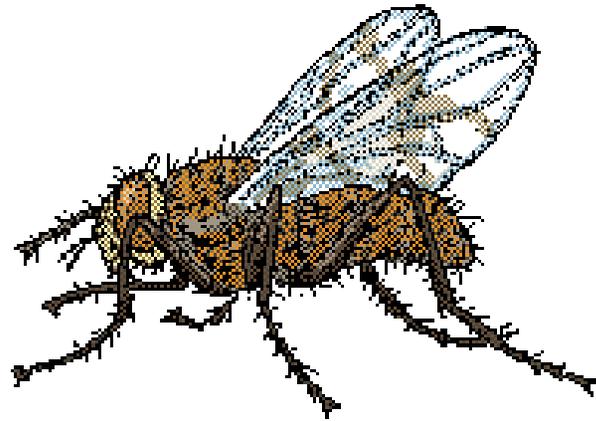# Systems genetics of complex traits in Drosophila melanogaster

Mikk Eelmets
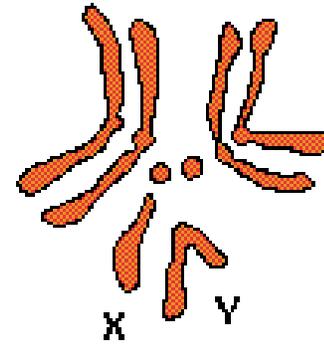
Journal Club

28.04.2009
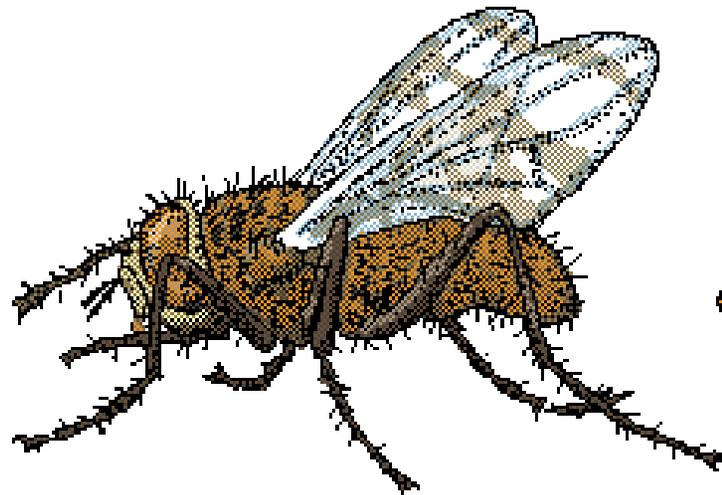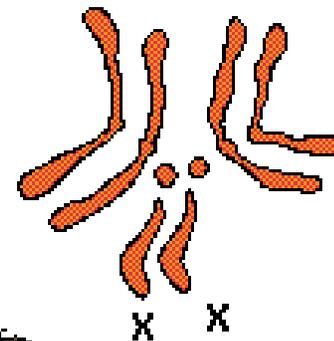
Drosophila melanogaster

Male

X Y

Female

X X

Drosophila melanogaster

Base Pairs: 168,736,537

Known protein-coding genes:14,141
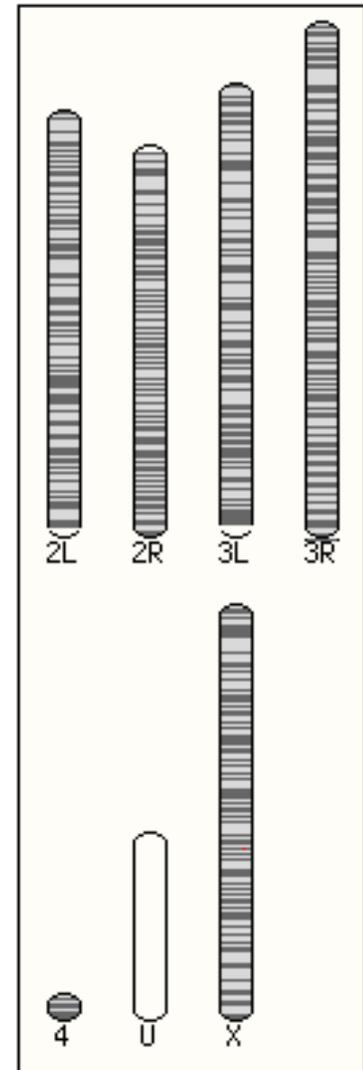
Pseudogenes:88

RNA genes:949

Gene exons:69,605

Gene transcripts:21,875

Genscan gene predictions:18,462

http://www.ensembl.org/Drosophila_melanogaster

Results
Data

- 40 highly inbred lines from the natural population of Releigh, North Carolina, USA

- Affymetrix Drosophila 2.0 array

- Of the **18800** transcripts on the array **14840 (78,9%)** were expressed in young adults.

Results

Correlated transcriptional modules.
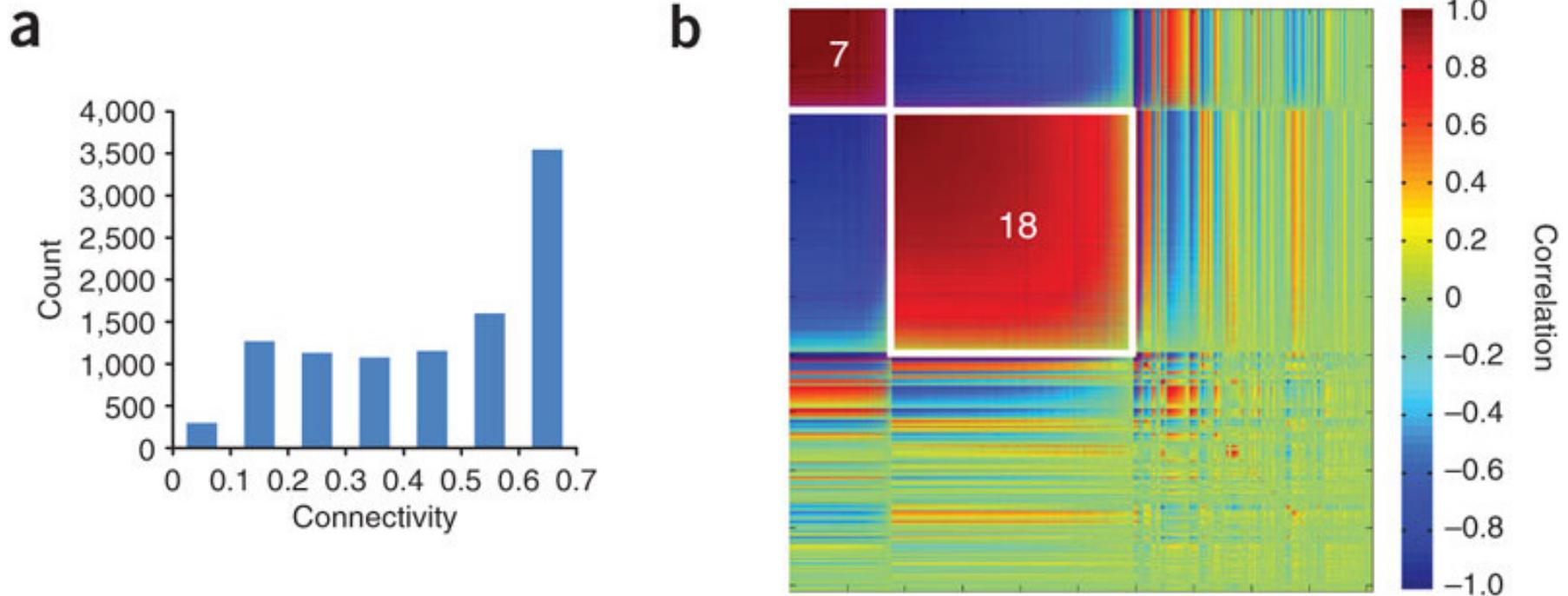
(**a**) Distribution of connectivity (average |*r*|) for the 10,096 genetically variable transcripts (line FDR < 0.001).
(**b**) Clustering of the genetically variable transcripts into 241 modules.

