

# What Is the Key Best Practice for Collaborating with a Computational Biologist?

## Go in with an Open Mind



**Anne Carpenter**  
Broad Institute of Harvard and MIT; CellProfiler

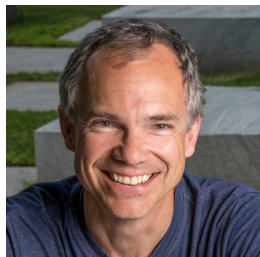
Go in with an open mind: you just might become computational yourself!

I certainly did. I made the transition from cell biologist to computational biologist out of necessity: my genome-wide screen for novel cell-size regulators was intractable using existing image analysis software. A search of the computer science literature revealed algorithms likely suitable... but the papers lacked raw code, much less user-friendly software. So I dove in, learned MATLAB, collaborated with a graduate student from MIT (Ray Jones), and the rest is history: CellProfiler was born, and I launched my own, entirely computational laboratory a few years later.

Might you be lured in a computational direction? After years of wrestling with cells that grow inconsistently or become contaminated, you just may fall in love with the satisfying reproducibility of computational work. Another attraction: the remarkable power. In image analysis, years of tedious visual examination become a few hours via an automated pipeline. Algorithms can also tease out subtleties that would elude manual analysis. Computational work is particularly attractive if you like a team approach to science, and enjoy playing an indispensable role as part of a larger effort. Seeing a biologist overjoyed at the results of computational analysis is a great job perk.

Any amount of computational literacy you gain will be useful, given the increasing scope and scale of experiments in modern biology.

## Make the Loop Tight



**Sean R. Eddy**  
Harvard University; HHMI

Try not to! Especially if by “collaborating,” you mean “hey, analyze my data for me.” Is there a computational research question in your project, or are you just putting off learning how to handle data on a computer?

*Experiments and data analysis are an iterative loop, not a straight line.*

I was recently reading a lovely passage in Richard Hamming’s book *The Art of Probability for Scientists and Engineers*, where he talks about a problem called Bertrand’s paradox. Draw a random chord in a circle, and inscribe an equilateral triangle in the same circle; what is the probability that the chord is longer than the triangle’s sides? There are three different correct answers (1/2, 1/3, and 1/4)—because it turns out that there are three different justifiable ways to randomly sample a chord, and it matters which way you choose. Hamming riffs on this to talk about the danger of decoupling the person asking the question from the person implementing the analysis. “All too often the actual randomness is left to some programmer who is interested in programming and not in the relevance of the model to reality.” Which choice was appropriate? Only the person asking the question knows that. You sharpen your questions in response to what you see in your data. Experiments and data analysis are an iterative loop, not a straight line. Make the loop tight. You may be surprised how easy it is to learn to analyze your own data.

## Partnership, Trust, Compromise



**Paul Flicek**  
European Molecular Biology Laboratory; European Bioinformatics Institute

Think long term and strive for collaboration as partnership.

The best projects have a computational lead from my lab and an experimental lead from another lab working together. They determine which experiments succeed or fail, go over data, suggest new ways forward, and puzzle out the interpretations. As partners.

Partnership takes work and takes compromise: if one of the experimental replicates is bad, it must be redone; if the computational approach is not appropriate, it needs to be replaced. Tasks or decisions may appear easier and faster if done alone or without consultation, but worthwhile projects don’t need these shortcuts because they are not done in a few days or a couple of weeks, they take months (or years). Over the course of that time the person doing the experiments and the person doing the analysis need to work seamlessly and they need to understand each other’s strengths and weaknesses. They need trust.

Of course, working with a computational biologist doesn’t always require such close coordination or months of work, but the concepts of partnership and trust do not change. When someone approaches me to collaborate, my goal is to find out if they are more interested in the science or more interested in some software that I can run for them. Do they care about my interpretation of the data or are they really just interested in the size of my computer cluster? While, we can assist by running software (and we do so regularly), true collaboration is something much more.

**Overcome the Language Barrier**

**Melissa Gymrek**  
University of California, San Diego

In a data heavy world, biologists can no longer ignore the UNIX terminal and must work as part of multidisciplinary teams with computational expertise. These collaborations may be challenging at first: computational and wet lab biologists speak different languages and have distinct scientific cultures.

