

Bioconductor Annual Report 2009

Patrick Aboyoun
Fred Hutchinson Cancer Research Center

May 2009

Contents

1	Summary of Core Tasks and Challenges	2
1.1	Automated package building and testing	2
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	4
4	The Bioconductor Website	5
5	Package Building and Testing	5
6	Accomplishments	5
6.1	Papers Citing Bioconductor	5
6.2	Bioconductor Courses	6
6.3	Sponsorships	7
6.4	BioC2008 Conference	7
7	Project Participants and Key Personnel	7
7.1	Gentleman Lab Members	7
7.2	Harvard Medical School Members	8
7.3	European Bioinformatics Institute Members	8
7.4	Johns Hopkins University School of Hygiene and Public Health Members	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 June, 2008 – May, 2009, over 84 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.

In order to synthesize data from the various public repositories, we must maintain automated tools that can parse the available information. Due to quickly changing data standards, the maintenance of the code used to produce the annotation packages requires constant attention.

We are also focusing resources on the underlying storage mechanism used for the annotation data packages. New high-throughput technologies such as SNP and exon arrays require significantly larger annotation libraries; the infrastructure requires improvement to support work with these emerging technologies.

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 world-wide bioinformatics community. There are currently 212 active developers and 325 contributed packages in Bioconductor’s development repository. The project also maintains 385 annotation data packages that aid in the analysis of data from microarray 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					
Package Count	214	233	260	294	320					

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. The `bioconductor` list is a forum for user questions, project announcements, and general discussion of interest to the Bioconductor community. As of May, 2009 the list has 2046 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 May, 2009 this list has 479 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 May, 2009 this list has 416 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 seven 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*	429	108

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

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>

4 The Bioconductor Website

The Bioconductor website, <http://bioconductor.org>, averaged over 14080 unique visitors and over 1584GB of content per month in the year from May, 2008 to April, 2009. The most active month during this period was April 2009, where the site served 2267GB of content of which 2244GB (99%) corresponded to package downloads. The Biobase package was downloaded by 34354 unique IP addresses between June, 2008 and May, 2009. This website 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, Solaris, 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
Linux 32-bit	2x dual-core Xeon 2.80 GHz	4 GB	11h
Windows 32-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 May, 2009, Google Scholar notes there are 1462 scientific documents that cite the groundbreaking *Genome Biology*

2004 paper **Bioconductor: open software development for computational biology and bioinformatics**. These widespread citations has made the 2004 Bioconductor paper the second most accessed article of all time from *Genome Biology*.

Bioconductor citations in leading scientific journals have increased from January, 2003 to May, 2009. Table 4 contains the results of PubMed searches for “bioconductor” over different timeframes. It shows there have been at least 192 journal citings from January, 2003 to May, 2009, with a nearly 50% (95) being made in *Bioinformatics*. A sample of 76 publications citing Bioconductor in 2008 or 2009 are listed in the bibliography of this report.

Publication	2003	2004	2005	2006	2007	2008	2009*
<i>Bioinformatics</i>	3	8	13	16	23	21	11
Other	4	5	6	14	21	30	17
Total	7	13	19	30	44	51	28

Table 4: PubMed searches for “bioconductor” on publications from January, 2003 – May, 2009.

6.2 Bioconductor Courses

Bioconductor courses have been held in the following worldwide locations in 2008 and early 2009:

1. **Introductory Bioconductor** – Seattle, WA – April 27-29, 2009.
2. **Bioconductor Basics 2009** – Boston, MA – February 23-24, 2009.
3. **Intermediate Bioconductor: Analyzing Microarray and Sequencing Experiments** – Seattle, WA – January 21-23, 2009.
4. **Using Bioconductor for ChIP-seq experiments** – Seattle, WA – November 12-14, 2008.
5. **Using Bioconductor for Array Quality Assessment** – Riva del Garda, Italy – September 4, 2008.
6. **Analysis of Short Reads from High-Throughput Sequencing using R and Bioconductor** – Riva del Garda, Italy – September 4, 2008.

