



**Application number/Title:** 27317 - Predict cancer risk based on genomic, environmental and lifestyle factors

**Applicant PI:** Professor Edwin Wang

**Application Institution:** University of Calgary, Biochemistry and Molecular Biology, Health Sciences Centre Room 1151, 3330 Hospital Drive NW  
Calgary AB T2N 4N1, Canada

**Keywords provided by the Applicant PI to describe the research project:**

Cancer; risk prediction, algorithm, machine learning

**Application Lay Summary:**

1a: Genetic and environmental factors and their interactions determine the cancer initiation and progression. We would like to find out which factors and their interactions are critical for cancer development and then will develop computational methods which enable to predict cancer risk of an individual based on the genetic and environmental factors.

1b: 1. We will use the health related data and genotypic data collected in the bioBank to find key factors for the cancer risk

2. We will develop computational algorithms that predict cancer risk (e.g., breast, lung, colon, bladder, melanoma, and prostate) risk for healthy people

1c: Here we use ER+ (estrogen receptor positive) breast cancer (ER+ breast cancer is a subtype of breast cancer (BC), ~70% of the BC patients are ER+ BC patients) as an example.

1. we will randomly pick up 60% of the ER+ BC patients and 5 times of age-matched non-cancer people

2. we will conduct analysis to compare the genomic and other factors between the 2 groups

3. We will develop predictive models to predict the ER+ BC risk using the factors mentioned above

4. We will validate the models using the remaining ER+ BC patients and non-cancer people' data

1d: All the cancer patients will be included

5-6 time of the age-sex-matched non-cancer people (case-control dataset required)

For finding the age-sex-matched samples, we should access the whole set

Inclusions

ICD codes for breast, lung, colon, bladder, melanoma, and prostate cancers  
age > 20 yrs old.

For obtaining medical history, we need include all the ICD codes for the selected populations

Exclusions

children