Practical Bioinformatics

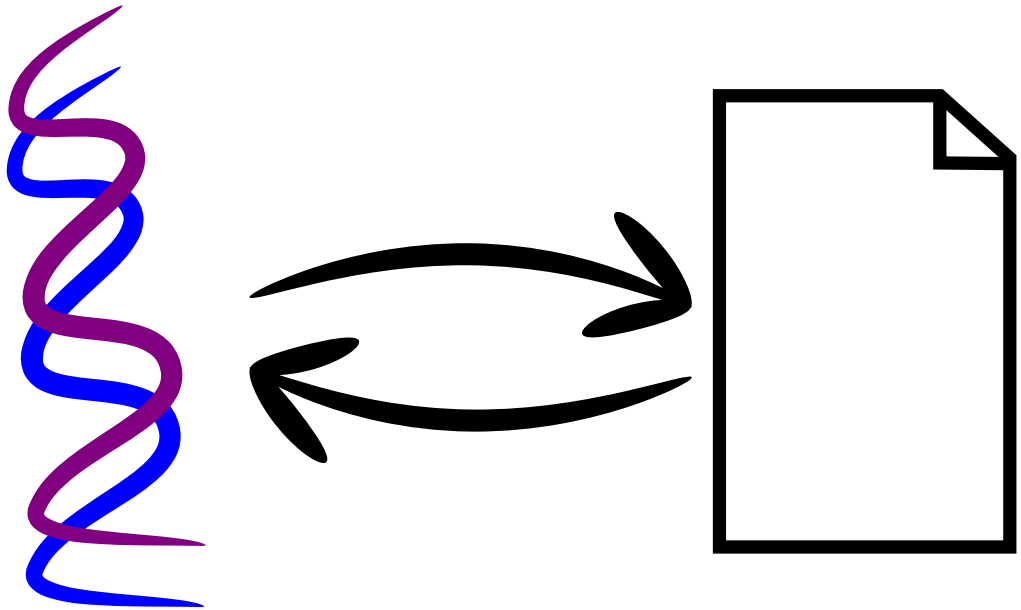
Mark Voorhies

5/29/2019

Habits are things you get for free, without requiring any special work.