The key best practice boils down to communication. Everyone should have a clear understanding of the biological question, computational challenges, and project goals. Both sides must feel there is a genuine collaboration, rather than a “core service” provided by one party. Design experiments and analyses together and prepare for multiple iterations. It helps to work in close proximity to promote frequent communication. For the wet lab biologist: learn how to translate a biological question into a computational challenge. For the computational biologist: write code that is well-documented and understandable by collaborators. Have open discussions on authorship policies, which vary between fields.

The importance of multidisciplinary science is rapidly growing, and we need to continue taking steps to facilitate this shift. We are training new scientists to straddle both worlds, who can design and perform experiments, write and understand code, and communicate with diverse team members. Moreover, publication policies are changing to better acknowledge important intellectual contributions of computational biologists, with more opportunities to publish and cite code and stronger emphasis on reproducible analyses.

**No Question Is Too Basic**

**Molly Hammell**  
Cold Spring Harbor Laboratory

I love the term computational biology over other terms that have cropped up to describe this field, like “bioinformatics” or others. The reason is that I see myself as a true biologist, with computational techniques being the particular approach I use for answering questions about biology. I hope the day will arrive soon when all biologists have a basic fluency in the mathematical, statistical, and programming skills that comprise computational biology. In the meantime, I have really enjoyed the collaborations between other “wet” labs and mine.

The best collaborations have been those where we talk openly and frequently about the project—even from the early planning stages. It’s so much easier to design an experiment to be well powered to detect a signal, rather than trying to come in at the later stages of a project and decide what questions one is powered to ask. This also gives the wet-bench students in the group a better understanding of the most important statistical considerations that need to be included in good overall experimental design.

Finally, one of the best parts of science involves just talking to one another, brainstorming about potential interpretations and proposing new avenues to explore. The key is to have those discussions in an open and supportive environment where no questions are considered too basic. Sometimes, in answering the simplest questions, we are reminded where the untested assumptions lie that might be blocking the next big discovery.

**Iterative Optimization**

**Khuloud Jaqaman**  
University of Texas Southwestern Medical Center

My experience is largely in the realm of quantitative analysis of experimental biological data, especially light microscopy data. Therefore, I will focus on that aspect of computational biology. I would say the key best practice for conducting a “quantitative microscopy” study is the iterative, back-and-forth optimization of both the experimental and computational approaches so that they build on each other’s strengths and complement each other’s limitations. With this, the combined approach is truly integrative and arguably stronger than the sum of its two parts.

The optimization and feedback go both ways: You collect initial data. You build initial software and apply it to the data. You determine where things fail, and then optimize both experiments and analysis software to improve things. You do this multiple times until things are “as good as it gets.” I have for example customized particle-tracking software to track the peculiar movement of a specific cell-surface receptor. But just like experiments have their realities and limitations, so do analysis tools. Therefore, my collaborators have also had to spend time altering image acquisition parameters to produce images that are analyzable by the available software. In the end it is always a balance, a compromise.

Needless to say, this iterative optimization requires a great deal of communication (frequent and regular!), patience and understanding of “the other side” in order to find common grounds.

### Trust and Mutual Co-education



**Jeremy L. Jenkins**  
Novartis Institutes for BioMedical Research

“This might be a naïve question, but...” Trust is often built on such humble statements. Nobody has mastered all of science; collaborations between “wet” and “dry” lab scientists flourish when each person understands enough of both disciplines to be conversant, yet leans on each other’s expertise. New hypotheses will arise by bleeding over into each other’s worlds through shared perspectives. To do so requires that the computational biologists are treated as collaborators, not just as project support. Data scientists are in high demand with many options for projects. It is therefore good practice to cultivate a partnership with a computational biologist where there are common interests and mutual co-education, with scientific credit flowing freely between.

The computational collaborator should be involved early so that experimental design effectively frames the resulting analysis. Sit together in front of a computer, iterating on data interpretation in order to create visualizations that tell a narrative. File formats that are primed for humans to read may not be good for computers. Consider the heavy burden of file munging, or reformatting of data for integration. Standardizing results and experimental metadata terms (with controlled vocabulary or reference ontologies) will save labor and time and facilitate future meta-analyses across projects. And finally, try to resist referring to computational analysis as “waving a wand”—it is science, not magic, after all!

