The field-dependent nature of PageRank values in citation networks

This manuscript (<u>permalink</u>) was automatically generated from <u>greenelab/indices manuscript@ad020b5</u> on November 7, 2022.

Authors

- · Benjamin J. Heil
 - © 0000-0002-2811-1031 · © ben-heil · 🎔 autobencoder

 Genomics and Computational Biology Graduate Group, Perelman School of Medicine, I

Genomics and Computational Biology Graduate Group, Perelman School of Medicine, University of Pennsylvania · Funded by The Gordon and Betty Moore Foundation (GBMF4552)

- Casey S. Greene

 ✓

Department of Pharmacology, University of Colorado School of Medicine; Department of Biochemistry and Molecular Genetics, University of Colorado School of Medicine · Funded by The Gordon and Betty Moore Foundation (GBMF4552); The National Human Genome Research Institute (R01 HG10067)

Abstract

Introduction

There are more academic papers than any human can read in a lifetime. Determining which papers are important and in doing so solving the resulting prioritization problem is hard.

Part of this difficulty is because "importance" is a subjective measure. Different metrics attempt to capture it through different approaches. Citation count assumes the number of citations determines a paper's importance, h-index and Journal Impact Factor focus more on secondary factors like author' or journals' track record, and graph-based methods like PageRank or disruption index use the context of the surrounding papers to evaluate a given article's relevance [1,2,3,4]. Each of these methods have their own strengths, as well as various permutations that shore up their weaknesses [5,6,7,8].

In particular, because there are differences between fields' citation practices [9], scientists have developed approaches like normalizing the number of citations based on nearby papers in the citation network, studying the relative performance of disruption index variants in a single field, rescaling fields' citation data to give more consistant PageRank results, and so on [5,10,11,12] Such approaches account for differences between fields, but they tend to focus on normalizing out the effects a field has on a metric.

We argue that this removal of field-specific signal obscures real differences in relevance. Intuitively, Nobel prize-winning work will stay at the bottom of a cancer biologist's paper pile if the article is about astrophysics. Trying to apply a universal ranking metric to papers or journals ignores that phenomenon and yields strange results. For example, Mason and Signh recently pointed out that depending on what field is evaluating it, the journal *Christian Higher Education* is either ranked as a Q1 (top quartile) journal or a Q4 (bottom quartile) journal [13].

The purpose of this paper is not to create a new metric that attempts to measure articles' importance. Instead, we aim to shed light on a less-studied aspect of prioritizing papers: differences between fields. To that end we show differences between PageRanks for the same paper in different field citation networks. We then demonstrate that some of the differences are field-specific by estimating the distribution of possible citation networks with a degree-preserving graph shuffling method. We also demonstrate that while it is possible to rank journals using standard metrics [14], their rank changes depending on field. Finally, we make our results more easily available by creating a web app that visualizes pairs of fields with overlapping papers.

Ultimately we show that there is great potential in analyzing articles while keeping field in mind. Going forward, we hope more field-aware methods of assessing article importance will be developed.

Results

Workflow

We built citation networks from pairs of MeSH headings corresponding to fields. We then created 100 shuffled versions of each combined network using a degree-preserving shuffling approach (see Methods). We split both the original and shuffled networks into papers of each constituent field, then calculated the PageRank for all networks. Finally, we created percentile scores by comparing the PageRanks for the original and shuffled networks, and made our results available at (TODO web server URL).

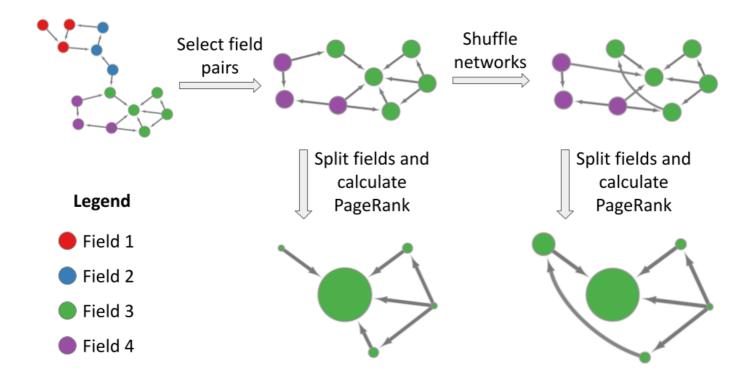


Figure 1: Schematic of the network analysis workflow. We construct networks of citations from pairs of MeSH headings, create shuffled versions of the graphs, split the networks into their component fields, and calculate the PageRank for each article in the networks.

The same paper will have different PageRanks in different fields

To demonstrate different relationships between fields' PageRanks, we visualized networks from four different field pairs. The first, nanotechnology/microscopy is a fairly average network in terms of interfield PageRank correlation (Fig. 2 A). The second, immunochemistry/anatomy has a high correlation but still has outlier points (Fig. 2 B). The third and fourth, proteomics/metabolomics and computational biology/human genetics show two fields with a larger degree of difference between PageRanks in each field (Fig. 2 C,D).

