Bioinformatics

RTI International's interdisciplinary bioinformatics program advances research by combining expertise in genomics, metagenomics, biomedical and evolutionary biology, and computational biology. Our scientists have expertise in the development of ontologies, analysis of genome-wide association studies (GWAS), data management, data standards, DNA sequence annotation, software development, and high-performance computing.

Overview

Bioinformatics applies computer science to basic and applied problems in life sciences research—including epidemiological, genomic, clinical, and translational studies. RTI's bioinformatics program produces innovative bioinformatics analysis and new software tools; the program's scientists have expertise in areas such as genomics, toxicology, data management, tool development, and software engineering. We frequently work with other collaborators to develop resources and bioinformatics tools that enable researchers to better understand and address complex research questions.

The group develops elegant systems that make research data FAIR (findable, accessible, interoperable, and reusable) by using our deep commitment to transparency in data handling and analysis as well as in scientific reproducibility. The work we do supports collaborative, cross-disciplinary research relevant to human health and disease.

Demonstrated Expertise

PhenX (Consensus Measures for Phenotypes and eXposures) (National Human Genome Research Institute [NHGRI], with co-funding from the National Institute on Drug Abuse, 2007–present). Under a cooperative agreement with NHGRI, RTI facilitates the expert consensus process and develops the online PhenX Toolkit (www.phenxtoolkit.org) to improve the consistency of data collection, facilitate cross-study analyses, and help investigators identify opportunities for collaborative biomedical research. The Toolkit currently contains 480 measures spanning 23 research domains. Initially focused on measures suitable for GWAS and common, complex diseases, the Toolkit has expanded to include measures for rare genetic conditions and measures suitable for a variety of study types. Toolkit users can browse or search the resource, add measures of interest to their Toolkit, and access tools to help integrate PhenX measures into their study design. Users can also request custom reports and data dictionaries compatible with the database of Genotypes and Phenotypes (dbGaP) and Research Electronic Data Capture (REDCap).

Molecular Atlas of Lung Development (LungMAP) Data Coordinating Center (National Heart, Lung, and Blood Institute, 2014–2019). The LungMAP consortium consists of four research centers (RCs), a human tissue core (HTC), and a data coordinating center (DCC). Its goal is to improve lung health by providing the research community with the data and tools necessary to support advanced investigations into the processes that regulate lung development. LungMAP researchers apply

innovative imaging and "-omics" technologies to a broad array of cell types in mice and humans to generate a novel map of where and when lung cells differentiate and alveoli form. RTI provides much of the technical expertise for the LungMAP DCC, leading development of the triple-store database that houses data generated by the LungMAP RCs and HTC and of web-based tools for viewing and interacting with the data. The program also coordinates data transfer, processing, loading, and retrieval. LungMAP makes it possible for researchers to consistently annotate images, which are a key component of the resource.

Ontology-based Application Development. We have expertise in development of ontology-based tools building on Open Biomedical Ontologies standards, Web Ontology Language (OWL), and automated reasoning. We apply semantic knowledge representation approaches to support computational data integration and knowledge discovery, inferring links across biomedical disease models, gene functions, environmental exposures, and phenotypic diversity. With the **Gene Ontology Consortium (National Institutes of Health [NIH]/NHGRI)**, we are developing tools supporting annotation of complex interactions between genes and biological processes. RTI's bioinformatics program leads software development for the **National Science Foundation–funded Phenoscape Knowledgebase** triple-store database and web application, which applies ontology-based approaches to make free-text descriptions of biodiversity more computable. As part of the **Biomedical Data Translator program (NIH/National Center for Advancing Translational Sciences)**, we will integrate environmental exposure data into a comprehensive, relational, N-dimensional Biomedical Data Translator that integrates multiple types of existing data sources, including objective signs and symptoms of disease, drug effects, and intervening types of biological data relevant to understanding pathophysiology.

Integrative "-Omics" (2006–present). Over a series of NIH-funded grants (e.g., R01 DA038632 and R01 DA042090), we conduct GWAS, integrate GWAS results with genome-wide RNA expression and DNA methylation data, and leverage numerous bioinformatics tools (e.g., HaploReg) and publicly available data (GTEx) to make novel discoveries of biological drivers of human disease. Starting with sequential integration of multiple omics, we move from GWAS discovery to putative mechanisms for gene variant association with diseases (e.g. CHRNA4 and nicotine dependence) and, in reverse direction, from gene regulation to functional variant associations with diseases (e.g., OPRM1 and heroin addiction). Current projects include concurrent analyses of multiple omics in disease-relevant tissue (e.g., Orthogonal Projections to Latent Structures) and functionally informed fine-mapping of disease associated loci (e.g., PAINTOR). The bioinformatics program works with statisticians and substantive researchers to bring the power of informatics to understanding human disease.

GxG Interactions in Genetics of Nicotine Dependence (2013–present). GWAS conducted by our investigators and others have found unequivocal evidence that variants in nicotinic acetylcholine receptor (e.g., CHRNA5 and CHRNA4) and metabolism genes are associated with risk of nicotine dependence. To better understand the functional importance of these and other nicotine dependence-associated variants, we are carrying out genome-wide studies that harness variant-by-variant as well as variant-by-sex interactions (NIH-funded grant number R01 DA035825). We are also comparing genome-wide DNA methylation and RNA expression in postmortem human brain samples from smokers versus nonsmokers to uncover gene regulatory mechanisms that underlie nicotine dependence (R01 DA042090). En route to applying these novel approaches, we have assembled the largest ever GWAS meta-analysis, containing nearly 45,000 European and African American ancestry participants, which has led to the discovery of additional genetic variant associations with nicotine dependence.

More Information

David B. Rice
Senior Business Development Specialist
Research Computing Division
919.316.3912
drice@rti.org

RTI International
3040 E. Cornwallis Road, PO Box 12194
Research Triangle Park, NC 27709-2194 USA

RTI 10349 R4 0321