

Bioconductor Annual Report 2010

Patrick Aboyoun
Fred Hutchinson Cancer Research Center

July 2010

Contents

1	Summary of Core Tasks and Challenges	1
1.1	Automated package building and testing	1
1.2	Package submission management	2
1.3	Annotation data package building	2
1.4	Other Tasks	3
2	Size of Project	3
3	Bioconductor Electronic Mail Lists	3
4	The Bioconductor Web Site	4
5	Package Building and Testing	5
6	Accomplishments	5
6.1	Papers Citing Bioconductor	5
6.2	Bioconductor Courses	7
6.3	BioC2009 Conference	8
7	Project Participants and Key Personnel	8

1 Summary of Core Tasks and Challenges

1.1 Automated package building and testing

The Bioconductor project provides access to its packages through package repositories hosted on `bioconductor.org`. One of the services provided to

the Bioconductor community is the automated building and testing of all packages.

Maintaining the automated build and test suite and keeping the published package repositories updated requires a significant amount of time on the part of the Seattle Bioconductor team. As the project has grown, the organizational and computational resources required to sustain the package build system have also increased.

1.2 Package submission management

The Bioconductor project relies on a peer-review process of candidate package add-ons to ensure it grows containing high-quality, scientifically-relevant software. It has achieved a virtuous cycle, where its success has brought in new scientific software developers, and they, in turn, have been contributing more and more to the Bioconductor project.

The Seattle Bioconductor team has been spending a considerable amount of time managing new contributions by previewing the software for quality, managing peers during the review process to ensure scientific relevance, and communicating with the software developers on what steps need to be taken for their contribution to be included within Bioconductor. From July, 2009 – June, 2010, over 73 software packages add-ons have been managed by the Seattle Bioconductor team, of which over 60 have been accepted for inclusion in Bioconductor.

1.3 Annotation data package building

The Bioconductor project synthesizes genomic and proteomic information available in public data repositories in order to annotate the probes of standard microarray chips. These annotation data packages are made available to the community and allow Bioconductor users to easily access meta data relating to their experimental platform. We maintain automated tools to parse the available information. Due to quickly changing data standards, the maintenance of the code used to produce the annotation packages requires constant attention.

Work during the recent release cycles has focused on flexible approaches to transitioning from gene-level annotations relevant for expression arrays to genome coordinate annotations that form the basis of sequence-based annotations.

1.4 Other Tasks

In addition to the tasks listed above, the Seattle Bioconductor team engages in the following auxiliary tasks:

1. Providing user and developer support on project mail lists.
2. Developing new functionality and improving architecture of key packages.
3. Orchestrating the Bioconductor releases that occur every six months.

2 Size of Project

The Bioconductor project is comprised of R packages contributed by a worldwide bioinformatics community. There are currently 268 active developers and 372 contributed packages in Bioconductor’s development repository. The project also maintains 501 annotation data packages that aid in the analysis of data from microarray and sequencing experiments. Table 1 tracks the growth of the project over the semi-annual releases.

Release	1.0	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9
Package Count	15	20	30	49	81	100	123	141	172	188
Release	2.0	2.1	2.2	2.3	2.4	2.5	2.6			
Package Count	214	233	260	294	320	352	389			

Table 1: Number of contributed packages included in each of the Bioconductor releases. Releases occur twice per year.

3 Bioconductor Electronic Mail Lists

The project maintains three email lists, `bioconductor`¹, `bioc-devel`², and `bioc-sig-sequencing`³.

¹<http://www.stat.math.ethz.ch/mailman/listinfo/bioconductor>

²<http://www.stat.math.ethz.ch/mailman/listinfo/bioc-devel>

³<http://www.stat.math.ethz.ch/mailman/listinfo/bioc-sig-sequencing>

1. The `bioconductor` list is a forum for user questions, project announcements, and general discussion of interest to the Bioconductor community. As of July, 2010 the list has 2441 subscribers (individuals who receive mail from the list).
2. The `bioc-devel` list is a forum for package contributors' questions and discussion relating to the development of Bioconductor packages. As of July, 2010 this list has 540 subscribers.
3. The `bioc-sig-sequencing` list is a forum for discussing the management and analysis of high-throughput short read data such as that from Solexa or 454 technologies. As of July, 2010 this list has 621 subscribers.

