Clinical trial design for microarray predictive marker discovery and assessment

L. Pusztai1* & K. R. Hess2

Departments of 1Breast Medical Oncology and 2Biostatistics, The University of Texas M. D. Anderson Cancer Center, Houston, TX, USA

Received 23 March 2004; revised 15 June 2004; accepted 18 June 2004

Transcriptional profiling technologies that simultaneously measure the expression of thousands of mRNA species represent a powerful new clinical research tool. Similar to previous laboratory analytical methods including immunohistochemistry, PCR and in situ hybridization, this new technology may also find its niche in routine diagnostics. Outcome predictors discovered by these methods may be quite different from previous single-gene markers. These novel tests will probably combine the information embedded in the expression of multiple genes with mathematical prediction algorithms to formulate classification rules and predict outcome. The performance of machine learning-algorithm-based diagnostic tests may improve as they are trained on larger and larger sets of samples, and several generations of tests with improving accuracy may be introduced sequentially. Several gene-expression profiling-technology platforms are mature enough for clinical testing. The most important next step that is needed for further progress is the development and validation of multigene predictors in prospectively designed clinical trials to determine the true accuracy and clinical value of this new technology. This manuscript reviews methodological and statistical issues relevant to clinical trial design to discover and validate multigene predictors of response to therapy.

Key words: clinical trials, microarrays, gene-expression profiling, predictive markers, multigene predictors

Introduction

Decades of extensive research have yielded few clinically useful single molecular markers predictive of response to chemotherapy in patients with cancer [1, 2]. With the advent of high-throughput genomic technologies it is now possible to survey the expression of a large number of genes simultaneously in cancer tissue [3–6]. It is hypothesized that the pretreatment gene expression profile of cancer holds information about sensitivity to chemotherapy, and that this information can be extracted with transcriptional profiling and multigene predictors of response can be developed through mathematical analysis of the data [7]. However, the true accuracy and clinical value of these novel predictive tests is yet to be defined in prospectively designed marker discovery and validation trials.

It may be useful to think of marker discovery studies as conceptually similar to clinical trials that lead to the introduction of new drugs. The hallmark of clinical drug development is the multistage trial process. A similar focused, prospective, multistage evaluation of genomic markers could facilitate the introduction of new diagnostic markers into the clinic [8]. Phase I–II marker discovery studies would be expected to show that a technology can be reliably and reproducibly applied to clinical specimens and that the estimated predictive accuracy of the proposed test falls within a range that is considered clinically useful. Phase III marker validation studies would then evaluate the predictor in a larger number of cases to demonstrate that clinical outcome is better when the new marker is used for decision making compared with the current standard, which may be another marker- or no marker-based recommendation.

Tissue sampling

Gene expression profiling with DNA microarrays are best performed on fresh or frozen tissues, because the accuracy of the results is dependent on good quality RNA. Many investigators use excisional biopsy specimens; however, core needle biopsies or fine needle aspiration of cancer can also yield sufficient amounts of RNA for microarray experiments [9–11]. Instantaneous RNA preservations can be achieved by collecting biopsy specimens into a one-step RNA-preserving reagent (RNAlater; Ambion, Austin, TX, USA). Recently, it has also been shown that reverse transcription-PCR (RT–PCR) and DNA microarray profiling can be performed on RNA