Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Behavioral Sciences

# Systems genetics of behavior: a prelude

Megan K Mulligan and Robert W Williams

Behaviors are among the most complex phenotypes — the end product of intertwined networks of genetic, molecular, cellular, physiological, environmental, and social factors. While reductionist methods that test contributions of single genes are crucial to define components of behavior, there are now complementary integrative methods that can model and predict behavior in a more global context. Systems genetics is a new field that exploits massive multiscalar data to study patterns of covariation in complex biological systems. As applied to behavioral genetics, this field looks inside the black box to understand hierarchies of molecular and cellular traits and connections that link gene variants and environment to behavior. Here we explore recent progress and challenges in this new field.

**Address**
Department of Genetics, Genomics and Informatics, University of Tennessee Health Science Center, Memphis, TN 38163, United States

Corresponding author: Mulligan, Megan K (mmulliga@uthsc.edu)

**Current Opinion in Behavioral Sciences** 2015, **2**:108–115

This review comes from a themed issue on **Behavioral genetics**

Edited by **William Davies** and **Laramie Duncan**

For a complete overview see the Issue and the Editorial

Available online 9th February 2015

http://dx.doi.org/10.1016/j.cobeha.2015.01.014



## Introduction

Recent advances in genomics and informatics have given rise to a powerful new integrative method — the systems genetics approach — to study behavioral variation. In this approach, behavior is systematically perturbed by thousands, or even millions, of sequence variants segregating among large cohorts of genetically distinct individuals. This variation can be profiled at many levels and used to identify linkage between gene variants, and progressively higher order molecular, cellular, physiological, and behavioral networks. Systems genetics has already been applied to several areas of biology, and for an outstanding introduction to the field see [1••]. In this review we focus on some of the first uses of these methods to study behavior.

Behavior is a complex product of genetic, developmental, and environmental factors. As a result, even seemingly straightforward and essential behaviors, such as feeding in invertebrates, are surprisingly complex and contingent on environmental conditions [2••,3••]. Over the last two decades, gene mapping studies across a menagerie of animal populations and human cohorts have demonstrated that most traits, including behavior, are modulated by many independent loci that individually account for well under 1–5% of trait variability [4]. The underlying variants modulating behavior often exhibit overlapping and multifactorial effects (pleiotropy) and complex non-linear interactions (epistasis) [5••,6•,7••,8•,9•]. This complex genetic architecture of behavior has often defied a reductionist approach; that is, attempting to dissect system function by analyzing individual components in isolation. For behavioral genetics this is often achieved by the use of single gene manipulations such as transgenic lines or viral-mediated manipulations. The problem with this approach is that there is a serious risk that results will not generalize across genetic backgrounds or environments [10••] or, even worse, will be due to off-target effects, hitchhiking genes [11], or compensation [12]. Some of these risks are revealed by an elegant study of the genetics of epilepsy by Klassen and colleagues [13••]. They sequenced more than 237 ion channels, many already linked to abnormalities in neuronal activity in epileptics and matched controls. Despite sophisticated analysis of coding variants in ion channel genes, epileptic cases and normal controls were almost indistinguishable and many variants strongly linked to seizure were also surprisingly common in normal subjects. Although it has been the norm in behavioral research, candidate gene studies have not been as effective at predicting individual differences in behavior as we had hoped. In contrast, a more holistic systems approach can reveal patterns of interactions among the constituents that underlie and drive behavior. However, this systems genetics approach usually involves complex genetic and experimental designs and typically requires greater resources than reductionist methods. Fortunately, technical improvements, including fast and inexpensive genotyping and sequencing, high throughput and high content behavioral assays [14], and large multicenter collaborations [15•,16,•17] now make a systems genetics approach to studying complex behavior much more practical.

