordinated downstream cascades (69,70). *Lar, aPKC,* the Fz target Kermit (71) and the motor proteins Dhc64C (nonsense SNP) and Khc-73, whose human ortholog is significantly associated with height, are implied in PCP establishment. Kermit and motor proteins act downstream in the PCP cascade and likely have specialized roles for this process, but PCP can itself impact on growth, as proper establishment of polarity provides the orientation of cell division, and loss of a PCP component in zebrafish causes a reduction in body length (72). Interestingly, *kermit* was a candidate interactor of EGFR, which acts in a combinatorial manner with Fz signaling in PCP (73), providing a biological basis for this interaction.

Metabolic genes are prime candidates for improving our understanding of growth, which depends on the amount of energy and precursors available for biosynthesis, and thus to metabolic coordination. The recent findings that the growth and PCP regulator Fat can couple growth and metabolism and mitochondrial proteins can causally affect growth pathway activity (74) underscore the importance of metabolic coordination. A missense SNP in the validated candidate *Cep89*, a gene involved in mitochondrial metabolism and growth in *Drosophila* and humans (75) was associated with most wing phenotypes. Elucidating the function of *Cep89* and other validated candidates with putative roles in metabolism, e.g. *CG3011*, *CG6084* and *Fbp2,* whose human ortholog has been linked to growth defects and cancer (*e.g.* 76), may provide further insight into this coordination.

Sexual dimorphism of size: Of the top 100 SNPs for each trait only 25% - 43% are shared between the sexes, a surprisingly small overlap given the high genetic and phenotypic correlations between sexes. With respect to genes, only 50% of the top 1000 SNPs in or near annotated genes are shared. Interestingly, an intronic SNP in the sex determination gene *dsx* had the lowest *p*-value in the female relative wing size GWAS but had a smaller effect size in males. Dsx is a transcription factor with sex-specific isoforms, and has many targets with sex- and tissue specific effects (77). These data suggest there may be marked differences in the genetic networks underlying size determination in males and females in natural populations, a possibility that is neglected in single gene studies.

Conclusions & Outlook: Growth control has been well studied, particularly in *Drosophila,* where many genes and pathways affecting growth have been documented by mutational analyses. However, such screens are far from saturation and do not scale well to investigating effects of combinations of mutations. Here we took advantage of naturally occurring, multifactorial perturbations genome-wide to identify novel genes affecting growth and to place them in genetic interaction networks. Rather than deepening our understanding of growth control, the identification of ever more growth regulators raises new questions about how all these loci interact to govern growth. The challenge for the future will be to shift our focus from studying genes in isolation towards investigating them in the context of developmental networks, and to assess the effects of network perturbations on intermediate molecular phenotypes of transcript, protein and metabolite levels.

ONLINE METHODS

Drosophila medium and strains

Fly food was prepared according to the following recipe: 100 g fresh yeast, 55 g cornmeal, 10 g wheat flour, 75 g sugar, 8 g bacto-agar and 1 liter tap water.

Experiments were performed with 149 of the DGRP lines. RNAi lines used are listed in Supplementary Table 7.

Standardized culture conditions

Lines were set up in duplicate vials, with five males and five females per vial. After seven days, the parental flies were removed. From the F_1, five males and five females were put together in duplicate vials and discarded after seven days of egg laying. From the F_2, thirty males and thirty females were mated in a laying cage with an apple juice agar plate plus a yeast drop as food source and allowed to acclimatize for 24 hours. A fresh plate of apple juice agar plus yeast drop was then applied and flies were left to lay eggs for another 24 hours. From this plate, F_3 L1 larvae were picked with forceps and distributed into three replicate vials, with 40 larvae per vial. The food surface in the vials was scratched and 100μl of ddH_2O added prior to larvae transfer. The adult F_3 flies were pooled from the three vials and frozen at -20°C approximately 1-2 days after eclosion. The whole experiment was performed in a dedicated incubator (DR-36VL, CLF Plant Climatics GmbH) with a 12-hour day-night cycle, constant humidity of 65 - 68% and constant temperature of 25.5°C +/- 1°C. Vials were shuffled every two days during the first and second round of mating but left at a fixed position in the incubator for the duration of the development of the F_3 generation.

For the parental generation, lines were all set up on the same day on the same foodbatch. For the F_1 matings, different foodbatches had to be used due to different developmental timing of the lines. F_2 matings were set up using the same batch of apple agar plates and yeast for all lines. F_3 larvae were distributed on four different foodbatches and the batch number was recorded for each line.

The control experiment (Supplementary Table 1) was performed using the same procedure as above, except that the same foodbatch was used for all flies of a generation. We used the DGRP lines 25176 (RAL 303), 25203 (RAL 732), 28220 (RAL 721) and 28263 (RAL 908) for this experiment because they had comparable generation times and set up ten replicates of each of these lines according to the standardized culture conditions.

Phenotyping and morphometric measurements

Depending on the number of flies available, between five and twenty-five flies per sex and line were measured for the dataset (median 25 flies per sex and line, mean 23 ($CS_{females}$,

CS_{males}, IOD_{males}) and 24 for $IOD_{females}$; exact numbers are given in Supplementary Table 2). For the experimental generation we distributed a total of 19200 larvae in four batches spaced throughout 1.5 weeks according to developmental timing, and the final dataset consisted of morphometric data of 6978 flies, 3500 females and 3478 males.

For the control experiment we phenotyped 25 flies per replicate, sex and line, resulting in a total of 2000 flies (1000 males, 1000 females).

Flies were positioned on a black apple agar plate and photographed using a VHX-1000 digital light microscope (KEYENCE). Morphometric body traits were measured manually using the VHX-1000 dedicated measurement software. If intact the right and otherwise the left wing was removed and mounted in water on a glass slide for wing image acquisition. Morphometric measurements were extracted from the wing images using WINGMACHINE (78) and MATLAB (MATLAB version R2010b Natick, Massachusetts: The MathWorks Inc.)

Centroid size was measured as the square root of the summed squared distances of 14 landmarks from the center of the wing (Fig. 1). Interocular distance was measured from eye edge to eye edge along the anterior edge of the posterior ocelli and parallel to the base of the head.

Quantitative genetic analysis

All analyses were performed in R Studio using the R statistical language version 2.15 (http://www.R-project.org).