All lists provide a means of disseminating project news and a space for members of the community to share their knowledge about use of Bioconductor packages and best practices for data analysis.

Table 2 lists the number of posts and number of unique authors as a monthly average over the past eight years.

Year	Posts/month	Authors/month
2002	59	13
2003	231	47
2004	320	60
2005	353	61
2006	348	59
2007	432	75
2008	424	83
2009	450	86

Table 2: Monthly average number of posts and number of unique authors for the `bioconductor` mail list from January, 2002 – December, 2009.

4 The Bioconductor Web Site

Software packages on the Bioconductor web site, <http://bioconductor.org>, were downloaded by 69098 unique IP addresses between August, 2009 and July, 2010. This is up from 59038 unique IP addresses between August, 2008 and July, 2009. In an average month within this time frame, the web

site experienced over 19958 unique visitors and over 2360GB of downloaded content. The most active month during this period was May 2010, where the site served 3673GB of content, of which 3632GB (99%) corresponded to package downloads. This web site is hosted on a dual-Xeon 3.0GHz server with 2GB of RAM from Dell.

5 Package Building and Testing

The Bioconductor project is committed to providing packages for all computing platforms common in the bioinformatics community. We currently provide source packages that can be installed on Linux and most UNIX-like variants, as well as binary packages for Windows and OS X.

To ensure that packages are consistently documented, easy to install, and functioning properly, we run a nightly build during which we test all packages in the development repository. Table 3 provides details on the systems we currently have available for the nightly build.

Platform	CPU	RAM	Build Time (Hours)
Linux 64-bit	2x dual-core Xeon 3.00 GHz	8 GB	7h + 1h
Windows 32-bit	2x dual-core Xeon 3.00 GHz	3 GB	9h + 3h
Windows 64-bit	2x dual-core Xeon 3.00 GHz	3 GB	9h + 3h
OS X 32-bit	2x dual-core Xeon 2.00 GHz	3 GB	6h30 + 1h30

Table 3: Servers used to build and test Bioconductor packages along with the number of hours required for a build/test cycle of all software packages (first number) and all experiment data packages (second number).

6 Accomplishments

6.1 Papers Citing Bioconductor

Bioconductor has become a vital software platform for the worldwide genomic research community. As of July, 2010, Google Scholar notes there are 2375 scientific documents that cite the ground-breaking *Genome Biology* 2004 paper **Bioconductor: open software development for computational biology and bioinformatics**. The 2004 Bioconductor paper is the second most accessed article of all time from *Genome Biology*.

Bioconductor citations in leading scientific journals have increased from January, 2003 to July, 2010. Table 4 contains the results of PubMed searches for “bioconductor” over different time frames. It shows there have been at least 233 journal citations from January, 2003 to July, 2010, with nearly 50% (115) being made in *Bioinformatics*. A sample of 84 publications citing Bioconductor in 2009 or 2010 are listed in the bibliography of this report.

Publication	2003	2004	2005	2006	2007	2008	2009	2010*
<i>Bioinformatics</i>	3	8	13	16	23	21	20	11
Other	4	5	6	14	21	31	42	16
Total	7	13	19	30	44	52	62	27

Table 4: PubMed searches for “bioconductor” on publications from January, 2003 – July, 2010.

Software packages within the Bioconductor project also have been cited in leading scientific journals. Table 5 contains citations captured in July, 2010 by Google scholar for select Bioconductor packages. The citations are either for the artical with the associated PubMed ID or is a chapter in the book *Bioinformatics and Computational Biology Solutions Using R and Bioconductor*, Springer, ISBN 978-0-387-25146-2.

