Context
Problem
Idea
Implementation
Performance
Discussion
Future

The R Genetics Project
Bioconductor foR Genetics

Gregory Warnes[1,2]   Scott Chasalow[4]   Giovanni Montana[5]
Michael O'Connell[6]   David Henderson[6]   Nitin Jain[1]   Weiliang Qiu[3]
Junsheng Cheng[7]   Ross Lazarus[3]

[1]Statistical Applications, Pfizer, Inc.
[2]Department of Computer Science, Yale
[3]Channing Laboratory, Harvard University
[4]Bristol-Myers Squibb
[5]Department of Mathematics, Imperial College
[6]Insightful, Inc.
[7]University of Chicago

Context
Problem
Idea
Implementation
Performance
Discussion
Future

Abstract

The R Genetics Project is a collaborative effort to develop a complete set of tools for storing, accessing, manipulating, and analyzing genetics data, from small candidate gene studies consisting of a few genetic markers to large whole genome studies containing hundreds of thousands of markers. The initial goal is to provide a foundation of efficient data structures and easy to use manipulation functions. We intend this foundation to allow methods developers to quickly and easily develop packages implementing their own techniques, while maintaining interoperability. This will reduce the burden on both method developers and applied data analysis, who must currently move data between numerous packages and data formats.

The foundation R Genetics packages, GENETICSBASE, has reached sufficient maturity for introduction to the R community. This talk will describe the R Genetics project, provide an outline of the data structures and features within GENETICSBASE. We will give a brief demo of some of these features, as well as mentioning several additional packages which are building upon this common base, including FBAT, and GENETICSPED.