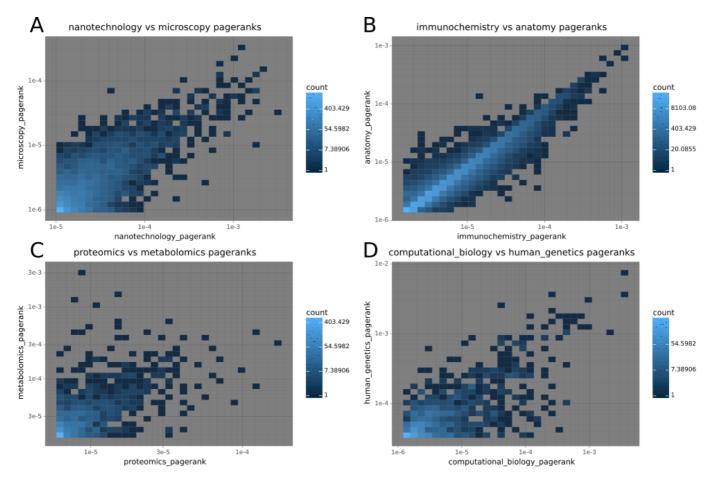


Figure 2: Heatmap of shared papers across fields

This variance is not entirely field independent

It's hard to say that a paper with a large difference between field-specific PageRanks is more important in one field than the other. It's possible that the scores are skewed by different field sizes, or caused by randomness in MeSH tagging.

To disentangle the random chance from the actual differences in network structure, we needed something to compare against. We decided on shuffling citation networks in a way that preserved their citation counts while randomizing which papers they cite. By doing so, we created 100 shuffled networks for each pair of fields where the degree distribution of the network is maintained but the global structure is removed from the networks. We then calculated a percentile score based on how many of the random networks had lower PageRank for a given paper than its true PageRank. That is to say that a paper with a PageRank higher than all the random networks would have a score of one, a paper with a PageRank less than all the random networks would have a zero, and the rest would fall somewhere in between. We used the difference in these scores between fields to determine the field-specific affinity for a given paper.

The nanotechnology/microscopy plot has a normal amount of variance, and clearly demonstrates the differences between fields (Fig. 3 A). In the center of the plot are papers that have similar percentile scores in both fields. Towards the right side most papers have a large positive nanotech-microscopy score, indicating that they are more highly valued in the field of nanotechnology than in microscopy. Such papers include "A robust DNA mechanical device controlled by hybridization topology" and "Bioadhesive poly(methyl methacrylate) microdevices for controlled drug delivery", while other papers such as "Turning the spotlight on cellular imaging" appropriately have a large negative nanotechnology-microscopy score despite falling far to the nanotech side of the distribution [15,16,17]. Meanwhile, the left side holds papers with strong negative scores, indicating the field of microscopy values the papers more. The left side has papers like "Photostable luminescent

nanoparticles as biological label for cell recognition of system lupus erythematosus patients" and "WSXM: a software for scanning probe microscopy and a tool for nanotechnology" [18,19]. The papers with a high PageRank and similar percentile scores between fields are likely to be influential in both fields. In nanotechnology/microscopy these papers are frequently nanoscopic imaging papers like "Imaging intracellular fluorescent proteins at nanometer resolution", "In vivo imaging of quantum dots encapsulated in phospholipid micelles", and "Water-Soluble Quantum Dots for Multiphoton Fluorescence Imaging in Vivo" [20,21,22].

The immunochemistry/anatomy fields have much higher correlation in their shared papers, but still have some outliers (Fig. 3 B). Examples of papers preferred in the immunochemistry field are "Immunoelectron microscopic exploration of the Golgi complex", "Immunocytochemical and electrophoretic analyses of changes in myosin gene expression in cat posterior temporalis muscle during postnatal development", and "Electron microscopic demonstration of calcitonin in human medullary carcinoma of thyroid by the immuno gold staining method" [23,24,25]. Papers preferred in anatomy include "Grafting genetically modified cells into the rat brain: characteristics of E. coli β-galactosidase as a reporter gene", "Vitamin-D-dependent calcium-binding-protein and parvalbumin occur in bones and teeth", and "Mapping of brain areas containing RNA homologous to cDNAs encoding the alpha and beta subunits of the rat GABAA gamma-aminobutyrate receptor" [26,27,28] Papers with high PageRank in both domains and a similar importance are "Studies of the HER-2/neu Proto-Oncogene in Human Breast and Ovarian Cancer", "Expression of c-fos Protein in Brain: Metabolic Mapping at the Cellular Level", and "Proliferating cell nuclear antigen (PCNA) immunolocalization in paraffin sections: An index of cell proliferation with evidence of deregulated expression in some, neoplasms" [29,30,31].

