Overview

In these tips, we focus on research projects that fall into the *Methods development* category. This means that the main contribution of the paper is often methodological; however, the results should include some novel domain-specific findings that were enabled by the new method.

There are many different types of methods development projects. For example:

- *Old data, new methods*: The data has been looked at already, maybe in many different ways. But you are developing methods that allow new signals to be identified in these data to find what others have not found yet.

- *New data, new methods*: You have a new, one-of-a-kind data set!

- *New data, old methods*: Perhaps no one has yet thought to apply this model to this problem in data analysis yet; you also present a new data set

- *Old data, old methods*: no one has thought to apply this model to these data yet, even though both have been analyzed. What will this show in these data that other methods haven't already?

Selecting a topic

With your data set chosen, you should now discuss with your group what you would like to look at in these data. Be very specific.

- *Comparing results from different available methods*: What software and models perform best for identifying underlying ancestral structure in genotype data?

- *Testing a specific hypothesis*: Do lincRNAs have ubiquitous, dosage-dependent trans-effects on transcription?

- *Identifying underlying structure*: What is the ancestry of individuals? Can I identify signals of selection in genotypes? Can I find protein-protein interactions, or identify drugs that work for disease for which they were not originally designed?

- *Searching for specific relationships*: Which genes are differentially expressed in smokers? What networks only exist among gene expression levels in smokers?

- *Develop a method to perform a specific task*: Can I write a model that will allow me to predict methylation status at a specific site?

Most importantly: can these data be used to support the chosen research topics? If there is not signal in the data, another topic or another data set should be chosen.

Finally, motivate this work. If you answer these questions, what will be the contribution to the project domain of the selected topics? What will be the contribution to computer science and computational biology? What will be the contribution to statistical genetics?

Selecting a Journal

Your choice of project should be decided on what interests you the most. When deciding on where to publish your work, however, a number of considerations come into play. There are many factors upon which a journal may be selected for a manuscript. These include:

- Topic: does this journal publish papers on related topics? Will their readers find your project interesting?

- Your career path: do you want a job in a stats department? Publish papers in stats journals.

- Open access: will your journal be put behind a paywall, or made public?

- Impact factor: do people read and cite articles in this journal?

- Review process: is this journal known for a careful but quick review process? Does it allow preprints?

- Overshooting: is this the type of work that belongs in this journal? If not– either in impact or in style – often it is better to go to a lower impact journal so you avoid resubmitting the manuscript many times.

Once you have selected a journal, the appropriate google search will allow you to find the most recent LaTeX templates for manuscript preparation and submission. Often the manuscript format for submission is suboptimal for preparation (e.g., figures at the end or submitted separately). Ignore the format and develop the manuscript in a way that is easy to read and prepare. At the time of submission, reread the submission format information.

Some scientific journals make the claim that they do not support LaTeXsubmissions. I have never found this to be true in practice. Everyone accepts PDFs, as far as I know.

Some journals that publish genomics work:

- *Science, Nature, Cell*

- *Nature Methods, Nature Genetics*

- *PLOS Computational Biology, PLOS Genetics*

- *Genome Research*

- *Genome Biology*

- *Bioinformatics*

- *PeerJ, eLife*

- *PLOS One*

- *Journal of Machine Learning Research (JMLR)*

- *Annals of Applied Statistics (AOAS), Biostatistics*

You might also consider sending a shorter manuscript to a conference that accepts papers. Be careful to get the submission dates right; these only happen once a year:

- RECOMB

- ISMB

- Pacific Symposium on Biocomputing

- NIPS

- ICML

- AI-Stats

- UAI

You have a project; now what?

Now you should decide with your group how to organize the work that will go into this project. At a minimum, you will need:

- Set up a git repository for your project. Include folders: data, code, docs

- Commit the .tex and .sty files for your manuscript (but not the LaTeX intermediate files, or the PDF!). Start a supplemental information document also in a readable format (e.g., I like the PLOS-one template for supplements; JMLR also has a good readable template).

- Put your data in the /data folder

- Start your iPython docs in the /code folder. You might separate these out into specific analyses: data preprocessing, visualization, methods application, simulations, and processing results.

- Write, in the Methods section, a subsection detailing your data set (where you got it, when, who developed it, on what platforms they developed it, and the numbers and features that it contains). More on writing a Methods section later; it is important to write this section as you go.

- Write a title. Acknowledge that it will most certainly change.

Project definitions resources

- Loman & Watson (2013) [So you want to be a computational biologist?] (http://www.nature.com/nbt/journal/v31/n11/full/nbt.2740.html)

- [Google Scholar] (https://scholar.google.com/): search the literature

- [PubMed] (http://www.ncbi.nlm.nih.gov/pubmed): search the scientific literature

- [SciReader] (http://scireader.org/): helps find relevant research to read about