Results

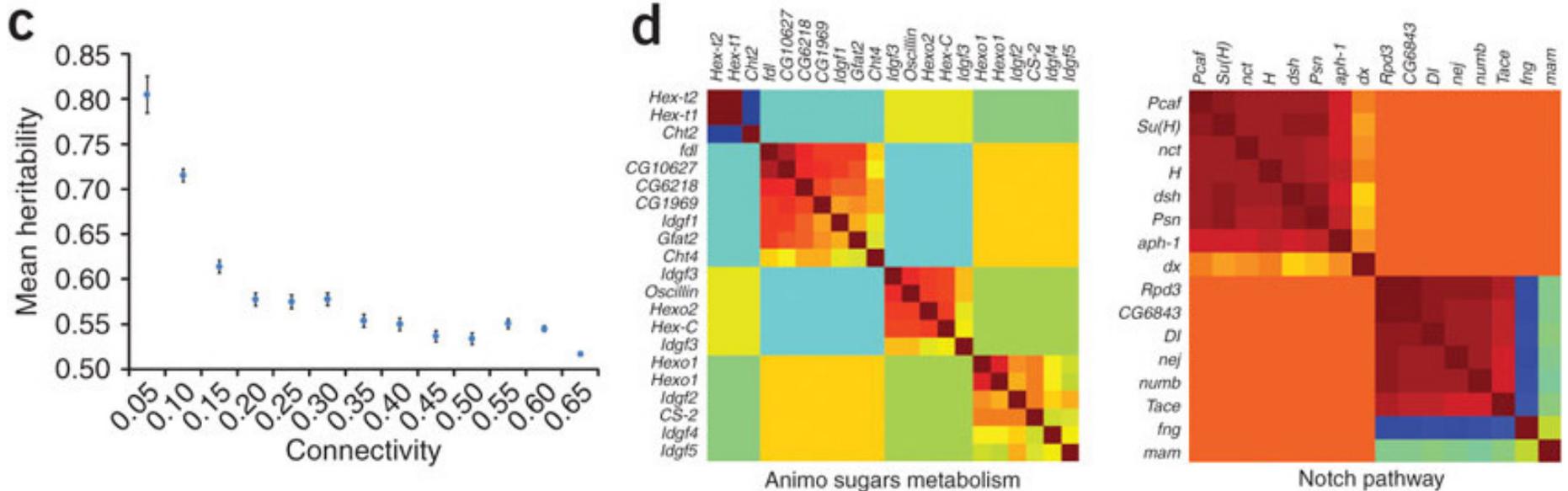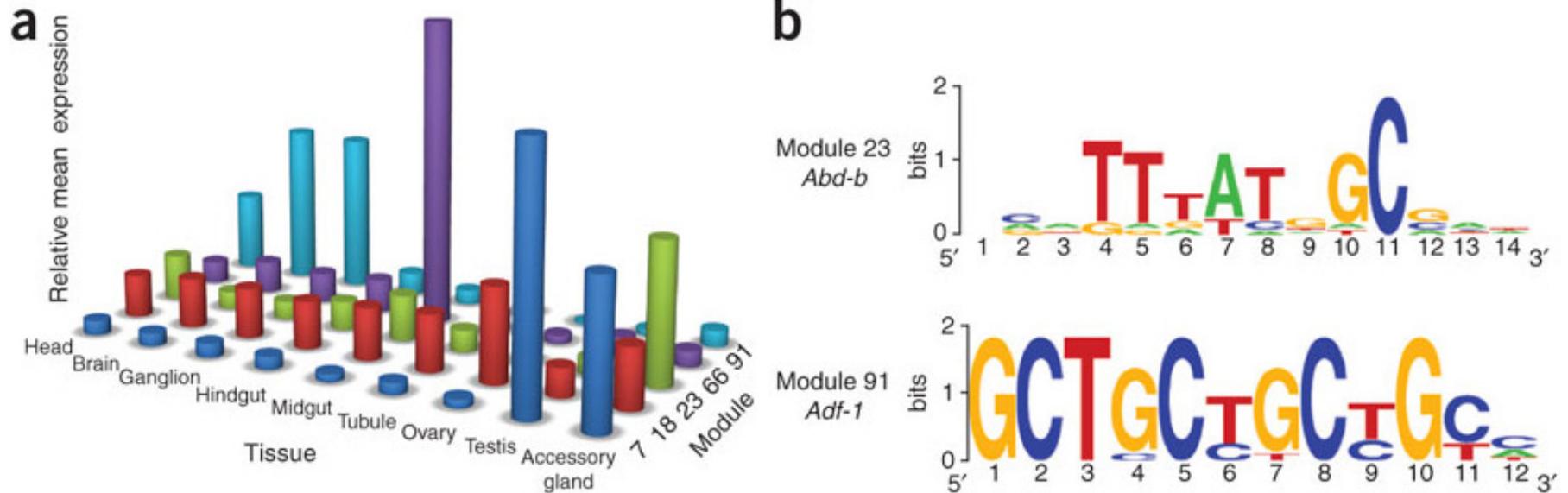## Correlated transcriptional modules.

(**c**) Relationship between transcript *H*2 and average connectivity. Error bars, s.e.m.

(**d**) Correlated transcriptional modules for genes in the amino sugars metabolism and Notch pathway KEGG ontologies. The colors on the off-diagonal represent the average cross-module correlations.

Results

Biology of transcriptional modules.

