|  |
| --- |
| I. Project Title and pCode |
| Whole exome sequencing analysis of lung cancer prognosis | **Date: Mar 31,** **2023** |
| **II. Project Leaders (List Names and Institutes)**  | **Primary Contact****Information** |
| Ruyang Zhang, David Christiani, Chris Amos, Rayjean Hung, MDAC PI, Baylor PI, IARC PI, Liverpool PI, MCCTH PI and other authors from consortium  | **Name:** David Christiani**Email:** dchris@hsph.harvard.edu **Tel:**617.432.3323 |
| III. Progress to date. *What has been done so far to move the project forward (data pooling, grant submission, etc), and when it was done* |
| We devise a three-phase study to identify exon variants associated with lung cancer prognosis. In the discovery phase, the sequenced or imputed (TOPMed) exon variants with significant main effects, gene-gene (GxG) interactions or gene-smoking (GxS) interactions will be screened out using samples from ILCCO-OncoArray study. Furthermore, these signals with *q*-FDR ≤ 0.05 will be confirmed in validation phase 1 using samples from TRICL study ESTHER study, Mayo study and TCGA, and in validation phase 2 using samples from Research Program on Genes, Environment and Health (RPGEH), PLCO and UK Biobank (UKB).To date, we have collected WES or GWAS data from all studies aforementioned. All genotype data has been prepared via a uniform quality control process. TOPMED reference panel 5b was used for imputation with Michigan Imputation Server. Pre-phasing using phased reference data from TOPMed release 5b was conducted using EAGLE 2.4. Imputation was conducted against the same reference panel using minimac4. Besides, we have collected databases of methylation quantitative trait loci (meQTL), expression QTL (eQTL), protein QTL (pQTL) and metabolite QTL (mQTL).The meQTL data comes from GoDMC[1], eQTL from eQTL Gen[2], pQTL from Iceland study[3],and mQTL from Metabolic-QTL Summary Data[4].Via integrative analysis of GWAS and meQTL, eQTL, pQTL and mQTL data, we also aim to detect significant associations between lung cancer overall survival and exon genetically determined DNA methylations (EWAS), gene expressions (TWAS), proteins (PWAS) and metabolites (MWAS). |
| **IV. Project Status and Preliminary Results**. *Summarize the current status of the project. Report preliminary results if there is any*  |
| We have already observed significant lung cancer survival associated exon variants based on WES and GWAS data. |
| **V. Justification of Extension** |
| With imputed exon data released from large-scale cohort studies and biobanks, it provides opportunities for systematically understanding WES variants and lung cancer survival. Originally, we only focus on several small-sample-size WES datasets. Currently, we enlarge sample size by more cohort studies and expand study model from genetic study to epigenetic, transcriptional, protein and metabolic study. Thus, we need more time to achieve our aims. |
| **VI. Future Plans & Timeline** |
| We request ONE more year for future data analysis.We need 3 months to perform GWAS analysis.We need 5 months to perform the EWAS, TWAS, PWAS and MWAS analysis.We need 2 months to build up prognostic prediction model.We need 2 months to draft a manuscript. |

[1] MIN J L, HEMANI G, HANNON E, et al. Genomic and phenotypic insights from an atlas of genetic effects on DNA methylation [J]. Nat Genet, 2021, 53(9): 1311-21.

[2] VõSA U, CLARINGBOULD A, WESTRA H J, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression [J]. Nat Genet, 2021, 53(9): 1300-10.

[3] FERKINGSTAD E, SULEM P, ATLASON B A, et al. Large-scale integration of the plasma proteome with genetics and disease [J]. Nat Genet, 2021, 53(12): 1712-21.

[4] SHIN S Y, FAUMAN E B, PETERSEN A K, et al. An atlas of genetic influences on human blood metabolites [J]. Nat Genet, 2014, 46(6): 543-50.