The total phenotypic variance in the control experiment was partitioned using the mixed model $Y = S + L + SxL + R(L) + \varepsilon$, where S is the fixed effect of sex, L is the random effect of line (genotype), SxL is the random effect of line by sex interaction, R is the random effect of replicate and ε is the within line variance. The brackets represent that replicate is implicitly nested within line. The total phenotypic variance in the dataset was partitioned using the mixed model $Y = S + L(F) + SxL(F) + F + \varepsilon$, where S is the fixed effect of sex, L is the random effect of line (genotype), SxL is the random effect of line by sex interaction, F is the random effect of foodbatch and ε is the within line variance. The random effects of line and line by sex are implicitly nested within foodbatch, as each line was raised only on one of the four foodbatches. Models of this form were fitted using the *lmer()* function in the *lme4* package in R. We also ran reduced models separately for males and females. The *rand()* function in the *lmerTest* package was used to assess significance of the random effects terms in the dataset.

Relative contributions of the variance components to total phenotypic variance (σ^2_P) was calculated as σ^2_i / σ^2_P where σ^2_i represents any of σ^2_L, σ^2_{LxS}, σ^2_F, σ^2_R, σ^2_E, and $\sigma^2_P = \sigma^2_L + \sigma^2_{LxS} + \sigma^2_C + \sigma^2_E$. σ^2_C stands for σ^2_R in the control dataset and for σ^2_F in the analysis of the

GWAS dataset. σ^2_L = variance due to genotype, σ^2_{LxS} = variance due to genotype by sex interactions, σ^2_F = variance due to food, σ^2_R = variance due to replicate and σ^2_E = residual (intra-line) variance. The broad sense heritability for each trait was estimated as $H^2 = \sigma^2_G / \sigma^2_P = (\sigma^2_L + \sigma^2_{LxS}) / (\sigma^2_L + \sigma^2_{LxS} + \sigma^2_C + \sigma^2_E)$. The cross-sex genetic correlation was calculated as $rMF = \sigma^2_L / (\sigma_{LF}\ \sigma_{LM})$ where σ^2_L is the variance among lines from the analysis pooled across sexes, and, σ_{LF} and σ_{LM} are, respectively, the square roots of the among line variance from the reduced models of females and males. Similarly, cross-trait genetic correlations were calculated as $r_{AB} = \sigma^2_{G(AB)}/(\sigma_{GA}\ \sigma_{GB})$ where $\sigma^2_{G(AB)}$ is the genetic covariance between traits A and B, and σ_{GA} and σ_{GB} are the square roots of the genetic variance for traits A and B, respectively. The phenotypic correlation between sexes was determined using the *cor()* function with method = "spearman" in R.

Phenotypes for GWAS

We found a large effect of foodbatch on CS, and inversions *In(2L)t* and *In(3R)Mo* were associated with IOD and to a lesser extent CS. We modeled these covariates using a mixed model. The foodbatch was modeled by a random effect and the rearrangements were coded as (0,1,2) depending on whether both, one or no inversion was present in the homozygous state. We did not observe correlation between Wolbachia infection status and any trait and thus did not include this as a covariate in the model.

Specifically, the models used were: $CS_{raw} = \alpha + X_1\beta_1 + X_2\beta_2 + Fu + \varepsilon$, where X_1 refers to the sex covariate, X_2 refers to the inversion covariate, $\varepsilon \sim N_n(0,\sigma_\varepsilon\ I_n^2)$ with n being the number of lines, $u \sim N_k(0, \sigma_u^2 I_k)$ with k being the number of foodbatches and F an (n,k)-indicator matrix, associating each line to its respective foodbatch. The GWAS analyses were performed using the estimated residual of this model ($CS = \varepsilon$).

To find loci that specifically affected variation in wing size unrelated to the overall body size variation we constructed an additional phenotype (rCS) that had the effect of IOD on CS removed via regression: $CS_{raw} = \alpha + IOD + X_1\beta_1 + Fu + \varepsilon$, where X_1 and Fu refer again to the sex-effect and the foodbatch effect. We did not model the inversions because the residuals of this model were not correlated with the inversions. The residuals ε from this regression were used as relative size phenotypes.

Association Analysis

We performed GWAS using male and female line means. Genotypes for 143 of the 149 lines were obtained from the DGRP Freeze 2 (http://dgrp2.gnets.ncsu.edu). Only SNPs that were missing in a maximum of ten lines and occurred in at least ten lines (7% of the measured

lines, 1,319,937 SNPs in total) were used. GWAS was performed using FaST-LMM (42) for separate sexes. Association results were visualized using the *manhattan()* function in the R package *qqman* (79). To determine correlation between SNPs for a given phenotype we extracted the genotype of the top n SNPs ($p<10^{-05}$) and calculated the correlation between genotypes at these loci across all DGRP lines used in the GWAS. We used the FaST-LMM SNP *p*-values to apply the sum of chi-squares VEGAS method (46) to calculate gene wise statistics. Gene boundaries were defined using annotation from popDrowser (http://popdrowser.uab.cat/gb2/gbrowse/dgrp/), but we included also SNPs lying within 1000 bp up- or downstream of these margins. The correlation matrix was calculated from the genotypes themselves.

GO annotation and interaction enrichment

To determine enrichment of functional classes, annotate genes with functions and curated interactions among our candidate genes, we used the functional annotation chart function from DAVID (44,45) and the functional annotation and protein interaction enrichment tools from STRING (43).

RNAi validation

SNPs with an association *p*-value $<10^{-05}$ lying in a gene region or +/- 1kb from a gene were mapped to that gene. From the gene based VEGAS analysis, we chose the top 20 genes from each list as candidates. RNAi lines for a subset of candidate genes for each wing phenotype were ordered from VDRC. For one gene, *chinmo*, there was no appropriate line available from VDRC and we instead tested two Bloomington lines (26777 and 33638, indicated in Supplementary Table 7 with (BL)). For the random control knockdowns we tested a set of 24 genes that did not contain a significant SNP in or within 1kb of their transcribed region. All lines are listed in Supplementary Table 7**.** For wing size candidates, validation was performed by crossing males of the respective RNAi line to virgin females carrying the *GAL4* transcriptional activator under the control of the *nubbin (nub)* promoter. The VDRC line 47097 containing a *UAS*-RNAi construct against the *CG1315* gene served as a negative control for the knockdowns. We decided to use this line as reference because it was in the same background as most of our tester lines, an essential factor to consider when assessing genes that presumably only have a small effect on size upon knockdown. The *CG1315* knockdown had so far never shown an effect in any setting and it allowed us to evaluate unspecific effects of RNAi knockdown on wing size. Prior to the experiment, driver lines were bred under controlled density to eliminate cross-generational effects of crowding on size. Wings were phenotyped as described above and wing area used as a phenotypic readout.