(**a**) Distribution of tissue-specific expression in modules 7, 18, 23, 66, 91. Module 7 is enriched for male-biased transcripts and expression in the testes and accessory glands. Module 18 is enriched for female-biased transcripts and expression in ovaries. Module 23 is enriched for transcripts affecting reproduction and gametogenesis that are highly expressed in ovaries and male accessory glands. Module 66 is enriched for transcripts in the Notch signaling pathway and nervous system development expressed in the midgut. Module 91 is enriched for transcripts affecting the function of the nervous system with high expression in the brain.

(**b**) Modules 23 and 91 are, respectively, enriched for the *Abd-b* ($P = 0.004$) and *Adf-1* ($P = 0.001$) transcription factor binding motifs. *Abd-b* has been implicated in genital disc development and *Adf-1* in memory and synaptogenesis, consistent with the inferred function of genes in these modules.

Results

Biology of transcriptional modules.
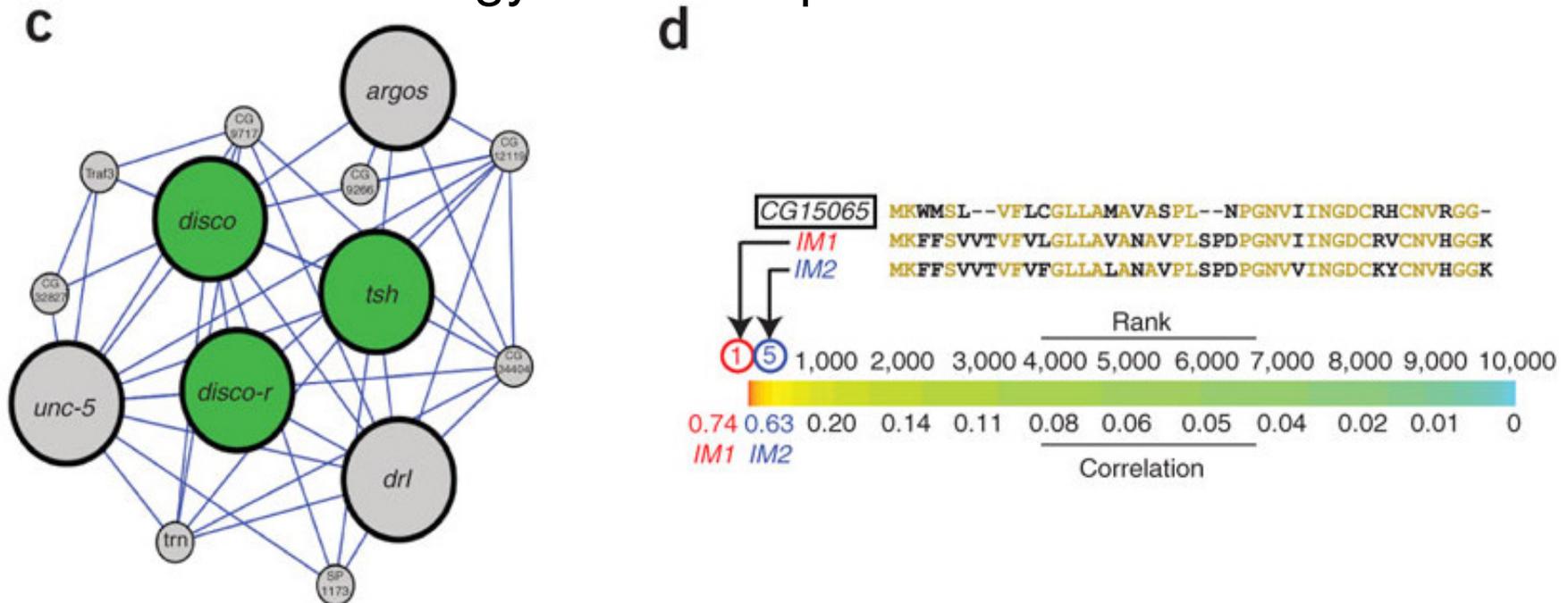
(**c**) Network representation of module 164, emphasizing the genetic correlations between adult transcripts for three transcription factors that interact during embryonic and larval development.

(**d**) Putative functional annotation of *CG15065* as a gene encoding an immune-induced molecule. Ranking all genetically variable transcripts according to their correlation to *CG15065* shows that *IM1* is the strongest transcriptional correlate (*r* = 0.74) and *IM2* is the fifth strongest (*r* = 0.63). The protein alignments of *CG15065*, *IM1* and *IM2* are highly conserved.

Results

Variation for organismal phenotypes among 40 wild-derived inbred lines.
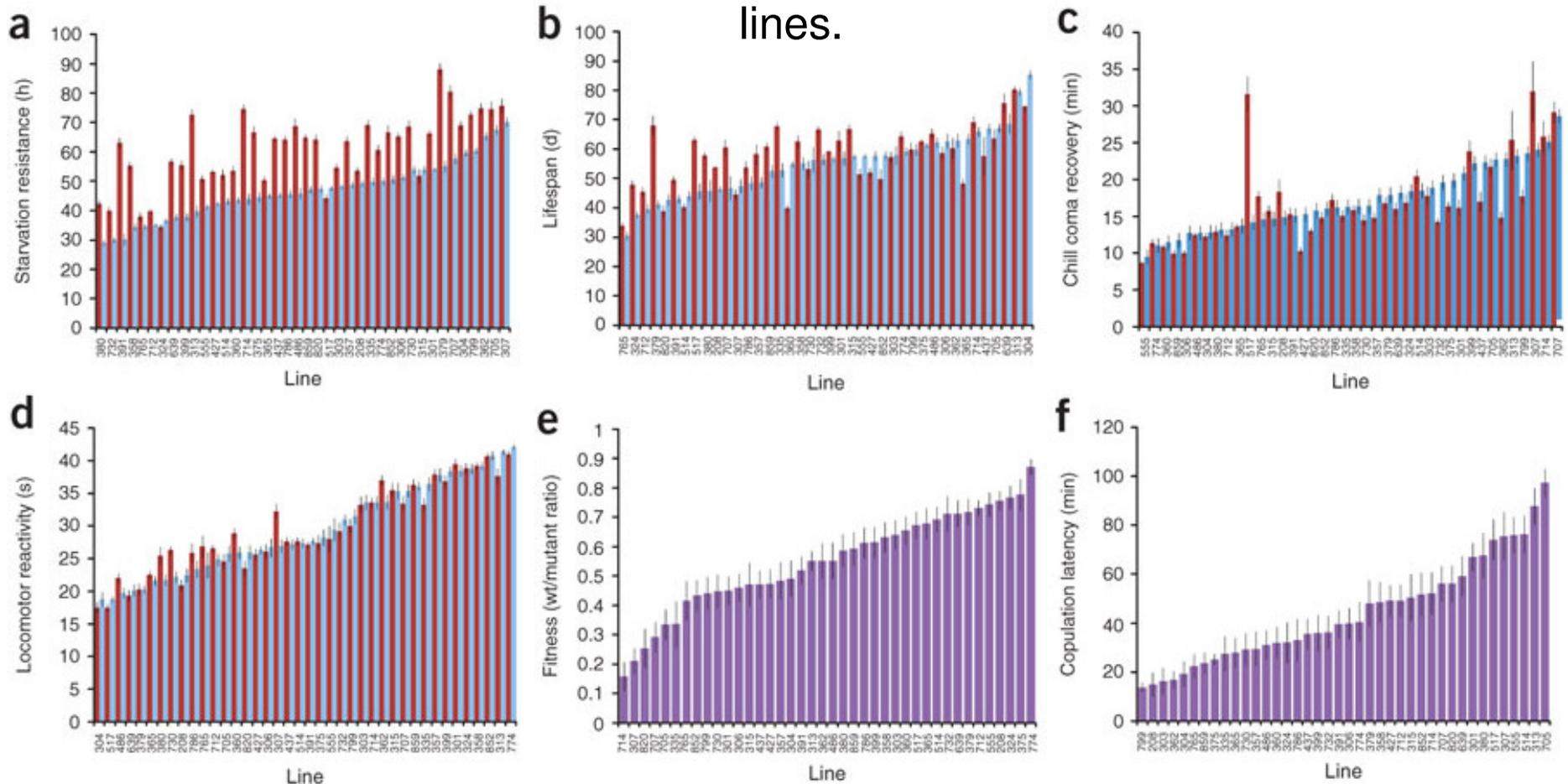
(**a–f**) Distributions of line means among 40 wild-derived inbred lines. The red and blue bars in panels **a–d** depict females and males, respectively. Sexes were not measured separately in panels **e–f**. Error bars, s.e.m. (**a**) Starvation stress resistance ($H2 = 0.56$). (**b**) Chill coma recovery ($H2 = 0.23$). (**c**) Life span ($H2 = 0.54$). (**d**) Locomotor reactivity ($H2 = 0.58$). (**e**) Copulation latency ($H2 = 0.25$). (**f**) Competitive fitness ($H2 = 0.32$).