Bioconductor package	PMID	Citations (Google scholar)
vsn	12169536	780
limma	BioC Book Chapter	681
affy	14960456	540
globaltest	14693814	256
xcms	16448051	242
affycomp	14960458	224
limmaGUI	15297296	177
aCGH	16159913	166
biomaRt	16082012	163
G0stats	17098774	152
eisa	12689096	137
MassSpecWavelet	16820428	82
cellHTS2	16869968	67
altcdfenvs	15310390	63
tilingArray	16787969	59
affyilmGUI	16455752	56
made4	15797915	55
lumi	18467348	54
affyPLM	BioC Book Chapter	50
beadarray	17586828	50

Table 5: Citations for select Bioconductor software packages as captured by Google scholar in July, 2010.

6.2 Bioconductor Courses

Courses with significant input from key Bioconductor personnel have been held in the following worldwide locations in 2009:

1. **Introduction to R and Bioconductor** – Genentech – November 23-24, 2009.
2. **High throughput sequence analysis tools and approaches with Bioconductor** – Seattle, WA – November 18-20, 2009.
3. **Bioconductor Basics** – Boston, MA – October 29-30, 2009.
4. **Gene expression based on sequencing technologies** – Copenhagen, Denmark – August 24-25, 2009.

5. **Statistical analysis of gene expression data with R and Bioconductor** – Copenhagen, Denmark – August 17-21, 2009.
6. **BioC2009 User Conference** – Seattle, WA, July 27-29, 2009.
7. **CSAMA09 BioConductor Course** – Brixen, Italy – June 14-19, 2009.
8. **EMBL09 BioConductor Course** – Heidelberg, Germany – June 8-10, 2009.
9. **Using Bioconductor for High Throughput Data** – Vancouver, BC – May 31, 2009.
10. **Introductory Bioconductor** – Seattle, WA – April 27-29, 2009.
11. **Bioconductor Basics 2009** – Boston, MA – February 23-24, 2009.
12. **Intermediate Bioconductor: Analyzing Microarray and Sequencing Experiments** – Seattle, WA – January 21-23, 2009.

6.3 BioC2009 Conference

The Bioconductor team organizes an annual conference to highlight current Bioconductor developments and to provide a forum for discussing the use and design of software for analyzing data arising in biology with a focus on Bioconductor and genomic data.

The **BioC2009: Where Software and Biology Connect** conference was held in Seattle at the Fred Hutchinson Cancer Research Center on July 27–28, 2009. Over 90 scientists attended. The conference consisted of 10 talks from leading researchers in computational biology and 16 hands-on lab sessions presented by Bioconductor package developers. We also provided travel expense and conference fee scholarships for attending the BioC2009 conference to 7 students.

BioC2010 will take place in Seattle, July 29–30, 2010.

7 Project Participants and Key Personnel

These individuals, all working at the Fred Hutchinson Cancer Research Center (FHCRC) in Seattle, Washington, played a central role in executing project objectives during 2009 and 2010.

Martin Morgan Principle investigator.

Patrick Aboyoun Developer.

Marc Carlson Developer.

Herve Pages Developer.

Nishant Gopalakrishnan Developer (flow cytometry focus).

Chao-Jen Wong Developer (flow cytometry focus).

Additional collaborations, sub-contracts, and leadership roles involve the following individuals:

Robert Gentleman Genentech.

Vincent Carey Harvard Medical School.

Rafael Irizarry Johns Hopkins University School of Hygiene and Public Health.

Wolfgang Huber European Bioinformatics Institute.

Sean Davis National Institutes of Health.

James MacDonald University of Michigan.

