Statistical Computing with Pathway Tools using RCyc

Tomer Altman taltman1@stanford.edu

Biomedical Informatics, Stanford University

R & BioConductor

- S: software community over 30 years of statistical computing, data mining, machine learning, and data visualization knowledge
- R: open-source S with a lazy Scheme interpreter at its heart (including closures, symbols, and even macros!)
- RCyc: an R package to allow the interaction between Pathway / Genome Databases and the wealth of biostatistics software in the R community

BioConductor



About Bioconductor

Bioconductor provides tools for the analysis and comprehension of highthroughput genomic data. Bioconductor uses the R statistical programming language, and is open source and open development. It has two releases each year, 610 software packages, and an active user community. Bioconductor is also available as an <u>Amazon</u> Machine Image (AMI).

Figure: BioConductor: Thousands of peer-reviewed biostatistics packages.

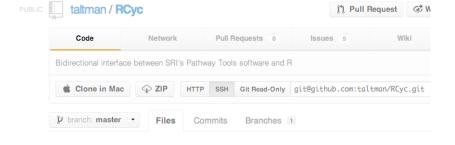
Software'R'-chitecture

- C code extension to R to allow Unix socket access
- Common Lisp code to hack in XML-based communication
 - Make the life of *Cyc API developers easier.
 - Currently supports exchange of numbers, strings, and lists
- R code and documentation
 - Provides utilities for starting PTools and marshaling data types
 - Assumes user is familiar with the PTools API: http://bioinformatics.ai.sri.com/ptools/api/
- All wrapped up in R package
 Easily installs via standard command-line R interface

Simple Example

```
callPToolsFn("so",list("'meta"))
callPToolsFn("get-slot-value",list("'proton", "'common-name"))
callPToolsFn("get-class-all-instances",list("'|Reactions|"))
```

Availability



- http://github.com/taltman/RCyc
- Linked from PTools website

Next Steps

- Dynamic instantiation of API functions in R
 - Coming next release (coordination with BRG)
 - Make development of *Cyc APIs easier, less boilerplate code
- Frame to Object import/export
 Provide "RCelot" functionality to slurp Ocelot frames directly into R S4 reference objects for direct data access
- Support for more exotic data types
 Symbols, hash tables, arrays, structures, etc.
- Submission of package to BioConductor repository Rigorous vetting process

End

Questions?