Where immunochemistry/anatomy served as an example of a highly correlated pair of fields, papers in proteomics/metabolomics have very different PageRanks between fields (Fig. 3 C). These fields illustrate one of the reasons why the difference in percentile scores is a useful metric: there is so much dispersion that it's hard to tell where a trend line should be. When coloring by the difference in scores, however, it is easy to see which papers are valued more in each field. The proteomics-heavy papers include "Proteomics Standards Initiative: Fifteen Years of Progress and Future Work", "Limited Environmental Serine and Glycine Confer Brain Metastasis Sensitivity to PHGDH Inhibition", and "A high-throughput processing service for retention time alignment of complex proteomics and metabolomics LC-MS data", [32,33,34] while the metabolomics papers have titles such as "MeltDB: a software platform for the analysis and integration of metabolomics experiment data", "In silico fragmentation for computer assisted identification of metabolite mass spectra", and "The Metabonomic Signature of Celiac Disease" [35,36,37]. The articles with high PageRank and importance in both fields include "Visualization of omics data for systems biology", "FunRich: An open access standalone functional enrichment and interaction network analysis tool", and "Proteomic and Metabolomic Characterization of COVID-19 Patient Sera" [38,39,40].

The final pair, computational biology/human genetics, is interesting because while there is a good degree of variance in the PageRank values across fields, the highest rated papers in both fields had very similar percentile scores (Fig. 3 D). The high percentile difference papers in computational biology include several database papers [41,42,43]. Their counterparts in human genetics were more on the policy paper side, likely because the human genetics MeSH subheading is distinct from the several other genetics subheadings including genetic research, population genetics, genomics, etc. [44,45,46]. Meanwhile, the papers with high PageRanks in both fields tend to have similar percentile scores and be related to the genome sequencing and mapping [47,48,49].

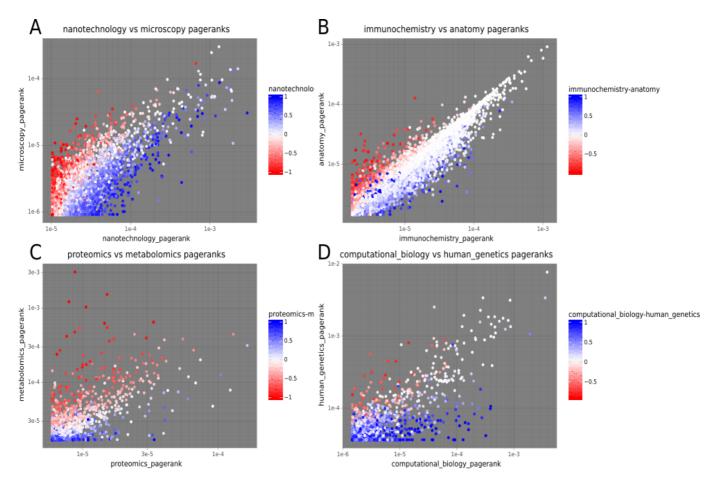


Figure 3: The difference between percentile scores for four field pairs.

Journals aren't a great proxy for field, as they often have papers present from multiple fields.

In addition to article metrics, papers are often prioritized by selecting prestigious journals. In an attempt to quantify the relative importance of journals, scientists have created rankings using metrics like Eigenfactor [50]. While such metrics return intuitively reasonable results, we discovered that the inherent aggregation that they apply masks differences between fields.

When considering the fields nanotechnology or microscopy, the journal *Science* is far and away the most influential journal based on median PageRank (Fig. 4 A). *Science*'s dominance isn't universal though — it ranks fifth behind journals like *Nature* and *Cell* when looking at immunochemistry and anatomy (Fig. 4 B). Further, while journals are sometimes used as a proxy measure for field [10], we find they often play host to interdisciplinary papers that may be more accurately assigned to a different field. For example, "The metabonomic signature of celiac disease" is a paper with an extremely high PageRank in the field of metabolomics, but it was actually published in the *Journal of Proteome Research* [37].

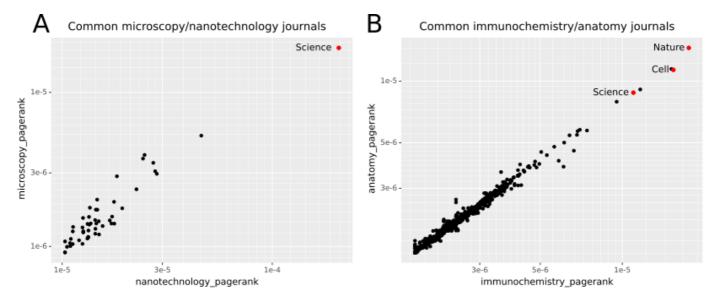


Figure 4: Differences between journal ranks across fields. Each point in the figure represents the median pagerank of all papers in the journal for the given fields.

Here's a web app

To demonstrate our findings, we have also set up a web server with all pairs of networks instead of solely the ones shown in this paper. The web server can be found at X

Methods

COCI

We used the March 2022 version of the COCI citation index [51] as the source of our citation data. This dataset contains around 1.3 billion citations from ~73 million bibliographic resources.

