Systems Genetics of Complex Traits in *Drosophila*

Trudy F. C. Mackay

Advances in animal and plant breeding, medicine and predicting adaptive evolution all depend on a detailed understanding of segregating genetic variation affecting complex, or quantitative, traits. Specifically, we need a complete accounting of the genetic variation segregating in a population of interest, and estimates of the homozygous, heterozygous, epistatic and pleiotropic effects of the variants on complex traits, in a range of ecologically relevant environments. Large genome-wide association studies for many human quantitative traits and common diseases have revealed an unexpected complexity of mapping genotype to phenotype. Many loci with individually small effects affect each trait, and in total only account for a small fraction of the total heritable variation. Studies in model organisms recapitulate this complexity, including the ubiquity of unexpected effects on multiple traits, the importance of gene-gene and gene-environment interactions, and sex-specific effects of segregating alleles. Furthermore, DNA variants do not affect complex phenotypes directly, but do so via networks of interacting transcripts, proteins and metabolites. Intermediate molecular phenotypes such as transcript abundance also vary genetically in populations, and are themselves quantitative traits for which we can map causal molecular variants. The variable transcripts cluster into modules of genetically correlated transcripts, or co-expression networks, which are in turn associated with organismal phenotypes. The emerging field of 'systems genetics' integrates DNA sequence variation, variation in transcript abundance (and other molecular phenotypes) and variation in organismal phenotypes in a linkage or association mapping population, to derive causal, biologically meaningful, co-expression networks affecting genetic variation in complex organismal phenotypes.

Optimal conditions for using systems genetics include: (1) a mapping population with an ultra-fine recombination map, for precise localization of causal variants; (2) complete genome sequence of each genotype in the mapping population; (3) a sufficiently large sample size to detect loci with moderately small effects; and (4) the ability to measure multiple complex traits, including whole genome transcript abundance, for all individuals in the population. At present, these criteria are best fulfilled using genetic model organisms. We have developed a new community resource for association mapping of complex traits, the *Drosophila melanogaster* Genetic Reference Panel (DGRP). The DGRP consists of 192 lines derived by 20 generations of full sib mating from wild-caught females from the Raleigh, USA population. Complete genome sequences for the DGRP lines are being obtained in collaboration with Drs. Stephen Richards and Richard Gibbs of the Baylor College of Medicine Human Genome Sequencing Center using Illumina next generation sequencing technology. The sequences are a mixture of 75 bp and 95 bp

Department of Genetics, North Carolina State University, Raleigh, NC 27695-7614, USA

paired end reads, at an average coverage of 15x per line. Linkage disequilibrium declines rapidly in *Drosophila* regions of normal recombination – a highly favorable scenario for using association mapping to identify genes and even causal polymorphisms affecting variation for quantitative traits when all polymorphisms are genotyped. The DGRP lines are inbred, therefore large numbers of genetically identical individuals can be reared under controlled environmental conditions, and the same lines can be interrogated for genetic variation in transcript abundance and a suite of complex organismal phenotypes. The *Drosophila* complex trait community is committed to quantifying a broad spectrum of complex traits on these lines, many of which are homologous to human and animal health-related phenotypes.

At the time of writing, sequences of 160 lines have been obtained and made publicly available. I will discuss the results of genome wide association mapping for a number of life history (longevity, starvation stress resistance, time to recover from a chill induced coma) and behavioral (aggression, sleep) traits. The key features of these results are as follows. (1) Many loci affect each trait. (2) The distribution of the absolute value of allelic effects is J-shaped, with many alleles with small effects and fewer with large effects. (3) There is a trend for the alleles with smaller effects to be at intermediate frequency, and those with larger effects to be less common. (4) The loci are largely novel and previously not implicated to affect the traits. (5) Most variants affecting the traits are intronic, and in many cases the variants lie in gene deserts. In addition to the sequence data, we have quantified genome-wide transcript abundance for 40 of the DGRP lines. We found 10,096 genetically variable transcripts at a conservative false discovery rate of 0.001, which formed 241 genetically correlated transcriptional modules that were enriched for transcripts in common pathways, gene ontology categories, tissue-specific expression, and transcription factor binding sites. The high intra-module genetic correlation greatly reduces the dimensionality of the transcriptome, such that each module is an effective unit of observation. The high transcriptional connectivity allowed us to infer not only genetic networks, but also the function of predicted genes based on annotations of other genes in the network. We integrated the transcriptional variation data with the sequence variation and phenotypic variation, to identify expression QTLs associated with the transcripts, and modules of correlated transcripts associated with the quantitative traits. I will discuss examples of these causal transcriptional networks and tests to confirm their biological reality.

Drosophila is a model system from two perspectives. First, general features of the genetic architecture of complex traits are likely to be conserved across taxa. Second, key genes and networks are evolutionarily conserved, such that many pathways first discovered in *Drosophila* are also biologically relevant in other animals (e.g., the Wnt and Notch signaling pathways). Thus, genetic networks and biological functions discovered in *Drosophila* will be invaluable for functionally annotating genes in animal species, from livestock to humans.