

# Open access publishing

## A researcher's perspective

Karl Broman

Biostatistics & Medical Informatics  
University of Wisconsin–Madison

[kbroman.org](http://kbroman.org)  
[github.com/kbroman](https://github.com/kbroman)  
[@kwbroman](https://twitter.com/kwbroman)

These are slides for a talk I will give on 30 Oct 2014, at a symposium on scholarly publishing.

I'm a statistician. My research focuses on genetics, and most of my papers are in genetics journals.

So in commenting on open access, I'm focusing on scientific publications, and perhaps more narrowly, on the biological sciences.

# Access in action

## Interesting reference

- [8] Kang C. and Speller R. The effect of region of interest selection on dual energy x-ray absorptiometry emasurements of the calcaneus in 55 post-menopausal women. *The british Journal of Radiology*, 72:864–871, 1999.
- [9] The 1000 Genomes Project Consortium. A map of human genome variation from population scale sequencing. *Nature*, 467:1061–1073, 2010.
- [10] John C.V., Mark D.A., Eugene W.M., Peter W.L., Richard J.M., Granger G.S., and Hamilton O.S. The sequence of the human genome. *Science*, 291:1304–1351, 2001.
- [11] Schwartz D.C. and Waterman M.S. [New generations: sequencing machines and their computational challenges](#). *Journal of Computer Science and Technology*, 25(1):3–9, 2010.
- [12] Church D.M., Goodstadt L., Hillier L.W., Zody M.C., Goldstein S., She X., Bult C.J., Agarwala R., Cherry J.L., DiCuccio M., Hlavina W., Kapustin Y., Meric P., Maglott D., Birtle Z., Marques A.C., Graves T., Zhou S., Teague B., Potamouisis K., Churas C., Place M., Herschleb J., Runnheim R., Forrest D., Amos-Landgraf J., Schwartz D.C., Cheng Z., Lindblad-Toh K., Eichler E.E., and Ponting C.P. Lineage-specific biology revealed by a finished genome assembly of the mouse. *PLoS Biology*, 7.5:e1000112, 2009.
- [13] Tor D.W., Matthew C.K., Steven C.L., and John J. Increased sensitivity in neuroimaging analyses using robust regression. *Neuroimage*, 26:99–113, 2005.

I’ll begin with an illustration of what I mean by access.

Some time back, I was reading a manuscript and saw an article of interest.

# Access in action

## Google Scholar

The screenshot shows a Google Scholar search result for the article "New Generations: Sequencing Machines and Their Computational Challenges" by David C. Schwartz and Michael S. Waterman. The search bar at the top contains the text "New generations: sequencing machines and their computational challenges". The article is from the "Journal of Computer Science and Technology", January 2010, Volume 25, Issue 1, pp 3-9. The authors are listed as "DC Schwartz, Schwartz DC, challenges, J ... New Gener". The article is cited by 13. There is a "Buy now" button for \$39.95 / €34.95 / £29.95. A "Get Access" button is also visible. The abstract is displayed below the article information. On the right side, there is a "Share" section with social media icons for Facebook, Twitter, and LinkedIn, and an "Other actions" section with links for "Export citations", "Register for Journal Updates", "About This Journal", and "Reprints and Permissions".

Google Scholar

New generations: sequencing machines and their computational challenges

Scholar

New gene: DC Schwartz, Schwartz DC, challenges, J ... New Gener Cited by 13

[BOOK] Evolt DB Fogel - 20 ... If the proces fundamentally beyond our ov Cited by 2874

[HTML] How C Trapnell, SL ... As a practic the computing investigators, Cited by 142

Journal of Computer Science and Technology  
January 2010, Volume 25, Issue 1, pp 3-9

### New Generations: Sequencing Machines and Their Computational Challenges

David C. Schwartz, Michael S. Waterman

Purchase on Springer.com  
\$39.95 / €34.95 / £29.95 \* Buy now

\* Final gross prices may vary according to local VAT.

Get Access

#### Abstract

New generation sequencing systems are changing how molecular biology is practiced. The widely promoted \$1000 genome will be a reality with attendant changes for healthcare, including personalized medicine. More broadly the genomes of many new organisms with large samplings from populations will be commonplace. What is less appreciated is the explosive demands on computation, both for CPU cycles and storage as well as the need for new computational methods. In this article we will survey some of these developments and demands.

Share

Facebook Twitter LinkedIn

Other actions

- » Export citations
- » Register for Journal Updates
- » About This Journal
- » Reprints and Permissions

3

If I paste the article title into Google Scholar, I immediately find the paper and can go directly to the journal.

But I was sitting at home on my couch.

And they charge \$40 for a 7 page paper!

