

## EDUCATION

## Ten simple rules for computational biologists collaborating with wet lab researchers

Mark D. Robinson<sup>1,2</sup>\*, Peiyong Cai<sup>1,2</sup>, Martin Emons<sup>1,2</sup>, Reto Gerber<sup>1,2</sup>, Pierre-Luc Germain<sup>1,2,3</sup>, Samuel Gunz<sup>1,2</sup>, Siyuan Luo<sup>1,2</sup>, Giulia Moro<sup>1</sup>, Emanuel Sonder<sup>1,2,3</sup>, Anthony Sonrel<sup>1,2</sup>, Jiayi Wang<sup>1,2</sup>, David Wissel<sup>1,2</sup>, Izaskun Mallona<sup>1,2\*</sup>

**1** Department of Molecular Life Sciences, University of Zurich, Zurich, Switzerland, **2** SIB Swiss Institute of Bioinformatics, Zurich, Switzerland, **3** D-HEST Institute for Neuroscience, ETH Zurich, Zurich, Switzerland

\* These authors contributed equally to this work.

\* [mark.robinson@mls.uzh.ch](mailto:mark.robinson@mls.uzh.ch) (MDR); [izaskun.mallona@mls.uzh.ch](mailto:izaskun.mallona@mls.uzh.ch) (IM)

## Abstract

Computational biologists are frequently engaged in collaborative data analysis with wet lab researchers. These interdisciplinary projects, as necessary as they are to the scientific endeavor, can be surprisingly challenging due to cultural differences in operations and values. In this Ten Simple Rules guide, we aim to help dry lab researchers identify sources of friction and provide actionable tools to facilitate respectful, open, transparent, and rewarding collaborations.



## Introduction

Many computational biologists (or biologists doing applied computational analysis, or statisticians making inferences from biological data, or physicists modeling biological processes) are often faced with “collaborative data analysis” projects. Here, we specifically use the term collaborative data analysis instead of support or service to better represent our role in the scientific endeavor. The perspective represented here is that of a dry lab research group primarily engaged in assessing and developing new computational tools to process, interpret, explore, and make inferences from complex molecular data. In theory, effective development of computational methods happens in harmony with the analysis of collaborators’ “wild caught” data and all the baggage that comes with cross-disciplinary multi-research-group collaborations (e.g., egos). In the ideal case, our contributions glue together robust streams of evidence for fascinating and complicated biological phenomena; in other cases, one could perceive such work as a rather unrewarding and pointless publish-or-perish endeavor.

Despite our primary role as methodologists and that we often spend only a minority of our time on active collaboration, this cross-culture working arrangement can sometimes be a greater source of tension than the “blissful” world of pure method development. Complementary to the advice given to experimentalists working with dry lab researchers [1], bioinformaticians working in facilities [2], or running cross-disciplinary collaborations including support [3,4], we have gathered these 10 simple tips. We aim to unpack not only the pain points, forking paths, and mundane though important logistics of everyday collaborative data analysis (Fig 1), but also to reveal the wild and random though nonetheless inspiring adventures that a day in the life of a collaborative computational biologist brings.

## OPEN ACCESS

**Citation:** Robinson MD, Cai P, Emons M, Gerber R, Germain P-L, Gunz S, et al. (2024) Ten simple rules for computational biologists collaborating with wet lab researchers. *PLoS Comput Biol* 20(6): e1012174. <https://doi.org/10.1371/journal.pcbi.1012174>

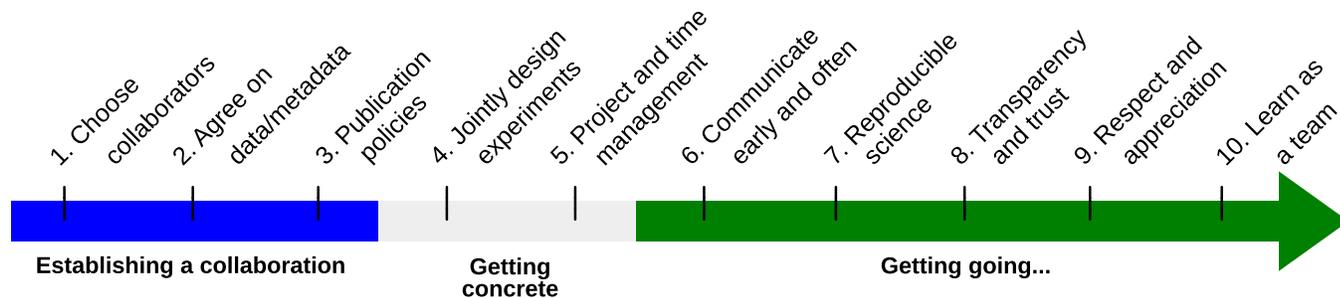
**Editor:** Scott Markel, Dassault Systemes BIOVIA, UNITED STATES

