Dear MVA students,

I am a third year PhD student in Electrical Engineering and Computer Sciences co-advised by Michael I. Jordan and Nir Yosef at UC Berkeley. I am writing this email as a call for applications to a research internship between the **Statistical Artificial Intelligence Lab** and the **Center for Computational Biology** of **University of California at Berkeley**.

We are looking for interns with a strong mathematical and statistical background; experience in designing and conceiving machine learning algorithms with scripting languages; ease with continuous integration and version control tools; and most importantly, eagerness to make an impact by applying advanced statistical methods to genomics data.

Some of my fellow friends from Ecole polytechnique went through your MVA program a few years ago and I also know some of the professors. I believe that you have a great profile to come have fun with us at Berkeley.

Please find more detailed information in the following pages,
Good luck for the applications, looking forward to seeing you in Berkeley !

Romain Lopez

# Research

Our research group focuses on the development of machine learning tools to interpret single-cell RNA sequencing datasets. Namely, we have built scVI, a Bayesian model thanks to which most of the important tasks in the field can be addressed via posterior sampling. We have developed a scalable approximate posterior inference procedure using stochastic variational inference and neural networks. Lately, we have worked on improving our model to integrate datasets from different laboratories and different technologies (which is literally the immediate hottest problem in the field), which we formulate as a domain adaptation / counterfactual inference problem. Our research is published / to appear in the top machine learning conferences and the top computational biology journals. Our algorithm is currently used by computational biology experts on real data to (hopefully) make new discoveries.

# Selected References

- codebase: https://github.com/YosefLab/scVI;
- our main computational Biology manuscript (accepted in principle to Nature Methods) https://www.biorxiv.org/content/early/2018/09/23/292037;
- an example of methodology work (to appear in NIPS 2018) https://arxiv.org/abs/1805.08672;
- another example of methodology work (20 min presentation at Facebook for the Bay Area Machine Learning symposium, from last year's interns!) https://arxiv.org/abs/1809.05957;

# Skills you will learn

A representative (non-exhaustive) skill set on which the internship would build on:
- graphical models, Bayesian modeling, stochastic variational inference, auto-encoding variational Bayes;
- basic machine learning knowledge https://people.eecs.berkeley.edu/~jrs/papers/machlearn.pdf;
- advanced programming capabilities for machine learning (Python / R / PyTorch / Tensorflow);
- experience with fitting neural networks / variational autoencoders or any other algorithms with a sensitive number of parameters to tune;
- experience with single-cell genomics data / reading computational biology papers;
- expertise in a field of applied statistics (Bayesian hypothesis testing, approximate posterior inference, counterfactual inference, etc.).

# Internship

During the research internship, you will be part of the scVI team (including one postdoctoral fellow, two doctoral students and some undergraduate students doing summer research) and participate in weekly group meetings. You will be working on key machine learning algorithm developpements for interpreting single-cell RNA sequencing data. We have funding to support your research.

Last year, two students (Edouard from MVA and Maxime from X) visited the lab for around five months. I proposed them a set of eight projects to choose from. They worked on key problems in the field (namely merging datasets across technologies) and we are following up with them to turn their projects into journal publications. Edouard and Maxime also worked on all the project codebase maintenance (built the github repo as it is now), and did significant work on the methodology side (adapting plenty of inference methods for these deep generative models). Most importantly, they left the office with plenty of funny memes.
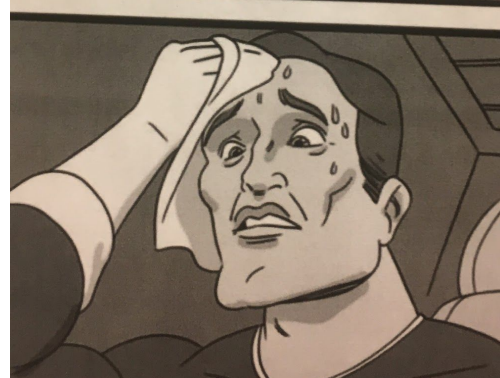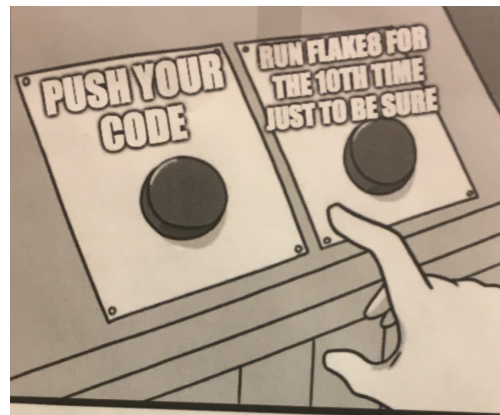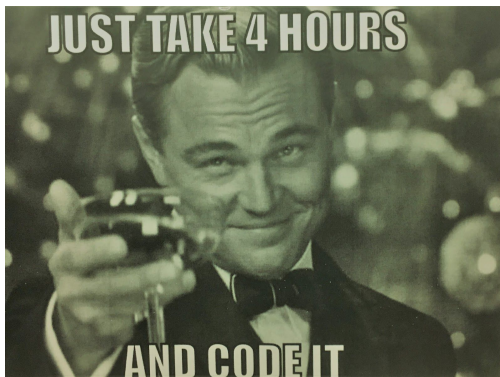
# Call for Applications

Please email me before **Monday October 22th 9AM PST** with your resume, transcripts (probably not important for making our decision) and github repo (if any). In your email, indicate why you would like to join us. As we believe that it is always more fun to work in a team and that so far we have two positions available, we will accept single or joint applications (either from a single student or a pair, no more than two).

I will contact you for a Skype interview so we can get to know each other. After that, I will design a little research project on which I will give you a few days to work on. There is nothing you need to review for this one, it will be original research content either proof-based, algorithmic-based or analytic-based. Last year, I asked some students to work out some details of our main manuscript probability bounds (that was at this time not online).

Good luck ! I'm looking forward to meeting you !