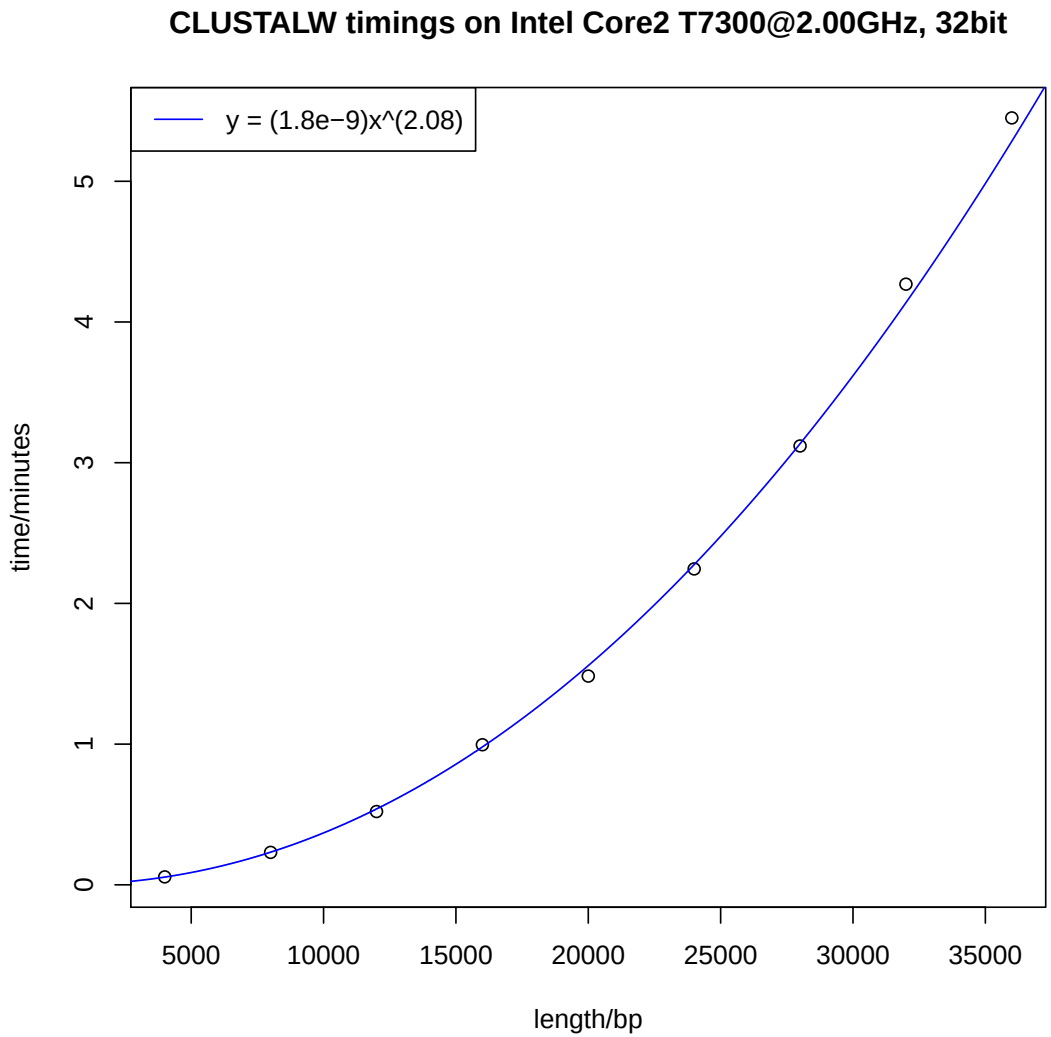
–Cory Doctorow *Advice to Writers, 4/5/2012*

Every object should have an isomorphism to a file

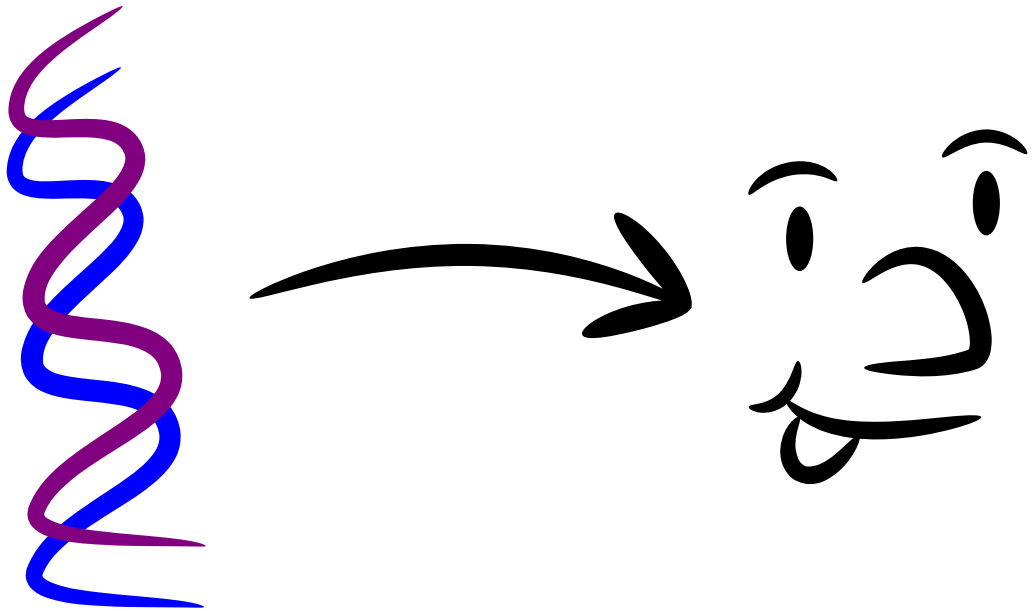


- Export, audit, edit, and import *independent* of a given program.
- Standard file formats for portability.
- Don't be afraid to look inside and hack on *your* data files.