Results

Modules of correlated transcripts associated with organismal phenotypes.
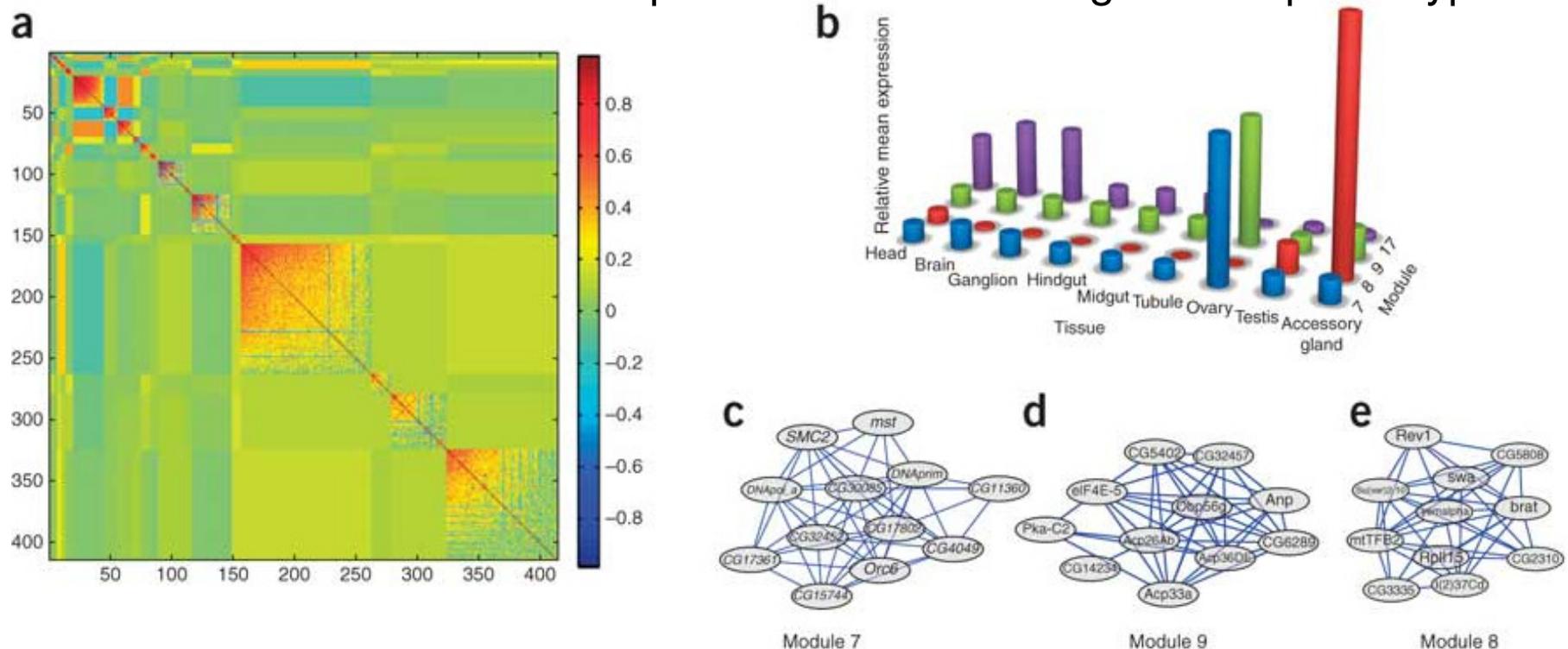
(a–e) Competitive fitness. (a) Clustering of the 414 transcripts significantly associated with variation in fitness into 20 modules. (b) Tissue-specific expression of transcripts in modules 7 and 9 (ovaries), module 8 (accessory glands and testes) and module 17 (head, brain and thoracicoabdominal ganglion). (c) Interaction network for module 7. Each node represents a gene and each edge the correlation between a pair of genes. Module 7 is enriched for female-biased transcripts and transcripts affecting DNA replication. (d) Interaction network for module 9. Module 9 is enriched for female-biased transcripts and transcripts affecting oogenesis and transcriptional regulation. (e) Interaction network for module 8. Module 8 is dominated by male-biased genes, and is enriched for genes involved in male-induced postmating behaviors, including three genes encoding accessory gland proteins (Acps).