**Published:** June 20, 2024

**Copyright:** © 2024 Robinson et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The authors received no specific funding for this work.

**Competing interests:** The authors have declared that no competing interests exist.



**Fig 1. Typical timeline of a collaboration between experimentalists and analysts.** The different steps are coarsely grouped in 3 phases on the timeline, but some steps are regular processes throughout the collaboration.

<https://doi.org/10.1371/journal.pcbi.1012174.g001>

## Rule 1: Choose your collaborators wisely

*“It is better to be alone than in bad company”*

–George Washington

Our experience has been that good collaborators are those that are engaged not only in their own domain, but also have a thirst for knowledge about computational aspects. In many ways, this mirrors our interest in collaborative data analysis: We are tasked with handling data analysis aspects, but we endeavor to understand as much as possible of the biological context of the data. Less attractive collaborators are those that simply want a table of *P*-values or a set of figures; worst-case scenario are those that perceive that we simply press a button to do so.

Choosing collaborators represents the first fundamental step for collaboration between wet lab and dry lab researchers. In particular, we suggest to carefully discuss scientific interests and values with potential collaborators; discrepancies are harder to navigate later on. Two questions should be considered. First, is there an adequate scientific match in the collaboration (e.g., for us, analysis of a certain type of omic data)? That is, do the groups complement each other well in terms of scientific interests and skills? Second, are (research) values between groups aligned? Meaning, do both groups have broadly similar ideas of how research is to be conducted day-to-day? For example, is excellence defined as a Cell/Nature/Science paper, or is it robust and high-quality science?

In practice, collaborations originate between researchers who know each other already (e.g., in the same institute). Thus, the scientific match is usually there or can be easily established. Ascertaining the alignment of research values is more difficult and sometimes requires months and years of working together to fully comprehend.

In addition, choosing collaborators can also be thought of as a process that repeats itself at natural stopping points, such as the finishing of a project (stage). It can be advantageous to evaluate at regular intervals whether all parties are getting what they had hoped out of the collaboration and re-discuss the collaboration if not.

In our experience, most friction arises from having different expectations that are not clearly expressed nor negotiated. To detect and smoothen possible misalignments (for this and other rules), we suggest both teams fill an expectations form (such as [S1 Table](#)), compare them, and discuss conflicting views early on. Explicit is better than implicit.

Actionable items: Openly discuss expectations ([S1 Table](#)); be transparent and agree on funding sources (i.e., ideologically loaded private foundations).

## Rule 2: Agree on data and metadata structure

*“With big data comes big responsibilities”*

–Kate Crawford

In omics data analysis, there are typically many stages to an analysis, each one having its own type of data (e.g., raw data, filtered data, normalized data, modeling, inference). How and what data are shared internally between the project members during the project should be agreed upon, i.e., how are data shared between collaborators, what (intermediate) data can be shared, data formats that can be easily read by all researchers. The FAIR principles [5] on data sharing offer good guidelines and can be useful later when publishing the open data of the project.

Sensitive data (frequently patient-related personal data) represent a special case as they require additional protection with data storage and sharing. A data sharing agreement should be determined between the collaborators. Analysts working on these datasets are obliged to take steps in accordance with the law and the signed agreement. Normally, the agreement specifies who is granted access to the data; the purpose of the data analysis; what measures need to be taken (e.g., access, logging) to ensure the privacy and integrity of the data.

Clear standards for metadata format and style are often lacking (even among some fields of computer science), which can often lead to misleading or erroneous data and ultimately to biased results. A good practice is to agree on a common and systematic way of sharing the metadata, which can both be easily accessible and modified by experimentalists, and easily parsed by analysts, and consistent enough to be integrated in an analysis workflow.

