A Bioinformatician, Computer Scientist, and Geneticist lead bioinformatic tool development - which one is better?

Paul P. Gardner[1]*

Abstract
The development of accurate bioinformatic software tools is crucial for the effective analysis of complex biological data. This study examines the relationship between the academic department affiliations of authors and the accuracy of the bioinformatic tools they develop. By analyzing a corpus of previously benchmarked bioinformatic software tools, we mapped bioinformatic tools to the academic fields of the corresponding authors and evaluated tool accuracy by field. Our results suggest that "Medical Informatics" outperforms all other fields in bioinformatic software accuracy, with a mean proportion of wins in accuracy rankings exceeding the null expectation. In contrast, tools developed by authors affiliated with "Bioinformatics" and "Engineering" fields tend to be less accurate. However, after correcting for multiple testing, none of the results are statistically significant ($p > 0.05$). Our findings reveal no strong association between academic field and bioinformatic software accuracy. These findings suggest that the development of interdisciplinary software applications can be effectively undertaken by any department with sufficient resources and training.

Keywords
Bioinformatics, Software Tool Development, Interdisciplinary Research, Departmental Affiliation, Software Accuracy, Medical Informatics, Academic Expertise, Benchmarking, Computational Biology

[1] *Department of Biochemistry, University of Otago, Dunedin, New Zealand.*
***Corresponding author**: paul.gardner@otago.ac.nz

Key Points

- The study investigates the relationship between academic department affiliations of authors and the accuracy of bioinformatic software tools they develop.
- Results indicate no statistically significant association between departmental affiliation and the accuracy of bioinformatic tools, challenging the assumption that departmental expertise predicts software quality.
- Tools from the "Medical Informatics" field showed a higher mean proportion of wins, while those from "Bioinformatics" and "Engineering" fields ranked lower, though these differences were not statistically significant after correcting for multiple-testing.
- The findings suggest that allocating sufficient resources for long term software development are more critical to accurate bioinformatic tools than departmental expertise.

Background

Much is made of departmental divisions within academia; These can denote research and teaching expertise [1], influence hiring decisions, access to funding, publishing and the training of students recruited for research projects [2]. However, interdisciplinary subjects such as bioinformatics break down the traditional barriers between departments and subject areas [3, 4, 5].

Bioinformatics, is an interdisciplinary field that fuses biology, computer science, and mathematics, and now plays a pivotal role in modern biological research [3]. The development of bioinformatic tools and software is critical for interpreting complex biological questions, such as what evolutionary, structural and functional analyses of genomic, transcriptomic, and proteomic data can tell us. The field of "bioinformatics" can include many overlapping research fields that include computational biology, biomathematics, biostatistics, medical informatics and other similar areas.

Bioinformatics gained significant traction with the advent of high-throughput sequencing technologies, which created a need for robust computational tools to manage the large volumes of data generated [3, 5]. In response, departments specializing in bioinformatics began to emerge from biology, computer science, and engineering faculties, each bringing unique contributions to the field's development and expansion.

Bioinformatics integrates methods from various disciplines, fostering novel insights that are unattainable within a single field [6]. This requires a deep understanding of biological sciences for interpreting data, and strong computational skills are required to develop algorithms and analyze the results at

scale.

Experts in **biological and health sciences**, such as genetics and molecular biology, can ensure that software tools are biologically relevant and accurate by identifying critical questions and guiding computational tool development. However, biology departments may lack the advanced computational expertise required for developing sophisticated software. In our analysis, we have grouped the biological and health fields under the category of "**domain experts**".

The **mathematical, engineering and computational sciences**, referred to here as the "**development experts**", contribute significantly by bringing expertise in algorithm development, mathematical modelling, statistics, and software engineering principles. These skills are essential for creating efficient, scalable, and robust bioinformatic tools. However, the challenge for development experts is gaining a deep understanding of biological domains, which is necessary to ensure software tools are both relevant and accurate.

Departmental differences can influence the development of bioinformatic software tools because they reflect the varying expertise, resources, and perspectives that different academic fields offer. Development experts may excel in areas such as algorithm efficiency, mathematical modeling, data handling, and software engineering, while domain experts may bring a deep understanding of biological questions, data interpretation, limitations, and potentially better curation of control datasets. Consequently, the success of a tool may rely more on the integration of diverse forms of expertise rather than the specific departmental affiliation of its developers. This blending of skills can mitigate potential disparities between departments, resulting in comparable outcomes regardless of the department of origin.