Selecting fields

To differentiate between scientific fields, we needed a way to map papers to fields. Fortunately, all the papers in Pubmed Central (https://www.ncbi.nlm.nih.gov/pmc/) have corresponding Medical Subject Headings (MeSH) terms. While MeSH terms are varied and numerous, the subheadings of the Natural Science Disciplines (H01) category fit our needs. However, MeSH terms are hierarchical, and vary greatly in their size and specificity. To extract a balanced set of terms we recursively traversed the tree and selected headings that have least 10000 DOIs and don't have multiple children that also meet the cutoff. Our resulting set of headings contained 45 terms, from "Acoustics" to "Water Microbiology".

Building single heading citation networks

The COCI dataset consists of pairs of Digital Object Identifiers (DOIs). To change these pairs into a form we could run calculations on, we needed to convert them into networks. To do so, we created 45 empty networks, one for each MeSH term we selected previously. We then iterated over each pair of DOIs in COCI, and added them to a network if the DOIs corresponded to two journal articles written in english, both of which were tagged with the corresponding MeSH heading.

Building combined networks

Because we were interested in the differences between fields, we also needed to build networks from pairs of MeSH headings. These networks were built via the same process, except that instead of keeping articles corresponding to a single DOI we added a citation to the network if both articles were in the pair of fields, even if the citation occurred across fields. Running this network-building process yielded 990 two-heading networks.

Shuffling networks

Sampling a graph from the degree distribution while preserving the distribution of degrees in the network turned out to be challenging. Because citation graphs are directed, it's not possible to simply swap pairs of edges and end up with a graph that is uniformly sampled from the space. Instead, a more sophisticated three-edge swap method must be used [52]. Because this algorithm had not been implemented yet in NetworkX [53], we wrote the code to perform shuffles and submitted our change to the library. With the shuffling code implemented, we created 100 shuffled versions of each of our combined networks to act as a background distribution to compare metrics against.

Splitting networks

Once we had a collection of shuffled networks, we needed to split them into their constituent fields. To do so, we reduced the network to solely the nodes that were present in the single heading citation network, and kept only citations between these nodes.

Running PageRank

We used the NetworkX implementation of PageRank with default parameters to evaluate paper importance within fields.

Percentile score

To determine the degree to which the papers' PageRank values were higher or lower than expected, we compared the PageRank values calculated for the true citation networks to the values in the shuffled networks for each paper. We then recorded the fraction of shuffled networks where the paper had a lower PageRank than in the true network to derive a single number that described these values. For example, if a paper had a higher PageRank in the true network than in all the shuffled networks it received a score of 1. Likewise, if it had a lower PageRank in the true network than in all the shuffled networks it received a score of 0. Papers in between the two extremes had fractional values, like .5 (a paper that fell in the middle of the pack) and so on.

Differences in percentiles

A convenient feature of the percentile scores is that they're directly comparable between fields. If a paper is present in two fields, the difference in scores between the two fields can be used to estimate its relative importance. For example, if a paper has a score of 1 in field A (indicating a higher PageRank in the field than expected given its number of citations and the network structure) and a score of 0 in field B (indicating a lower than expected PageRank), then the large difference in scores indicates the paper is more highly valued in field A than field B. If the paper has similar scores in both fields, it indicates that the paper is similarly valued in the two fields.

Hardware/runtime

The analysis pipeline was run on the RMACC Summit cluster. The full pipeline, from downloading the data to analyzing it to vizualizing it took about a week to run. However, that number is heavily dependent on details such as the number of CPU nodes available and the network speed.

Server details

Our webserver is built by visualizing our data in Plotly (https://plotly.com/python/plotly-express/) on the Streamlit platform (https://streamlit.io/). The field pairs made available by the frontend are those with at least 1000 shared papers after filtering out papers with more than a 5% missingness level of their PageRanks after shuffling. The journals available for visualization are those with at least 25 papers for the given field pair.

Discussion/Conclusion

We analyzed hundreds of pairwise citation networks in order to determine the effects of field on citation metrics. In doing so we found that there are, in fact, differences in PageRanks between fields that warrant some form of normalization when making comparisons. However, we also showed that these effects aren't solely driven by differences in citation practices that should be regressed out. Instead, there appear to be meaningful differences between fields for individual papers that exist in both fields. Normalizing out this variation or creating a single global ranking obscures meaningful signal about how papers are perceived in different communities.

That is not to say that our analysis is perfect, however. For one thing, there are inherent issues with the concept of ranking papers. While we point out that not using field information when prioritizing papers can lose some of the nuance of the data, we're still just discussing different ways of measuring proxy variables. One's goal is often to read the papers most relevant to them, or the most influential papers in their field. However, since the ground truth of relevance or influence is difficult or impossible to quanitfy we end up using PageRank, citation count, Journal Impact Factor, or some other metric. While these proxy variables are hopefully correlated with the characteristics we care about, its important to acknowledge that they're not the same thing.

