On behalf of the NCI I would like to begin by thanking the Institute of Medicine (IOM) for organizing this committee to address the important issues facing us regarding the development and use of tests based on “omics” technologies in clinical trials and ultimately in routine clinical care. The IOM staff have assembled a committee of distinguished scientists and stakeholders who are generously contributing their time and expertise to deliberate on these matters and develop recommendations for the processes by which tests based on omics technologies should be evaluated for readiness for use in clinical studies. We thank the committee members for their commitment to this important effort.

I will take just a moment to introduce myself to give you a sense for my perspective and why I am representing NCI today. I am a statistician at NCI who specializes in statistical methods for the development and evaluation of biomarkers (including biomarker profiles obtained via high-dimensional omics technologies), and in the design of clinical trials using biomarkers and omics profiles. I review protocols for biomarker correlative science studies and biomarker-based clinical trials, including studies involving genomic profiling. In addition, I conduct my own research relating to the development of statistical methods for the analysis of biomarker and genomic data. I have been a statistical collaborator in many correlative science research projects, and I have been involved in the development of reporting guidelines for tumor marker studies.

I have witnessed the birth of many omics technologies and remain excited about their potential for providing important biological insights and their potential to lead to clinical tests that might improve care for cancer patients. It is important, however, that we understand the challenges and potential pitfalls that can be encountered with use of these technologies. Some
unfortunate events at Duke University involving the use of genomic predictors in cancer clinical trials were a major impetus for the formation of this committee. We need to take a step back to evaluate the process by which tests based on omics technologies are developed and determined to be fit for use as a basis for clinical trial designs in which they may be used to determine patient therapy.

In order to provide some specific examples that might stimulate and inform the committee’s discussions, I will spend the next few minutes to describe some of NCI’s involvement and findings with regard to genomic predictors developed by the research group led by Dr. Anil Potti and Dr. Joseph Nevins at Duke University. The committee has been provided with documents giving extensive details about the situations I will describe briefly here.

In 2009, one of NCI’s cooperative trials groups, CALGB, submitted a proposal for the study CALGB-30702 in which advanced stage lung cancer patients would be randomized to receive standard (physician choice) chemotherapy versus chemotherapy directed by the use of six different genomic chemosensitivity predictors that had been developed by the Potti/Nevins group. NCI reviewers noted dramatic differences between the protocol descriptions of the predictors (cell lines used and genes involved) and descriptions provided for those same predictors in published papers that were cited as having provided validations of the predictors to be used in the trial. Additional concerns included lack of validation of the predictors on human lung tumors and poor reproducibility of some the predictors. Simultaneously NCI became aware of a paper by MD Anderson statisticians Keith Baggerly and Kevin Coombes that had been recently published online in the journal Annals of Applied Statistics. The nature and scope of the inconsistencies in the Duke data were troubling.

The CALGB-30702 protocol mentioned a few trials that were already in progress at Duke University using some of these genomic predictors to determine patient therapy. A search of ClinicaTrials.gov identified several trials that appeared to be using these predictors for which NCI now had significant concerns. NCI (CTEP) contacted Duke University with its concerns at the end of September 2009. This led to suspension of three trials and initiation of an investigation led by the Duke IRB with assistance from two external statisticians experienced in
genomic data analysis and hired by Duke. NCI assisted by providing names of several statisticians with the relevant expertise in analysis of genomic data and helped make the initial contact to the individuals Duke decided to approach, but NCI had no knowledge of the details of the conduct of the review and had no role in the review or access to the data or computer code provided to the statistical reviewers.

With questions mounting about multiple Duke genomic predictors, NCI decided to revisit a large CALGB trial, CALGB-30506, that involved another gene expression microarray-based predictor developed by the Potti/Nevins group called the Lung Metagene Score (LMS) predictor. CALGB-30506 is a large (> 1000 patients) trial in which stage I lung cancer patients are randomized to standard care (observation) versus chemotherapy. Until a few months ago the LMS predictor was being used as a stratification factor to ensure balance of treatment arms within each LMS-predicted risk group. The LMS predictor did not determine patient therapy, but its prognostic and predictive value would have been evaluated retrospectively. I will explain a little history about this trial, and then I will describe the results of our re-analysis of the LMS predictor that led to our removing the LMS predictor from the trial.