I could get the article through the UW Libraries web site, but it's a bit of a hassle.

# What's the deal with the prices?

Broman K, Speed T, Tigges M (1996) Estimation of antigen-responsive T cell frequencies in PBMC from human subjects. <i>J Immunol Meth</i> 198:119–132	\$39.95
Broman KW, Weber JL (1999) Method for constructing confidently ordered linkage maps. <i>Genet Epidemiol</i> 16:337–343	\$35.00
Broman KW, Feingold E (2004) SNPs made routine. <i>Nat Methods</i> 1:104–105	\$18.00
Broman KW (2005) Mapping expression in randomized rodent genomes. <i>Nat Genet</i> 37:209–210	\$18.00

4

I went back to some of my early papers, and found these outrageous prices.

\$18 for a 2-page paper?

I understand that the publishing industry has a long history of screwy pricing, but you'd have to be either desperate or stupid to pay this.

And for that 1999 Genetic Epidemiology article, published by Wiley, you have to register in order to find out that it's \$35 for just 24 hours of access.

# Access in action

There's also PubMed

The screenshot shows a PubMed search result for the article "New Generations: Sequencing Machines and Their Computational Challenges" by Schwartz DG and Waterman MS. The article is from the journal *J Comput Sci Technol*, 2010, 25(1):1-9. The abstract discusses the challenges of sequencing large genomes and the need for computational methods. A prominent banner for "Free in PMC" (PubMed Central) is overlaid on the right side of the page, indicating that the full-text article is available for free. Below the abstract, there are sections for "Grant Support", "LinkOut - more resources", "PubReader", and "Related citations in PubMed".

5

If I'd used PubMed rather than Google Scholar, I could have gotten to the published paper in just a few clicks, because the manuscript is in PubMed Central.

PubMed Central is only for federally-funded research, has a one year embargo, and (as here) might not include the published version of the paper.

PubMed Central is a good thing, but one generally can't wait a year, it's unfortunate that the published versions aren't always included, and from an author's point of view it can be a real hassle.

# Another example

PubMed Resources How To

PubMed is open possible, and the

Display Settings: C

J Dent Res, 2013 Jan; 92(1):32

**Clustering tooth surface: Biologically informative clustering of tooth surfaces**  
Shaffer JR, Feingold F, Wang X, Weyant R, McNell D, W. McNell

Department of Human

**Abstract**  
Dental caries affect uniformly across all ages. Therefore, we used a clustering algorithm to group tooth surfaces that differ in their posterior non-pit and non-fissure surface features. We replicated in a number of carious teeth (i.e., under genetic control) some cluster-based toothbrushing habit surfaces that are used in Health in Appalachia through hierarchical clustering representing the number of affected teeth. Survey, the second: PMID: 23064960 [PubMed]

Impact Factor: 3.826 | Ranking: 2/82 in Dentistry, Oral Surgery & Medicine | Eigenfactor Score: 0.02325 | 5-Year Impact Factor: 4.286

This item requires a subscription to Journal of Dental Research.

**Full Text (PDF)**  
**Clustering Tooth Surface: Biologically Informative C**  
J DENT RES January 2013 published on October 11, 2013

To view this item, select one of the options

- IADR Member Sign In  
IADR Members, please sign in at IADR to access the journal online.
- Sign In  
Already an individual subscriber? If so, please sign in to SAGE Journals with your User Name and Password.  
User Name:  Password:  Sign In  
 Remember my user name & password. Forgot your user name or password?
- Purchase Short-Term Access  
Pay per Article - You may purchase this article your purchase, which is yours to keep, within 30 days.  
Regain Access - You can regain access to a non-purchase period has not yet expired.

DOI: 10.1177/0022041813505222

Received  
Accepted  
A supplement  
Internationally

32

RESEARCH REPORTS  
Clinical

Institution: University of Wisconsin-Madison

SAGE journals

Search all journals Advanced Search Search History Browse

JDR JOURNAL OF DENTAL RESEARCH

Home OnlineFirst All Issues Subscribe RSS Email Alerts

Return to Search Results | Edit My Last Search

Search this journal

Impact Factor: 3.826 | Ranking: 2/82 in Dentistry, Oral Surgery & Medicine | Eigenfactor Score: 0.02325 | 5-Year Impact Factor: 4.286

**Current Issue**  
October 2013, 92 (10)

Alert me to new issues of Journal of Dental Research

Submit a Manuscript  
Free Sample Copy  
Email Alerts  
RSS feed  
More about this journal  
Other IADR Journals  
Advances in Dental Research  
Critical Reviews in Oral Biology & Medicine  
International and American Associations for Dental Research

Interested in article-level

As another example, I was interested a paper from the Journal of Dental Research.