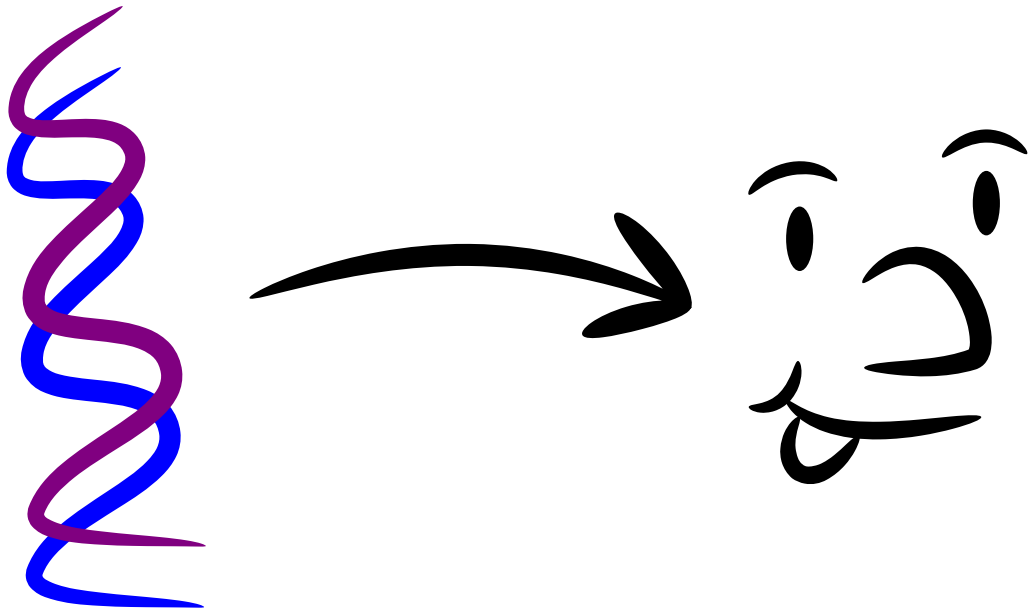
Run times are predictable and measurable