Change in median wing area relative to the control was tested with a Wilcoxon rank sum test (function *wilcoxon.test()* in R) for each line and for separate sexes. If possible, 25 flies per cross and sex were phenotyped for statistical analysis, however sometimes the number of progeny was lower. The number of phenotyped flies per cross and sex is given in Supplementary Table 7. We used the *fisher.test()* function in R to determine if the proportion of validated genes was different among candidates and random lines, and the *wilcoxon.test()* function to test for a difference in median *p*-value between candidates and random lines. Both these analyses were done in females exclusively as only this sex was measured for the random lines. Only genes not previously implied in wing development or growth control were included in the analysis, which excluded *chinmo*, *aPKC*, *tws* and *Ilp8* from the candidates and *EloA* and *spz5* from the random list. Furthermore, we only used one line (the more significant one) per gene where more than one line was tested for a given candidate gene.

Epistatic analysis

We explored epistatic interactions between SNPs lying within and 1kb around genes that were previously found to be involved in growth or wing development in *Drosophila* against all DGRP SNPs with missingness <11 and present in at least 10% of the lines. We compiled a list of SNPs within and 1kb up- or downstream of genes that were previously known to play a role in growth control (14,137 SNPs) or wing development (43,498 SNPs) and used these as focal SNPs (Supplementary Table 8). All phenotypes were normalized to follow a standard normal distribution. We used FasT-Epistasis (47) calculating interactions for all pairs between the focal SNPs and the set of all SNPs satisfying the above criteria (1,100,811 SNPs). Bonferroni corrected significance would thus require $p<7.9\times10^{-13}$ (0.05 divided by total number of tests ((14,137+43,498)*1,100,811 = 63,445,241,985). Interactions were visualized using Circos (80). To calculate significance for the overlap between genes found via epistasis and a given gene list, we first positionally indexed all n SNPs that were used in the epistasis analysis. We recorded the set of indices of SNPs with $p<10^{-09}$) yielding set *K*: *K* = {k: SNP_k is an epistasis hit}. We then generated random samples.

For random sample *j*, do:

For all elements in *K*, add a random integer r_j between *0* and *n-1*. Define new index as the modulo *n*: $k_i^j = mod(k_i + r_j, n)$, which yields $K^j = \{k_i^j; j = 1, .., m\}$. Given the shifted positions K^j, we look up the SNP positions P_{Kj}. For a given gene list, we record the number x_j of gene regions that overlap a position in P_{Kj}. Let *x* be number of gene regions overlapping an epistasis hit. Our p-value estimate is then $P_{approx} \approx 1/m \sum 1_{\{xj \geq x\}}$.

Intergenic element enrichment analysis

We determined the number of SNPs from each GWA candidate list and the overall number of SNPs that located within modENCODE elements annotated with Histone 3 lysine 4 monomethylation (H3K4Me1) or Histone 3 lysine 27 acetylation (H3K27Ac) or lincRNA loci. For the H3K4Me1/H3K27Ac enrichments we restricted ourselves to three developmental stages (L2, L3, pupae), which we considered to be the most relevant interval for gene activity affecting growth of imaginal discs. We obtained a table with lincRNAs in the *Drosophila* genome from the study of Young *et al.* (54) and searched for enrichment of SNPs located in those lincRNA loci. Enrichment was tested using a hypergeometric test (function *phyper()*) in R).

BLAST alignment

We downloaded the sequence of the region 10 kb upstream of the annotated transcription start site of the *expanded* locus (*2L*: 421227..431227) from FlyBase (81), as well as the sequence of the same relative region for seven of the twelve *Drosophila* species (82), which contained the ortholog of the *expanded* gene in the same orientation in the genome. We performed multiple sequence alignment using the discontiguous megablast option on NCBI BLAST (57).

Annotation with human orthologs

We combined candidate genes from GWAS for all phenotypes and searched for orthologs in humans using DIOPT-DIST (58). Enrichment of GWAS candidates for genes with human orthologs associated with height (10) was determined with a hypergeometric test (function *phyper()* in R). We determined *Drosophila* orthologs of gene annotations of all associated SNPs (total 697), resulting in 374 ortholog pairs supported by at least 3 prediction tools, and searched for overlap of these orthologs with the 62 of our GWAS candidate genes that had a human ortholog supported by at least 3 prediction tools, which resulted in 12 matches. Of those, only five matches were supported by three or more prediction tools (score >=3) and we used only those for enrichment calculation. Population size was determined as the total number of *Drosophila*-Human ortholog relationships (= 28,605) (58).

REFERENCES

1. Oldham, S., Bohni, R., Stocker, H., Brogiolo, W. & Hafen, E. Genetic control of size in Drosophila. *Phil. Trans. R. Soc. B: Biological Sciences* **355,** 945-952 (2000).

2. Johnston, L.A. & Gallant, P. Control of growth and organ size in Drosophila. *Bioessays* **24**, 54-64 (2002).

3. Mirth, C.K. & Riddiford, L.M. Size assessment and growth control: how adult size is determined in insects. *Bioessays* **29,** 344-355 (2007).

4. Shingleton, A.W. The regulation of organ size in Drosophila: physiology, plasticity, patterning and physical force. *Organogenesis* **6,** 76-87 (2010).

5. Oldham, S. & Hafen, E. Insulin/IGF and target of rapamycin signaling: a TOR de force in growth control. *Trends Cell Biol.* **13**, 79-85 (2003).

6. Pan, D. Hippo signaling in organ size control. *Genes Dev.* **21,** 886-897 (2007).

7. Tumaneng, K., Russell, R.C. & Guan, K.-L. Organ size control by Hippo and TOR pathways. *Curr. Biol.* **22,** R368-R379 (2012).

8. Gockel, J., Robinson S.J.W., Kennington, W.J., Goldstein, D.B. & Partridge, L. Quantitative genetic analysis of natural variation in body size in *Drosophila melanogaster*. *Heredity* **89,** 145-153 (2002).