Results

Modules of correlated transcripts associated with organismal phenotypes.
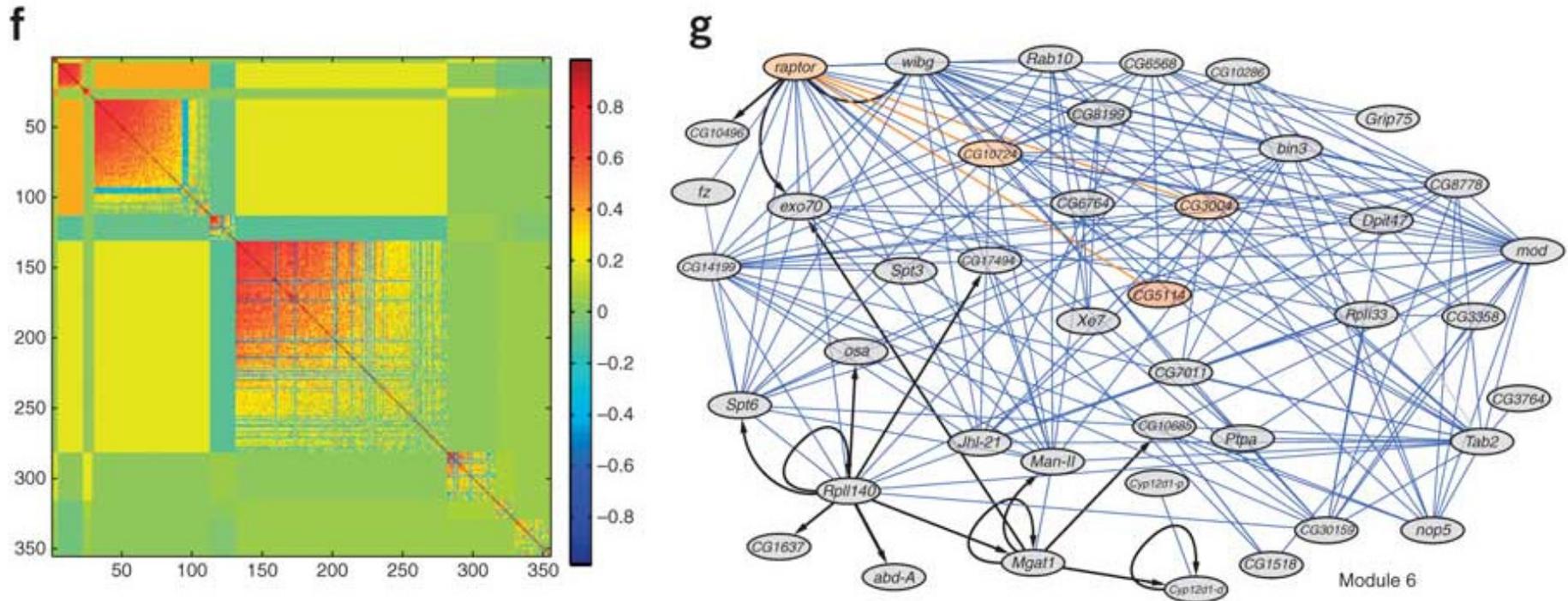
(**f–g**) Starvation stress resistance. (**f**) Clustering of the 355 transcripts significantly associated with variation in starvation resistance into 11 modules. (**g**) Interaction network for module 6. The black arrows indicate SFP variants in a probe set that are associated with variation in expression of the other probes in that probe set (*cis*-acting variants) and with variation in another transcript (*trans*-acting variants). The orange nodes indicate genes with a WD40 protein domain.

Results
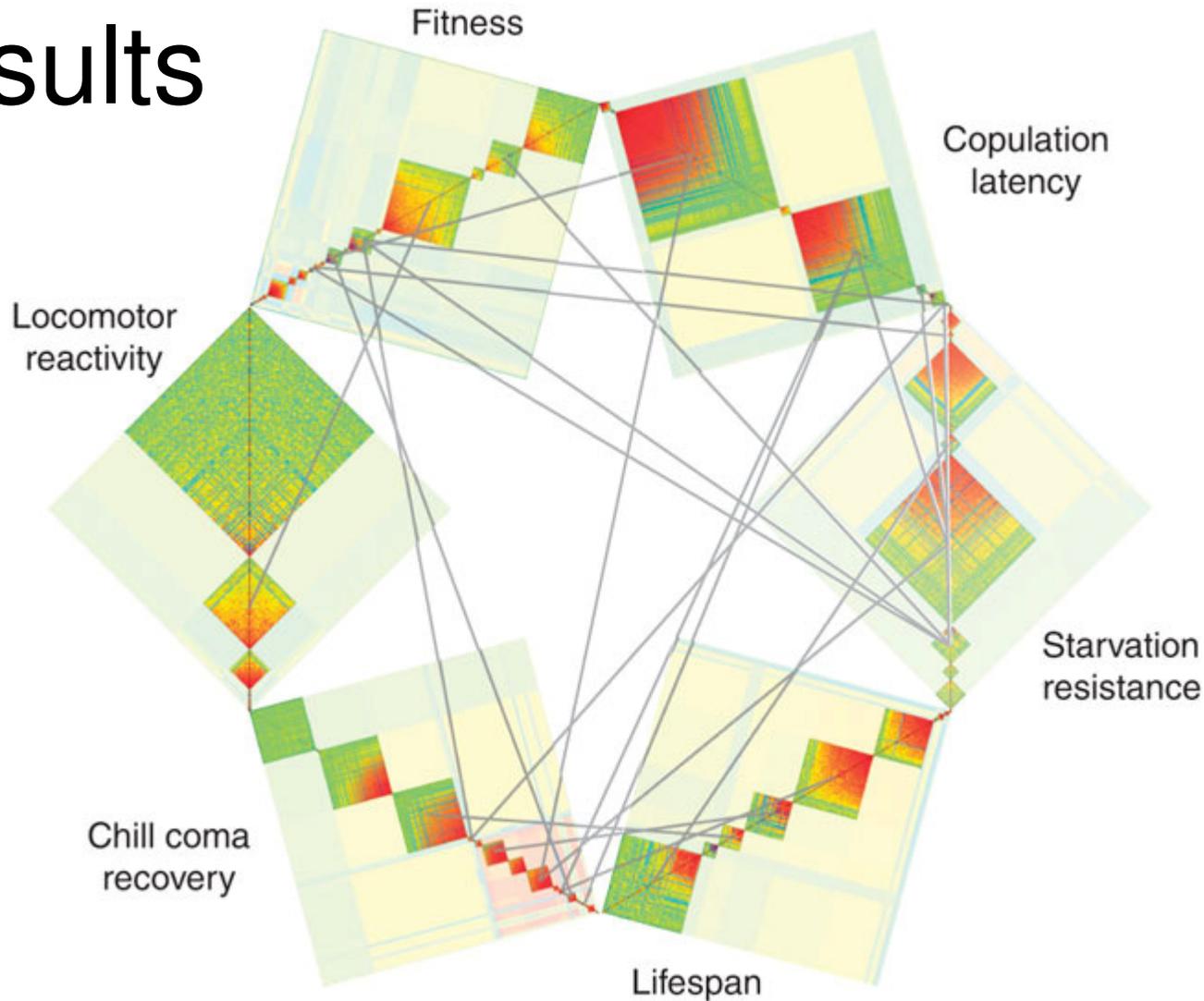
Pleiotropy between phenotypic modules.