Phenotype is more diverse than Genotype

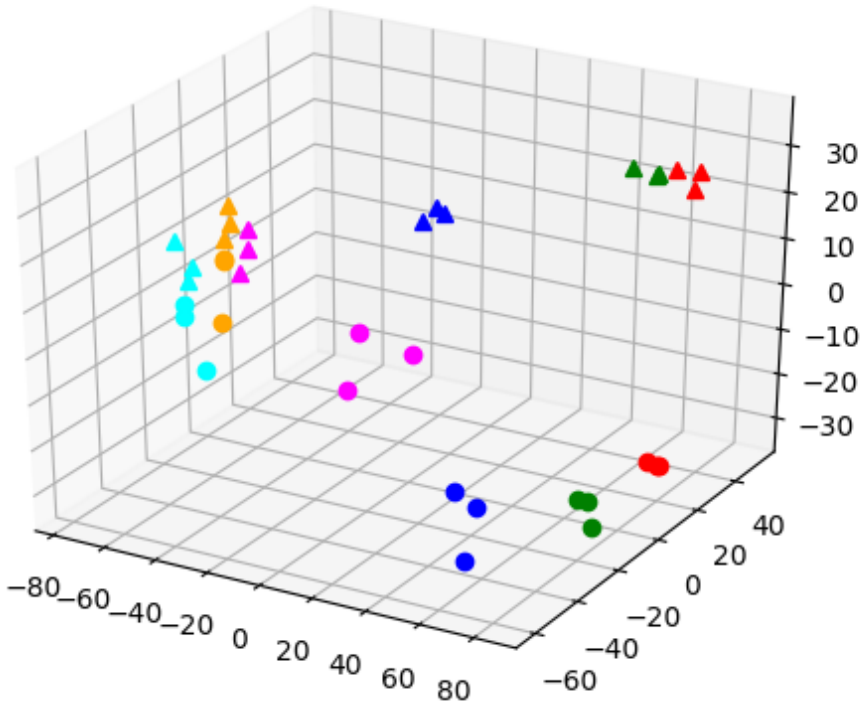


Phenotype is more diverse than Genotype

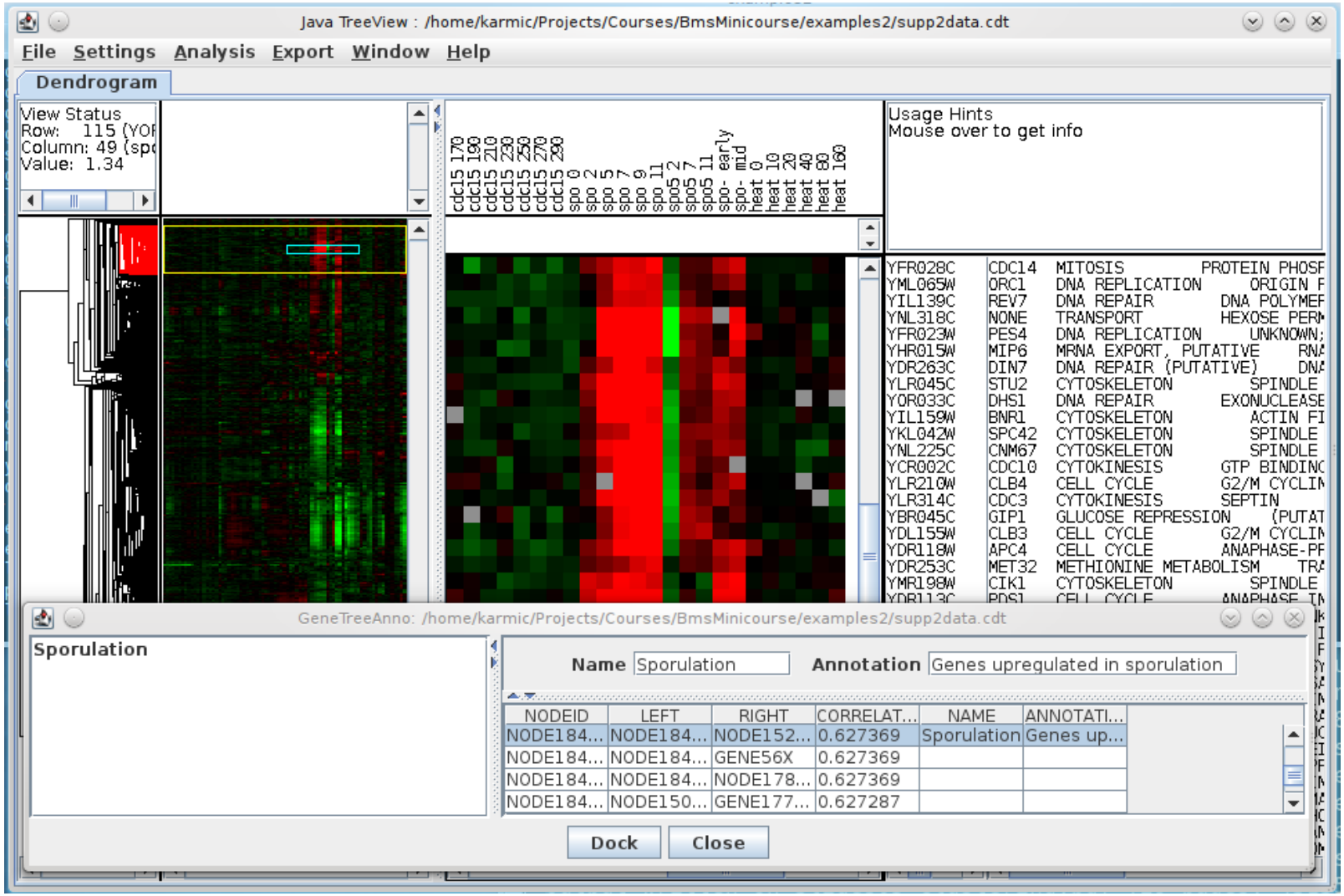


- Make sure you know what you are measuring
- Nucleic acid sequences are especially easy to address
- Many phenotypes can be analyzed by common numerical methods