9. Lango Allen, H. *et al.* Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature* **467,** 832–838 (2010).

10. Wood, A. R. *et al.* Defining the role of common variation in the genomic and biological architecture of adult human height. *Nat. Genet.* **46**, 1173-1186 (2014). doi:10.1038/ng.3097

11. Falconer, D. S., & Mackay, T. F. C. *Introduction to Quantitative Genetics (Edition 4)*. Longmans Green, Harlow, Essex, UK (1996).

12. Lynch, M. & Walsh, B. *Genetics and Analysis of Quantitative Traits.* Sinauer Associates, Inc., Sunderland, MA, USA (1998).

13. Womack, J. E., Jang, H.-J. & Lee, M. O. Genomics of complex traits. *Ann. N. Y. Acad. Sci.* **1271,** 33–36 (2012).

14. Korte, A. & Farlow, A. The advantages and limitations of trait analysis with GWAS- a review. *Plant Methods* **9,** p. 29 (2013).

15. Hirschhorn, J. N. & Daly, M. J. Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6,** 95–108 (2005).

16. McCarthy, M. I. *et al.* Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat. Rev. Genet.* **9,** 356–369 (2008).

17. Bergelson, J. & Roux, F. Towards identifying genes underlying ecologically relevant traits in Arabidopsis thaliana. *Nat. Rev. Genet.* **11,** 867–879 (2010).

18. Meijón, M., Satbhai, S.B., Tsuchimatsu, T. & Busch, W. Genome-wide association study using cellular traits identifies a new regulator of root development in Arabidopsis. *Nat. Genet.* **46,** 77-81 (2013).

19. Jumbo-Lucioni, P. *et al.* Systems genetics analysis of body weight and energy metabolism traits in *Drosophila melanogaster*. *BMC Genomics* **11,** 297 (2010).

20. Jumbo-Lucioni, P. *et al.* Nuclear genomic control of naturally occurring variation in mitochondrial function in *Drosophila melanogaster*. *BMC Genomics* **13**, p. 659 (2012).

21. Swarup, S., Huang, W., Mackay, T. F. C. & Anholt, R. R. H. Analysis of natural variation reveals neurogenetic networks for Drosophila olfactory behavior. *Proc. Natl. Acad. Sci. USA* **110**, 1017-1022 (2013).

22. Flint, J. & Eskin, E. Genome-wide association studies in mice. *Nat. Rev. Genet.* **13,** 807–817 (2012).

23. Huang, X. *et al.* Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nat. Genet.* **44,** 32-39 (2011).

24. Lipka, A. E. *et al.* Genome-wide association study and pathway-level analysis of tocochromanol levels in maize grain. *G3 (Bethesda)* **3,** 1287–1299 (2013).

25. García-Gámez, E. *et al.* GWA Analysis for Milk Production Traits in Dairy Sheep and Genetic Support for a QTN Influencing Milk Protein Percentage in the LALBA Gene. *PLoS ONE* **7,** e47782 (2012).

26. Makvandi-Nejad, S. *et al*. Four Loci Explain 83% of Size Variation in the Horse. *PLoS ONE* **7,** e39929 (2012).

27. Maxa, J., Neuditschko, M., Russ, I., Förster, M. & Medugorac, I. Genome-wide association mapping of milk production traits in Braunvieh cattle. *Journal of Dairy Science* **95,** 5357–5364 (2012).

28. Hwan Lee, S. *et al.* Genome-Wide Association Study Identifies Major Loci for Carcass Weight on BTA14 in Hanwoo (Korean Cattle). *PLoS ONE* **8,** e74677 (2013).

29. Minozzi, G. *et al.* Genome Wide Analysis of Fertility and Production Traits in Italian Holstein Cattle. *PLoS ONE* **8,** e80219 (2013).

30. Yang, J. *et al.* Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42,** 565–569 (2010).

31. Yang, J. *et al.* Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43,** 519–525 (2011).

32. Sutter, NB. *et al.* A Single IGF1 Allele Is a Major Determinant of Small Size in Dogs. *Science* **316,** 112-115 (2007).

33. Oksenberg, J. R., Baranzini, S. E., Sawcer, S. & Hauser, S. L. The genetics of multiple sclerosis: SNPs to pathways to pathogenesis. *Nat. Rev. Genet.* **9,** 516–526 (2008).

34. Thomas, D. Gene–environment-wide association studies: emerging approaches. *Nat. Rev. Genet.* **11,** 259–272 (2010).

35. Vilhjálmsson, B. J. & Nordborg, M. The nature of confounding in genome-wide association studies. *Nat. Rev. Genet.* **14,** 1–2 (2013).

36. Valdar, W. *et al.* Genetic and Environmental Effects on Complex Traits in Mice. *Genetics* **174,** 959-984 (2006).

37. Mackay, T. F. C. *et al.* The *Drosophila melanogaster* Genetic Reference Panel. *Nature* **482,** 173–178 (2012).

38. Huang, W. *et al.* Natural variation in genome architecture among 205 *Drosophila melanogaster* Genetic Reference Panel lines. *Genome Res.***24,** 1193–1208 (2014).

39. Ayroles, J.F. *et al.* Systems genetics of complex traits in *Drosophila melanogaster*. *Nat. Genet.* **41,** 299-307 (2009).

40. Massouras, A. *et al.* Genomic Variation and Its Impact on Gene Expression in *Drosophila melanogaster*. *PLoS Genet.* **8**, p. e1003055 (2012).

41. Nijhout, H. F. et al. The developmental control of size in insects. *WIRES Dev. Biol.* **3**, 113-134 (2014).

42. Lippert, C. *et al.* FaST linear mixed models for genome-wide association studies. *Nat. Methods* **8,** 833–835 (2011).

43. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41,** D808–D815 (2012).

44. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4,** 44–57 (2009).

45. Huang, D. W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37,** 1–13 (2009).

46. Liu, J. Z. *et al.* A Versatile Gene-Based Test for Genome-wide Association Studies. *Am. J. Hum. Genet.* **87,** 139–145 (2010).

47. Schüpbach, T., Xenarios, I., Bergmann, S. & Kapur, K. FastEpistasis: a high performance computing solution for quantitative trait epistasis. *Bioinformatics* **26,** 1468–1469 (2010).