### An Untapped Mode of Thinking



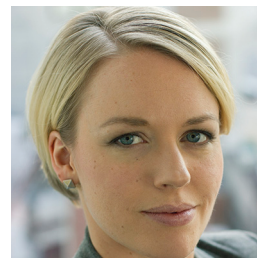
**Daphne Koller**  
Coursera

The advent of high-throughput experimental protocols have transformed biology into a discipline where massive amounts of data are produced constantly. These data are a treasure trove of information, much of which lies untapped, because there is just way too much to assimilate without powerful computational tools. Thus, a computational biologist is critical in extracting novel, meaningful insights from these data.

*Computing...is a mode of thinking.... Much greater benefit can come from viewing a computational biologist as a thought partner.*

Often, a computational biologist is viewed as someone who merely executes a well-defined analytic process specified by the biologist. However, this role considerably under-leverages a true computational expert. Computing is not just a set of tools—it is a mode of thinking that can provide new ways of looking at the data, and even suggest new questions that one could answer. Much greater benefit can come from viewing a computational biologist as a thought partner: give them context as deep as possible about the biological setting; explain the data that you have collected and the kinds of insights it might provide; and provide them with guidance on what might make for an impactful biological finding. Then, with that framing, let them loose to innovate on analysis and methods. Importantly, this process must be iterative: alternative models and results will need to be evaluated for biological plausibility, novelty, and impact. But when done well, such a collaboration will result in new types of analysis that provide insights of a type that you might never have expected.

### Crystal Clear Questions



**Tuuli Lappalainen**  
New York Genome Center; Columbia University

Biology as a field has very diverse array of methods and approaches, but all biologists share the passion to tackle biological questions founded on hypotheses and models of how living systems work. Thus, when experimental and computational researchers want to build a successful collaboration, the most important step is to make sure that the biological model and the fundamental questions of the study are crystal clear to everyone. This is not always easy: computational biologists working on large dataset often think of generalizations that can be applied to specific examples, whereas the molecular biology approach is often the opposite. But you should spend as much time as necessary on bidirectional communication on the null and alternative biological hypotheses, the underlying mechanisms and assumptions, how they would manifest themselves in the data, and biases and artifacts. Statistical models are guided by biological models and properties of the data, and thus molecular biologists need to communicate these aspects as clearly as possible to computational biologists. Finally, in all collaborations, mutual respect, willingness to ask and answer many questions, and shared intellectual curiosity are essential for good and innovative science and a healthy scientific community.

**Partnership from the Beginning**

**Alicia Oshlack**  
Royal Children's Hospital

In this era of high-throughput biology, major advances are being made through the collaboration of computational biologists and laboratory-based scientists. In my field, new genomic technologies are driving a rapid transformation in our understanding of many aspects of biology with breakthroughs that rely heavily on bioinformatics.

The key for having the best collaborations, that make the most of the technologies, is to establish a partnership between research groups from the beginning of the project. One aspect that is often underappreciated is that computational biologists have a depth of knowledge that can aid in refining the biological questions that can be addressed with a given technology. A partnership from inception will also ensure that the design of the experiment is relevant and robust and has the potential to provide maximum insight from the generated data. A partnership relies on clear communication and establishing a common language by recognizing that you can learn from each other's expertise. Once a partnership is established it is important to respect the knowledge, effort and experience a computational biologist brings to your project through prominent authorship, citation and funding.

**The Fruits of Understanding**

**Ron Shamir**  
Tel Aviv University

**Question:** You are a biomedical researcher and you need the help of a computational biologist. What should you do?

**Answers:** Define your goals quantitatively. Experimentalists often have a clear general agenda but no concrete quantifiable goals in mind. While today most computational biologists “speak biology,” they are unlikely to have the deep context you have on your favorite problem. Take time together to understand what is computationally doable and practical. Method/tool development is a process that takes realistic concrete goals, and often lots of trial and error.

Design the experiment together. Start discussing the project with your computational colleague long before the experiment. This will prevent pitfalls of insufficient statistical power, lack of validation data, etc.

Describe your data generation in detail, including possible biases, limitations, and any suspect confounding factors. Tons of times (and sometimes whole projects) can be lost rediscovering bits of information mistakenly thought unimportant.

Dialog! Maintain an ongoing discussion with your computational colleague throughout the project. If you just send over your data or problem description and wait for the solution—most likely it will be inadequate. Create a framework for meeting regularly and discussing experimental and computational progress. A joint student is a great facilitator here.

Understand the key computational steps in the project. Make your colleague explain them to you until you feel confident with them. Only if each side fully understanding the other, your collaboration will bear the expected fruits.