Actionable items: Follow the FAIR guidelines (<https://www.go-fair.org/fair-principles/>); define file naming policies; be aware of sensitive data handling regulations in your country (e.g., GDPR in Europe [6,7], PIPEDA in Canada [8], HIPAA in the US [8]); follow good practices for spreadsheets [9].

## Rule 3: Define publication policies

*“The only good ideas are the ones I can take credit for”*

–Richard Stevens

Explicitly ask your collaborator’s expectations on research dissemination, openly express yours and reach consensus. Expectations to discuss include strategies on paper publishing, conferences, and other deliverables, such as software; in particular, basic ground rules for author ordering should be broached.

For academic papers, discuss potential target and no-go journals, (posting and updating) preprints, and open access. For other deliverables, particularly software and data analysis workflows, discuss intellectual property, copyright, open software licenses, and whether these deliverables can be disseminated independently of the main collaboration paper or not. A common model that our research group accomplishes is 2 manuscripts: one primarily biological where we are second and second-last authors; and the other primarily methodological, where we are first and last authors.

Discuss author order expectations early, particularly for first author, senior, and corresponding roles. Also consider co-contributions, co-correspondence, and flexibility for future updates (i.e., new people joining the team). Do not neglect the ordering within co-authors, if applicable. Be aware of possible sources of biases during this negotiation, including gender,

PI's favoritism, seniority, or tradition (i.e., wet labs taking precedence over dry lab). The same considerations apply to senior and corresponding authors.

Actionable items: Embrace author contribution taxonomies like CRediT [10] (<https://credit.niso.org/>), re-evaluate the policy in case of major change of contributions or because of new collaborators joining while the project is ongoing.

#### **Rule 4: Jointly design the experiments**

*“No one believes an hypothesis except its originator but everyone believes an experiment except the experimenter”*

–William Ian Beardmore Beveridge

Under experimental design, we are referring to the process of defining the general workflow to reach a certain outcome such as testing a hypothesis, answering a biological question, or demonstrating a new experimental assay. This encompasses primarily the wet lab experiments but can also include data analysis. Ideally, the project members would define the core aspects of the experimental design prior to collecting data. As a first step, the goal of the research project should be clearly formulated and agreed upon. Both teams should then be involved in the definition of all steps of the experimental design. This may happen at different levels depending on their area of expertise, but all members should have a basic knowledge on all steps of the workflow, including wet lab and dry lab. After defining the experimental design, you should agree with your collaborators on a rough analysis plan. After defining the main steps of the experimental design and analysis plan, a rough timeline should also be generated (see also Rule 5: Agree on project and time management). Finally, the minimal desired outcome (such as initial data analysis, final figures, etc.) should also be defined.

The agreed-upon timeline may be reevaluated after obtaining the initial results. In this case, any changes in the experimental design and/or analysis should be discussed and agreed upon by members of both the wet lab and dry lab.

Often, dry labs are approached by wet labs to initiate a collaboration (see Rule 1: Choose your collaborators wisely) after the research question has already been defined and initial wet lab experiments have been performed (and sometimes after primary bioinformatics analysis has been conducted). Starting a collaboration at this stage is often risky and harder than being engaged from the start of the project, since strong expectations for the project may already be set by the wet lab researchers. You should then evaluate whether the proposition from the wet lab users makes sense. If needed, they should suggest additional controls/experiments that would be helpful for downstream data analysis or to reach the final aim of the project.

Actionable items: Include test and pilot experiments in your experimental design.

#### **Rule 5: Agree on project and time management**

*“If it's your job to eat a frog, it's best to do it first thing in the morning. And If it's your job to eat two frogs, it's best to eat the biggest one first”*

–Mark Twain

Define together individual tasks and responsibilities, the days or working hours allocated to that project (for communication and urgent tasks to perform), as well as time horizons (in calendar months) for these and for the collaboration more generally. This is also important to get a grasp of

the complexity of tasks beyond your expertise that one might underestimate. Plan with buffer time for unforeseen complications. Explicitly check your expectations are aligned (S1 Table).

