

# **The Center for Predictive Computational Phenotyping**

**The University of Wisconsin – Madison**

PI: Mark W. Craven

Grant Number: 1-U54AI117924-01

The biomedical sciences are being radically transformed by advances in our ability to monitor, record, store and integrate information characterizing human biology and health at scales that range from individual molecules to large populations of subjects. This wealth of information has the potential to substantially advance both our understanding of human biology and our ability to improve human health. Perhaps the most central and general approach for exploiting biomedical data is to use methods from machine learning and statistical modeling to infer predictive models. Such models take as input observable data representing some object of interest, and produce as output a prediction about a particular, unobservable property of the object. This approach has proven to be of high value for a wide range of biomedical tasks, but numerous significant challenges remain to be solved in order for the full potential of predictive modeling to be realized. To address these challenges, we propose to establish The Center for Predictive Computational Phenotyping (CPCP). Our proposed center will focus on a broad range of problems that can be cast as computational phenotyping. Although some phenotypes are easily measured and interpreted, and are available in an accessible format, a wide range of scientifically and clinically important phenotypes do not satisfy these criteria. In such cases, computational phenotyping methods are required either to (i) extract a relevant phenotype from a complex data source or collection of heterogeneous data sources, (ii) predict clinically important phenotypes before they are exhibited, or (iii) do both in the same application. PUBLIC HEALTH RELEVANCE: We will develop innovative new approaches and tools that are able to discover, and make crucial inferences with large data sets that include molecular profiles, medical images, electronic health records, population-level data, and various combinations of these and other data types. These approaches will significantly advance the state of the art in wide range of biological and clinical investigations, such as predicting which patients are most at risk for breast cancer, heart attacks and severe blood clots.