Grey lines connect modules with a significant overlap of greater than four genes between gene lists, as determined by Fisher's exact tests.

Reference

Ayroles J.F., Carbone M.A., et al. (2009).

" Systems genetics of complex traits in Drosophila melanogaster."

Nat Genet **41**(3): 299-307.

# THANK YOU

Affymetrix Genome 2.0

Critical Specifications	
Number of arrays in set	One
Number of transcripts	~18,500
Number of probe sets	18,880
Feature size	11 µm
Oligonucleotide probe length	25-mer
Probe pairs/sequence	14
Array format	100
Control sequences included:	
Hybridization controls:	*bioB, bioC, bioD* from *E. coli* and *cre* from P1 bacteriophage
Poly-A controls:	*dap, lys, phe, thr, trp* from *B. subtilis*
Housekeeping/Control genes:	Actin (Actin 42A), GAPDH (Glyceraldehyde 3 phosphate dehydrogenase 2), Eif-4a (Eukaryotic initiation factor 4a)
Detection sensitivity	1:100,000*

*As measured by detection in comparative analysis between a complex target containing spiked control transcriptions and a complex target with no spikes.
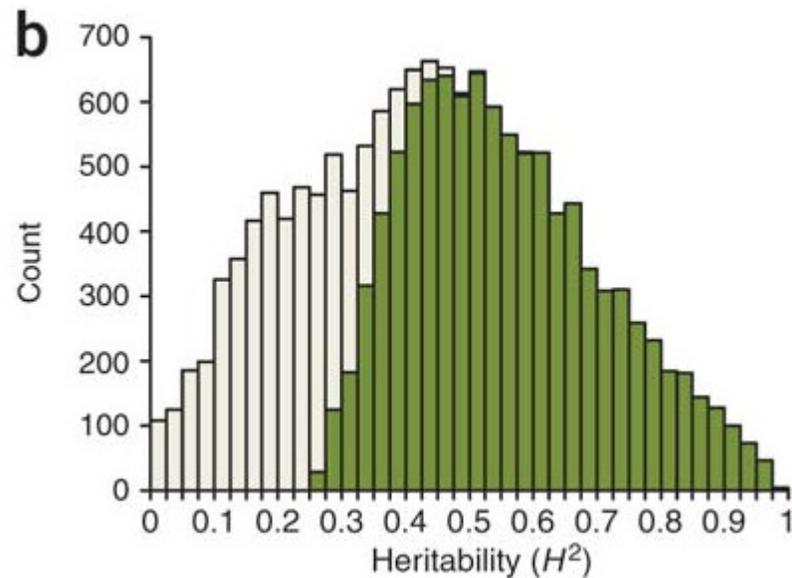
Results

Variation in transcript abundance among 40 wild-derived inbred lines.
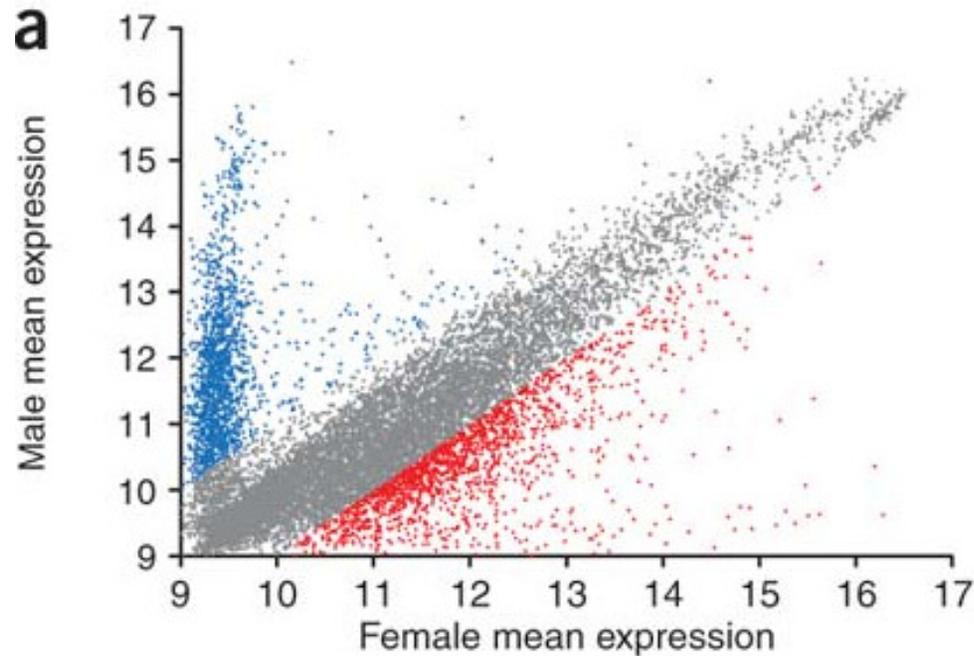
Sex bias for gene expression. Blue and red dots represent genes showing a twofold difference in gene expression between males and females, respectively.
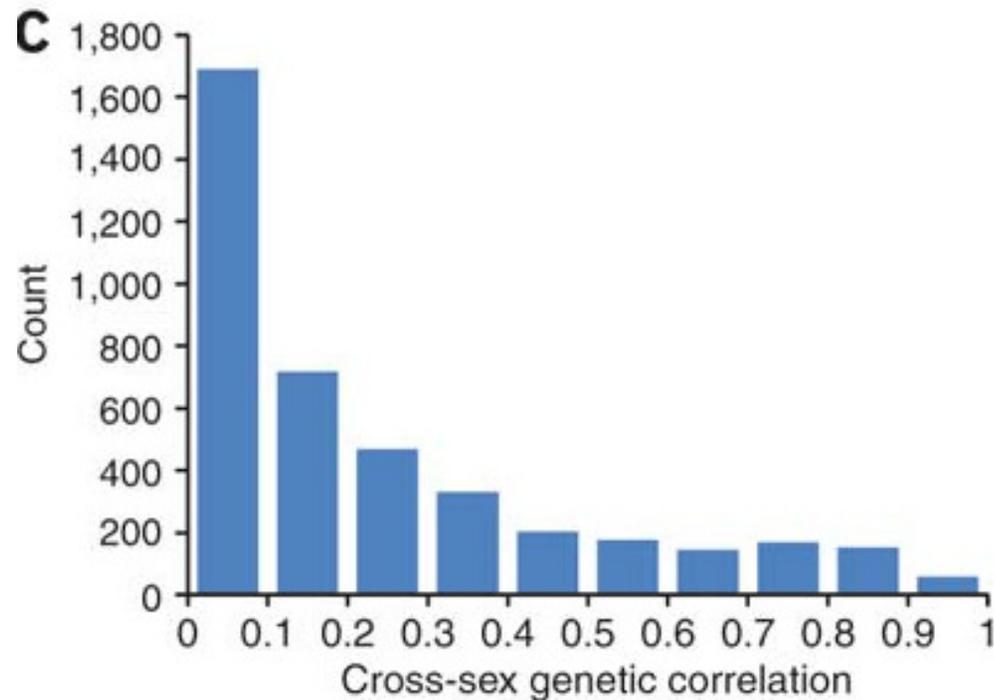
Results

Variation in transcript abundance among 40 wild-derived inbred lines.