The primary objective of this study is to investigate whether the academic department affiliation of a corresponding author is associated with the accuracy of the bioinformatic software tools they develop. We aim to determine if tools developed by authors from certain academic fields, such as "Bioinformatics", "Computer Science" or "Genetics" show higher or lower accuracy compared to those from other fields. To achieve this, we analyzed a benchmarked corpus of bioinformatic software tools and evaluated their accuracy based on the academic affiliations of their developers.

Results

We are interested in exploring the relationship between the accuracy of bioinformatic software tools and the academic fields of study of their developers. Using a published corpus of accuracy rankings for bioinformatic software tools, we mapped these tools to their respective academic fields and evaluated how software accuracy correlates with the developers' academic affiliations.

We used a published supplement [7] to obtain previously benchmarked accuracy rankings and mapped corresponding authors' addresses to a standardized list of specific 'fields of study' [8]. These were further grouped into broader general fields and areas of expertise. Figure 1 illustrates the number of tools representing each general and specific field ($N \geq 10$). Most bioinformatic software is developed by corresponding authors who list Genetics, Bioinformatics, Computer Science, or similar departments as their primary address (Figure 1A). Among the general fields, the Biological Sciences produce the most software tools, followed by the Computer Sciences (Figure 1B).

The mean proportion of wins (e.g., if tool 'A' outranks tool 'B', this counts as a win for tool 'A') and the corresponding Z-scores provide a method to rank fields of study based on the relative accuracy of bioinformatic software benchmarked between 2005 and 2020, compared to the expected number of wins for random groupings (i.e., $wins = 0.5$). A higher proportion of wins and a greater $(-1) * Z - score$ indicate that tools from a particular department more frequently outperformed competing tools in independent benchmarks.

The specific and general field that outperformed all others is "Medical Informatics", a branch of "Technologies", with 16 software tools categorized under both fields. Among these, five are different parameter options for the MAFFT multiple sequence alignment tool [9]. These Medical Informatic tools have a mean proportion of wins of 0.70 and 95% confidence interval of $0.53 - 0.85$, which excludes the null value of 0.5. The corresponding Z-score is -1.88, and after correcting for multiple testing, the P-value is 0.29.

At the other end of the spectrum is "Bioinformatics". Ironically, authors that list "Bioinformatics" in their address field tend to produce less accurate software for bioinformatic applications. The mean proportion of wins was 0.43, with a 95% confidence interval of $0.33 - 0.53$. The corresponding Z-score is 1.20 and after correcting for multiple testing the P-value is 0.46.

The general field of "Engineering" also ranked low, with a mean proportion of wins of just 0.34 and a 95% confidence interval of $0.16 - 0.60$. The corresponding Z-score is 1.25. This general field comprises several smaller specific fields, such as "Bioengineering and Biomedical Engineering," "Computer Engineering," and "Electrical and Electronics Engineering." However, since none of these specific fields individually had more than ten corresponding software tools, they were excluded from the more detailed analysis.

The remaining general and specific academic fields have confidence intervals that include the null value of 0.5, and relatively modest Z-scores that range from -0.49 to 0.96. The P-values for each were greater than 0.05.

For the highest level field classifications of software development expert, biological domain expert or interdisciplinary expert each had similar mean proportions of wins (0.51, 0.49 and 0.46 respectively). The interdisciplinary experts had a lower Z-score of -0.87, which corresponds to $P = 0.46$ after correcting for multiple testing.