**Forethought, Full Integration**

**Mona Singh**  
Princeton University

Scientists today use high-throughput experiments to investigate the inner workings of cells at the systems level. In many cases, experiments at the cutting edge require novel computational methods to analyze the resulting large-scale data, arguing for an approach where the computational methods are designed in concert with the experiments themselves. Experimental and computational biologists can collaborate effectively by getting together early on to discuss their scientific agendas, agree upon the most important research questions, and cooperate in the experimental design. All methods have inherent strengths and limitations, and these are best understood by the experts who develop and apply them. An experimentalist will know what measurements are feasible, and may have insights into the sources and level of technical and biological noise present in the data. Likewise, a computational biologist can ascertain what types of conclusions may be possible from analyses of these data, how much data is necessary and what types of new algorithms are needed. As the research progresses, she can also help assess how to best leverage existing datasets and perform preliminary analyses to uncover which new data will be the most useful in advancing their shared research goals. In the most successful collaborations, computation should not be regarded as an afterthought or as a savior for an experimental project, but as a fully integrated and critical component.



### Respect and Reciprocation



**Sarah Teichmann with Tzachi Hagai and Mike Stubbington**  
Wellcome Trust Sanger Institute; EMBL-EBI

First, anticipate in advance of carrying out the experiments that you would like to collaborate, and involve the collaborators from the beginning. You might be surprised at the extent to which the computational biologist can contribute to designing experiments, which will also engage them more deeply in the joint project and lead to better scientific outcomes.

Second, it should go without saying that collaborators need to respect each other. It can be difficult for experimental biologists to understand that anything beyond the most basic analyses will require quite a lot of bespoke, specific code to be written. Think of each new way of analyzing the data as performing an experiment which will take some time (rather than clicking a button and getting an instant result).

*...be open to the idea that you might one day provide experimental support to a computational biologist with a cool idea....*

Third and finally, reciprocate the favor. If a computational biology group has supported one of your projects, then be open to the idea that you might one day provide experimental support to a computational biologist with a cool idea, even if it is a little tangential to your interests. As the saying goes, what goes around, comes around.

### Joint Investment Early On



**Fabian J. Theis**  
Technical University Munich

“We have a stunning new data set, with [favorite omics]—can you please model this?” Collaboration requests like these are well-known to computational biologists, indeed they used to be the start of interactions for quite some time. They stem from the belief that if we observe a biological process with high enough accuracy and ideally the latest measurement technique, something will be found. In practice, however, this something is often not more than a set of weak correlations, which, even though drawn in a pretty graph, are not satisfying.

In recent times this has changed, due to many small steps improving communication between the experimental and computational researcher, which often determines success of a project. However in my opinion the most useful best practice is the instrument of joint supervision. By deciding to invest into a computational collaboration, e.g., by co-funding a shared PhD student, both the theoretical as well as the experimental partner are forced to interact right from the start. There will be a joint project plan even before data is being generated, there will be regular project meetings, and there will be thesis advisory meetings where the student and the PIs join to unify their goals. This ensures that generated data is useful for the modeler and that developed models result in biological insight. Eventually, a common language is found in both labs, and computational biology will simply be one more tool to do biology.

### Value Field-Specific Thinking



**Olga Troyanskaya**  
Princeton University; Simons Foundation

The key to successful collaborations between biomedical researchers and computational biologists is open, very frequent interactions and communication. Starting as early as possible in the project, regular check-ins and discussions, whether in person or video conferencing, are essential.

In these interactions, both sides should be flexible about goals and generous (and patient!) with explanations and discussions, keeping in mind that each side is not a “native speaker” in the other’s field. Each must educate the other about each field’s unique thought processes, challenges, and approaches—through multiple explanations with concrete examples, the reasoning behind decisions, and the intuition underlying important methods. Openness and give-and-take are critical, as well as understanding the broader context of the specific biological challenge. While “analyze my data” requests are most common, such collaborations are maximized through discussions that lead to deeper thinking about the data, important follow-up experiments, and ideally, development of methods that have an impact in the biomedical community beyond the specific problem at hand. Successful collaborations require patience, openness to questions beyond the biologist’s immediate goal, and willingness of the bioinformatician to work on many iterations of specific, often computationally “routine” analyses. Perhaps most importantly, both sides should never be afraid to ask “basic” questions and bring up potentially “naïve” ideas—such discussions can unearth hidden problems in analyses, lead to exciting discoveries, and facilitate future collaborations.