7. **Microarray Analysis with R & BioConductor** – Riverside, CA – August 28, 2008.
8. **BioC2008 Conference** - Seattle, WA – July 28-29, 2008.
9. **Introduction to R** – Riverside, CA – July 30, 2008.
10. **Laussane Bioconductor Developer Meeting** – Laussane, Switzerland – April 24-25, 2008.
11. **Boston Bioconductor Intermediate Training** – Boston, MA – March 5-7, 2008.
12. **Advanced R for Bioinformatics** – Seattle, WA – February 13-15, 2008.

6.3 Sponsorships

- We provided travel expense and conference fee scholarships for attending the BioC2008 conference to ??? BioC package developers and ??? students for a total of ??? scholarships.

6.4 BioC2008 Conference

The Gentleman Lab organized a 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 **BioC2008: Where Software and Biology Connect** conference was held in Seattle at the Fred Hutchinson Cancer Research Center on July 28–29, 2007. Over 100 scientists attended. The conference consisted of 10 talks from leading researchers in computational biology and 15 hands-on lab sessions presented by Bioconductor package developers.

BioC2009 will take place in Seattle, July 27–28, 2009.

7 Project Participants and Key Personnel

7.1 Gentleman Lab Members

These individuals, all working in the Gentleman Lab at the Fred Hutchinson Cancer Research Center in Seattle, Washington, played a central role in executing project objectives during 2008 and 2009.

Patrick Aboyoun Scientific programmer, build and test manager.

Marc Carlson Developer in charge of annotation data packages.

Martin Morgan Developer in charge of Biobase package.

Herve Pages Developer in charge of Biostrings package.

7.2 Harvard Medical School Members

Vincent Carey Co-investigator.

7.3 European Bioinformatics Institute Members

Wolfgang Huber Co-investigator.

7.4 Johns Hopkins University School of Hygiene and Public Health Members

Rafael Irizarry Co-investigator.

References

- 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.*, May 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.
- M. Barenco, E. Papouli, S. Shah, D. Brewer, C. J. Miller, and M. Hubank. rHVD: an R package to predict the activity and targets of a transcription factor. *Bioinformatics*, 25:419–420, Feb 2009.
- G. Barton, J. Abbott, N. Chiba, D. W. Huang, Y. Huang, M. Krznaric, J. Mack-Smith, A. Saleem, B. T. Sherman, B. Tiwari, C. Tomlinson, T. Aitman, J. Darlington, L. Game, M. J. Sternberg, and S. A. Butcher. EMAAS: an extensible grid-based rich internet application for microarray data analysis and management. *BMC Bioinformatics*, 9:493, 2008.

- J. M. Cairns, M. J. Dunning, M. E. Ritchie, R. Russell, and A. G. Lynch. BASH: a tool for managing BeadArray spatial artefacts. *Bioinformatics*, 24:2921–2922, Dec 2008.
- 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, J. Gentry, R. Sarkar, D. Gentleman, and S. Ramaswamy. SGDI: system for genomic data integration. *Pac Symp Biocomput*, pages 141–152, 2008.
- 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.
- R. Castelo and A. Roverato. Reverse engineering molecular regulatory networks from microarray data with qp-graphs. *J. Comput. Biol.*, 16:213–227, Feb 2009.
- L. S. Chen and J. D. Storey. Eigen-R2 for dissecting variation in high-dimensional studies. *Bioinformatics*, 24:2260–2262, Oct 2008.
- 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.
- T. Chiang, N. Li, S. Orchard, S. Kerrien, H. Hermjakob, R. Gentleman, and W. Huber. Rintact: enabling computational analysis of molecular interaction data from the IntAct repository. *Bioinformatics*, 24:1100–1101, Apr 2008.
- H. Cho, Y. J. Kim, H. J. Jung, S. W. Lee, and J. W. Lee. OutlierD: an R package for outlier detection using quantile regression on mass spectrometry data. *Bioinformatics*, 24:882–884, Mar 2008.
- V. T. Chu, R. Gottardo, A. E. Raftery, R. E. Bumgarner, and K. Y. Yeung. MeV+R: using MeV as a graphical user interface for Bioconductor applications in microarray analysis. *Genome Biol.*, 9:R118, 2008.
- L. Corradi, M. Fato, I. Porro, S. Scaglione, and L. Torterolo. A Web-based and Grid-enabled dChIP version for the analysis of large sets of gene expression data. *BMC Bioinformatics*, 9:480, 2008.