Our percentile score is also an area for potential improvement. We decide to make three times as many swaps as our network has edges, but it is unclear what the ideal number of swaps is to ensure a network is drawn uniformly at random from the distribution of possible networks with the same degree distribution. Additionally, sampling graphs at random from degree-preserving network swaps isn't perfectly representative of what a potential alternate universe's citation network would look like. For that we'd need a way to avoid edge swaps that caused "back-in-time" citations, a method to swap citations to a paper based on what similar papers had cited, and the ability to add or subtract citations to and from a given paper, as the number of citations for a given paper is partly due to random chance. While outside of the scope of this paper, implementing these features into a citation graph randomization method would be an interesting avenue of future research.

A third limitation of our paper comes from the way we selected MeSH terms. Our threshold number of papers required to include a field was chosen so that we could investigate a diverse set of fields while still keeping a reasonable amount of papers in each combination of fields. However, it is possible that a more (or less) granular field selection approach would give clearer results.

Similar approaches to ours may be helpful for an expert in one field learning a new one. For example, a expert in computational biology might get a better idea of the difference in perspective across fields by reading articles that have a high percentile score in genetics and a low score in computational biology. While such papers are important in genetics, the computational biologist would be unlikely to be familiar with them due to the field difference.

Ultimately we hope to see more analyses going forward use field information for evaluating the influence of scientific articles. We believe this to be particularly important in the case of interdisciplinary papers that exist between fields by their very nature. A paper largely ignored in one field may be groundbreaking in another, and that dynamic should be accounted for when attempting to make sense of citation networks.

References

1. An index to quantify an individual's scientific research output

JE Hirsch

Proceedings of the National Academy of Sciences (2005-11-07) https://doi.org/cbq6dz
DOI: 10.1073/pnas.0507655102 · PMID: 16275915 · PMCID: PMCID: PMC1283832

2. New Tools for Improving and Evaluating The Effectiveness of Research

Irving H Sher, Eugene Garfield Research Program Effectiveness (1965-06-27)

3. The PageRank Citation Ranking: Bringing Order to the Web.

Lawrence Page, Sergey Brin, Rajeev Motwani, Terry Winograd *Stanford InfoLab* (1999)

4. Large teams develop and small teams disrupt science and technology

Lingfei Wu, Dashun Wang, James A Evans

Nature (2019-02) https://doi.org/gfvnb9

DOI: <u>10.1038/s41586-019-0941-9</u> · PMID: <u>30760923</u>

5. Relative Citation Ratio (RCR): A New Metric That Uses Citation Rates to Measure Influence at the Article Level

Blan Hutchins, Xin Yuan, James M Anderson, George M Santangelo *PLOS Biology* (2016-09-06) https://doi.org/f88zk2

DOI: 10.1371/journal.pbio.1002541 · PMID: 27599104 · PMCID: PMC5012559

6. Measuring contextual citation impact of scientific journals

Henk F Moed

Journal of Informetrics (2010-07) https://doi.org/dpbgj9

DOI: <u>10.1016/j.joi.2010.01.002</u>

7. Collective topical PageRank: a model to evaluate the topic-dependent academic impact of scientific papers

Yongjun Zhang, Jialin Ma, Zijian Wang, Bolun Chen, Yongtao Yu *Scientometrics* (2017-12-23) https://doi.org/gc4b2s

DOI: 10.1007/s11192-017-2626-1

8. Are disruption index indicators convergently valid? The comparison of several indicator variants with assessments by peers

Lutz Bornmann, Sitaram Devarakonda, Alexander Tekles, George Chacko *Quantitative Science Studies* (2020-08) https://doi.org/gg2ts5

DOI: 10.1162/qss a 00068

9. Citation patterns in economics and beyond

Matthias Aistleitner, Jakob Kapeller, Stefan Steinerberger *Science in Context* (2019-12) https://doi.org/gq62s8
DOI: 10.1017/s0269889720000022 · PMID: 32202238

10. Disruptive papers published in Scientometrics: meaningful results by using an improved variant of the disruption index originally proposed by Wu, Wang, and Evans (2019)

Lutz Bornmann, Sitaram Devarakonda, Alexander Tekles, George Chacko

Scientometrics (2020-03-14) https://doi.org/ggzzxd

DOI: <u>10.1007/s11192-020-03406-8</u>

11. Quantifying and suppressing ranking bias in a large citation network

Giacomo Vaccario, Matúš Medo, Nicolas Wider, Manuel Sebastian Mariani Journal of Informetrics (2017-08) https://doi.org/gbzjdh

DOI: 10.1016/j.joi.2017.05.014

12. Quantitative evaluation of alternative field normalization procedures

Yunrong Li, Filippo Radicchi, Claudio Castellano, Javier Ruiz-Castillo *Journal of Informetrics* (2013-07) https://doi.org/f48tvt