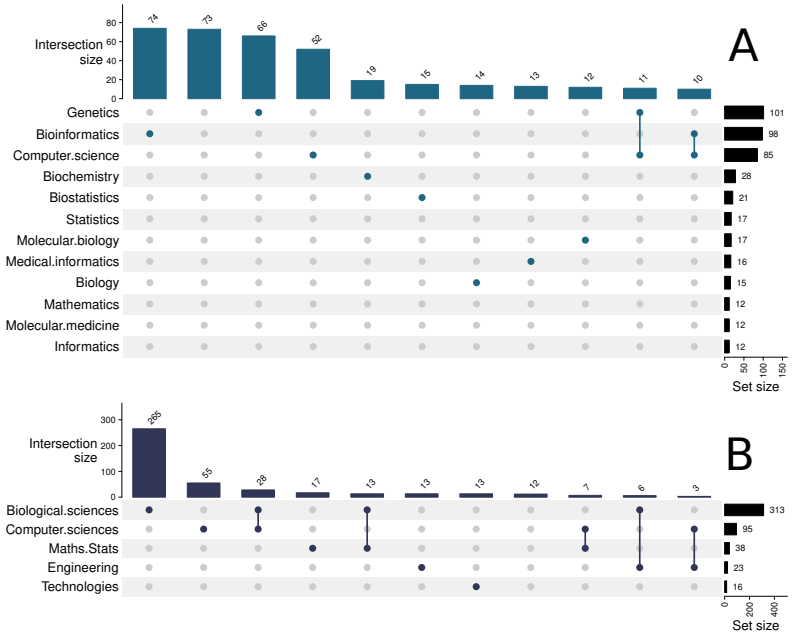


Figure 1. The number and intersection of general (**A**) or specific (**B**) fields that have contributed to bioinformatic software tools included in recent benchmark studies. The number and intersection of specific fields.

Conclusions and Limitations

We tested the assumption that the speciality of academic departments reflect the quality of the research software they produce. After correcting for multiple testing, we found no significant association between academic expertise type and the accuracy of bioinformatic tools, suggesting that academic department affiliation does not correlate with software quality. Similarly, both general and specific research fields showed no significant associations (Figure 2).

Our earlier paper found that a long-term commitment to keeping software updated was the primary factor associated with accurate software tools. This current study complements that finding by demonstrating that citation metrics (e.g. journal impact factors and author H-index), tool age, tool speed, and now academic fields of inquiry are **not** associated with software accuracy.

We focus on software tool accuracy here [10]. While speed, usability and some features of software tools are important, in our opinion the primary concern for bioinformatic software tools is the accuracy of the results they produce. As poor predictions may have long-term consequences for our general research field.

Medical Informatics, under the broader category of "Technologies," is identified as the top-performing group in developing accurate bioinformatic software tools. The tools include a number of methods for structural variation detection, single-cell profiling, long-read assembly, multiple sequence alignment and are derived from several different research teams.

Bioinformatics and Engineering ranked lower in terms of software accuracy. Tools developed by authors who affiliated with "Bioinformatics" typically had slightly lower accuracy than that of other fields. However, this was not a statistically significant finding. In addition (further bad news for this author), "Biochemistry" was similarly ranked, again this was not a statistically significant finding.

This leads us to conclude that an individual's host department are not reflective of the quality of software that they are capable of producing ($p > 0.05$ in all instances). As a consequence, the academic department should not be used as a proxy for judging the potential of software development projects.

Limitations: Some benchmarks rank multiple tool options, which can introduce modest to large effects as these options may not be independent. Additionally, the accuracy metrics used are diverse, and some may be flawed under certain conditions; for example, "accuracy" can be misleading with large class-imbalances [11], and the N50 metric for sequence assembly has been criticized by some commentators [12]. Moreover, some benchmarks are relatively small, so even small changes in rank can have a substantial impact on the proportion of wins for intermediate rankings. Lastly, the cohort of benchmarks has not been updated to include more recent results.

There is often a disconnect between a developer's training and their departmental affiliation, as evidenced by this author whose diverse academic and professional background spans mathematics, bioinformatics, computer science, molecular biology, and genomics, yet currently lists their affiliation with a Biochemistry Department. The "Biochemistry" departmental label does not accurately reflect the author's training or recent