To ensure mutual awareness and help monitor progress, it can be useful to keep an up-to-date project plan where everyone can see what was done and what is currently being worked on by whom. If there are digressions or changes of plan, discuss them with the collaborators ahead of engaging in many hours of work. Since most people involved will also have a number of other engagements (and holidays!) with loads that vary over time, it is important to inform each other, ahead of time if possible, of the varying availability for collaboration. Regular meetings, either planned or in response to new data or emerging problems, are also critical to a smooth collaboration (see Rule 6: Communicate early, openly, and often enough).

Actionable items: Use online project planners (i.e., Trello, GitHub projects); use multiuser online text editors (i.e., Google Docs, Overleaf) to draft manuscripts and other documents.

### Rule 6: Communicate early, openly, and often enough

*“The single biggest problem in communication is the illusion that it has taken place”*

–George Bernard Shaw

Establishing a productive and respectful communication strategy is essential for fruitful and low-friction collaborations. To ensure clarity, a mutual expectation agreement is also needed regarding project length, cost sharing, preferred communication channels, and frequency, as well as expected response times. Regular meetings, prompt feedback (for a mutually agreed definition of prompt), and attending relevant meetings all help maintain alignment on communication. Communication should not only take place when you need something concrete from your collaborators, but sharing and feedback require continuous cycling. However, all collaborators should be empathetic and flexible in meeting plans or project progress due to other responsibilities (teaching duties, other projects, personal reasons) and with mutual respect to others' work hours (see Rule 5: Agree on project and time management). Especially after long or complicated discussions, or when multiple people were involved, it can be helpful to prepare a summary of important discussion points and decisions to be circulated in written form (e.g., Google Docs or via a Slack or Mattermost channel), to represent a common understanding and have a memento of the conversations.

Actionable items: Adhere to effective meetings [11] and written interactions guidelines [12].

### Rule 7: Follow open and reproducible science guidelines

*“The way of the world is to make laws, but follow custom”*

–Michel de Montaigne

Project members should follow reproducible research practices (e.g., use FAIR file formats, version control, free and open source software, good coding practices, and reproducible reporting [5,13,14]) to ensure that data analysis is transparent and everyone can recreate your results and findings. Before a project starts, agree on the desired degree of public data availability (see also Rule 2: Agree on data and metadata structure) and how (and if) to report summary statistics of sensitive data. Typically, journal policies require authors to describe their data and code availability; this leaves room to share anything from raw data to intermediate results, as well as unstructured code to analysis reports to fully reproducible pipelines. Make sure your

collaborators agree with your high commitment to transparency and reproducibility. As a project goes on, it becomes increasingly important to keep all data, software environments, online code, and reports tractable. Ensuring code and results (figures and tables) are in accessible states and versioned enables everyone to stay updated on discussions, failed attempts, and outcomes. Similarly, literate programming [15], e.g., generating reports with both code and results in open formats (HTML, PDF), facilitates scientific reproducibility and open access. Reproducible computational research is crucial for guaranteeing the consistency and transparency of scientific discoveries. Conversely, a lack of reproducibility threatens all findings.

Since the expertise of both sides differ, it is important to agree on an adequate strategy for sharing results and adapt it to expectations and technical abilities. Although sharing the code among analysts is (or should be) a standard, reports shared with experimentalists could be easier to digest when limited to figures, main results, and conclusions. The format to share results should also be clearly defined (meetings, static reports, dynamic reports, etc.) and always discussed between the 2 parts to promote a clear understanding of the collaboration's outputs.

Actionable items: Use version control (git) and push frequently to remotes shared with your collaborators (i.e., GitHub, BitBucket, GitLab); include reproducible and versioned software installation steps in your code repositories; create an analysis workflow that can easily be rerun after a change in the data or in the analysis.