- P. J. De Groot, C. Reiff, C. Mayer, and M. Mller. NuGO contributions to GenePattern. *Genes Nutr*, 3:143–146, Dec 2008.
- M. Dondrup, S. P. Albaum, T. Griebel, K. Henckel, S. Jnemann, T. Kahlke, C. K. Kleindt, H. Kster, B. Linke, D. Mertens, V. Mittard-Runte, H. Neuweger, K. J. Runte, A. Tauch, F. Tille, A. Phler, and A. Goesmann. EMMA 2—a MAGE-compliant system for the collaborative analysis and integration of microarray data. *BMC Bioinformatics*, 10:50, 2009.
- C. C. dos Santos, D. Okutani, P. Hu, B. Han, E. Crimi, X. He, S. Keshavjee, C. Greenwood, A. S. Slutsky, H. Zhang, and M. Liu. Differential gene profiling in acute lung injury identifies injury-specific gene expression. *Crit. Care Med.*, 36:855–865, Mar 2008.
- P. Du, W. A. Kibbe, and S. M. Lin. lumi: a pipeline for processing Illumina microarray. *Bioinformatics*, 24:1547–1548, Jul 2008.
- S. Durinck. Pre-processing of microarray data and analysis of differential expression. *Methods Mol. Biol.*, 452:89–110, 2008.
- N. A. Ettinger and M. E. Wilson. Macrophage and T-Cell Gene Expression in a Model of Early Infection with the Protozoan *Leishmania chagasi*. *PLoS Negl Trop Dis*, 2:e252, 2008.
- J. M. Fleming, E. L. Long, E. Ginsburg, D. Gerscovich, P. S. Meltzer, and B. K. Vonderhaar. Interlobular and intralobular mammary stroma: genotype may not reflect phenotype. *BMC Cell Biol.*, 9:46, 2008.
- L. Folkersen, D. Diez, C. E. Wheelock, J. Z. Haeggstrm, S. Goto, P. Eriksson, and A. Gabrielsen. GeneRegionScan: a Bioconductor package for probe level analysis of specific, small regions of the genome. *Bioinformatics*, Apr 2009.
- J. Frelinger, T. B. Kepler, and C. Chan. Flow: Statistics, visualization and informatics for flow cytometry. *Source Code Biol Med*, 3:10, 2008.
- H. Frhlich, T. Beissrth, A. Tresch, D. Kostka, J. Jacob, R. Spang, and F. Markowetz. Analyzing gene perturbation screens with nested effects models in R and bioconductor. *Bioinformatics*, 24:2549–2550, Nov 2008a.
- H. Frhlich, M. Fellmann, H. Sltmann, A. Poustka, and T. Beissbarth. Estimating large-scale signaling networks through nested effect models with intervention effects from microarray data. *Bioinformatics*, 24:2650–2656, Nov 2008b.