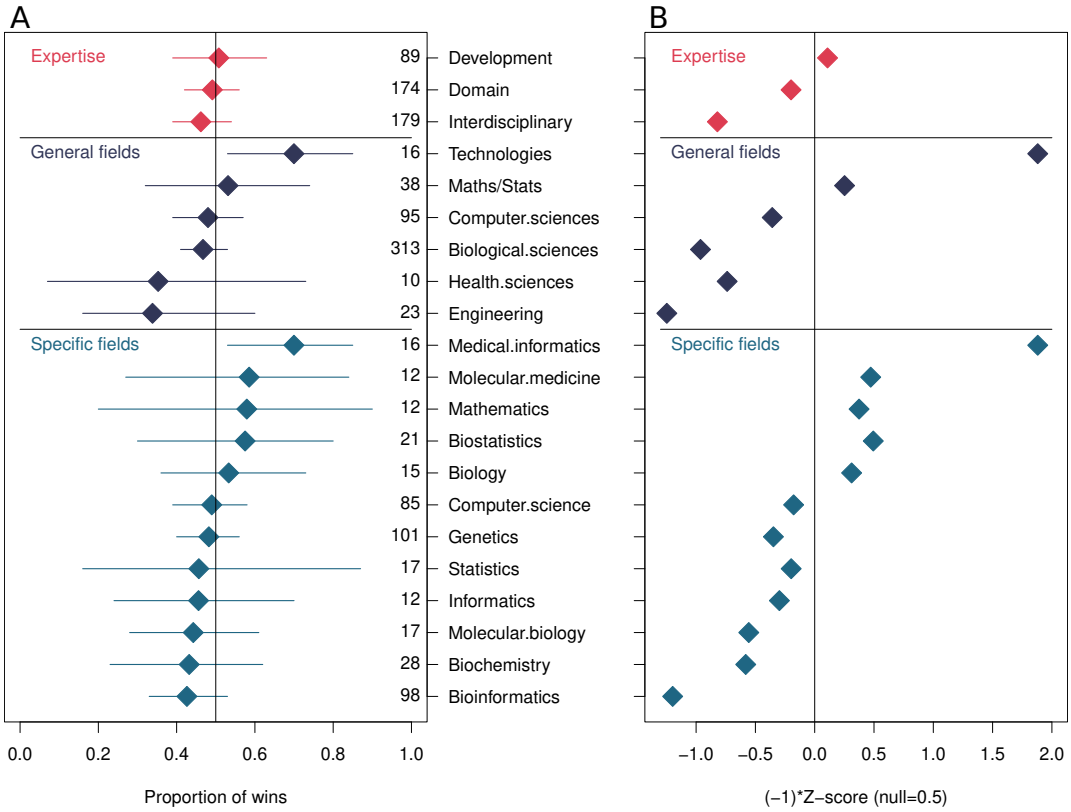


Figure 2. (**A**) A forest plot, illustrating the mean and 95% confidence intervals of the proportion of times software tools published by a given field "win" in pairwise comparisons. Confidence intervals and the mean was determined using a bootstrapping procedure. Within each field the entries have been sorted by the mean number of wins. The sample size for each field is indicated by the column of numbers on the right of the figure. (**B**) A Z-score was computed for each distribution of bootstrap samples for each field. The expected proportion of wins for randomly selected groups of tools was used as "x" (i.e. null=0.5).

publications. In fact, the author admits to not recalling key details of glycolysis, the citric acid cycle, or the structures of amino acids and proteins.

The last, or corresponding, author is typically the principal investigator who leads a project. They may have limited direct involvement in the development of any software tool, while their role is mainly to provide resources and overall direction for the project. However, there is likely to be a significant overlap between the department of the first author, who is usually the primary developer of the tool, and the department of the last author (though this was not tested in this study). Therefore, we expect that the results will be broadly similar if first-author departments were analyzed instead.

In conclusion, this study does not find strong evidence that the academic department affiliation of authors is associated with the accuracy of bioinformatic software tools. This challenges the assumption that departmental expertise correlates with software quality. Future research could explore other factors, such as the nature of interdisciplinary collaborations or the training of developers, to better understand what factors contribute to the development of high-quality bioinformatic tools.

Methods

Pre-registration: This study's desired sample size, included variables, hypotheses, and planned analyses were pre-registered on the Open Science Framework prior to any unpublished data being collected [7].

Benchmarking data: software ranks from previously gathered benchmarks are publically available [13], these include data from 68 publications that rank the accuracy of different sets of 498 distinct software tools.

Mapping tools to academic field: For each software tool, the corresponding publication(s) were identified, and the addresses of the primary corresponding author were manually extracted when available. If an author listed multiple addresses, only the first two were used. In cases with multiple corresponding authors, the last corresponding author was chosen.