48. Muller, P., Kuttenkeuler, D., Gesellchen, V., Zeidler, M. P. & Boutros, M. Identification of JAK/STAT signaling components by genome-wide RNA interference. *Nature* **436**, 871-875 (2005).

49. Wang, H., Chen, X., He, T., Zhou, Y. & Luo, H. Evidence for tissue-specific Jak/STAT target genes in *Drosophila* optic lobe development. *Genetics* **195**, 1291-1306 (2013).

50. Yang, L., Meng, F., Ma, D., Xie, W. & Fang, M. Bridging Decapentaplegic and Wingless signaling in *Drosophila* wings through repression of naked cuticle by Brinker. *Development* **140,** 413–422 (2012).

51. Madan, L. L. *et al.* Modulation of Catalytic Activity in Multi-Domain Protein Tyrosine Phosphatases. *PLoS ONE* **6,** e24766 (2011).

52. Carter, G. W. Inferring gene function and network organization in *Drosophila* signaling by combined analysis of pleiotropy and epistasis. *G3* **3**, 807-814 (2013).

53. Celniker, S. E. *et al.* Unlocking the secrets of the genome. *Nature* **459,** 927–930 (2009).

54. Young, R. S. *et al.* Identification and Properties of 1,119 Candidate LincRNA Loci in the *Drosophila melanogaster* Genome. *Genome Biol. Evol.* **4,** 427–442 (2012).

55. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genetics* **9,** e1003569 (2013).

56. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* **473,** 43–49 (2011).

57. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.***215,** 403–410 (1990).

58. Hu, Y. *et al.* An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12,** 357 (2011).

59. Durham, M. F., Magwire, M. M., Stone, E. A. & Leips, J. Genome-wide analysis in *Drosophila* reveals age-specific effects of SNPs on fitness traits. *Nat. Commun.* **5**, p. 4338 (2014).

60. Carrington, J. C. & Ambros, V. Role of microRNAs in plant and animal development. *Science* **301**, 336-338 (2003).

61. Inui, M., Martello, G. & Piccolo, S. MicroRNA control of signal transduction. *Nat. Rev. Mol. Cell Biol.* **11**, 252-263 (2010).

62. Fatica, A. & Bozzoni, I. Long non-coding RNAs: new players in cell differentiation and development. *Nat. Rev. Genet.* **15**, 7-21 (2014).

63. Madan, L. L. *et al.* Modulation of Catalytic Activity in Multi-Domain Protein Tyrosine Phosphatases. *PLoS ONE* **6,** e24766 (2011).

64. Schertel, C. *et al.* Systematic Screening of a *Drosophila* ORF Library *in vivo* Uncovers Wnt/Wg Pathway Components. *Dev. Cell* **25,** 207–219 (2013).

65. Unckless, R. L., Rottschaefer, S. M. & Lazzaro, B. P. A Genome-Wide Association Study for Nutritional Indices in *Drosophila*. *G3*, doi: 10.1534/g3.114.016477 (2015).

66. Iida, H., Nakamura, H., Ono, T., Okumura, M. S. & Anraku, Y. MID1, a novel *Saccharomyces cerevisiae* gene encoding a plasma membrane protein, is required for Ca2+ influx and mating. *Mol Cell Biol* **14,** 8259–8271 (1994).

67. Levin, D. E. & Errede, B. The proliferation of MAP kinase signaling pathways in yeast. *Curr. Opin. Cell Biol.* **7,** 197–202 (1995).

68. Syed, Z. A., Härd, T., Uf, A. & van Dijk-Härd, I. F. A potential role for *Drosophila* mucins in development and physiology. *PLoS ONE* **3**, e3041 (2008).

69. Povelones, M. Genetic Evidence That *Drosophila* frizzled Controls Planar Cell Polarity and Armadillo Signaling by a Common Mechanism. *Genetics* **171,** 1643–1654 (2005).

70. Parsons, L. M., Grzeschik, N. A., Allott, M. & Richardson, H. Lgl/aPKC and Crb regulate the Salvador/Warts/Hippo pathway. *Fly* **4,** 288–293 (2010).

71. Lin, C. & Katanaev, V. L. Kermit Interacts with Gαo, Vang, and Motor Proteins in *Drosophila* Planar Cell Polarity. *PLoS ONE* **8,** e76885 (2013).

72. Hatakeyama, J., Wald, J. H., Printsev, I., Ho, H. Y. H. & Carraway, K. L. Vangl1 and Vangl2: planar cell polarity components with a developing role in cancer. *Endocrine Related Cancer* **21,** R345–R356 (2014).

73. Weber, U., Pataki, C., Mihaly, J. & Mlodzik, M. Combinatorial signaling by the Frizzled/PCP and Egfr pathways during planar cell polarity establishment in the *Drosophila* eye. *Dev. Biol.* **316,** 110–123 (2008).

74. Sing, A. *et al.* The Atypical Cadherin Fat Directly Regulates Mitochondrial Function and Metabolic State. *Cell* **158,** 1293–1308 (2014).

75. van Bon, B. W. M. *et al.* CEP89 is required for mitochondrial metabolism and neuronal function in man and fly. *Hum. Mol. Genet.* **22,** 3138–3151 (2013).

76. Pereira, C. *et al.* Genetic variability in key genes in prostaglandin E2 pathway (*COX-2*, *HPGD*, *ABCC4* and *SLCO2A1*) and their involvement in colorectal cancer development. *PLoS ONE* **9**, e92000 (2014).

77. Clough, E. *et al.* Sex- and Tissue-Specific Functions of *Drosophila* Doublesex Transcription Factor Target Genes. *Dev. Cell* **31**, 761-773 (2014).

78. Houle, D., Mezey, J., Galpern, P. & Carter, A. Automated measurement of *Drosophila* wings. *BMC Evol. Biol.* **3**, p.25 (2003).

79. Turner, S. D. *qqman:* an R package for visualizing GWAS results using Q-Q and manhattan plots*. biorXiv* DOI: 10.1101/005165., http://cran.r-project.org/web/packages/qqman/index.html

80. Krzywinski, M. *et al.* Circos: An information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).

81. St Pierre, S. E., Ponting, L., Stefancsik, R., McQuilton, P., FlyBase Consortium. FlyBase 102--advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* **42,** D780–8 (2014).

