New meta-analysis tools reveal common transcriptional regulatory basis for multiple determinants of behavior

Seth A. Ament1,2, Charles A. Blatti3,1, Cedric Alaux4,5,1, Marsha M. Wheeler6, Amy L. Toth7, Yves Le Conte8, Greg J. Hunt1, Ernesto Guzmán-Novoa9, Gloria DeGrandi-Hoffman9, Jose Luis Uribe-Rubio10, Gro V. Amdam1,11, Robert E. Page, Jr.,1 Sandra L. Rodriguez-Zas1,1, Gene E. Robinson1,5,1, and Saurabh Sinha1,3,1

*Neuroscience Program, Departments of 1Computer Science, 2Entomology, and 3Animal Sciences, and 4Institute for Genomic Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801; 5Unité de Recherche (UR) 406 Abeilles et Environnement, Institut National de la Recherche Agronomique (INRA), Avignon cedex 9, France; 6Department of Entomology, Purdue University, West Lafayette, IN 47907; 7Department of Environmental Biology, University of Guelph, Guelph, ON, Canada N1G 2W1; 8Carl Hayden Bee Research Center, US Department of Agriculture–Agricultural Research Service, Tucson, AZ 85719; 9Centro Nacional de Investigación en Fisiología Animal, Instituto Nacional de Investigaciones Forestales, Agrícolas y Pecuarias, Ajuchitlán, Qro 76280, Mexico; 10School of Life Sciences, Arizona State University, Tempe, AZ 85287; and 11Department of Chemistry, Biotechnology and Food Science, Norwegian University of Life Sciences, 1432 Aas, Norway

AUTHOR SUMMARY

Complex phenotypes are shaped by a multitude of heritable and environmental determinants (1). A key challenge is to learn whether different determinants of a phenotype are subserved by common or independent molecular machinery. Because many heritable and environmental determinants influence gene expression (2), it is logical to search for shared mechanisms at the level of transcriptional regulation, which leads to the following question: do the transcriptional regulatory networks used by apparently distinct determinants of a phenotype rely on shared components or modules, and if so, what are they? We developed tools to address this question, Metalysis and cis-Metalysis (Fig. P1). We applied these tools to brain gene expression profiles from ~400 individual honey bees (Apis mellifera), revealing potential transcriptional regulatory mechanisms common to 11 different determinants of a complex behavioral trait: the age at which worker bees switch from hive work to foraging.

Current informatics and statistical tools are, to our knowledge, lacking in rigorously addressing this question. Despite substantial literature (3, 4), there is apparently no rigorous treatment of the following intuitive task: for sets of differentially expressed genes derived from related experimental conditions, determine if a given gene module or cis-regulatory motif shows significant statistical association with the gene sets in all or most of the experimental conditions studied.

Metalysis employs an analytical approach to search for biological properties that are shared among differentially expressed genes from multiple experiments. cis-Metalysis identifies statistically significant relationships between single or combinations of predefined cis-regulatory elements and multiple patterns of gene expression. It combines the statistical power of the Metalysis framework with state-of-the-art binding site prediction algorithms that have been successful in annotating several arthropod and vertebrate genomes. Metalysis and cis-Metalysis are broadly relevant to any study where a complex phenotype is probed at the transcriptomic level under a variety of conditions, an increasingly common approach in systems biology.

Fig. P1. Metalysis and cis-Metalysis reveal shared transcriptional regulatory mechanisms for determinants of honey bee behavioral maturation. (A) The cis-Metalysis procedure. Step 1: Scan promoters of genes (G) for conserved cis-regulatory motifs (M). Step 2: Compute statistical associations between motif presence and differential gene expression across multiple conditions (C). Step 3: Combine P values to characterize meta-associations across some or all conditions. (B) Combinations of cis-regulatory motifs shared by 11 determinants of fast vs. slow honey bee behavioral maturation (from hive work to foraging). A wiring diagram shows the most significant combinatorial rules involving motifs recognized by the transcription factors mtf7-A and USF. The roles of individual motifs differed between conditions. “And” and “not” logic operations are depicted by rounded rectangles and triangles, respectively.


The authors declare no conflict of interest. This is a Contributed submission.

Freely available online through the PNAS open access option.

Data deposition: The microarray datasets reported in this paper have been deposited in the ArrayExpress Archive, http://www.ebi.ac.uk/arrayexpress (accession nos. E-TABM-953, E-TABM-952, E-MEXP-3079, E-MTAB-507, and E-MTAB-490).

1To whom correspondence may be addressed. E-mail: generobi@illinois.edu or sinhas@illinois.edu.

See full research article on page E1801 of www.pnas.org.

Cite this Author Summary as: PNAS 10.1073/pnas.1205283109.
We illustrate here the insights that the tools deliver with an analysis of honey bee behavioral maturation. Adult worker honey bees shift from hive work to foraging outside for nectar and pollen midway through their 4- to 6-wk lifespan. Many genotypic, environmental, and endocrine determinants of the age at onset of foraging (2) are known, but whether (and to what extent) these determinants influence common molecular mechanisms to exert their shared effects on behavior is not known. Fig. P1B shows Boolean rules by which the presence or absence of two cis-regulatory motifs (recognized by the transcription factors mTTF-A and USF) together predict gene expression responses to 11 known determinants of behavioral maturation. cis-Metalysis revealed 16 such motif combinations. In all cases, motifs were associated with fast maturation in some conditions and slow maturation in others (as indicated at right in Fig. P1B), suggesting that shared transcription factors are used in distinct ways by different determinants. Many of these motifs are recognized by transcription factors that are known to mediate cellular processes already linked to maturation [e.g., neuronal plasticity (Creb), energy metabolism (mTTF-A), immunity (NF-κB), and juvenile hormone signaling (Broad)]. Our findings go beyond prior demonstrations of the strong relationship between behavioral plasticity and brain gene expression, and they provide a glimpse into its complex, combinatorial regulatory basis.

The tools developed here represent a major technological advance in the meta-analysis of gene expression datasets. They rely on a statistic (the meta-P value) that we derived to address a fundamental statistical problem: testing the significance of the aggregate of a number of statistical tests. This statistical problem, previously tackled by Fisher, among others (5), has become very relevant today in the fields of genomics and evolutionary biology. Compared with the few published approaches to this problem, our statistic is more sensitive to the scenario where only a subset of the aggregated tests carries evidence for a metalevel association. Another major advance afforded by cis-Metalysis is that it searches for sets of cis-regulatory motifs determined by multiple experimental conditions while being flexible about precisely how the motif set correlates with expression in different conditions. This finding allows us to determine common elements of the regulatory networks underlying distinct transcriptomic states while recognizing the fact that the common elements may be wired differently from one condition to another.

In summary, we solved a fundamental statistical question in meta-analysis, built on it to develop informatics tools that analyze sequence and expression data, and used these tools to reveal a flexible cis-regulatory code underlying a complex behavioral trait: honey bee behavioral maturation. As new technologies make large-scale transcriptomic studies ever more affordable, our tools will help scientists conduct integrative analyses of the multitude of datasets that they are generating and guide them to biological insights that can be revealed only through systematic meta-analysis.