- 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.
- D. Glez-Pea, R. Alvarez, F. Daz, 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. Goeman and U. Mansmann. Multiple testing on the directed acyclic graph of gene ontology. *Bioinformatics*, 24:537–544, Feb 2008.
- F. Hahne, A. Mehrle, D. Arlt, A. Poustka, S. Wiemann, and T. Beissbarth. Extending pathways based on gene lists using InterPro domain signatures. *BMC Bioinformatics*, 9:3, 2008.
- 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.
- A. D. Hershey, D. Burdine, C. Liu, T. G. Nick, D. L. Gilbert, and T. A. Glauser. Assessing quality and normalization of microarrays: case studies using neurological genomic data. *Acta Neurol. Scand.*, 118:29–41, Jul 2008.
- J. Hill, M. Hambley, T. Forster, M. Mewissen, T. M. Sloan, F. Scharinger, A. Trew, and P. Ghazal. SPRINT: a new parallel framework for R. *BMC Bioinformatics*, 9:558, 2008.
- H. M. Hsueh, H. C. Kuo, and C. A. Tsai. Multispectra CWT-based algorithm (MCWT) in mass spectra for peak extraction. *J Biopharm Stat*, 18:869–882, 2008.
- M. Hummel, R. Meister, and U. Mansmann. GlobalANCOVA: exploration and assessment of gene group effects. *Bioinformatics*, 24:78–85, Jan 2008.
- S. M. Hunter, F. C. Mansergh, and M. J. Evans. Optimization of minuscule samples for use with cDNA microarrays. *J. Biochem. Biophys. Methods*, 70:1048–1058, Apr 2008.
- R. Jaksik, J. Polaska, 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.*, May 2009.

- A. Kauffmann, R. Gentleman, and W. Huber. arrayQualityMetrics—a bioconductor package for quality assessment of microarray data. *Bioinformatics*, 25:415–416, Feb 2009.
- A. Kuhn, R. Luthi-Carter, and M. Delorenzi. Cross-species and cross-platform gene expression studies with the Bioconductor-compliant R package 'annotationTools'. *BMC Bioinformatics*, 9:26, 2008.
- N. Lama and M. Girolami. Vbmp: variational Bayesian Multinomial Probit Regression for multi-class classification in R. *Bioinformatics*, 24:135–136, Jan 2008.
- M. Lawrence, R. Gentleman, and V. Carey. rtracklayer: an R package for interfacing with genome browsers. *Bioinformatics*, May 2009.
- J. T. Leek. The tspair package for finding top scoring pair classifiers in R. *Bioinformatics*, 25:1203–1204, May 2009.
- S. M. Lin, P. Du, W. Huber, and W. A. Kibbe. Model-based variance-stabilizing transformation for Illumina microarray data. *Nucleic Acids Res.*, 36:e11, Feb 2008.
- 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, May 2009.
- R. Lu, G. C. Lee, M. Shultz, C. Dardick, K. Jung, J. Phetsom, Y. Jia, R. H. Rice, Z. Goldberg, P. S. Schnable, P. Ronald, and D. M. Rocke. Assessing probe-specific dye and slide biases in two-color microarray data. *BMC Bioinformatics*, 9:314, 2008.
- 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. Muoz-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.
- P. G. Martini, D. M. Taylor, J. Bienkowska, J. Jackson, G. McAllister, H. Keilhack, and R. K. Campbell. Comparative expression analysis of four breast cancer subtypes versus matched normal tissue from the same patients. *J. Steroid Biochem. Mol. Biol.*, 109:207–211, Apr 2008.

