COLUMBIA | ALLIANCE
COLUMBIA UNIVERSITY | ÉCOLE POLYTECHNIQUE
SCIENCES PO | PARIS 1 PANTHÉON-SORBONNE

20 Years Ans

ALLIANCE RESEARCH INTERNSHIP PROGRAM

COLUMBIA UNIVERSITY

ACADEMIC YEAR 2023-2024

DEADLINE TO APPLY:

5 DECEMBER 2023

COLUMBIA | ALLIANCE
COLUMBIA UNIVERSITY | ÉCOLE POLYTECHNIQUE
SCIENCES PO | PARIS 1 PANTHÉON-SORBONNE

20 Years Ans

Created in 2002, the Alliance Program is an innovative joint venture between Columbia University, the École Polytechnique, Sciences Po, and Paris 1 Panthéon-Sorbonne University. Every year, Columbia University offers several student internships in scientific disciplines open to École Polytechnique students. The process for applying to these internships is outlined below.

I. Internship Description

- Students work with a faculty member, who acts as an academic advisor and supervises their research project.
  Internships will start in March/April 2024. The duration, objectives and tasks of the internship will be discussed with the supervisor at the host center or department.

- If a stipend is offered, it will be specified in the internship offer.

- Students are responsible for finding housing.

- *All students are required to apply for a J1 Visa to conduct an internship in the United States*.

II. Applications requirements

- Applicants must include: a CV, a cover letter (1 page), and a letter of recommendation.

- Students must send their application to the Alliance Program: alliance@columbia.edu

- All materials must be submitted in English.

DEADLINE – 5 December 2023
All applications must be sent to alliance@columbia.edu

IRVING INSTITUTE FOR CANCER DYNAMICS (IICD)

1. Faculty Sponsor:

Professor Simon Tavaré (Director, Irving Institute for Cancer Dynamics [IICD]; Professor of Statistics and of Biological Sciences)
Co-advisor: Khanh N. Dinh, PhD

2. **Number of interns:** Two (2)

3. **Type of support offered:**

- ✓ Stipend – $11,856 for four (4) months
- ✓ Access to campus services and facilities
- ✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

Mathematical modeling and parameter inference for genomic alterations during cancer evolution

5. **Description:**

DNA sequencing data has demonstrated an increased level of genomic alterations in cancer compared to normal tissue, in the form of point mutations, copy number aberrations (CNAs), or structural changes. There are many open questions in interpreting the data however, such as the role of whole-genome duplication in altering mutation rates and the fitness landscape, the opposing forces between driver and passenger mutations, ordering of mutational events, and the functions of different CNA mechanisms.

The intern will have an opportunity to establish a mathematical model describing tumorigenesis as driven by genomic changes, and construct a parameter inference method based on Approximate Bayesian Computation. The intern will then apply the model to existing bulk and single-cell DNA-sequencing data, with the goal to advance the understanding of cancer biology.

6. Skills:

Background in mathematics, statistics, computer science or biological sciences with a strong quantitative component. Experience in R. Prior experience in cancer research is useful, but not required.

7. Additional Information:

The IICD is an interdisciplinary institute located on the Morningside Heights campus of Columbia University and focused on the interplay between the mathematical sciences and cancer research, collaborating across disciplinary boundaries to develop tools and methods that can improve our understanding of cancer biology, origins, treatment and prevention. Our website, at cancerdynamics.columbia.edu, gives an overview of our research teams and our current projects.

COLUMBIA | ALLIANCE
COLUMBIA UNIVERSITY | ÉCOLE POLYTECHNIQUE
SCIENCES PO | PARIS 1 PANTHÉON-SORBONNE

20 Years Ans

IRVING INSTITUTE FOR CANCER DYNAMICS (IICD)

1. Faculty Sponsor: Elham Azizi

2. **Number of interns:** One (1)

3. **Type of support offered:**

   - ✓ Stipend – $11,856 for four (4) months
   - ✓ Access to campus services and facilities
   - ✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

Machine learning framework for studying spatial organization of gene regulation.

5. **Description:**

Similar to the evolution of the single-cell genomics technologies in the past decade, a new wave of multi-modal spatial technologies is emerging with the capability to measure molecular features in the context of the tissue. Despite these exciting technological advancements, there are currently no computational techniques for integrating spatial data modalities. This project aims to develop a machine learning framework for integration of spatial transcriptomics and epigenomics which can lead to insight into spatially-varying regulators and mechanisms underlying intercellular interactions. For instance, the spatial organization of gene regulation can identify novel targets for treating heterogeneous and resistant tumors.
We aim to use deep generative modeling to learn regulation dynamics across and within regions of the tumor. More specifically, we are interested in how gene regulation signals diffuse across spatial locations. Such dynamics can be modeled by either establishing neural ordinary differential equations or parameterizing a time dependent distribution through diffusion-nosing models/normalizing flows.

