Title: Bioconductor with Python, What else ?

Authors: <u>Laurent Gautier</u>

Affiliations: CBS/DMAC, Department of Systems Biology, Technical University of Denmark, laurent@cbs.dtu.dk

URL (project): `http://pypi.python.org/pypi/rpy2-bioconductor-extensions/`

URL (downloads): `http://pypi.python.org/pypi/rpy2-bioconductor-extensions/`

License: AGPLv3.0

Abstract: The Bioconductor project has become a reference for the numerical processing and statistical analysis of data coming from high-throughput assays, providing a rich selection of methods and algorithms to the research community. Within the same time, Python has matured as a reliable platform for prototype development and data handling, with for example the biopython project providing the later for bioinformatics.

Building atop a bridge that allows the use of the R libraries from Python, we present an interface to Bioconductor data structures and function that uses an object-oriented paradigm familiar to Python users. This allows a seamless integration of the bioconductor project into existing Python infrastructure, be it for bioinformatics or web development with frameworks or GUI development.

The current implementation proposes Python-class representations for a large fraction of the core bioconductor infrastructure packages for the analysis of microarray and next-generation sequencing, and any other bioconductor package, or R package, is otherwise available through the underlying Python-R bridge *rpy2*. The design philosophy adopted allows to follow up with bioconductor as it continues developping, and we demonstrate with examples the benefits of the approach. In the face of exploding volumes for biological data generated, agile development can help link skills such as data analysis, data representation, user interfaces and prototype development, and we believe that Python will continue growing in that space.

Taking Python as a possible glue language, and now embedding what is probably the largest library of functions for data analysis and statistics in bioinformatics data, we outline perspectives for bioinformatics analysis frameworks in the age of omnipresent data.