References

- J. D. Allen, M. Chen, and Y. Xie. Model-Based Background Correction (MBCB): R Methods and GUI for Illumina Bead-array Data. *J Cancer Sci Ther*, 1:25–27, 2009.
- A. Ameer, A. Rada-Iglesias, J. Komorowski, and C. Wadelius. Identification of candidate regulatory SNPs by combination of transcription-factor-binding site prediction, SNP genotyping and haploChIP. *Nucleic Acids Res.*, 37:e85, Jul 2009.
- M. J. Aryee, J. A. Gutierrez-Pabello, I. Kramnik, T. Maiti, and J. Quackenbush. An improved empirical bayes approach to estimating differential gene expression in microarray time-course data: BETR (Bayesian Estimation of Temporal Regulation). *BMC Bioinformatics*, 10:409, 2009.
- A. L. Asare, Z. Gao, V. J. Carey, R. Wang, and V. Seyfert-Margolis. Power enhancement via multivariate outlier testing with gene expression arrays. *Bioinformatics*, 25:48–53, Jan 2009.

- B. Ballester, N. Johnson, G. Proctor, and P. Flicek. Consistent annotation of gene expression arrays. *BMC Genomics*, 11:294, 2010.
- N. L. Barbosa-Morais, M. J. Dunning, S. A. Samarajiwa, J. F. Darot, M. E. Ritchie, A. G. Lynch, and S. Tavare. A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data. *Nucleic Acids Res.*, 38:e17, Jan 2010.
- M. Barenco, E. Papouli, S. Shah, D. Brewer, C. J. Miller, and M. Hubank. rHVDM: an R package to predict the activity and targets of a transcription factor. *Bioinformatics*, 25:419–420, Feb 2009.
- A. L. Boulesteix and M. Slawski. Stability and aggregation of ranked gene lists. *Brief. Bioinformatics*, 10:556–568, Sep 2009.
- V. J. Carey and R. Gentleman. Interpreting genetics of gene expression: integrative architecture in Bioconductor. *Pac Symp Biocomput*, pages 380–390, 2009.
- V. J. Carey, A. R. Davis, M. F. Lawrence, R. Gentleman, and B. A. Raby. Data structures and algorithms for analysis of genetics of gene expression with Bioconductor: GGtools 3.x. *Bioinformatics*, 25:1447–1448, Jun 2009.
- B. S. Carvalho, T. A. Louis, and R. A. Irizarry. Quantifying uncertainty in genotype calls. *Bioinformatics*, 26:242–249, Jan 2010.
- R. Castelo and A. Roverato. Reverse engineering molecular regulatory networks from microarray data with qp-graphs. *J. Comput. Biol.*, 16:213–227, Feb 2009.
- T. Chiang and D. Scholtens. A general pipeline for quality and statistical assessment of protein interaction data using R and Bioconductor. *Nat Protoc*, 4:535–546, 2009.
- G. Csardi, Z. Kutalik, and S. Bergmann. Modular analysis of gene expression data with R. *Bioinformatics*, 26:1376–1377, May 2010.
- A. Cuadros-Inostroza, C. Caldana, H. Redestig, M. Kusano, J. Lisec, H. Pena-Cortes, L. Willmitzer, and M. A. Hannah. TargetSearch—a Bioconductor package for the efficient preprocessing of GC-MS metabolite profiling data. *BMC Bioinformatics*, 10:428, 2009.
- D. Diez, A. M. Wheelock, S. Goto, J. Z. Haeggstrom, G. Paulsson-Berne, G. K. Hansson, U. Hedin, A. Gabrielsen, and C. E. Wheelock. The use

of network analyses for elucidating mechanisms in cardiovascular disease. *Mol Biosyst*, 6:289–304, Feb 2010.

