

## **EXPECTATION-MAXIMIZATION METHOD FOR RECONSTRUCTING TUMOR PHYLOGENIES FROM SINGLE-CELL DATA**

G. Pennington

Computer Science Department, Carnegie Mellon University  
Pittsburgh, PA 15213, USA

C. A. Smith and S. Shackney

Allegheny Singer Research Institute, Allegheny General Hospital  
Pittsburgh, PA 15212, USA

R. Schwartz

Department of Biological Sciences, Carnegie Mellon University  
Pittsburgh, PA 15213 USA

Recent studies of gene expression in cancerous tumors have revealed that cancers presenting indistinguishable symptoms in the clinic can represent substantially different entities at the molecular level. The ability to distinguish between these different cancers makes possible more accurate prognoses and more finely targeted therapeutics. Making full use of this knowledge, however, requires characterizing commonly occurring cancer sub-types and the specific molecular abnormalities that produce them. Computational approaches to this problem to date have been hindered by the fact that tumors are highly heterogeneous masses typically containing cells at multiple stages of progression from healthy to aggressively malignant. We present a computational approach for taking advantage of tumor heterogeneity when characterizing tumor progression pathways by inferring those pathways from single-cell assays. Our approach uses phylogenetic algorithms to infer likely evolutionary sequences producing cell populations in single tumors, which are in turn used to create a profile of commonly used pathways across the patient population. This approach is combined with expectation maximization to infer unknown parameters used in the phylogeny construction. We demonstrate the approach on a set of florescent in situ hybridization (FISH) data measuring cell-by-cell gene and chromosome copy numbers in a large sample of breast cancers. The results validate the proposed computational methods by showing consistency with several previous findings on these cancers. They also provide novel insights into the mechanisms of tumor progression in these patients.