Start from an unbiased view

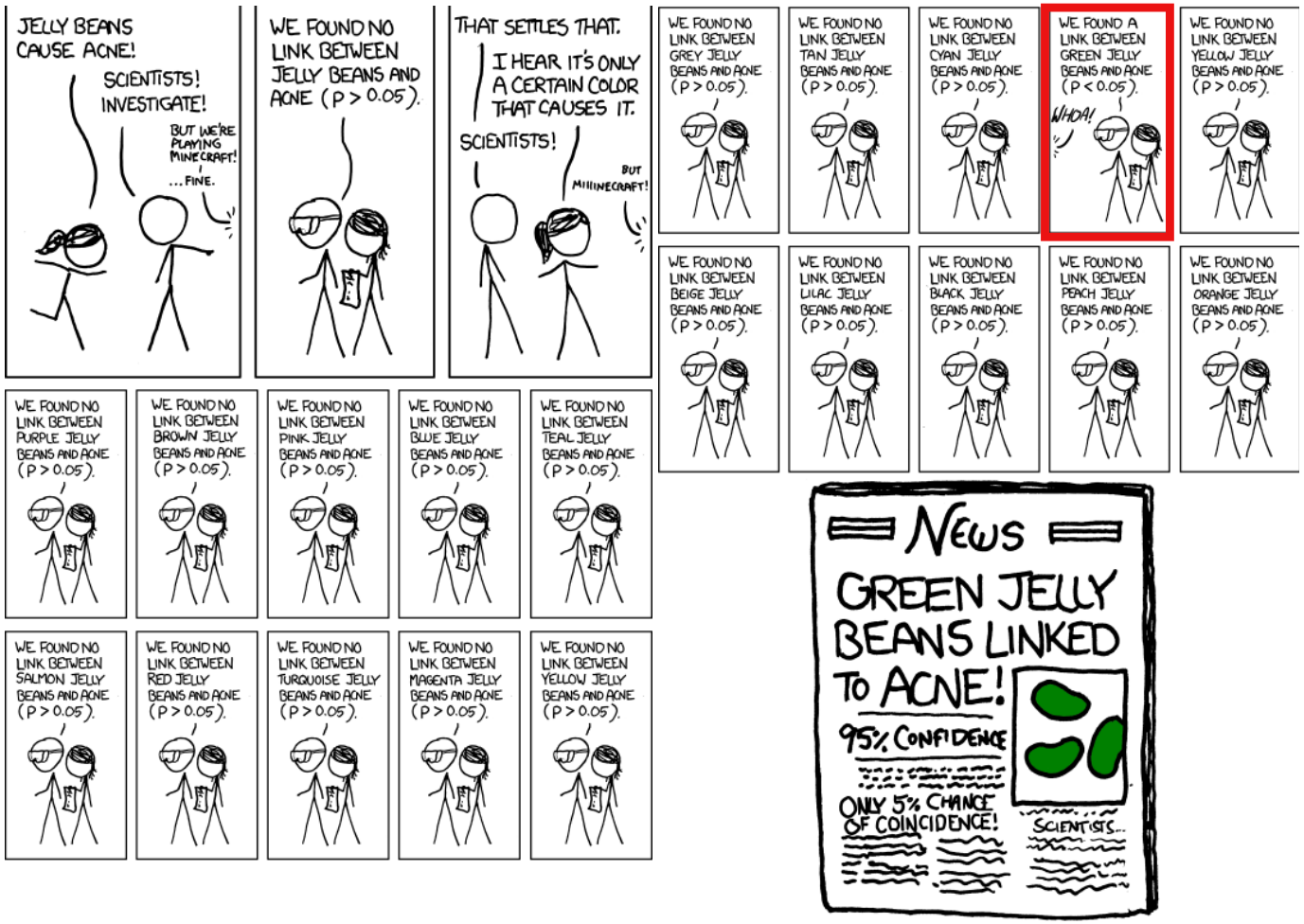


Tools should support aggregation and annotation



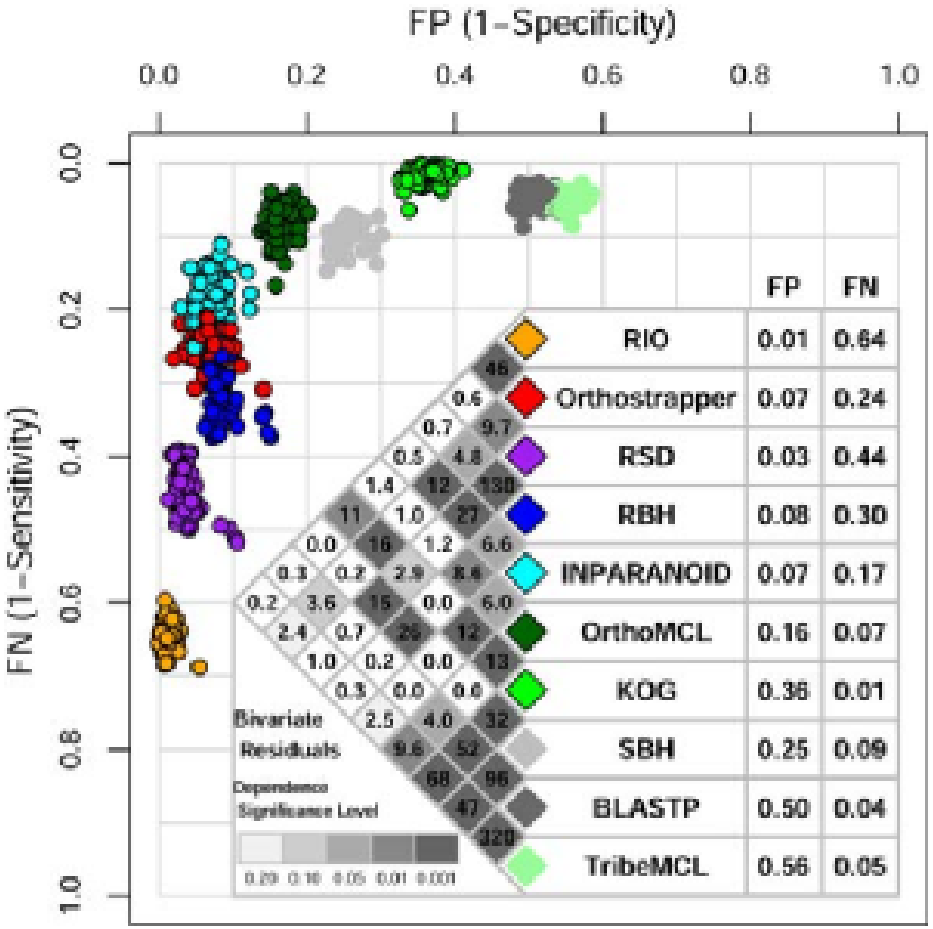
See also: IGV, JalView, and CytoScape

Account for Multiple Hypothesis Testing



http://xkcd.com/882/

Choose appropriate statistical tools



PLoS ONE 2:e383

- Solid theoretical basis
- Reasonable power
- Estimates significance *and* effect size
- When in doubt, find positive and negative controls

Science is a Conversation

- Follow computational methods as they evolve (Web of Science; RSS: PubMed, GEO, arXiv, ...)