## Exploiting genetic variation

At the heart of every systems genetics study is a genetically diverse population in which variation at the level of the genome is accompanied by heritable variation at multiple levels — from transcriptome to phenome. Most current work takes advantage of animal populations derived by crossing a few inbred parental lines whose genomes have been fully sequenced [18–24]. Common

strategies to create these genetically diverse individuals from progenitors include backcrosses, F2 intercrosses, recombinant inbred lines, and outcrosses. In most cases, inbred parental strains of type A and B are mated to produce genetically homogenous F1 generations. The F1 progeny can then either be mated to one of the parents to produce a backcross or mated to isogenic siblings to create an F2 intercross. Diversity and genetic complexity can be enhanced even further in an outcross or by increasing the number of progenitor strains. This increases the number of parental alleles and the number of segregating genetic variants (also known as polymorphisms). Creating these more elaborate crosses is expensive, time-consuming, and generates unique progeny that cannot be resampled. Recombinant inbred (RI) strains circumvent this problem and allow for limitless reanalysis of individual genotypes. RIs are created from F2 or outcross populations by inbreeding progeny for 20 or more generations to full homozygozity. Families of RI strains and other similar genetic reference panels (GRPs) create a stable but diverse set of individuals which are particularly amenable to the systems genetics approach for the simple reason that data sets and phenotypes are cumulative. In addition to these strategies, a few studies have used a more targeted approach to create populations with mutations in a handful of candidate genes [9•].

Thanks to large-scale sequencing projects over the past two decades, the position and sequence of every gene and the location of common sequence variants are known for most model organisms. For example, the entire *Drosophila* GRP made up of ~200 isogenic lines has been sequenced [25,26], as have all the parental lines of major mouse and rat GRPs [16•,21,27]. Each GRP is highly polymorphic and, like wild and human populations, these GRPs are segregating up to 5 or 50 million single nucleotide polymorphisms (SNPs). Fully sequenced and well annotated genomes now enable far more efficient generation of genotypes, transcriptomes, and epigenomes, which are being mined to study behavior at multiple levels [16•,28•,29,30] or perform reverse genetic studies of behavior as demonstrated by Carniero [31], Li [32], Turner [33] and colleagues, and Anholt [34,35] for the serotonin transporter, catechol-O-methyltransferase, the arginine vasopressin receptor 1A, and odorant binding proteins, respectively.

## Multiscalar data generation

For each individual in a systems genetics study a well annotated and extensible stack of omics data at different levels — genotypes, RNA and protein measurements, metabolic profiles, epigenetic modifications, and behavioral responses — must be generated. Each of these traits needs to be acquired in a systematic way across large numbers of genetically defined subjects, often with replication across environment or treatments. This multiscalar genetic design enables linkages and associations to be made between sets of data. Association can be as simple as a correlation coefficient or can be more complex and causal and depend on shared genetic control as shown by gene mapping experiments.

Although still a bottleneck, behavioral profiling is becoming more automated and higher throughput. Increased use of automated tracking systems, improved computer speed and memory, and better algorithms translate to higher density behavioral phenotypes for each individual. In contrast to screening many different behaviors in one genetic background with a single gene mutation, as in the reductionist approach, the goal of systems genetics is to collect behavioral profiles across many individuals. This requires fast, reproducible, and reliable acquisition of behavioral responses using tests with high validity. The latter is especially important with increasing model organism complexity [36]. In zebrafish and nematode, live animals can be sorted into multi-well plates and kinematic behavior can be automatically tracked and analyzed under a variety of conditions [37,38]. Automated home cage tracking systems in rodent models [17,39,40] record behavior across circadian cycles and under different experimental conditions and generate massive amounts of data. Recent progress in *in vivo* brain imaging and recording [41–43] provides insight into the brain circuitry and connectivity underlying behavior. New touchscreen assays that measure cognitive behavior have high validity and translational relevance [44]. However, many of these techniques — especially optogenetics — are not yet well suited for probing variation in brain activity across populations and not all assays will be amenable for all model organisms.

Generating genotypes and expression data sets is now becoming easier than generating behavioral responses. This is due to automated, high throughput and low cost molecular assays. Array technology and mass spectrometry can provide quantitative omics data across the whole genome on large numbers of subjects. Next-generation methods are essential for high coverage sequencing and can be useful for joint genotyping and expression analysis. Current RNA-sequencing methods still struggle with low abundance transcript measurements and transcript isoform detection at a cost comparable to that of arrays [45], but technology is advancing at a rapid rate and longer read lengths and higher coverage at lower cost will improve this situation quickly. Similarly, proteomic methods are now much more sensitive, quantitative, and comprehensive, and can readily be integrated into systems studies [46•].

Acquiring deep and broad multiscalar data sets — even for small cohorts consisting of hundreds of individuals — can be a burdensome requirement of a systems approach. As a result there is strong motivation to assemble these omics resources for model organisms using genetically