The department names of the authors were mapped to the closest associated "fields of study" as defined by the National

Science Foundation [8]. We analysed these fields at three hierarchical levels: first, specific fields (e.g. "genetics", "computer science", "bioinformatics" etc), which were then mapped to broader general fields (e.g. "biological sciences", "computer sciences" etc). Thirdly, we categorized them into three types of expertise: **development experts**, **domain experts** and **interdisciplinary experts**. Development experts, from fields such as computer science, mathematics, and engineering, are expected to bring relevant expertise in software engineering and the mathematical modeling of biological problems. Domain experts, from the biological and health sciences, are anticipated to possess detailed knowledge of their subject area and to be invested in producing high-performing software for their research needs. Interdisciplinary experts come from fields such as bioinformatics, biostatistics, and biomathematics, and also include researchers who list both development and domain expertise (e.g. "Computer Science" and "Genetics"). We have treated some fields as synonymous; for example, "Computational Biology" was mapped to "Bioinformatics", and "Genomics" is mapped to "Genetics".

We restricted all subsequent analyses to fields that contain at least 10 software tools in our benchmark corpus. This mitigated against potential issues due to small sample sizes.

Statistical analysis: The accuracy data is derived from benchmarks using a diverse number of metrics that include sensitivity, specificity, PPV, FDR, error rates, AUROC, MCC and others [10]. The number of tools ranked in any benchmark ranged from 3 to 50. In order to obtain a representative measure of accuracy for a field that accounts for the diversity in accuracy measures and number of ranked tools, we employed a rank-based and bootstrapping strategy. We randomly sampled, with replacement, sets of 200 tools from the total of 498 tools. For each tool, a corresponding benchmark was selected at random, and the number of times the tool "won" against another tool was recorded, along with the total number of pairwise comparisons made. These counts of wins and total comparisons were then assigned to the corresponding specific, general, and expertise areas. This process was repeated 1,000 times to estimate the mean proportions of wins for each field, along with a 95% confidence interval for these values (Figure 2A). Additionally, we calculated a Z-score for each field to determine the number of standard deviations the mean number of wins deviates from the expected null value of 0.5 for randomly grouped tools (Figure 2B).

$$Z - score = \frac{x-\mu}{\sigma}$$

Where μ is the mean, σ is the standard deviation, x is the raw value. In this case we set $x = 0.5$ as this is the null expectation for the proportion of wins for randomly grouped sets of tools. For the purposes of illustration we plot $(-1) * z$ so that the direction is the same as for the "proportion of wins" forest plot (Figure 1).

P-values are computed from the absolute value of the Z-scores to evaluate if any field is significantly distinguished from the null i.e. $P[X > x]$. The P-values are corrected for multiple testing by controlling the false discovery rate method [14].

Data and analysis scripts availability

The data, scripts, figures and manuscript draft files are availble at the GitHub repository:
https://github.com/ppgardne/departments-software-accuracy

Funding: This research was supported by the MBIE data science platform "Beyond Prediction: explanatory and transparent data science", Genomics Aotearoa, and Marsden Grants 19-UOO-040 and 20-UOA-282.

Authors' contributions: this work was conceived by PPG, designed by PPG, the analysis was carried out by PPG, interpretation of the results was by PPG, the manuscript was drafted by PPG. All authors revised and edited the final manuscript. All authors approve the submitted version.

References

[1] Ben R. Martin. What's happening to our universities? *Prometheus*, 34:7–24, 2016.

[2] Paul Bourke and Linda Butler. Institutions and the map of science: matching university departments and fields of research. *Research Policy*, 26(6):711–718, 1998.

[3] C A Ouzounis and A Valencia. Early bioinformatics: the birth of a discipline–a personal view. *Bioinformatics*, 19(17):2176–90, Nov 2003.

[4] S R Eddy. "antedisciplinary" science. *PLoS Comput Biol*, 1(1):e6, Jun 2005.

[5] Paulien Hogeweg. The roots of bioinformatics in theoretical biology. *PLoS computational biology*, 7(3):e1002021, 2011.

[6] Fulvio Mazzocchi. Scientific research across and beyond disciplines: Challenges and opportunities of interdisciplinarity. *EMBO reports*, 20(6):e47682, 2019.

[7] P P Gardner. Pre-registration: Which department makes the best software?, 2024. `https://doi.org/10.17605/OSF.IO/92PTZ` [10 July 2024].

[8] IPEDS Completions Survey; National Center for Science Department of Education, National Center for Education Statistics and Survey of Earned Doctorates. Engineering Statistics. Classification of fields of study, 2014. `https://ncsesdata.nsf.gov/sere/2018/html/sere18-dt-taba001.html` [Accessed: July 2024].