### **Rule 8: Establish the transparency and trust required for constructive feedback**

*“Far better an approximate answer to the right question, which is often vague, than an exact answer to the wrong question, which can always be made precise”*

–John Tukey

Project teams should jointly engage in thorough sanity checks throughout the research. This involves regularly questioning and testing various aspects of the project, in both wet and dry lab domains (see Rule 7: Follow open and reproducible science guidelines). It might include repeating experiments to confirm their consistency and using different analysis approaches to explore robustness of results to analysis choices. Both sides should be transparent about the specifics and the output of their work, including unsuccessful attempts, failed controls, irreproducible, or not-straightforward-to-interpret results. The reporting of unsuccessful results not only helps to build trust and transparency between both sides, but also helps to discuss alternatives and improve the (wet lab or dry lab) process. Results should be shared fully, e.g., raw western blots, raw imaging data, and all quality control checks. The goal of such a continuous policy of critical thinking and sanity checking is to maintain integrity and build trust among collaborators and to confirm that both the experimental setups and computational analyses are functioning as expected. By routinely challenging the integrity and interpretability of wet lab and analysis results, potential misalignments can be avoided.

Actionable items: Share all raw and processed data; listen to and acknowledge dissenting opinions.

### **Rule 9: Be respectful and show appreciation**

*“By mutual confidence and mutual aid great deeds are done, and great discoveries made”*

–Homer

Acknowledge that everyone who has agreed to take part in the collaboration has their validity and represents a different aspect of the scientific endeavor. There is no role that is per se more important than the others, no matter how niche it is. Mutual respect of each other's work and (domain-specific) knowledge should be a given. In turn, this does not rule out but rather includes the questioning and critical assessment of each other's results on a scientific level in case of doubt (see Rule 8: Establish the transparency and trust required for constructive feedback). Limitations of both scientific approaches and also personal availability for the project should be clearly stated and respected by the other parties involved (see Rule 6: Communicate early, openly, and often enough). Polite communication between collaborators is required at all times, be it when discussing potentially unsatisfying results or agreeing on a time schedule for future steps. Be kind: do not forget to regularly acknowledge and thank others for their work, and explicitly remind you and others the project could not be carried out alone.

Actionable items: Be respectful; be aware of common cognitive biases, including the impostor syndrome [16] and Dunning–Kruger effect [17].

### Rule 10: Learn as a team

*“Question everything. Learn something. Answer nothing”*

–Euripides

The more you know about what the other side is doing and why, the easier the collaboration. It is of great help if everyone involved has both a basic understanding of the different experimental and computational components involved as well as the willingness to gain a deeper understanding of each others methodology as the collaboration proceeds. What can we really interpret from a PCA plot or how does the library preparation protocol influence sequencing results? This common knowledge is not a requirement at the start of a collaboration but can be established during the project. Treat knowledge gaps as an opportunity to learn something new. This should be practiced at all levels, irrespective of level of seniority. It will make the common work and communication a lot easier. For example, it will save the analyst a lot of time if the metadata are machine-readable, in turn a wet lab collaborator might gain more independence if the analysis scripts and outputs are easily accessible to them and well documented. Take advantage of the diversity of backgrounds as it provides different perspectives on a problem that can offer new approaches to solve problems.

Actionable items: Add explanatory comments to the analysis code; do not shy away from asking what you do not understand; commit to answering basic and advanced questions about the analysis.

### Conclusions

In any situation when multiple (scientific) cultures collide, there can be tension and misunderstandings at some stage and there is no simple formula to navigate all the personalities and egos involved. Let us not forget that along the way, all members of a collaboration need to absorb a vast assortment of private as well as work pressures, power differentials, deadlines, and commitments. Nonetheless, these differences and cultural quirks are a reality to be observed, studied, and understood such that we can embrace the diversity of scientific mind-sets. We also note that the goal is not necessarily to pursue fully harmonious collaborations, because stumbling through our misunderstandings helps us crystalize norms for collaborations and define our core values of doing science. This may lead to lifelong collaborations or also to those that will never happen again.

Collaborative data analysis is not our primary role, but in many cases becomes the springboard to new methodological projects. The 10 simple tips discussed here are largely about being open, organized, empathetic, professional, practical, systematic, and fair.

## Supporting information

**S1 Table. Collaboration expectations form.** The idea here is that there are no good answers (well, sometimes there are), but that each party fills the form and then compares their answer, to hopefully align expectations. Some questions are phrased rhetorically, but meant to explicitly prepare all members for a collaboration's potential tension points. To fill put an "x" in the appropriate column, where 3 is "agree equally with both statements," 1 is "strongly agree with statement on left," and 5 is "strongly agree with statement on right."