Distribution of broad-sense heritabilities (*H*2). Dark green denotes significant *H*2 estimates (line FDR < 0.001) and grey indicates nonsignificant *H*2 estimates.
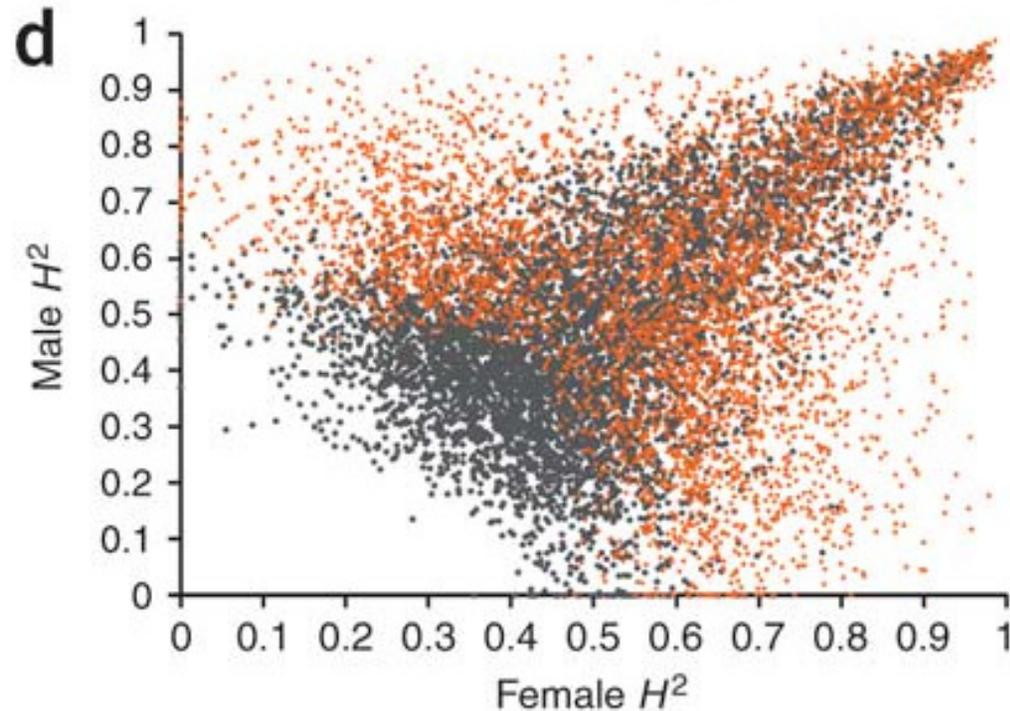
Results

Variation in transcript abundance among 40 wild-derived inbred lines.

Distribution of cross-sex genetic correlations for transcripts showing significant variation in sexual dimorphism (significant sex   line interaction variance at FDR < 0.001).
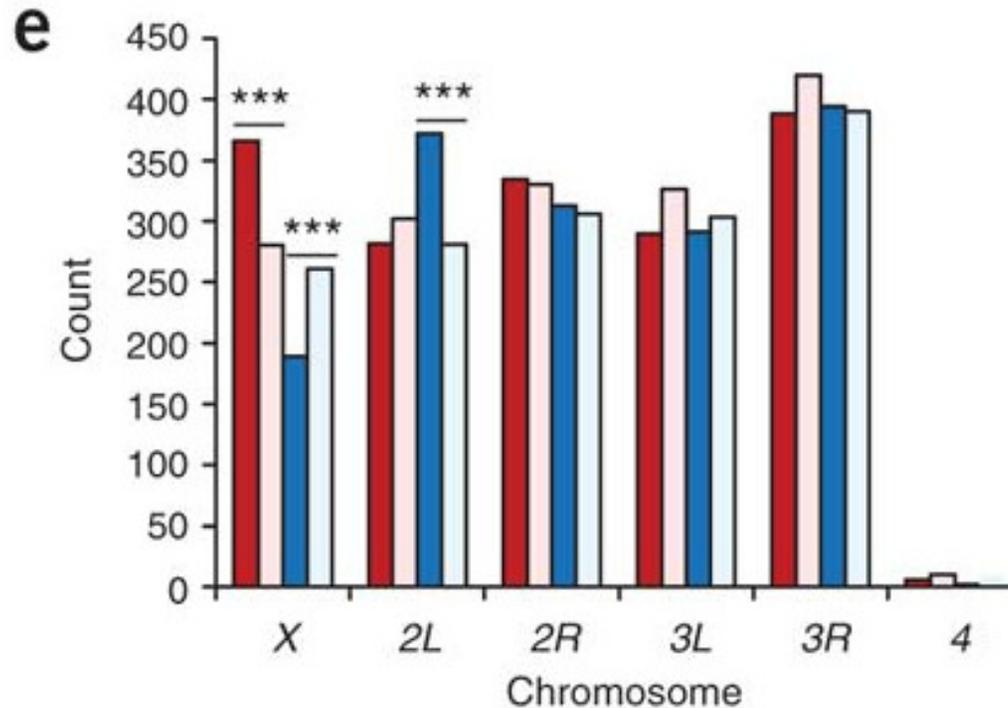
Results

Variation in transcript abundance among 40 wild-derived inbred lines.

Bivariate plot of $H2$ estimates in males and females. Orange dots indicate significant line-by-sex interaction variance.

Results

Variation in transcript abundance among 40 wild-derived inbred lines.


Chromosomal distribution of sex-biased gene expression. The dark blue and red bars are observed male and female counts, respectively, and the light blue and red bars are the expected numbers of male and female transcripts, respectively. Asterisks indicate significant deviation of observed from expected values ($P < 0.001$).