[9] Kazutaka Katoh and Hiroyuki Toh. Recent developments in the mafft multiple sequence alignment program. *Briefings in bioinformatics*, 9(4):286–298, 2008.

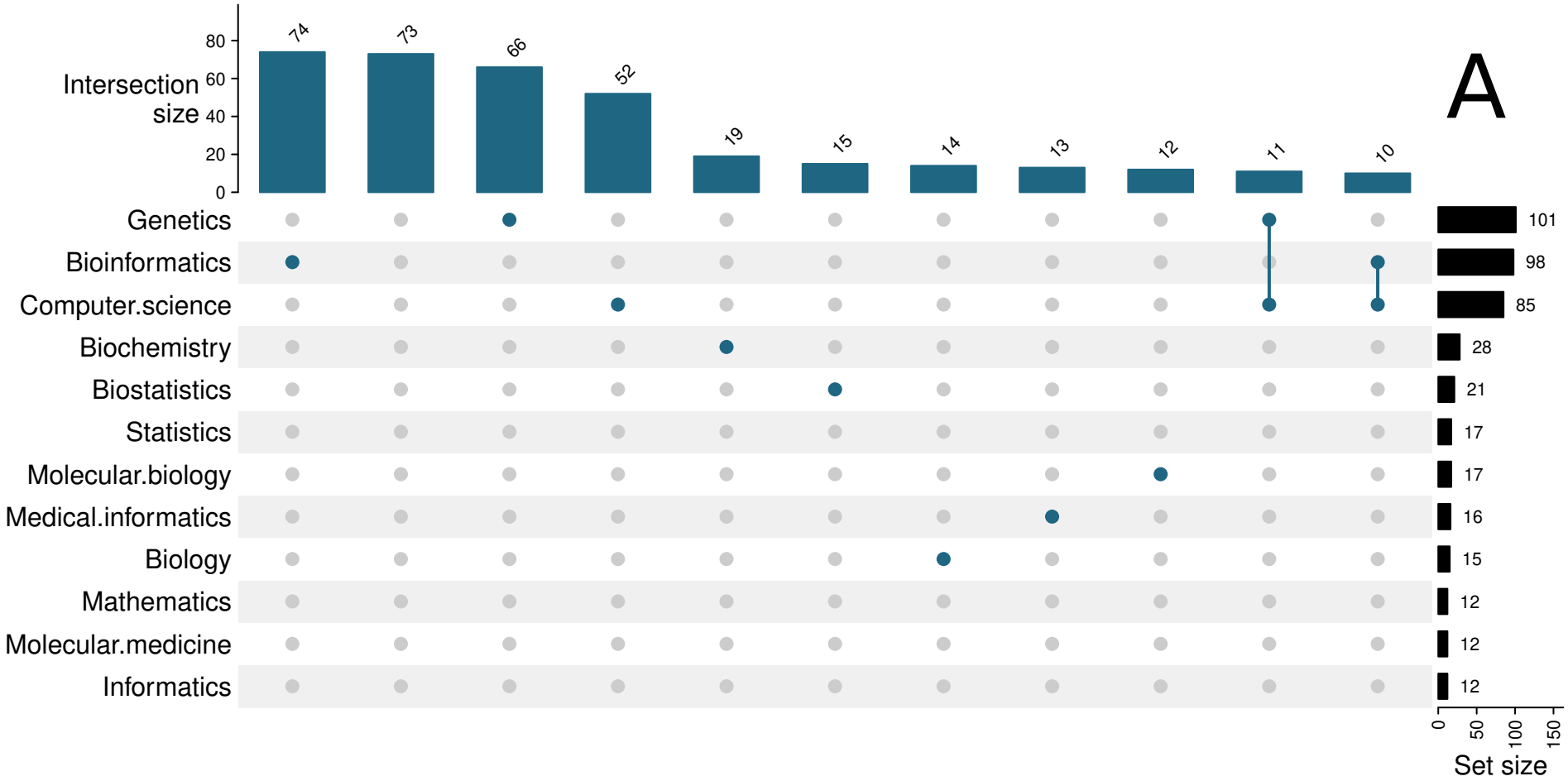
[10] Lukas M Weber, Wouter Saelens, Robrecht Cannoodt, Charlotte Soneson, Alexander Hapfelmeier, Paul P Gardner, Anne-Laure Boulesteix, Yvan Saeys, and Mark D Robinson. Essential guidelines for computational method benchmarking. *Genome biology*, 20:1–12, 2019.

