# Supervisor & Project Information Form

Please complete and return via email ONLY to gdip.hres@utoronto.ca by Monday November 2, 2020

**Supervisor Information**

*MUST have unrestricted SGS appointment (appointment to supervise graduate students)*

<table>
<thead>
<tr>
<th><strong>SGS Department</strong></th>
<th>Email: <a href="mailto:geoffrey.liu@uhn.ca">geoffrey.liu@uhn.ca</a></th>
</tr>
</thead>
<tbody>
<tr>
<td>Dalla Lana School of Public Health (Dept of Epidemiology), Medicine, Department of Medical Biophysics, Institute of Medicine Science</td>
<td><strong>Field of Research</strong>: Pharmacoepidemiology; clinical and molecular oncology</td>
</tr>
<tr>
<td><strong>Research Institution affiliation (if applicable)</strong>: Princess Margaret Cancer Centre (University Health Network)</td>
<td><strong>Location of Work</strong>: MaRS, 101 College St, Toronto, ON</td>
</tr>
<tr>
<td><strong>Student contact time (number of hours per week YOU are available to the student for any concerns or to review progress)</strong>: 1h/wk in-person meeting + available by email as needed; student will have additional support from the clinical and biostatistics groups and a scientific associate in the Liu Lab</td>
<td></td>
</tr>
</tbody>
</table>
**Project Information (will be posted on GDipHR website for student access)**

**TITLE:** Machine Learning and Biostatistics for Analysis of Symptoms and Toxicity (MBLAST) in Lung Cancer

**DESCRIPTION (MAX 500 WORDS):**

This innovative project intersects the use of complex, longitudinal clinical data with machine learning and biostatistical methods. The use of machine learning and AI for data analysis and pattern recognition has revolutionized the world we live in, enabling personalized marketing, facial recognition and driver-less cars. We aim to apply similar methods to analyze longitudinal symptoms and toxicities data from Princess Margaret thoracic cancer patients to identify prognostic/predictive symptom clusters and improve patient management and survival.

Our unparalleled access to data is the backbone of this project. Our lab is helping build the Princess Margaret Digital Health Platform (DHP) as a pilot project in lung cancer, a data lake which contains our medication ordering data (OPIS program), laboratory data, radiation data (MOSAIQ), patient-reported symptom and other clinical data (see 2BLAST catalyst project on [http://pmdatascience.ca/projects/](http://pmdatascience.ca/projects/)). We are also co-leading the lung cancer pilot for Canadian Personalized Health Innovation Network (CPHIN, [https://www.cphin.ca/lung-kick-start-program/](https://www.cphin.ca/lung-kick-start-program/)), through which we are building a comprehensive lung cancer TOSCA database (Thoracic Oncology Site Consolidated Ambispective database). Both structured and unstructured clinical data from the electronic health record of thoracic cancer patients, as well as various study and external databases including CTGA clinical trial data ([https://gicancer.org.au/clinical-trial/ncic-co-17/](https://gicancer.org.au/clinical-trial/ncic-co-17/ and others)) will be integrated and analyzed. All of this complex, multi-dimensional, longitudinal data allows for and requires innovative analysis.

This project is well-suited for a student who has strong inherent interest in oncology (medical, radiation, surgical), clinical research, programming and data science. The student will receive support from the clinical and biostatistics groups but will need to be comfortable with programming independently; Python experience is strongly encouraged.

The student will work closely with our data science team to describe, build and iterate algorithms for identifying symptoms and toxicity clusters and attributions to specific therapies and clinicodemographics. The student will also work closely with our clinical teams to contextualize the algorithms and results within the clinical framework of lung cancer, making the student a bridge between the clinical teams and our data scientists. The student will therefore be learning clinical research, oncology patient management, in combination with innovative machine learning methods. Co-Supervisors Dr. Liu and Xu have a longstanding history of supervising medical students locally, nationally, and internationally (see [www.uhncombiel.com](http://www.uhncombiel.com)) with students presenting their work in abstracts and published.
If human subjects are involved, have the appropriate Research Ethics Board approvals been obtained?

☐ Yes  ☐ No  ☐ Application Submitted (Date: ________________)

Do you expect this work will be published within the 20 months?

☐ Yes  ☐ No  ☐ Uncertain / Other

**Student Roles & Responsibilities (please be as specific as possible)**

Please indicate who will serve as the student’s direct report for daily oversight (PI, PhD student, technician, etc...)

The GDipHR student will be responsible for overseeing data management and integration of large clinical and trial datasets. The student will also be responsible for developing and applying machine learning and biostatistical methods, under the supervision of Dr. Xu, to identify clusters of symptoms and toxicities through time, their relationships to molecular alterations, treatment regimen, etc., and the predictive capability of these patterns. A small-scale proof of concept will first be completed before expanding the scope of this project. The student will meet regularly with Dr. Liu and the clinical teams to ensure that the work is grounded in clinical practice, and relevance. The student will be working with a member of the Princess Margaret biostatistics team to design, conduct and interpret analysis, as well as meeting regularly with the machine learning team to discuss results, and receive direction and guidance. The GDipHR student will be directly supervised by Cathi Brown (scientific associate), Benjamin Grant (Project Manager) in the Liu Lab and will be teamed with a post-doc in Dr. Xu’s lab. This work will be published as a foundational work from which to build out and justify future large-scale data analysis efforts through CPHIN/TOSCA.