- M. Dondrup, S. P. Albaum, T. Griebel, K. Henckel, S. Junemann, T. Kahlke, C. K. Kleindt, H. Kuster, B. Linke, D. Mertens, V. Mittard-Runte, H. Neuweger, K. J. Runte, A. Tauch, F. Tille, A. Puhler, and A. Goemann. EMMA 2—a MAGE-compliant system for the collaborative analysis and integration of microarray data. *BMC Bioinformatics*, 10:50, 2009.
- A. Droit, C. Cheung, and R. Gottardo. rMAT—an R/Bioconductor package for analyzing ChIP-chip experiments. *Bioinformatics*, 26:678–679, Mar 2010.
- S. Durinck, P. T. Spellman, E. Birney, and W. Huber. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc*, 4:1184–1191, 2009.
- H. Dvinge and P. Bertone. HTqPCR: high-throughput analysis and visualization of quantitative real-time PCR data in R. *Bioinformatics*, 25:3325–3326, Dec 2009.
- G. Feng, P. Du, N. L. Krett, M. Tessel, S. Rosen, W. A. Kibbe, and S. M. Lin. A collection of bioconductor methods to visualize gene-list annotations. *BMC Res Notes*, 3:10, 2010.
- G. Finak, A. Bashashati, R. Brinkman, and R. Gottardo. Merging mixture components for cell population identification in flow cytometry. *Adv Bioinformatics*, page 247646, 2009.
- L. Folkersen, D. Diez, C. E. Wheelock, J. Z. Haeggstrom, S. Goto, P. Eriksson, and A. Gabrielsen. GeneRegionScan: a Bioconductor package for probe-level analysis of specific, small regions of the genome. *Bioinformatics*, 25:1978–1979, Aug 2009.
- H. Frohlich, O. Sahin, D. Arlt, C. Bender, and T. Beissbarth. Deterministic Effects Propagation Networks for reconstructing protein signaling networks from multiple interventions. *BMC Bioinformatics*, 10:322, 2009.
- R. Gaujoux and C. Seoighe. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics*, 11:367, Jul 2010.
- P. Geeleher, D. Morris, J. P. Hinde, and A. Golden. BioconductorBuntu: a Linux distribution that implements a web-based DNA microarray analysis server. *Bioinformatics*, 25:1438–1439, Jun 2009.

- C. S. Gillespie, G. Lei, R. J. Boys, A. Greenall, and D. J. Wilkinson. Analysing time course microarray data using Bioconductor: a case study using yeast2 Affymetrix arrays. *BMC Res Notes*, 3:81, 2010.
- D. Glez-Pena, R. Alvarez, F. Diaz, and F. Fdez-Riverola. DFP: a Bioconductor package for fuzzy profile identification and gene reduction of microarray data. *BMC Bioinformatics*, 10:37, 2009.
- J. J. Gosink, G. D. Means, W. A. Rees, C. Su, and H. A. Rand. Bridging the Divide between Manual Gating and Bioinformatics with the Bioconductor Package flowFlowJo. *Adv Bioinformatics*, page 809469, 2009.
- F. Hahne, N. LeMeur, R. R. Brinkman, B. Ellis, P. Haaland, D. Sarkar, J. Spidlen, E. Strain, and R. Gentleman. flowCore: a Bioconductor package for high throughput flow cytometry. *BMC Bioinformatics*, 10:106, 2009.
- S. Hochreiter, U. Bodenhofer, M. Heusel, A. Mayr, A. Mitterecker, A. Kasim, T. Khamiakova, S. Van Sanden, D. Lin, W. Talloen, L. Bijens, H. W. Gohlmann, Z. Shkedy, and D. A. Clevert. FABIA: factor analysis for bicluster acquisition. *Bioinformatics*, 26:1520–1527, Jun 2010.
- W. A. Hofmann, A. Weigmann, M. Tauscher, B. Skawran, T. Focken, R. Burman, L. U. Wingen, B. Schlegelberger, and D. Steinemann. Analysis of array-CGH data using the R and Bioconductor software suite. *Comp. Funct. Genomics*, page 201325, 2009.
- R. Jaksik, J. Polańska, R. Herok, and J. Rzeszowska-Wolny. Calculation of reliable transcript levels of annotated genes on the basis of multiple probe-sets in Affymetrix microarrays. *Acta Biochim. Pol.*, 56:271–277, 2009.
- A. Kauffmann and W. Huber. Microarray data quality control improves the detection of differentially expressed genes. *Genomics*, 95:138–142, Mar 2010.
- A. Kauffmann, R. Gentleman, and W. Huber. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25:415–416, Feb 2009a.
- A. Kauffmann, T. F. Rayner, H. Parkinson, M. Kapushesky, M. Lukk, A. Brazma, and W. Huber. Importing ArrayExpress datasets into R/Bioconductor. *Bioinformatics*, 25:2092–2094, Aug 2009b.