It's less than a year old, so it's not available in PubMed Central.

I ordered a copy by inter-library loan, but it didn't include the supplemental methods, and those are behind a paywall at the journal!

# Twitter is useful (for venting and more)



## RESEARCH REPORTS

Clinical

J.R. Shaffer<sup>1,2\*</sup>, E. Feingold<sup>1,3</sup>, X. Wang<sup>2,4,5</sup>,  
D.E. Weeks<sup>1,3</sup>, R.J. Weyant<sup>2,6</sup>, R. Crout<sup>1,7</sup>,  
D.W. McNeil<sup>8</sup>, and M.L. Maraziti<sup>1,2,4,5,9</sup>

<sup>1</sup>Department of Human Genetics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; <sup>2</sup>Center for Oral Health Research in Appalachia, University of Pittsburgh, Pittsburgh, PA 15261, USA; <sup>3</sup>West Virginia University, Morgantown, WV 26506, USA; <sup>4</sup>Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261, USA; <sup>5</sup>Center for Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15219, USA; <sup>6</sup>Department of Oral Biology, School of Dental Medicine, University of Pittsburgh, Pittsburgh, PA 15261, USA; <sup>7</sup>Department of Dental Public Health and Information Management, University of Pittsburgh, School of Dental Medicine, Pittsburgh, PA 15261, USA; <sup>8</sup>Department of Periodontics, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; <sup>9</sup>Dental Practice and Rural Health, West Virginia University School of Dentistry, Morgantown, WV 26506, USA; and <sup>10</sup>Clinical and Translational Science Institute, and Department of Psychiatry, School of Medicine, University of Pittsburgh, Pittsburgh, PA, USA; \*corresponding author, jrs1@pitt.edu

J Dent Res DOI: 10.1177/0022034512463241

## Clustering Tooth Surfaces into Biologically Informative Caries Outcomes

### APPENDIX

#### COHRA Population Characteristics

Descriptive characteristics of the COHRA sample are shown in the Appendix Table. Compared with the general US population, this Appalachian sample from rural West Virginia and Pennsylvania is poorer, less educated, and underserved with respect to oral health care.

#### Cross-validation and Defining Separate Clusters

Determining the number of separate clusters is the principal challenge in the field of cluster analysis, and there is currently very little statistical theory in this area. Defining *a priori* minimum distances between clusters is arbitrary and may lead to over-fitting. Therefore, we instead performed two-fold cross-validation (e.g., Salvador and Chan, 2004), a commonly used approach for determining the number of clusters. Cross-validation allows us to determine the sensitivity of our clusters to perturbations of the input data and to identify what level of within-cluster similarity and between-cluster dissimilarity defines stable clusters. We randomly divided our sample into two halves and performed hierarchical clustering on each half. This process was repeated for 10 random halves. By comparing dendrograms, we determined the maximum number of separate clusters that were consistently observed across all random halves. Overall, tooth surfaces were distributed across 5 very stable clusters. Example cluster results for two complementary halves (which total to the full sample) are shown in Appendix

Fig. 1. Similarly, to assess any effects of the inclusion of biological relatives in our sample, we repeated hierarchical clustering in the maximal subset of unrelated individuals (see Appendix Fig. 2), which were nearly identical to the total COHRA and NHANES 1999-2000 samples.

In both the total COHRA sample and in the NHANES 1999-2000 sample, the fifth cluster was subdivided into maxillary and mandibular components (indicating possibly 6 instead of 5 clusters). However, these subdivisions were not consistently observed *via* cross-validation of the COHRA sample. That is, in some randomly chosen halves of the COHRA sample, the sixth cluster did not distinguish maxillary and mandibular components of the fifth cluster. Therefore, we have conservatively shown results for the 5 stable clusters, as well as the maxillary and mandibular subdivisions of the fifth cluster.

Overwhelmingly, the cluster results were stable within the COHRA sample, and consistent between the COHRA sample and NHANES 1999-2000 sample. That being stated, there were subtle differences observed among cluster results. For example, left-right asymmetry was observed, albeit rarely, in the random halves (e.g., Appendix Fig. 1A, tooth #20 and tooth #29). Likewise, some tooth surfaces physically positioned on the border between two adjacent clusters shifted membership (e.g., Appendix Figs. 1A and 2, tooth #21 and tooth #28). In many cases, the result of such shifts in cluster membership echoed the subtle differences between COHRA and NHANES 1999-2000. We speculate that the all-or-nothing nature of our clustering approach represents an oversimplification of the relationships among tooth surfaces with respect to dental caries risk factors

1

7

I was reduced to venting on twitter.

But then I got the appendix I wanted by email (twice!), within an hour of my tweet. (Thanks, MM and KW!)

