Computational inference of tumor heterogeneity for cancer phylogenetics

While cancer can in theory develop from a seemingly infinite variety of combinations of mutations, in practice most tumors seem to fall into a relatively small number of recurring sub-types characterized by roughly equivalent sequences of genetic abnormalities by which healthy cells progress into increasingly aggressive tumors. This observation raises the hope that identifying these common sub-types and their defining genetic features will lead to new prognostic markers and drug targets. One promising approach to this problem is "tumor phylogenetics": treating tumors as evolving populations and analyzing their likely evolutionary pathways through phylogenetic algorithms. Two main variants of this approach have been proposed: a tumor-by-tumor approach, in which one treats each observed tumor in a population as a possible end state in a phylogenetic tree or network; and a cell-by-cell approach, in which one examines differences between individual cells in a tumor sample to build trees explaining variation both within and between tumors. The latter approach has the advantage of allowing one access to information about within-tumor heterogeneity that can provide important clues about conserved pathways of tumor progression, but at the cost of allowing one to examine only a few markers of state per cell versus the genome-wide markers sets one can apply to samples of whole tumors or significant sub-regions thereof.

Here, we will examine recent work intended to give us many of the advantages of each of the two approaches to tumor phylogenetics. This work uses computational inferences of within-tumor heterogeneity to infer cell-by-cell progression from tissue-wide measures of tumor state. The approach builds on the use of "unmixing" methods that allow us to treat each tumor as a mixture of fundamental cell states and computationally infer the states and their usage in each tumor. We will see how one can pose tumor unmixing as a problem in computational geometry. We will then examine algorithms for solving this problem in the presence of noisy, high-dimensional assays of tumor state. Finally, we will see how the resulting mixture models can be applied for phylogenetic studies on development. The methods will be illustrated by application to genome-wide expression and DNA copy number data sets from lung and breast tumors. The results of these studies show the promise of computational inferences as a way of gaining the advantages of both genome-wide assays of tumor state and within-tumor heterogeneity to develop a more complete picture of the common mechanisms of tumor progression.

Joint work with Stanley Shackney, Ayshwarya Subramanian, David Tolliver, and Charalampos Tsourakakis