学术报告会

报告题目：How Much Can Blocking and Randomization Improve Molecular Biomarker Discovery?

– A Block Randomized Study of microRNAs in Gynecologic Tumors

报告人： Li-Xuan Qin, PhD,

Memorial Sloan Kettering Cancer Center, New York，USA

时间：2016 年 7 月 19 日（周二） 上午 10：00~11：30

地点： 电信学院 2-406

邀请： 沈红斌教授

Abstract:

Blocking and randomization have the potential to prevent the negative impacts of non-biological effects on molecular biomarker discovery. Their use in practice, however, has been scarce. To demonstrate the logistic feasibility and scientific benefits of blocking and randomization, we conducted a microRNA study of endometrial tumors (n=96) and ovarian tumors (n=96) using a blocked randomization design to control for non-biological effects; we profiled the same set of tumors for a second time using no blocking or randomization. We assessed empirical evidence of differential expression in the randomized study and the non-randomized study. There was moderate and asymmetric differential expression (10%=351/3,523) between endometrial and ovarian tumors in the randomized dataset. Array effects were observed in the non-randomized dataset and 1,934 markers (55%) were called differentially expressed (DE). Among them, 181 were deemed DE (181/351, 53%) and 1,749 non-DE (1,749/1,934, 90%) in the randomized dataset. We further conducted simulation studies to evaluate the benefits of various forms of blocking and randomization on the accuracy of biomarker detection. In the simulation study, when randomization was applied to all samples at once or within each of multiple batches balanced in sample groups, blocking improved the true positive rate (TPR) from 0.95 to 0.97 and the false positive rate (FPR) from to 0.02 to 0.002; when sample batches are unbalanced, randomization within each batch is associated with a 0.92 TPR and a 0.10 FPR regardless of blocking. Normalization improved the detection of true positive markers often at the price of increased false positive markers. Through empirical and simulated studies, we showed that blocking and randomization (across all samples or within balanced batches) can effectively improve the accuracy of detecting disease markers. Blocking and randomization should be used to more fully reap the benefits of genomics technologies.