6. **Skills:**

Strong programming skills in Python
Machine learning
Probability and Statistics
Linear Algebra

COLUMBIA | ALLIANCE
COLUMBIA UNIVERSITY | ÉCOLE POLYTECHNIQUE
SCIENCES PO | PARIS 1 PANTHÉON-SORBONNE

20 Years Ans

COLUMBIA SURGERY – PEDIATRIC/CONGENITAL CARDIAC SURGERY

1. Faculty Sponsor: David Kalfa MD, PhD

   Director, Pediatric Heart Valve Center
   Surgical Director, Initiative for Pediatric Cardiac Innovation
   Director, Kalfa research lab

2. Number of interns: Two (2)

3. Type of support offered:

   - ✓ Stipend: $1,530 per month for four (4) to six (6) months
   - ✓ Access to campus services and facilities
   - ✓ Immigration and visa assistance/sponsor

4. Internship Title:

Innovation to treat congenital cardiac malformations

5. Description:

Projects related to the development of growth accommodating devices, innovative biomaterials for heart valve repairs in children and adults, storage and rehabilitation of living heart valves in bioreactors, computational model-based patient specific predicting tools for decision making in the care of neonates with congenital heart disease.

6. Skills:

Mechanical engineering
Polymer science
Computational model
Design optimization
Cell culture
Manuscript and grant writing.
Native English speakers are particularly appreciated

DEPARTMENT OF COMPUTER SCIENCE

1. Faculty Sponsor: Roxana Geambasu

Associate Professor of Computer Science

2. **Number of interns:** Two (2)

3. **Type of support offered:**

- ✓ Stipend: Between $1,530 to $3,000 for six (6) months (depending on the number of students selected)
- ✓ Access to campus services and facilities
- ✓ Immigration and visa assistance/sponsor

4. **Internship Title:**

Privacy-preserving analytics platform engineer

5. **Description:**

We are a group of computer science researchers at Columbia University and the University of British Columbia working to democratize access to advanced privacy technologies by incorporating support for them into popular data processing systems. One such advanced privacy technology is differential privacy, which we believe is an essential technology to increase privacy in today's data-driven world, in which users' data is avidly collected and processed through a variety of machine learning and analytics workloads aimed at improving products, targeting ads, and more.

We posit that across all of these workloads, user privacy is a critical computing resource that is being inherently consumed, but whose consumption is not being accounted for, constrained, managed, or paid for in any way. These risks exposing user data to unintended parties, such as privacy-transgressing internal employees who may dig into user databases to spy on family and friends, or external hackers who may inspect machine learning models to glean details of user data used to train these models.

Our goal is to incorporate privacy as a first-order resource into data processing infrastructure systems so it can be carefully accounted for, tracked, and constrained in ways that will give assurances of privacy to the users whose data is being collected. Differential privacy (DP) gives us the theoretical and algorithmic building blocks for defining such a privacy resource.

Using it, we've incorporated privacy as a resource into the Kubernetes orchestrator, the Tensorflow-Extended ML training platform, and most recently, into the caching components of the Tumult DP data analytics system that was used in the recent deployment of DP in the 2020 U.S. Census. Every time we incorporate the DP privacy resource into one of these infrastructure systems, we observe that its use and operation becomes substantially easier, leading to an increased chance to popularize this essential privacy technology across the numerous companies and organizations now accumulating user data. Our project website contains details of our vision, how we hope to popularize this privacy technology, and research papers and public code releases we have made so far.

We are seeking two interns to help us with the development of a new data analytics platform that will extend the already existing Tumult DP analytics engine with key privacy management components that are currently missing from that engine, but which we believe will be vital to making the system truly operational and widely adoptable.

These components include: a scheduler that arbitrates access to the privacy resource by multiple, competing tasks to ensure the fair and near-optimal use of the privacy resource, which in DP systems is a very limited resource; caching components that conserve the use of the privacy resource; a workload optimization component that identifies and eliminates redundant portions of computations among tasks for the purpose of optimizing the use of the privacy resource; and privacy budget management components that will track the use of the privacy resource so dev-ops can identify tasks that over-use this limited privacy resource. We've already developed most of these components separately, and we've shown that they can help address some very sticky problems with DP, which have dogged this technology's adoption for years (see our web website for details). It is now time to integrate these components into one operational data analytics platform, and this is what the proposed project aims to do. We are committed to releasing our data analytics platform open-source and free of charge, and we have multiple industry partners interested in it for their own deployments of DP.
Thus, if the project is successful we expect significant chance of societal impact.

The interns will work with us to design, implement, and evaluate the data analytics platform that will integrate these components and make Tumult (and DP in general) more operational and easier to deploy. A significant portion of the work will involve system design, as the appropriate architecture for the system we aim to build is not yet established, but also implementation, testing, and evaluation so we can provide artifacts that work well. Finally, we expect that theoretical analysis of the system, whose privacy properties are mathematical and must be proven formally, will comprise a sizeable portion of the work.

From interns' perspective, we see this as a rich training opportunity to participate in a multi-faceted project that builds upon their systems design, development, and theoretical skills to develop an open-source artifact that has the potential for high societal impact.

6. Skills:

Solid background in math and statistics
Strong coding skills (Python, optionally Go)
Some prior system design experience a plus

7. Additional Information:

Contact Roxana Geambasu for further information.
roxana@cs.columbia.edu