- M. N. McCall and R. A. Irizarry. Consolidated strategy for the analysis of microarray spike-in data. *Nucleic Acids Res.*, 36:e108, Oct 2008.
- D. J. McCarthy and G. K. Smyth. Testing significance relative to a fold-change threshold is a TREAT. *Bioinformatics*, 25:765–771, Mar 2009.
- 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.
- P. E. Meyer, F. Lafitte, and G. Bontempi. minet: A R/Bioconductor package for inferring large transcriptional networks using mutual information. *BMC Bioinformatics*, 9:461, 2008.
- H. Miao, L. Chen, S. M. Riordan, W. Li, S. Juarez, A. M. Crabb, T. J. Lukas, P. Du, S. M. Lin, A. Wise, O. A. Agapova, P. Yang, C. C. Gu, and M. R. Hernandez. Gene expression and functional studies of the optic nerve head astrocyte transcriptome from normal African Americans and Caucasian Americans donors. *PLoS ONE*, 3:e2847, 2008.
- C. Murie and R. Nadon. A correction for estimating error when using the Local Pooled Error Statistical Test. *Bioinformatics*, 24:1735–1736, Aug 2008.
- E. F. Murphy, G. J. Hooiveld, M. Mller, 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*, Mar 2009.
- J. Noirel, G. Sanguinetti, and P. C. Wright. Identifying differentially expressed subnetworks with MMG. *Bioinformatics*, 24:2792–2793, Dec 2008.
- M. J. Okoniewski and C. J. Miller. Comprehensive analysis of affymetrix exon arrays using BioConductor. *PLoS Comput. Biol.*, 4:e6, Feb 2008.
- A. P. Oron, Z. Jiang, and R. Gentleman. Gene set enrichment analysis using linear models and diagnostics. *Bioinformatics*, 24:2586–2591, Nov 2008.
- W. Raffelsberger, Y. Krause, L. Moulinier, D. Kieffer, A. L. Morand, L. Brino, and O. Poch. RReportGenerator: automatic reports from routine statistical analysis using R. *Bioinformatics*, 24:276–278, Jan 2008.

- 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.
- G. Rigaille, P. Hup, A. Almeida, P. La Rosa, J. P. Meyniel, C. Decraene, and E. Barillot. ITALICS: an algorithm for normalization and DNA copy number calling for Affymetrix SNP arrays. *Bioinformatics*, 24:768–774, Mar 2008.
- D. Sarkar, N. Le Meur, and R. Gentleman. Using flowViz to visualize flow cytometry data. *Bioinformatics*, 24:878–879, Mar 2008.
- D. Scholtens, T. Chiang, W. Huber, and R. Gentleman. Estimating node degree in bait-prey graphs. *Bioinformatics*, 24:218–224, Jan 2008.
- H. Schwender and K. Ickstadt. Empirical Bayes analysis of single nucleotide polymorphisms. *BMC Bioinformatics*, 9:144, 2008.
- S. H. Shah and J. A. Pallas. Identifying differential exon splicing using linear models and correlation coefficients. *BMC Bioinformatics*, 10:26, 2009.
- M. Slawski, M. Daumer, and A. L. Boulesteix. CMA: a comprehensive Bioconductor package for supervised classification with high dimensional data. *BMC Bioinformatics*, 9:439, 2008.
- 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.
- R. Tautenhahn, C. Bttcher, and S. Neumann. Highly sensitive feature detection for high resolution LC/MS. *BMC Bioinformatics*, 9:504, 2008.
- J. Toedling and W. Huber. Analyzing ChIP-chip data using bioconductor. *PLoS Comput. Biol.*, 4:e1000227, Nov 2008.
- K. Vanhoutte, C. de Asmundis, A. Francesconi, J. Figys1, G. Steurs, T. Boussy, M. Roos, A. Mueller, L. Massimo, G. Paparella, K. Van Caelenberg, G. B. Chierchia, A. Sarkozy, P. B. Y 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.
- W. Wang, B. Carvalho, N. D. Miller, J. Pevsner, A. Chakravarti, and R. A. Irizarry. Estimating genome-wide copy number using allele-specific mixture models. *J. Comput. Biol.*, 15:857–866, Sep 2008.

- T. Yates, M. J. Okoniewski, and C. J. Miller. X:Map: annotation and visualization of genome structure for Affymetrix exon array analysis. *Nucleic Acids Res.*, 36:D780–786, Jan 2008.
- C. Zeller, H. Frhlich, 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.
- K. Q. Zhu, G. J. Carrougher, O. P. Couture, C. K. Tuggle, N. S. Gibran, and L. H. Engrav. Expression of collagen genes in the cones of skin in the Duroc/Yorkshire porcine model of fibroproliferative scarring. *J Burn Care Res*, 29:815–827, 2008a.
- Y. Zhu, S. Davis, R. Stephens, P. S. Meltzer, and Y. Chen. GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus. *Bioinformatics*, 24:2798–2800, Dec 2008b.