

Cell Lines, Chemotherapy Response, and the Need for Reproducible Research

Kevin Coombes

The University of Texas MD Anderson Cancer Center
kcoombes@mdanderson.org

Keywords: Duke University, forensic bioinformatics

In November 2006, researchers at Duke University published the first in a series of papers that claimed that (1) microarray and drug response data from cancer cell lines could be used to develop genomic signatures of response to specific chemotherapies, and (2) these signatures successfully predicted patient responses. Duke later began running clinical trials based on this work. We attempted to reproduce their analyses on publicly available data, and were unsuccessful. We identified a series of errors in data provenance and analysis, which we published as letters to editors and as an article in a statistical journal. As a result of our efforts, four clinical trials have been terminated and at least eight papers have been retracted. In this talk, I will describe some of the errors we found and some of the “forensic bioinformatics” methods used to discover them. I will also discuss tools in R and beyond that promote and enable “reproducible research.”