- D. J. Klinke and K. M. Brundage. Scalable analysis of flow cytometry data using R/Bioconductor. *Cytometry A*, 75:699–706, Aug 2009.
- K. G. Kugler, L. A. Mueller, and A. Graber. MADAM - An open source meta-analysis toolbox for R and Bioconductor. *Source Code Biol Med*, 5:3, 2010.
- M. Lawrence, R. Gentleman, and V. Carey. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, 25:1841–1842, Jul 2009.
- K. Lee, F. Hahne, D. Sarkar, and R. Gentleman. iFlow: A Graphical User Interface for Flow Cytometry Tools in Bioconductor. *Adv Bioinformatics*, page 103839, 2009.
- J. T. Leek. The tspair package for finding top scoring pair classifiers in R. *Bioinformatics*, 25:1203–1204, May 2009.
- S. Lin, P. Du, N. Jafari, and T. Ouchi. Using Free and Open-Source Bioconductor Packages to Analyze Array Comparative Genomics Hybridization (aCGH) Data. *Curr. Genomics*, 10:60–63, Mar 2009.
- K. Lo, F. Hahne, R. R. Brinkman, and R. Gottardo. flowClust: a Bioconductor package for automated gating of flow cytometry data. *BMC Bioinformatics*, 10:145, 2009.
- M. Lohse, A. Nunes-Nesi, P. Kruger, A. Nagel, J. Hannemann, F. M. Giorgi, L. Childs, S. Osorio, D. Walther, J. Selbig, N. Sreenivasulu, M. Stitt, A. R. Fernie, and B. Usadel. Robin: an intuitive wizard application for R-based expression microarray quality assessment and analysis. *Plant Physiol.*, 153:642–651, Jun 2010.
- J. C. Mar, Y. Kimura, K. Schroder, K. M. Irvine, Y. Hayashizaki, H. Suzuki, D. Hume, and J. Quackenbush. Data-driven normalization strategies for high-throughput quantitative RT-PCR. *BMC Bioinformatics*, 10:110, 2009.
- V. Martin-Requena, A. Munoz-Merida, M. G. Claros, and O. Trelles. PreP+07: improvements of a user friendly tool to preprocess and analyse microarray data. *BMC Bioinformatics*, 10:16, 2009.
- D. J. McCarthy and G. K. Smyth. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*, 25:765–771, Mar 2009.