Twitter is useful

#icanhazpdf

8

If you search twitter for #icanhazpdf, you'll find lots of people asking for copies of articles. Quite effective.

# It's all about money

(Costs in scientific publishing)

- ▶ Research
- ▶ Writing
- ▶ Peer review, editorial oversight
- ▶ Journal administration
- ▶ Copy editing, typesetting
- ▶ Distribution
- ▶ Profit

9

Open access is all about money.

Most of the costs behind a research paper are paid by grants or institutional funds. For most journals, peer review and editorial oversight are unpaid.

There are real costs associated with journals, but in the end they are all paid from the same sources (grants and institutional funds).

Do we really want to give away the product of our research and then buy it back repeatedly, at great profit to the publishers?

And shouldn't the literature be available generally and not just to those with access to well-funded research libraries?

## It's not about

- ▶ Peer review
- ▶ Predatory publishing
- ▶ **Impact factors**
- ▶ **Evaluating researchers**  
(for grants & promotions)

Well, it sort of is...

10

The Open Access discussion often gets tied up with discussion about peer review, predatory publishing, and journal impact factors.

But to me, it is a completely separate issue, whether we want stringent peer review before publication or instead leave the evaluation entirely to post-publication review.

On the other hand, the current culture is to evaluate researchers based on the perceived quality of the journals in which they've published. This makes it difficult to change to open access.

If everyone's still going to send their best work to Science, Nature, & Cell, then that work will continue to be locked up behind pay walls.

# Paying for it

- ▶ Traditional approach
  - subscriptions
  - page charges
- ▶ Open access
  - bigger page charges
  - submission charges?
- ▶ Endowments
- ▶ Direct grants to journals

11

The usual way in which publishing costs are paid are through a combination of subscriptions (both institutional and individual) and direct charges to the author.

In the new open access model, the page charges are increased in order to eliminate the subscription fees. One might have a fee for all submitted manuscripts and not just those accepted for publication.

I've not seen much discussion of other alternatives, but I would prefer to see endowments established, particularly for society journals. Alternatively, journals might be funded directly through grants.

# Invoice

## GENETICS

### Review Invoice

Article Information	
Publisher:	Genetics Society Of America
Title:	Genetics
Issue:	Volume 192, Number 1
Manuscript Title:	Mapping Quantitative Trait Loci onto a Phylogenetic Tree
Manuscript Number:	142448
Article Type:	Regular Research Papers
Corr. Author Name (e-mail addr.):	Karl W Broman (kbroman@biostat.wisc.edu)
Membership Status:	Member

### Review Estimated Publication Charges

Items	Unit Price	Quantity	Amount
Page Charges		\$70.00 13	\$910.00
Figure Charges		\$40.00 6	\$240.00
Supplemental Files (six pages or greater)		\$500.00 1	\$500.00
Open Access Option		\$1,200.00 1	\$1,200.00
Author Alterations		\$2.55 16	\$40.80
		<b>Subtotal:</b>	<b>\$2,890.80</b>

12

Here's an invoice for a paper I published in 2012.

The charges would have been “just” \$1700, but I paid an additional \$1200 to have it freely available (otherwise it would have been behind a pay wall for one year).

# Choices for young investigators

- ▶ Pay for open access
- ▶ Support young open access journals

OR

- ▶ Let subscribers pay & do more experiments
- ▶ Continue to go after Science, Nature, & Cell

13

The page charges, and the continued reliance on impact factors, lead to difficult choices, particularly for young investigators.

Should I pay for open access, or should I let the subscribers pay and use the savings to do more experiments?

Should I support open access journals, or should I continue to go after Science, Nature, & Cell?

The best scientists may confidently maintain their pure publication record.

But more mediocre scientists, who may be just scraping by, probably don't feel they have that luxury. A Nature paper can "make you."

# What can we do?

- ▶ Send our best work to open access journals
- ▶ Support junior faculty to keep their papers open
- ▶ Pay attention to the quality of the work  
(not the impact factor of the journal)
- ▶ Raise endowments for trusted journals
- ▶ Reform copyright law

14

We need to send our best work to open access journals.

We need to find ways to support our junior colleagues, so that they may do so as well.

We need to evaluate people based on their work and not by the name of the journal in which it appeared. We all may say, “Science and Nature are often crap and there are lots of fantastic papers that appear elsewhere.” But somehow when we see Nature or Cell on someone’s CV, we still have an immediate, positive reaction.

I would like to see endowed journals, open forever.

The quickest way to free the product of federally funded research would be to reform copyright law. If the product of our research were forced open by law, the publishing industry would figure out how to pay for it in short order.

But given the state of politics in the US, I’m not too optimistic about that.