A New England Journal of Medicine paper published in 2006 by Dr. Potti and co-authors presented the LMS predictor and reported that it showed dramatically good performance in distinguishing between early stage lung cancer patients who had good versus poor prognosis. When the CALGB-30506 trial concept was first presented to NCI for review, the study design would have used the LMS predictor to partly determine treatment. Patients predicted to be low risk would have been directed to observation, whereas patients predicted to be high risk would have been randomized between observation and chemotherapy.

NCI/CTEP required that the LMS predictor be tested in a blinded “pre-validation” study before a decision would be made on the trial. Microarray data were supplied to Dr. Potti by NCI, and Dr. Potti and his group were kept blinded to the clinical outcome data. The initial pre-validation attempt failed with the predictor predicting in the wrong direction, almost reaching statistical significance. However, subsequent exploratory analyses (after the outcome data were partly unblinded) suggested that the predictor might have some promising performance after the
microarray data had been corrected for having been processed in two different laboratories. During this pre-validation attempt, all microarray data preprocessing and risk prediction calculations were performed by Dr. Potti or members of his group. Neither NCI nor the CALGB Statistical Center had access to the computer software for running the predictor, but the Potti/Nevins group assured all parties that no changes had occurred in the form of the predictor between observation of the initial failed results and observation of the subsequent promising results. Due to the apparent sensitivity of the LMS predictor to laboratory effects, and as a condition for approval of the trial, NCI insisted on a trial design change in which the LMS predictor results would remain blinded so that they could not be used in any way to determine patient therapy. All patients would be randomized to standard care (observation) versus chemotherapy following surgery, and the LMS predictor would be used only as a stratification factor to ensure balance of treatment arms within each LMS-predicted risk group.

By Fall 2009 when numerous data irregularities had been documented in published papers and inconsistent predictor descriptions had been observed in protocols submitted to NCI, NCI had increasing concerns about the pre-validation study that had been conducted as part of the approval process for CALGB-30506. In November 2009, NCI requested, through CALGB, that Dr. Nevins and Dr. Potti provide copies of the computer code and data preprocessing instructions to NCI so that NCI could reproduce the earlier LMS predictor pre-validation results in its own hands. Disappointingly, we were unable to reproduce the promising results obtained 2 years earlier using the computer code and instructions provided to us by Drs. Potti and Nevins. An additional surprising and important finding was that the computer code used in 2007 to make LMS predictions apparently produced predictions with a substantial random component. When exactly the same data were entered into the computer software and it was run using exactly the same settings, the predictions produced on two separate occasions would agree only about 75% of the time. The predictor model was changing every time the software was run apparently because the prediction algorithm was dependent on a random number generator for which the software did not allow one to fix the starting seed. Although the LMS predictor was apparently “locked down” to something when the CALGB 30506 trial was initiated, there was no way to trace a clear path from a predictor that had been pre-validated to
the version of the predictor being used in the trial. Based on these findings NCI and CALGB mutually agreed that the LMS predictor should be removed from the trial. This removal required that the co-primary aim for evaluation of the LMS predictor be eliminated and that the risk predictions would no longer be used for randomization stratification.

In early January 2010, while NCI’s review of the LMS predictor was still underway, the Duke IRB sent to NCI the report from its investigation into the three Duke clinical trials that had been suspended. Due to the incomplete information in the report and the fact that NCI had no access to the data provided to the reviewers, NCI could not make a judgment on whether the concerns about the predictors used in the Duke trials had been adequately addressed. However, Duke’s interpretation was that the outcome of the review was positive and had cleared the way for reopening the trials.