- B. A. McGivney, S. S. Eivers, D. E. MacHugh, J. N. MacLeod, G. M. O’Gorman, S. D. Park, L. M. Katz, and E. W. Hill. Transcriptional adaptations following exercise in thoroughbred horse skeletal muscle highlights molecular mechanisms that lead to muscle hypertrophy. *BMC Genomics*, 10:638, 2009.
- B. H. Meacham, P. S. Nelson, and J. D. Storey. Supervised normalization of microarrays. *Bioinformatics*, 26:1308–1315, May 2010.
- K. H. Mellits, I. F. Connerton, M. F. Loughlin, P. Clarke, J. Smith, E. Dillon, P. L. Connerton, F. Mulholland, and C. J. Hawkey. Induction of a chemoattractant transcriptional response by a *Campylobacter jejuni* boiled cell extract in colonocytes. *BMC Microbiol.*, 9:28, 2009.
- Q. Mo and F. Liang. A hidden Ising model for ChIP-chip data analysis. *Bioinformatics*, 26:777–783, Mar 2010.
- M. Morgan, S. Anders, M. Lawrence, P. Aboyoun, H. Pages, and R. Gentleman. ShortRead: a bioconductor package for input, quality assessment and exploration of high-throughput sequence data. *Bioinformatics*, 25:2607–2608, Oct 2009.
- E. F. Murphy, G. J. Hooiveld, M. Muller, R. A. Calogero, and K. D. Cashman. The effect of trans-10, cis-12 conjugated linoleic acid on gene expression profiles related to lipid metabolism in human intestinal-like Caco-2 cells. *Genes Nutr*, 4:103–112, Jun 2009.
- H. Nie, P. B. Neerincx, J. van der Poel, F. Ferrari, S. Bicciato, J. A. Leunissen, and M. A. Groenen. Microarray data mining using Bioconductor packages. *BMC Proc*, 3 Suppl 4:S9, 2009.
- G. Pau, F. Fuchs, O. Sklyar, M. Boutros, and W. Huber. EBImage—an R package for image processing with applications to cellular phenotypes. *Bioinformatics*, 26:979–981, Apr 2010.
- R. D. Pearson, X. Liu, G. Sanguinetti, M. Milo, N. D. Lawrence, and M. Rattray. puma: a Bioconductor package for propagating uncertainty in microarray analysis. *BMC Bioinformatics*, 10:211, 2009.
- O. Pelz, M. Gilsdorf, and M. Boutros. web cellHTS2: a web-application for the analysis of high-throughput screening data. *BMC Bioinformatics*, 11:185, 2010.

- T. Phang. R and Bioconductor solutions for alternative splicing detection. *Hum. Genomics*, 4:131–135, Dec 2009.
- S. Pounds, C. Cheng, X. Cao, K. R. Crews, W. Plunkett, V. Gandhi, J. Rubnitz, R. C. Ribeiro, J. R. Downing, and J. Lamba. PROMISE: a tool to identify genomic features with a specific biologically interesting pattern of associations with multiple endpoint variables. *Bioinformatics*, 25:2013–2019, Aug 2009.
- J. Reimand, J. M. Vaquerizas, A. E. Todd, J. Vilo, and N. M. Luscombe. Comprehensive reanalysis of transcription factor knockout expression data in *Saccharomyces cerevisiae* reveals many new targets. *Nucleic Acids Res*, Apr 2010.
- N. Rieber, B. Knapp, R. Eils, and L. Kaderali. RNAither, an automated pipeline for the statistical analysis of high-throughput RNAi screens. *Bioinformatics*, 25:678–679, Mar 2009.
- M. E. Ritchie, B. S. Carvalho, K. N. Hetrick, S. Tavaré, and R. A. Irizarry. R/Bioconductor software for Illumina’s Infinium whole-genome genotyping BeadChips. *Bioinformatics*, 25:2621–2623, Oct 2009.
- M. D. Robinson, D. J. McCarthy, and G. K. Smyth. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26:139–140, Jan 2010.
- W. T. Rogers and H. A. Holyst. FlowFP: A Bioconductor Package for Fingerprinting Flow Cytometric Data. *Adv Bioinformatics*, page 193947, 2009.
- R. S. Savage, K. Heller, Y. Xu, Z. Ghahramani, W. M. Truman, M. Grant, K. J. Denby, and D. L. Wild. R/BHC: fast Bayesian hierarchical clustering for microarray data. *BMC Bioinformatics*, 10:242, 2009.
- R. B. Scharpf and I. Ruczinski. R classes and methods for SNP array data. *Methods Mol. Biol.*, 593:67–79, 2010.
- M. Schmidberger, E. Vicedo, and U. Mansmann. affyPara—a Bioconductor Package for Parallelized Preprocessing Algorithms of Affymetrix Microarray Data. *Bioinform Biol Insights*, 3:83–87, 2009.
- S. H. Shah and J. A. Pallas. Identifying differential exon splicing using linear models and correlation coefficients. *BMC Bioinformatics*, 10:26, 2009.