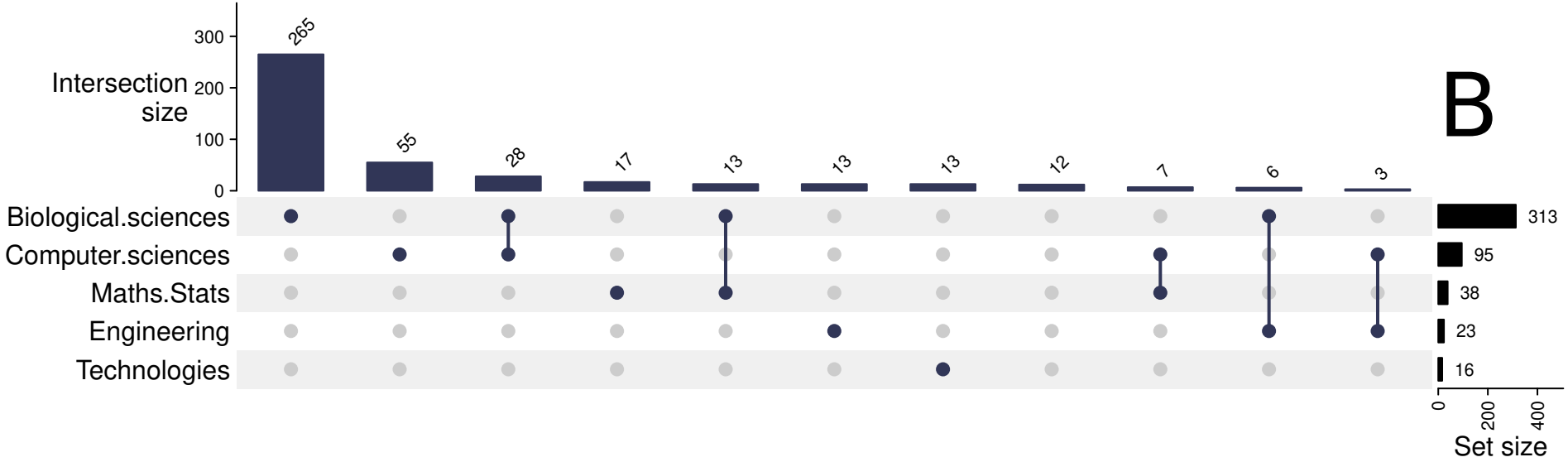
[11] Amalia Luque, Alejandro Carrasco, Alejandro Martín, and Ana de Las Heras. The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91:216–231, 2019.

