

Submitter Name: Gary Peltz  
Submitted email: [gpeltz@stanford.edu](mailto:gpeltz@stanford.edu)

## **Computational Genetic Analysis of 8300 Biomedical Phenotypes and of Responses to Drugs of Abuse**

Ahmed Arslan<sup>1</sup>, Yuan Guan<sup>1</sup>, Xinyu Chen<sup>1</sup>, Robin Donaldson<sup>2</sup>, Wan Zhu<sup>1</sup>, Madeline Ford<sup>1</sup>, Manhong Wu<sup>1</sup>, Ming Zheng<sup>1</sup>, David L. Dill<sup>2</sup>, S.M. Mooney-Leber<sup>3</sup>, L. R. Seemiller<sup>3</sup>, L. R. Goldberg<sup>3</sup>, P. Smith<sup>4</sup>, A. Patterson<sup>5</sup>, T. J. Gould<sup>3</sup>, David Jentsch<sup>6</sup> and Gary Peltz<sup>1</sup>

<sup>1</sup>Department of Anesthesiology, Perioperative, and Pain Medicine, Stanford University School of Medicine; <sup>2</sup>Department of Computer Science, Stanford University; <sup>3</sup>Department of Biobehavioral Health, Penn State University; <sup>4</sup>Eberly College of Science, Penn State University; <sup>5</sup>Department of Veterinary and Biomedical Sciences, Penn State University; <sup>6</sup>Department of Psychology, SUNY-Binghamton University

Haplotype-based computational genetic mapping (HBCGM) is a genetic analysis method that has identified genetic factors affecting many biomedical traits in mice. It previously was used to analyze one dataset at a time; but now, a vast amount of mouse phenotypic data is available. Therefore, HBCGM and a whole genome SNP map covering 43 inbred strains was used to analyze 8300 publicly available datasets of biomedical responses that were measured in panels of inbred mouse strains. Causative genetic factors affecting susceptibility for eye, metabolic and infectious diseases were identified when automated methods were used to analyze the output. One analysis identified a novel genetic effector mechanism; allelic differences within the mitochondrial targeting sequence affected the subcellular localization of a protein. We also found allelic differences within the mitochondrial targeting sequence of many murine and human proteins, and these could affect a wide range of biomedical phenotypes. These initial results indicate that genetic factors affecting biomedical responses could be identified through analysis of very large datasets and provide an early indication of how artificial intelligence-like methodology can facilitate genetic discovery. These methods were also used to analyze inbred strain responses to cocaine and nicotine. For example, male mice from 21 inbred strains received either chronic saline or nicotine and were tested for learning deficits using hippocampal-dependent contextual fear conditioning. There markedly different responses in commonly used inbred mouse strains, were analyzed using the above methodology to identify novel genetic factors affecting nicotine withdrawal.