Results

Distribution of SFP effects.
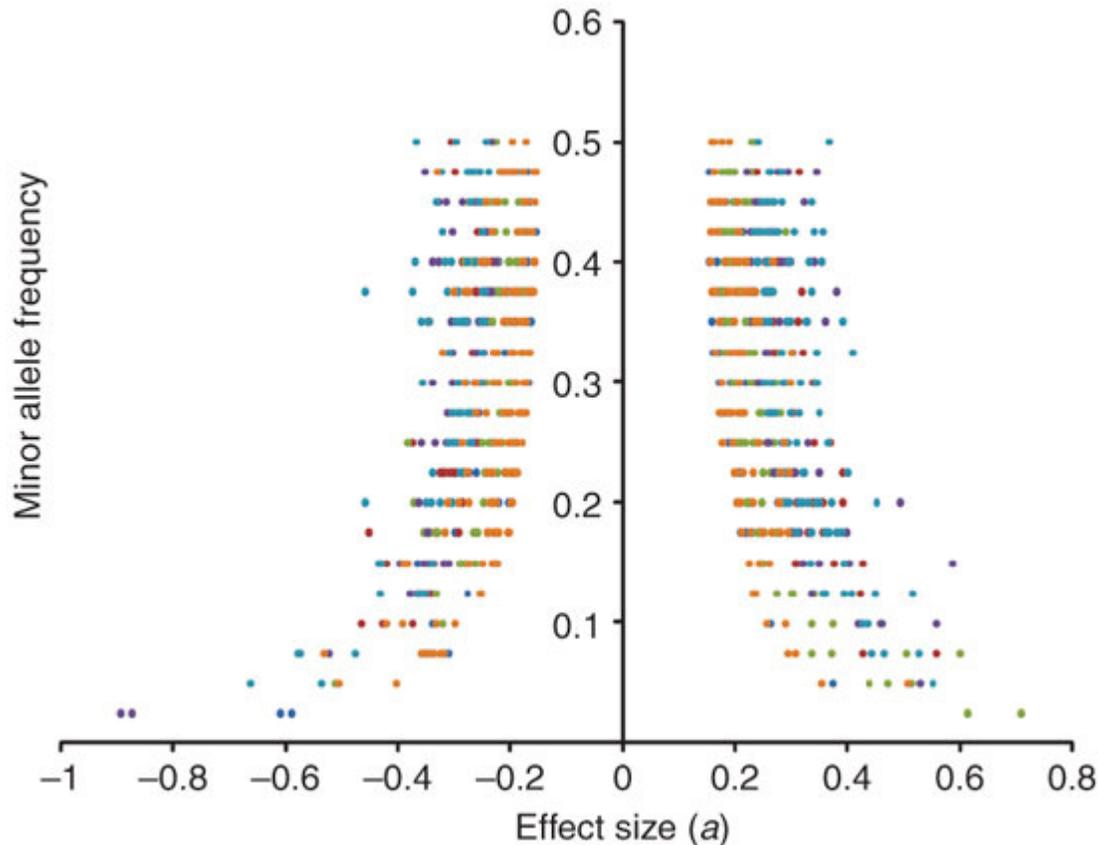
The *x* axis is the SFP allele effect, *a*/sigmaG, where *a* is one half the difference in trait mean between the SFP alleles and sigmaG is the genetic standard deviations of each trait. The *y* axis is the minor allele frequency. The traits are color-coded: chill coma recovery (dark blue), starvation resistance (red), fitness (green), lifespan (purple), locomotor reactivity (turquoise) and copulation latency (orange).
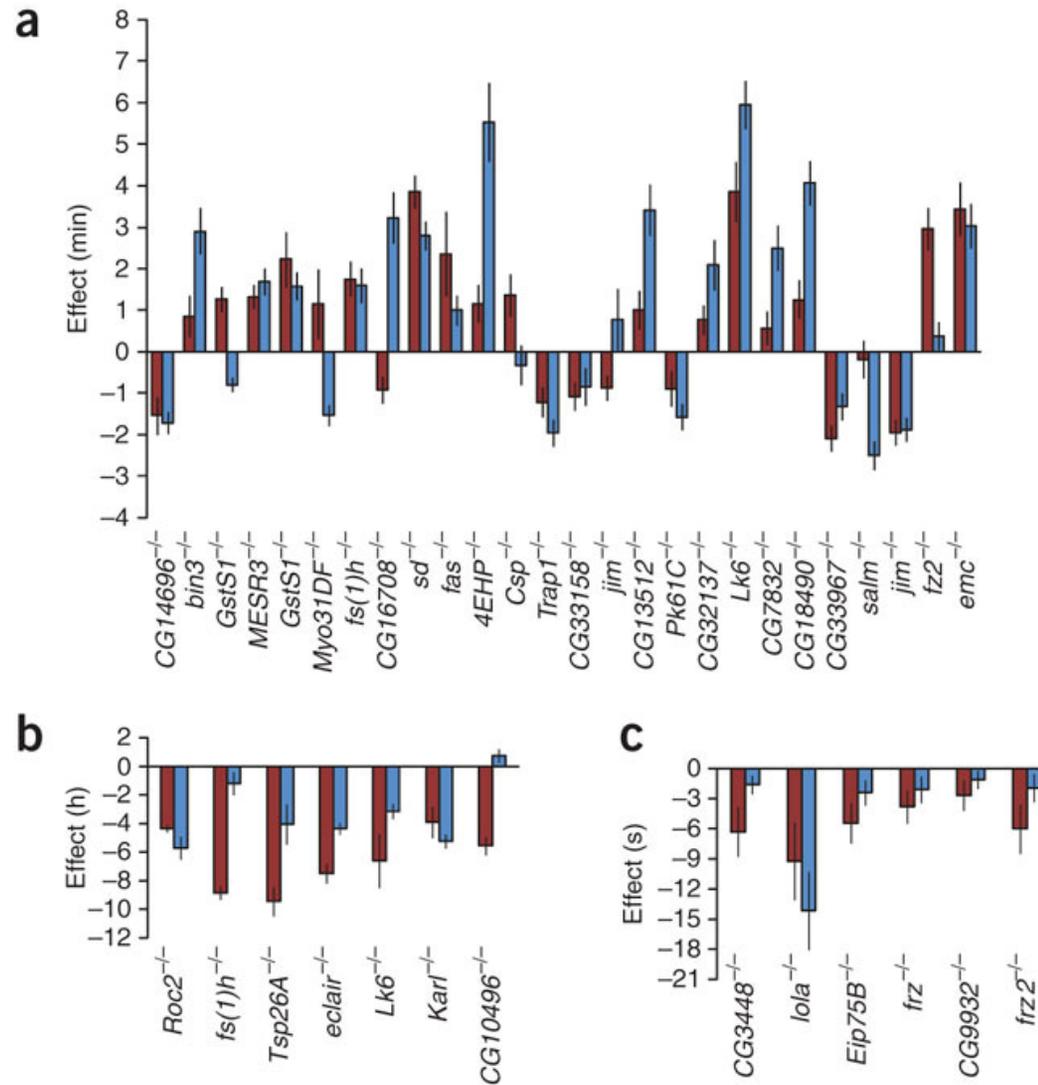
Effects of P-element mutations in candidate genes affecting quantitative traits.
Mutational effects are given as deviations from the co-isogenic control line. Red and blue bars represent males and females, respectively. Mutations in all genes shown have significant effects in one or both sexes (Supplementary Table 7). Error bars, s.e.m. (**a**) Chill coma recovery time. (**b**) Starvation stress resistance. (**c**) Locomotor reactivity (data from ref. 27).