DOI: 10.1016/j.joi.2013.06.001

13. When a journal is both at the 'top' and the 'bottom': the illogicality of conflating citation-based metrics with quality

Shannon Mason, Lenandlar Singh

Scientometrics (2022-05-25) https://doi.org/gq2468

DOI: 10.1007/s11192-022-04402-w

14. Eigenfactor: Measuring the value and prestige of scholarly journals

Carl Bergstrom

College & research libraries news (2007)

15. A robust DNA mechanical device controlled by hybridization topology

Hao Yan, Xiaoping Zhang, Zhiyong Shen, Nadrian C Seeman

Nature (2002-01) https://doi.org/czh8hg

DOI: <u>10.1038/415062a</u> · PMID: <u>11780115</u>

16. Bioadhesive poly(methyl methacrylate) microdevices for controlled drug delivery

Sarah L Tao, Michael W Lubeley, Tejal A Desai

Journal of Controlled Release (2003-03) https://doi.org/c7fpg4

DOI: <u>10.1016/s0168-3659(03)00005-1</u>

17. Turning the spotlight on cellular imaging

Peter Mitchell

Nature Biotechnology (2001-11) https://doi.org/dr77vj

DOI: <u>10.1038/nbt1101-1013</u> · PMID: <u>11689841</u>

18. Photostable Luminescent Nanoparticles as Biological Label for Cell Recognition of System Lupus Erythematosus Patients

Xiaoxiao He, Kemin Wang, Weihong Tan, Jun Li, Xiaohai Yang, Shasheng Huang, Dan Xiao *Journal of Nanoscience and Nanotechnology* (2002-07-01) https://doi.org/dcj5cg

DOI: <u>10.1166/jnn.2002.105</u> · PMID: <u>12908257</u>

19. WSXM: A software for scanning probe microscopy and a tool for nanotechnology

I Horcas, R Fernández, JM Gómez-Rodríguez, J Colchero, J Gómez-Herrero, AM Baro *Review of Scientific Instruments* (2007-01)

DOI: <u>10.1063/1.2432410</u>

20. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution

Eric Betzig, George H Patterson, Rachid Sougrat, OWolf Lindwasser, Scott Olenych, Juan S Bonifacino, Michael W Davidson, Jennifer Lippincott-Schwartz, Harald F Hess *Science* (2006-09-15) https://doi.org/dd8dkd

DOI: 10.1126/science.1127344 · PMID: 16902090

21. In Vivo Imaging of Quantum Dots Encapsulated in Phospholipid Micelles

Benoit Dubertret, Paris Skourides, David J Norris, Vincent Noireaux, Ali H Brivanlou, Albert Libchaber

Science (2002-11-29) https://doi.org/dd6sqp

DOI: 10.1126/science.1077194 · PMID: 12459582

22. Water-Soluble Quantum Dots for Multiphoton Fluorescence Imaging in Vivo

Daniel R Larson, Warren R Zipfel, Rebecca M Williams, Stephen W Clark, Marcel P Bruchez, Frank W Wise, Watt W Webb

Science (2003-05-30) https://doi.org/cn9j76

DOI: <u>10.1126/science.1083780</u> · PMID: <u>12775841</u>

23. Immunoelectron microscopic exploration of the Golgi complex.

JW Slot, HJ Geuze

Journal of Histochemistry & District Manager (1983-08) https://doi.org/dxxzxg

DOI: 10.1177/31.8.6863900 · PMID: 6863900

24. Immunocytochemical and electrophoretic analyses of changes in myosin gene expression in cat posterior temporalis muscle during postnatal development

JFY Hoh, S Hughes, C Chow, PT Hale, RB Fitzsimons

Journal of Muscle Research and Cell Motility (1988-02) https://doi.org/d72278

DOI: 10.1007/bf01682147 · PMID: 3392187

25. Electron microscopic demonstration of calcitonin in human medullary carcinoma of thyroid by the immuno gold staining method

J Dämmrich, W Ormanns, R Schäffer

Histochemistry (1984) https://doi.org/ct253c

DOI: <u>10.1007/bf00514331</u> · PMID: <u>6511490</u>

26. Grafting genetically modified cells into the rat brain: characteristics of E. coli β-galactosidase as a reporter gene

S Shimohama, MB Rosenberg, AM Fagan, JA Wolff, MP Short, XO Breakefield, T Friedmann, FH Gage

Molecular Brain Research (1989-06) https://doi.org/dptnbm

DOI: <u>10.1016/0169-328x(89)90061-2</u>

27. Vitamin-D-dependent calcium-binding-protein and parvalbumin occur in bones and teeth

MR Celio, AW Norman, CW Heizmann

Calcified Tissue International (1984-12) https://doi.org/fdnfdg

DOI: 10.1007/bf02405306 · PMID: 6423230

28. Mapping of brain areas containing RNA homologous to cDNAs encoding the alpha and beta subunits of the rat GABAA gamma-aminobutyrate receptor.