82. Clark, A. G. *et al.* Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450,** 203–218 (2007).

ACKNOWLEDGEMENTS

We thank Anna Troller, Benjamin Schlager and Anni Strässle for manual support during experiments. This work was funded by grant SXRTX0-123851 from SystemsX.ch, the Swiss National Science Foundation grant 31003AB_135699 and financial support from ETH Zurich to EH.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

AUTHOR CONTRIBUTIONS

EH and TM conceived the approach. SV designed and performed experiments and SV and DL performed statistical analyses. SV wrote the manuscript and SV, DL, EH, SB and TM edited the manuscript.

FIGURE LEGENDS

Figure1. Standardized *Drosophila* culture conditions for the quantification of morphometric traits. a) The protocol extends over three generations and efficiently controls known covariates of size, such as temperature, humidity, day-night-cycle and crowding.

Additionally, effects of other environmental covariates, such as intra-vial environment, light intensity and incubator position, are randomized. b) Illustration of the two phenotypes used for GWAS: interocular distance (IOD) and the 14 landmarks that were used in the calculation of centroid size (CS).

Figure 2. Phenotypic variation in the DGRP for two size traits. Plots show mean phenotypic values for a) centroid size and b) interocular distance. Each dot represents the mean phenotype per line of males (black) and corresponding females (red), with error bars denoting one standard deviation. Lines are ordered on the x-axis according to male trait value, from lowest to highest, meaning that the order of lines is different for each plot. Line means are listed in Supplementary Table 2.

Figure 3. GWAS for wing and body size shows nominally associated loci ($p<10^{-05}$) are distributed across the genome and most abundant in intergenic and regulatory regions. a) Manhattan plot of the SNP *p*-values from the IOD GWAS in females shows that nominally associated SNPs are distributed over all chromosomes. Negative log10 *p*-values are plotted against genomic position, the black line denotes the nominal significance threshold of 10^{-05} and the circle marks the cluster of Bonferroni-significant SNPs upstream of *kek1 on 2L*. b) Correlation between SNPs nominally associated with female IOD. The cluster of Bonferroni-significant SNPs on *2L* shows high correlation among individual SNPs over a larger region, whereas most other SNPs except a few in a narrow region on 3L represent individual associations. Blue = No correlation, orange = complete correlation. Pixels represent individual SNPs and black lines divide chromosomes. c) Overlap in the number of nominally associated SNPs for different wing traits in females. The overlap is bigger between the absolute wing size phenotypes and only a few SNPs are candidates for all traits. d) Nominally associated SNPs are most abundant in the intergenic space and in regulatory regions. Boxes show the distribution of negative log10 *p*-values of the SNPs nominally associated to rCS in females among site classes. Numbers of SNPs belonging to each site class are denoted above the boxes. As a SNP can fall into multiple classes, the sum of SNPs from all site classes is higher than the total number of nominally associated SNPs.

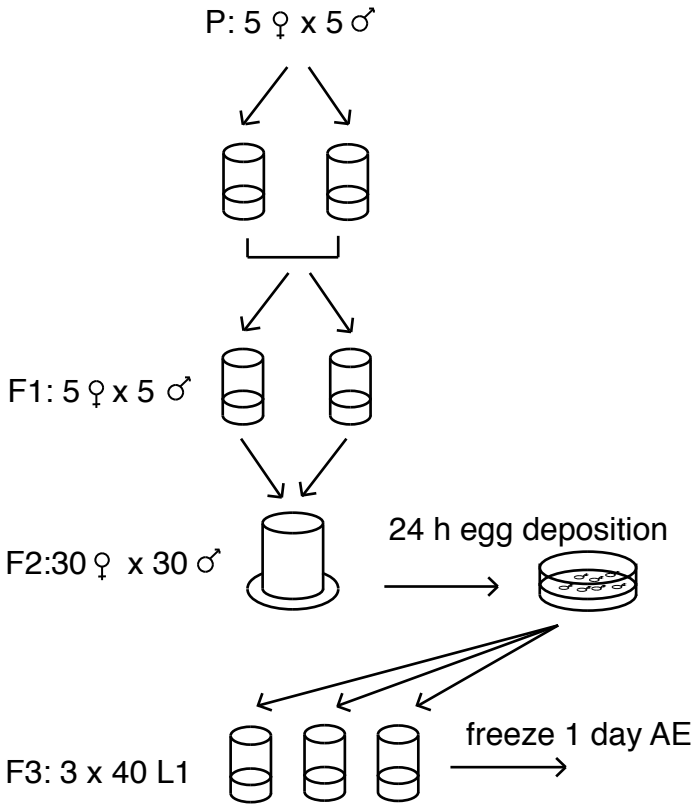
Figure 4. Associated SNPs overlap 33 functionally diverse novel candidate genes for wing size determination and localize within putative enhancer elements. a) Percent change in median wing area compared to *CG1315* RNAi upon wing-specific knockdown of candidate genes in females. Only the lines yielding a significant wing size change ($p<0.001$, Wilcoxon rank sum test) are depicted. b) Alignment of the 2kb region on chromosome arm *2L*

upstream of the *D. melanogaster ex* locus that shows sequence conservation across *Drosophila* species. The position of the SNP is indicated by the vertical blue line. The *D. melanogaster* sequence is represented by the dark grey bar at the top ("Sequence"). The respective sequences of each compared species are represented below. Light grey regions are matches to the *D. melanogaster* sequence, red regions are mismatches, gaps in the alignment are denoted by horizontal red lines and insertions by black lines and arrows.

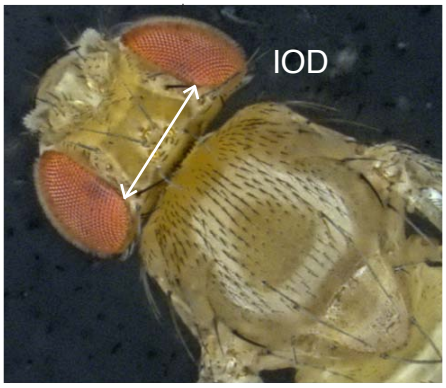
Figure 5. Pairwise interactions between focal genes and DGRP SNPs for male wing size (rCSM). The plot shows the focal genes annotated in black and the interactors in red. Interaction lines are colored according to the chromosome the focal gene is located on and the thick black lines denote Bonferroni-significant interactions. The outer circle demarks the chromosome arms (*2L* = orange, *2R* = yellow, *3L* = green, *3R* = purple, *X* = blue). The colored bars inside the inner circle demark the locations of cosmopolitan inversions (orange: *In(2L)t*; yellow: *In(2R)NS*; purple: *In(3R)K*, *In(3R)P*, *In(3R)Mo*).