- D. Smedley, S. Haider, B. Ballester, R. Holland, D. London, G. Thorisson, and A. Kasprzyk. BioMart—biological queries made easy. *BMC Genomics*, 10:22, 2009.
- J. R. Stevens and G. Nicholas. metahtdep: meta-analysis of hierarchically dependent gene expression studies. *Bioinformatics*, 25:2619–2620, Oct 2009.
- E. Strain, F. Hahne, R. R. Brinkman, and P. Haaland. Analysis of High-Throughput Flow Cytometry Data Using plateCore. *Adv Bioinformatics*, page 356141, 2009.
- L. Ting, M. J. Cowley, S. L. Hoon, M. Guilhaus, M. J. Raftery, and R. Cavicchioli. Normalization and statistical analysis of quantitative proteomics data generated by metabolic labeling. *Mol. Cell Proteomics*, 8:2227–2242, Oct 2009.
- B. P. van Houte, T. W. Binsl, H. Hettling, and J. Heringa. CGHnormaliter: a Bioconductor package for normalization of array CGH data with many CNAs. *Bioinformatics*, 26:1366–1367, May 2010.
- M. van Iterson, P. A. 't Hoen, P. Pedotti, G. J. Hooiveld, J. T. den Dunnen, G. J. van Ommen, J. M. Boer, and R. X. Menezes. Relative power and sample size analysis on gene expression profiling data. *BMC Genomics*, 10:439, 2009.
- K. Vanhoutte, C. de Asmundis, A. Francesconi, J. Figysl, G. Steurs, T. Boussy, M. Roos, A. Mueller, L. Massimo, G. Paparella, K. Van Caelenberg, G. B. Chierchia, A. Sarkozy, P. B. Terradellas, and M. Zizi. Leaving out control groups: an internal contrast analysis of gene expression profiles in atrial fibrillation patients—a systems biology approach to clinical categorization. *Bioinformatics*, 3:275–278, 2009.
- M. D. Wilkerson and D. N. Hayes. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics*, 26:1572–1573, Jun 2010.
- Y. Yang, E. J. Kort, N. Ebrahimi, Z. Zhang, and B. T. Teh. Dual KS: Defining Gene Sets with Tissue Set Enrichment Analysis. *Cancer Inform*, 9:1–9, 2010.
- G. Yu, F. Li, Y. Qin, X. Bo, Y. Wu, and S. Wang. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics*, 26:976–978, Apr 2010.

- B. Zacher, P. F. Kuan, and A. Tresch. Starr: Simple Tiling ARRay analysis of Affymetrix ChIP-chip data. *BMC Bioinformatics*, 11:194, 2010.
- C. Zeller, H. Frohlich, and A. Tresch. A Bayesian network view on nested effects models. *EURASIP J Bioinform Syst Biol*, page 195272, 2009.
- J. D. Zhang and S. Wiemann. KEGGgraph: a graph approach to KEGG PATHWAY in R and bioconductor. *Bioinformatics*, 25:1470–1471, Jun 2009.
- Y. Zhang, J. Szustakowski, and M. Schinke. Bioinformatics analysis of microarray data. *Methods Mol. Biol.*, 573:259–284, 2009.
- Z. Zhang, D. L. Gasser, E. F. Rappaport, and M. J. Falk. Cross-platform expression microarray performance in a mouse model of mitochondrial disease therapy. *Mol. Genet. Metab.*, 99:309–318, Mar 2010.
- L. J. Zhu, C. Gazin, N. D. Lawson, H. Pages, S. M. Lin, D. S. Lapointe, and M. R. Green. ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. *BMC Bioinformatics*, 11:237, 2010.