JM Séguier, JG Richards, P Malherbe, GW Price, S Mathews, H Möhler

Proceedings of the National Academy of Sciences (1988-10) https://doi.org/fv2p49

DOI: 10.1073/pnas.85.20.7815 · PMID: 2845424 · PMCID: PMC282284

29. Studies of the HER-2/ <i>neu</i> Proto-Oncogene in Human Breast and Ovarian Cancer

Dennis J Slamon, William Godolphin, Lovell A Jones, John A Holt, Steven G Wong, Duane E Keith, Wendy J Levin, Susan G Stuart, Judy Udove, Axel Ullrich, Michael F Press

Science (1989-05-12) https://doi.org/cngtqx

DOI: <u>10.1126/science.2470152</u> · PMID: <u>2470152</u>

30. Expression of c- <i>fos</i> Protein in Brain: Metabolic Mapping at the Cellular Level

SM Sagar, FR Sharp, T Curran

Science (1988-06-03) https://doi.org/b39h2t

DOI: <u>10.1126/science.3131879</u> · PMID: <u>3131879</u>

31. Proliferating cell nuclear antigen (PCNA) immunolocalization in paraffin sections: An index of cell proliferation with evidence of deregulated expression in some, neoplasms

PA Hall, DA Levison, AL Woods, CC-W Yu, DB Kellock, JA Watkins, DM Barnes, CE Gillett, R Camplejohn, R Dover, ... DP Lane

The Journal of Pathology (1990-12) https://doi.org/cntmbr

DOI: 10.1002/path.1711620403 · PMID: 1981239

32. Proteomics Standards Initiative: Fifteen Years of Progress and Future Work

Eric W Deutsch, Sandra Orchard, Pierre-Alain Binz, Wout Bittremieux, Martin Eisenacher, Henning Hermjakob, Shin Kawano, Henry Lam, Gerhard Mayer, Gerben Menschaert, ... Andrew R Jones

Journal of Proteome Research (2017-09-15) https://doi.org/gbw99d

DOI: <u>10.1021/acs.jproteome.7b00370</u> · PMID: <u>28849660</u> · PMCID: <u>PMC5715286</u>

33. Limited Environmental Serine and Glycine Confer Brain Metastasis Sensitivity to PHGDH Inhibition

Bryan Ngo, Eugenie Kim, Victoria Osorio-Vasquez, Sophia Doll, Sophia Bustraan, Roger J Liang, Alba Luengo, Shawn M Davidson, Ahmed Ali, Gino B Ferraro, ... Michael E Pacold *Cancer Discovery* (2020-09-01) https://doi.org/ghf85j

DOI: 10.1158/2159-8290.cd-19-1228 · PMID: 32571778 · PMCID: PMC7483776

34. A high-throughput processing service for retention time alignment of complex proteomics and metabolomics LC-MS data

Isthiaq Ahmad, Frank Suits, Berend Hoekman, Morris A Swertz, Heorhiy Byelas, Martijn Dijkstra, Rob Hooft, Dmitry Katsubo, Bas van Breukelen, Rainer Bischoff, Peter Horvatovich *Bioinformatics* (2011-02-23) https://doi.org/cxsszv

DOI: <u>10.1093/bioinformatics/btr094</u> · PMID: <u>21349866</u>

35. MeltDB: a software platform for the analysis and integration of metabolomics experiment data

Heiko Neuweger, Stefan P Albaum, Michael Dondrup, Marcus Persicke, Tony Watt, Karsten Niehaus, Jens Stoye, Alexander Goesmann

Bioinformatics (2008-09-02) https://doi.org/fds6vt

DOI: 10.1093/bioinformatics/btn452 · PMID: 18765459

36. In silico fragmentation for computer assisted identification of metabolite mass spectra

Sebastian Wolf, Stephan Schmidt, Matthias Müller-Hannemann, Steffen Neumann *BMC Bioinformatics* (2010-03-22) https://doi.org/d7gpf5

DOI: <u>10.1186/1471-2105-11-148</u> · PMID: <u>20307295</u> · PMCID: <u>PMC2853470</u>

37. The Metabonomic Signature of Celiac Disease

Ivano Bertini, Antonio Calabrò, Valeria De Carli, Claudio Luchinat, Stefano Nepi, Berardino Porfirio, Daniela Renzi, Edoardo Saccenti, Leonardo Tenori

Journal of Proteome Research (2008-12-11) https://doi.org/c6sdnp

DOI: <u>10.1021/pr800548z</u> · PMID: <u>19072164</u>

38. Visualization of omics data for systems biology

Nils Gehlenborg, Seán I O'Donoghue, Nitin S Baliga, Alexander Goesmann, Matthew A Hibbs, Hiroaki Kitano, Oliver Kohlbacher, Heiko Neuweger, Reinhard Schneider, Dan Tenenbaum, Anne-Claude Gavin