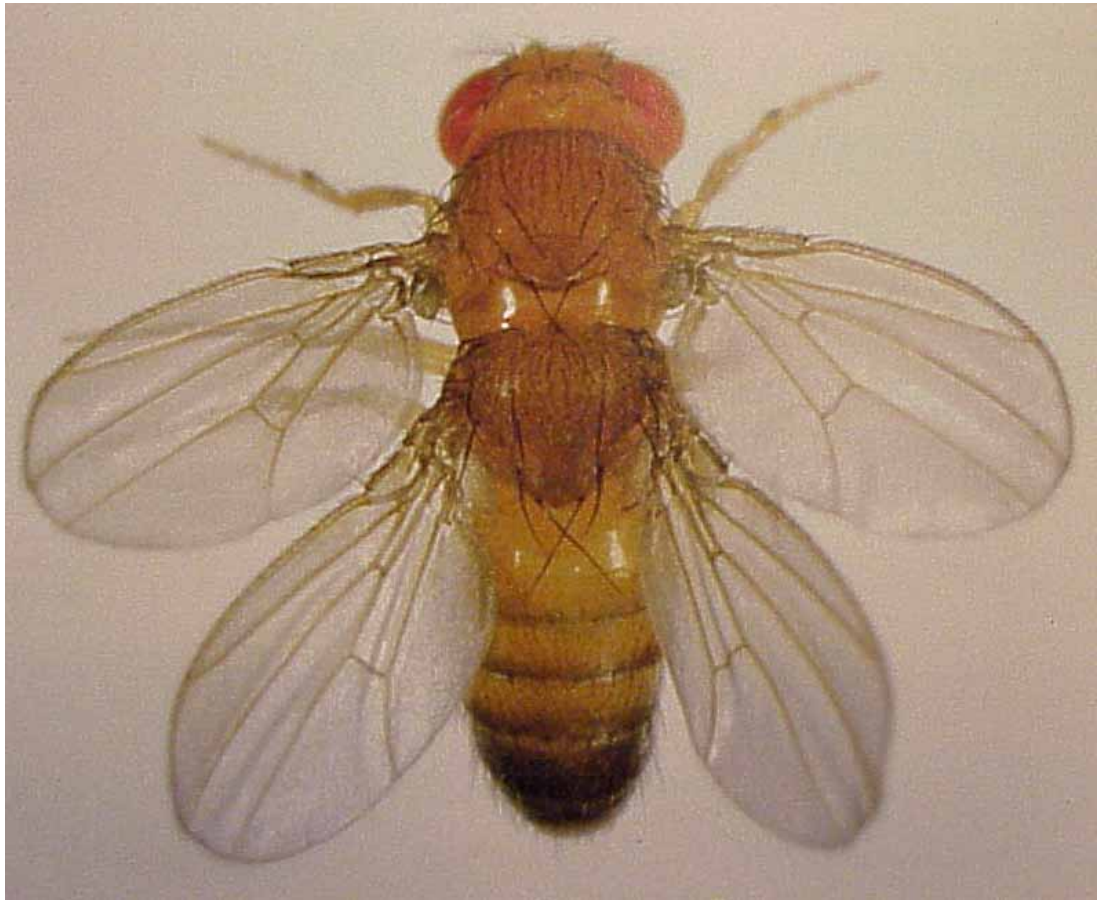
Science is a Conversation

- Follow computational methods as they evolve (Web of Science; RSS: PubMed, GEO, arXiv, ...)
- As a reviewer, insist on availability of data and source code

Science is a Conversation

- Follow computational methods as they evolve (Web of Science; RSS: PubMed, GEO, arXiv, ...)
- As a reviewer, insist on availability of data and source code
- Draw on your classmates' expertise

We understand systems by breaking them



Source: Peter A. Lawrence via http://www.bio.davidson.edu/courses/molbio/ubx/ubx.html