Challenges in identifying cancer genes in the face of inter and intra tumor heterogeneity

Massively parallel sequencing has permitted an unprecedented examination of cancer genomes, leading to predictions that all genes important to cancer may soon be identified by genetic analysis of tumours from sufficiently large patient cohorts. In this presentation I will explore evidence suggesting this promise may have been premature. I will present our evaluation of the ability of state-of-the-art sequence analysis methods to recover known cancer genes. While some cancer genes are identified by analysis of recurrence, spatial clustering, or predicted impact of somatic mutations, many remain undetected due to lack of power to discriminate driver mutations from background. Furthermore, cancer genes not detected by mutation recurrence tend to be missed by other types of analysis of patient cohorts. Nonetheless, undetected genes are implicated by other experiments such as functional genetic screens and expression profiling. I will examine ways by which such genes elude detection due to inter and intra patient heterogeneity; first, due to gene dependency effects in inherited germline and somatic mutations, and second due to transcriptional heterogeneity within tumors. Finally, I will present preliminary findings from unbiased transcriptional profiling of single cells from multiple colon cancer tumors. Our examination of expression subtypes derived from bulk RNA expression analysis of colon cancer patient cohorts, showed that different subpopulations of cells within a single tumor have high expression of genes related to different subtypes, and a single tumor may be comprised of a heterogenous mixture of subtypes. These results are suggestive that current monolithic classification of cancer into types may be unable to capture the transcriptional heterogeneity in cancer, presenting an additional challenge to comprehensive ‘omic charting of individual tumors.