Prognostic biomarker identification and subgroup analysis using high dimensional inference in CAR-T cell immunotherapy trial

Qian Wu

Fred Hutch
E-mail: wuqian7@gmail.com

Abstract: Chimeric antigen receptor (CAR) T cell immunotherapy has shown remarkable efficacy in patients with relapsed CD19+ B-cell non-Hodgkin lymphoma (NHL). Durable response has been observed in a subset of patients with lymphodepletion chemotherapy followed by infusion of T cells. One of the important objectives is to identify patients' subgroups defined by one or multiple factors associated with progression-free-survival (PFS) and overall survival (OS), though most of the immunotherapy trials are still in the stage of phase I/II with limited sample size but large number of covariates, such as cytokine and manufactory biomarkers. Univariate Cox regression showed many covariates with marginal significance and they are highly correlated. Traditional variables selection approach, such as forward/backward/stepwise regression, cannot deal with large p and small n (p>>n) and might not be good to select variables highly correlated. Penalized regression, such as LASSO and Elastic Net, can deal with p>>n and address multicollinearity, though output provides biased coefficient without p-value and confidence interval, where we cannot control false discovery rate with high-dimensional data. We also observed un-stable results by penalized regression with small sample size n (~50) but large number of covariates p (p>100). Recently, high dimensional inference (Fang et al., 2017) proposed de-correlated approach for Cox regression (CoxHDI), which is flexible to choose different penalized regression, stable for the choice of tuning parameter, provide de-biased coefficient, p-value and confidence interval. We proposed a stability approach (freqNet) based on Elastic Net but run 100 times and selected three biomarkers (pre-lymphodepletion serum LDH, MCP-1, IL-7) with frequency > 85% (Hirayama et al., 2019). Extensive simulation is conducting to compare the performance between CoxHDI and freqNet, and we demonstrate that the low dimensional empirical type I error rate is controlled when p >> n, though observed type I error inflation when p << n for both methods. We further applied CoxHDI on the same NHL dataset, and observed the same three biomarkers with p-value < 0.05 and one more biomarker (TNFRp75) identified, which also have good biological explanation associated with PFS.