In April 2010, NCI determined that it was providing partial funding through an R01 grant to Dr. Potti for some of the operational costs for one of the Duke (not CALGB) lung cancer clinical trials that had been suspended and then re-opened on the basis of Duke’s IRB review. That trial used the cisplatin chemosensitivity predictor to determine patient therapy. In addition to partial funding for the trial, the grant also supported further work to refine the chemosensitivity predictor for another drug, pemetrexed. The pemetrexed chemosensitivity predictor was already being used in another one of the three Duke trials that had been reviewed by the Duke IRB. The supporting information provided in the R01 grant included the paper by Hsu et al (JCO 2007) that reported validations of the cisplatin and pemetrexed chemosensitivity predictors. The connection of the R01 grant to the Duke trial gave NCI justification for requesting data and computer code to verify the supporting information provided in the grant.

Dr. Potti provided to NCI data and computer code in response to NCI’s request to reproduce the results published in the Hsu et al paper. The computer code and data provided by Dr. Potti produced results with a few broad similarities to the results in the Hsu et al paper for the cisplatin predictor, but there were some critical differences suggesting that there had been changes in the data used to build the predictor and in the data used to validate the predictor.
Insufficient information was provided by Dr. Potti for NCI to attempt to reproduce the pemetrexed predictor results.

A meeting was held in the NCI/CTEP offices in Rockville on June 29, 2010 with several NCI staff, Drs. Potti and Nevins and their statistician, and two Duke officials in attendance. The purpose of the meeting was to allow discussion of NCI’s continued concerns and to provide the Duke investigators with an opportunity to clarify their research methods. The meeting concluded with NCI remaining unconvinced of the validity of the Duke predictors. Further, NCI directed that a search of original laboratory and computer records be initiated to produce evidence of the correct versions of the data and computer code with the expectation that this task would be performed in expedited fashion. Less than 3 weeks later, there were reports of irregularities noted in Dr. Potti’s CV. The three Duke trials were once again suspended. On October 22, 2010 NCI was notified by Duke officials that errors had been identified in the data used to build and validate several of the chemosensitivity predictors, including the cisplatin predictor. The process of retracting the Hsu et al paper (JCO 2007) was initiated. The investigation is ongoing.

Reflection on NCI’s experiences with the Duke genomic predictors identifies several issues to be considered as the committee begins its deliberations. Investigators conducting studies using omics technologies face many challenges. Frequently investigators must acquire additional expertise themselves or find collaborators with the relevant expertise to address these challenges that include handling high volumes of data and use of new and complex bioinformatic, computational and statistical tools. Massive amounts of data are publicly available, and data analysis software is often freely shared. Errors can be introduced in data handling, and poorly documented data can be misinterpreted. Computer software might be “research-grade” and highly complex and can be misunderstood or used inappropriately. There has to be a level of trust in the competence, carefulness, and integrity with which individual members of the research team carry out their responsibilities because it is a rare individual who possesses all of the required types of expertise to carefully monitor and fully understand all aspects of a project.
Sometimes the glamour of the technology or the sheer volume of omics data seem to make investigators forget basic scientific principles. In addition, mistakes or other unfortunate events can occur, and their occurrence is harder to detect when the data sets are bigger and the data and analysis methods are more complex. If we are going to move clinical tests based on omics technologies into clinical trials where they will have an impact on patient treatment and outcome, we need to instill more rigor into the development and validation process. When we conduct clinical trials of new therapeutics, we would not accept situations in which data sources could not be verified or trusted, drug formulations were not clearly specified or documented, and drug delivery mechanisms only delivered the right dose 75% of the time. We don’t assume that a drug that has been studied in ovarian cancer only would automatically work in lung cancer. We have long understood the valuable role that blinding can play and the importance of pre-specified analysis plans in therapy trials. There are many common sense principles that could be applied but have not been consistently applied in the development of omics-based tests.

This committee is about to take on the challenge of developing recommendations regarding criteria important for analytical validation, qualification, and utilization of omics-based tests. Such criteria could be applied to determine the readiness of a test to be used in a clinical study. The recommendations developed will have to be practical and reasonable so as not to stifle important advances in the development of clinically useful omics-based tests. At the same time the recommendations need to effectively address the weaknesses and problems in our current approaches. Adoption and adherence to such recommendations should allow us to more rapidly distinguish useful tests from useless tests and offer patients greater protection from use of tests that are unlikely to lead to clinical benefit.

I thank you for your kind attention, and I would be happy to answer any questions from the committee members or from the Institute of Medicine staff.