Figure 1

a



b



14 landmarks for CS

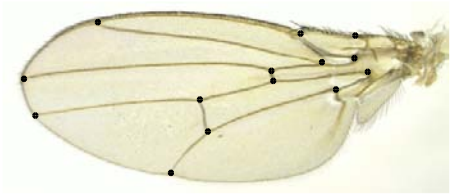


Figure 2

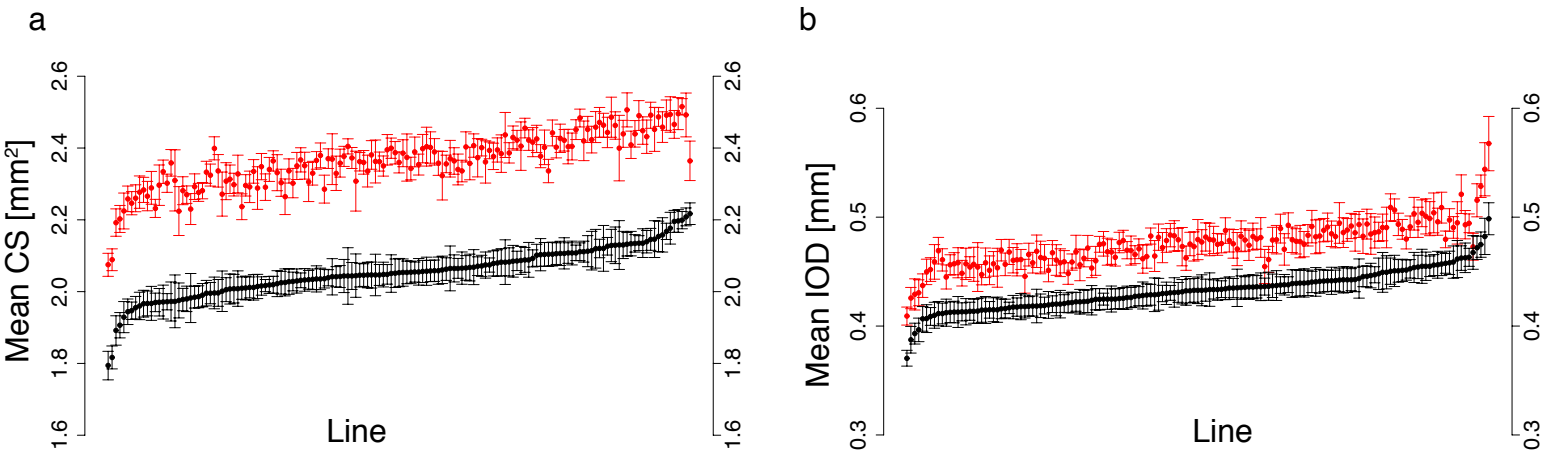
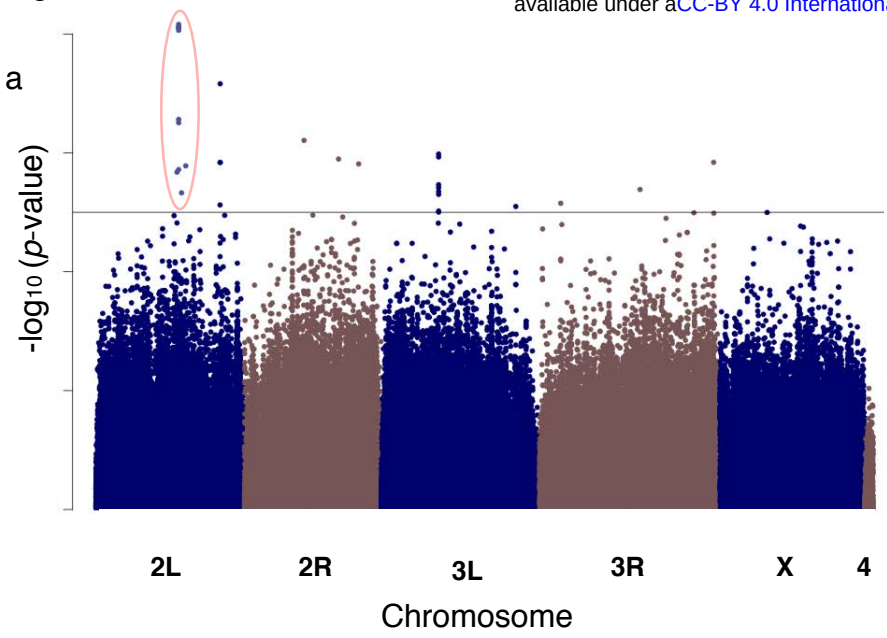
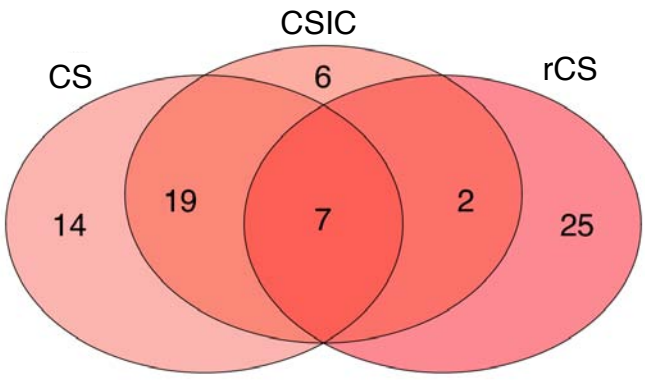
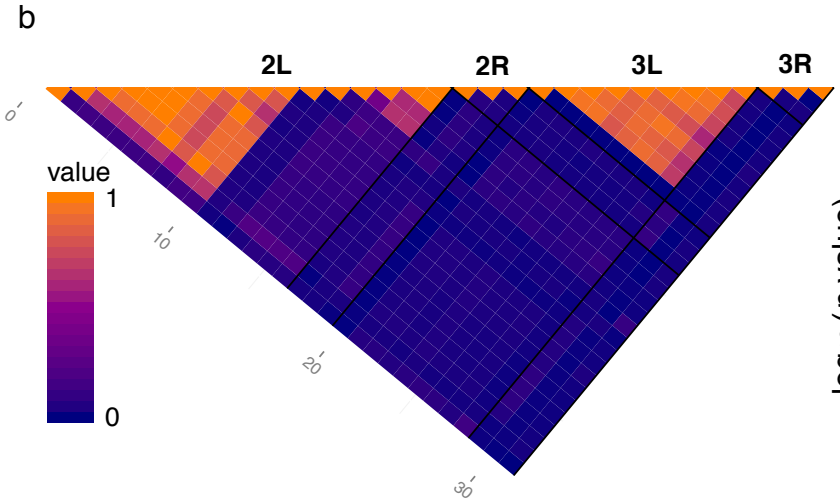