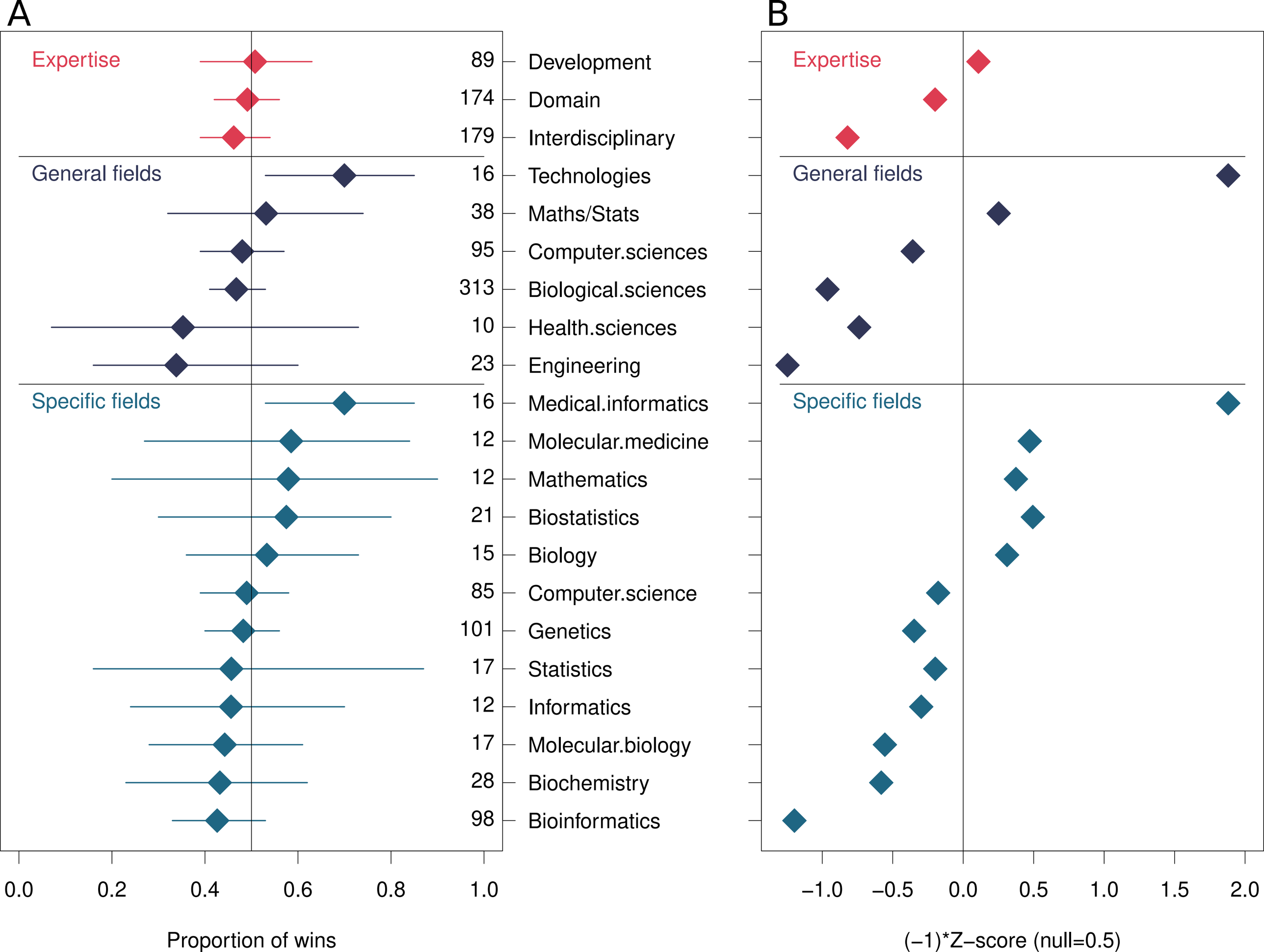
[12] Luyu Xie and Limsoon Wong. Pdr: a new genome assembly evaluation metric based on genetics concerns. *Bioinformatics*, 37(3):289–295, 2021.

[13] P P Gardner, J M Paterson, S McGimpsey, F Ashari-Ghomi, S U Umu, A Pawlik, A Gavryushkin, and M A Black. Sustained software development, not number of citations or journal choice, is indicative of accurate bioinformatic software. *Genome Biol*, 23(1):56, Feb 2022.

[14] Yoav Benjamini and Yosef Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57(1):289–300, 1995.


A
Intersection size
0
20
40
60
80
74
73
66
52
19
15
14
13
12
11
10
Genetics
Bioinformatics
Computer.science
Biochemistry
Biostatistics
Statistics
Molecular.biology
Medical.informatics
Biology
Mathematics
Molecular.medicine
Informatics
101
98
85
28
21
17
17
16
15
12
12
12
0
50
100
150
Set size


B
Intersection size
0
100
200
300
265
55
28
17
13
13
13
12
7
6
3
Biological.sciences
Computer.sciences
Maths.Stats
Engineering
Technologies
313
95
38
23
16
0
200
400
Set size


A
B
Expertise
General fields
Specific fields
89 Development
174 Domain
179 Interdisciplinary
16 Technologies
38 Maths/Stats
95 Computer.sciences
313 Biological.sciences
10 Health.sciences
23 Engineering
16 Medical.informatics
12 Molecular.medicine
12 Mathematics
21 Biostatistics
15 Biology
85 Computer.science
101 Genetics
17 Statistics
12 Informatics
17 Molecular.biology
28 Biochemistry
98 Bioinformatics
0.0
0.2
0.4
0.6
0.8
1.0
Proportion of wins
−1.0
−0.5
0.0
0.5
1.0
1.5
2.0
(−1)*Z−score (null=0.5)