(PDF)

## Acknowledgments

We thank our past collaborators for their often-pleasurable, sometimes-painful contributions that led to to distilling these rules.

The content of this manuscript was brainstormed during a lab retreat in the scenic town of Aeschi bei Spiez, Switzerland.

## References

1. <https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1008281> (Ten simple rules for biologists initiating a collaboration with computer scientists).
2. Aron S, Jongeneel CV, Chauke PA, Chaouch M, Kumuthini J, Zass L, et al. Ten simple rules for developing bioinformatics capacity at an academic institution. *PLoS Comput Biol*. 2021; 17(12):e1009592. <https://doi.org/10.1371/journal.pcbi.1009592> PMID: 34882684
3. Knapp B, Bardenet R, Bernabeu MO, Bordas R, Bruna M, Calderhead B, et al. Ten simple rules for a successful cross-disciplinary collaboration. *PLoS Comput Biol*. 2015; 11(4):e1004214. <https://doi.org/10.1371/journal.pcbi.1004214> PMID: 25928184
4. Kumuthini J, Chimenti M, Nahnsen S, Peltzer A, Meraba R, McFadyen R, et al. Ten simple rules for providing effective bioinformatics research support. *PLoS Comput Biol*. 2020; 16(3):e1007531. <https://doi.org/10.1371/journal.pcbi.1007531> PMID: 32214318
5. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016; 3(1):160018. <https://doi.org/10.1038/sdata.2016.18> PMID: 26978244
6. Shabani M, Marelli L. Re-identifiability of genomic data and the GDPR: Assessing the re-identifiability of genomic data in light of the EU General Data Protection Regulation. *EMBO Rep*. 2019; 20(6):e48316. <https://doi.org/10.15252/embr.201948316> PMID: 31126909
7. Data Protection Working Party. Advice paper on special categories of data ("sensitive data"). Article 29 of Directive 95/46/EC, Ares 444105. 2011.
8. Phillips M. International data-sharing norms: from the OECD to the General Data Protection Regulation (GDPR). *Hum Genet*. 2018; 137:575–582. <https://doi.org/10.1007/s00439-018-1919-7> PMID: 30069638
9. Broman KW, Woo KH. Data organization in spreadsheets. *Am Stat*. 2018; 72(1):2–10.
10. Allen L, Scott J, Brand A, Hlava M, Altman M. Publishing: Credit where credit is due. *Nature*. 2014; 508(7496):312–313. <https://doi.org/10.1038/508312a> PMID: 24745070
11. LeBlanc LA, Nosik MR. Planning and leading effective meetings. *Behav Anal Pract*. 2019; 12(3):696–708. <https://doi.org/10.1007/s40617-019-00330-z> PMID: 31976280
12. Gruber J, Somerville LH, van Bavel JJ. A scientist's guide to email etiquette. 2020. <https://doi.org/10.1126/science.caredit.abb2664>
13. Sandve GK, Nekrutenko A, Taylor J, Hovig E. Ten simple rules for reproducible computational research. *PLoS Comput Biol*. 2013; 9(10):e1003285. <https://doi.org/10.1371/journal.pcbi.1003285> PMID: 24204232

14. Stawarczyk B, Roos M. Establishing effective cross-disciplinary collaboration: Combining simple rules for reproducible computational research, a good data management plan, and good research practice. *PLoS Comput Biol.* 2023; 19(4):e1011052. <https://doi.org/10.1371/journal.pcbi.1011052> PMID: [37104248](https://pubmed.ncbi.nlm.nih.gov/37104248/)
15. Knuth DE. Literate programming. *Comput J.* 1984; 27(2):97–111.
16. Clance PR, Imes SA. The imposter phenomenon in high achieving women: Dynamics and therapeutic intervention. *Psychother Theory Res Pract.* 1978; 15(3):241.
17. Kruger J, Dunning D. Unskilled and unaware of it: how difficulties in recognizing one's own incompetence lead to inflated self-assessments. *J Pers Soc Psychol.* 1999; 77(6):1121. <https://doi.org/10.1037//0022-3514.77.6.1121> PMID: [10626367](https://pubmed.ncbi.nlm.nih.gov/10626367/)