Nature Methods (2010-03) https://doi.org/cp9zgi

DOI: 10.1038/nmeth.1436 · PMID: 20195258

39. FunRich: An open access standalone functional enrichment and interaction network analysis tool

Mohashin Pathan, Shivakumar Keerthikumar, Ching-Seng Ang, Lahiru Gangoda, Camelia YJ Quek, Nicholas A Williamson, Dmitri Mouradov, Oliver M Sieber, Richard J Simpson, Agus Salim, ... Suresh Mathivanan

PROTEOMICS (2015-06-17) https://doi.org/f278rp
DOI: 10.1002/pmic.201400515 · PMID: 25921073

40. Proteomic and Metabolomic Characterization of COVID-19 Patient Sera

Bo Shen, Xiao Yi, Yaoting Sun, Xiaojie Bi, Juping Du, Chao Zhang, Sheng Quan, Fangfei Zhang, Rui Sun, Liujia Qian, ... Tiannan Guo

Cell (2020-07) https://doi.org/gg2cck

DOI: 10.1016/j.cell.2020.05.032 · PMID: 32492406 · PMCID: PMC7254001

41. Development of Human Protein Reference Database as an Initial Platform for Approaching Systems Biology in Humans

Suraj Peri, JDaniel Navarro, Ramars Amanchy, Troels Z Kristiansen, Chandra Kiran Jonnalagadda, Vineeth Surendranath, Vidya Niranjan, Babylakshmi Muthusamy, TKB Gandhi, Mads Gronborg, ... Akhilesh Pandey

Genome Research (2003-10) https://doi.org/bc8cnv

DOI: <u>10.1101/gr.1680803</u> · PMID: <u>14525934</u> · PMCID: <u>PMC403728</u>

42. A database for post-genome analysis

Minoru Kanehisa

Trends in Genetics (1997-09) https://doi.org/cfgb98

DOI: 10.1016/s0168-9525(97)01223-7

43. Use of mass spectrometry-derived data to annotate nucleotide and protein sequence databases

Matthias Mann, Akhilesh Pandey

Trends in Biochemical Sciences (2001-01) https://doi.org/ch565r

DOI: <u>10.1016/s0968-0004(00)01726-6</u>

44. Genetic Discrimination: Perspectives of Consumers

EVirginia Lapham, Chahira Kozma, Joan O Weiss

Science (1996-10-25) https://doi.org/df7k88

DOI: 10.1126/science.274.5287.621 · PMID: 8849455

45. Committee Opinion No. 690: Carrier Screening in the Age of Genomic Medicine

Obstetrics & amp; Gynecology

Ovid Technologies (Wolters Kluwer Health) (2017-03) https://doi.org/f92g56

DOI: 10.1097/aog.000000000001951 · PMID: 28225425

46. Public health genomics: The end of the beginning

Muin J Khoury

Genetics in Medicine (2011-03) https://doi.org/bjsxzk

DOI: 10.1097/gim.0b013e31821024ca · PMID: 21311338

47. Initial sequencing and analysis of the human genome

, Eric S Lander, Lauren M Linton, Bruce Birren, Chad Nusbaum, Michael C Zody, Jennifer Baldwin, Keri Devon, Ken Dewar, Michael Doyle, ...

Nature (2001-02-15) https://doi.org/bfpgjh

DOI: 10.1038/35057062 · PMID: 11237011

48. An STS-Based Map of the Human Genome

Thomas J Hudson, Lincoln D Stein, Sebastian S Gerety, Junli Ma, Andrew B Castle, James Silva, Donna K Slonim, Rafael Baptista, Leonid Kruglyak, Shu-Hua Xu, ... Eric S Lander

Science (1995-12-22) https://doi.org/ftf

DOI: 10.1126/science.270.5244.1945 · PMID: 8533086

49. A New Five-Year Plan for the U.S. Human Genome Project

Francis Collins, David Galas

Science (1993-10) https://doi.org/fwkrnb

DOI: 10.1126/science.8211127 · PMID: 8211127

50. Eigenfactor: Measuring the value and prestige of scholarly journals

Carl Bergstrom

College & Devis (2007-05-01) https://doi.org/gf24tg

DOI: 10.5860/crln.68.5.7804

51. Software review: COCI, the OpenCitations Index of Crossref open DOI-to-DOI citations

Ivan Heibi, Silvio Peroni, David Shotton

Scientometrics (2019-09-14) https://doi.org/ggzz8b

DOI: <u>10.1007/s11192-019-03217-6</u>

52. A simple Havel-Hakimi type algorithm to realize graphical degree sequences of directed graphs

Péter L Erdős, István Miklós, Zoltán Toroczkai *arXiv* (2010-01-21) https://arxiv.org/abs/0905.4913

53. Exploring Network Structure, Dynamics, and Function using NetworkX

Aric A Hagberg, Daniel A Schult, Pieter J Swart Proceedings of the 7th Python in Science conference (2008)