ILCCO/TRICL New Proposal Form ***(version May 2011)***

Investigators who would like to propose a new research project, please complete the form
For ILCCO, please send the form back to Dr. Rayjean Hung (rayjean.hung@lunenfeld.ca)
For TRICL, please send the form back to Dr. Chris Amos (Christopher.I.Amos@dartmouth.edu)

I. Project Title

Genetic landscape analysis of short tandem repeats on the risk and survival of lung cancer

pCode: XXX

Date:
Dec 11, 2019

II. Project Group Investigators

Project leaders: Mulong Du, David C. Christiani; Harvard
Investigators: ILLCO contributing study/ data PI's

III. Background

Short tandem repeats (STRs), also known as microsatellites, consist of repeating motifs of 1–6 base pairs (bp) and comprise about 3% of the human genome. STRs polymorphism might regulate gene expression and function involved in the etiology of complex disease including cancers. We propose to investigate STR and potential functional consequences on the development and progression of lung cancer.

IV. Specific Aims

To globally evaluate the association of STRs with lung cancer risk and survival.

V. Methods

(A) Imputation of STR: use Beagle to impute STR from Gymrek lab's reference panel into whole-genome genotype data of lung cancer (*Saini et al., Nat Commun. 2018 Oct 23;9(1):4397.*).

(B) Association evaluation:

a. use logistic regression analysis to assess the risk effect of STR on lung cancer susceptibility.
b. use Cox regression analysis to assess the effect of STR on lung cancer survival.

(C) Function analysis: impute STR into genotype data matched with gene expression (databases from in-house BLCS and public GTEx and TCGA), then calculate the relationship between STR and gene to identify eQTL-STR.

VI. Materials or variables needed from the study PIs

(A) Demographic and clinical descriptions of study samples

For risk:

a. Number of lung cancer cases and controls (currently largest sample size: 29,266 cases and 56,450 controls in European).
b. Age, gender, race, smoking status (never/ever/current), package-year of smoking, histology (LUAD/LUSC/others) and tumor stage (1/2/3/4).

For survival:

a. Number of lung cancer cases (need the largest sample size in European).
b. Age, gender, race, smoking status (never/ever/current), package-year of smoking, histology (LUAD/LUSC/others), tumor stage (1/2/3/4), vital status (death/live) and follow-up time.

(B) Whole-genome genotyping data of lung cancer (if possible, all imputed genotyping data using 1KG phase 3).

VII. Time line

Once we receive the datasets from PIs, we will take 3 months to do the association analysis, and an additional 3 months to write the manuscript.

VIII. Other remarks (e.g. publication plan, etc)