Figure 3







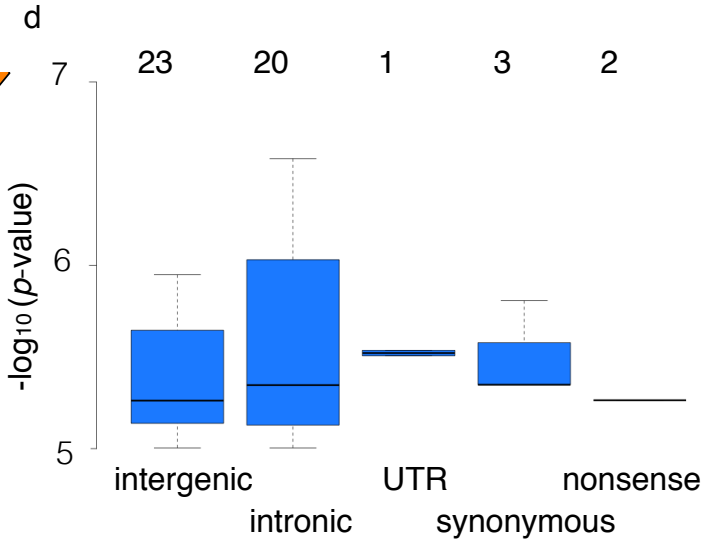
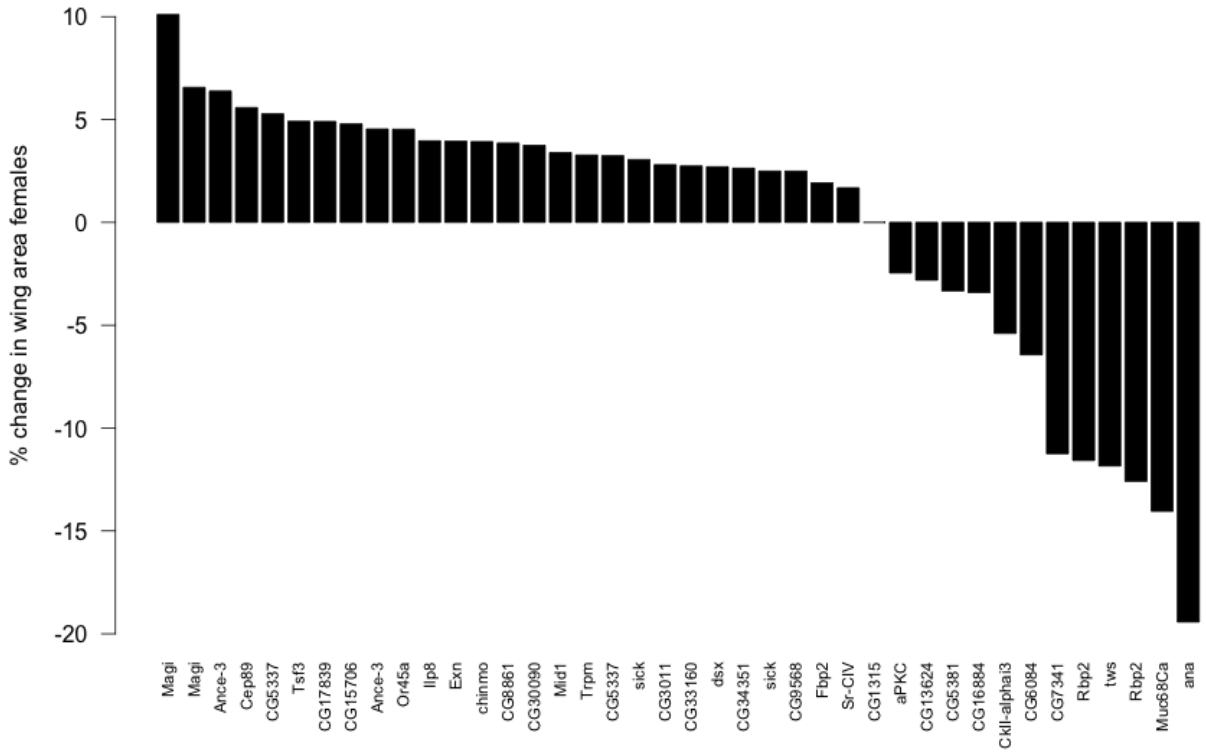


Figure 4

a



b

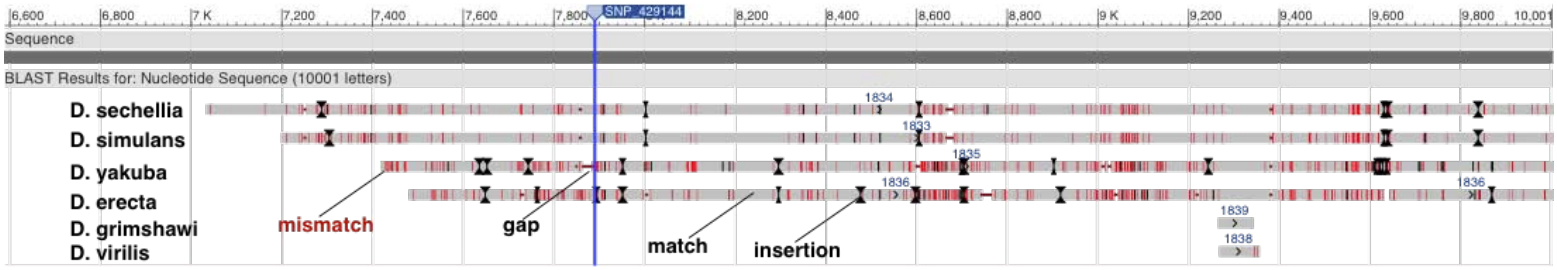


Figure 5