Summary of comments of Dr. Lisa McShane (NCI) before the IOM Discovery of Process Working Group Meeting on June 21, 2011

The IOM Discovery of Process Working Group (DPWG) expressed interest in meeting with Dr. Lisa McShane (NCI) to discuss her June 29, 2010 slide set, entitled "Discussion of Genomic Predictors Developed at Duke University" (NCI_Meeting_Slides_29June2010-IOM.doc). The purpose of this meeting was to prepare for Dr. McShane's appearance before the full committee on June 30, 2011. This document, along with many others, was released by NCI to the IOM prior to the December 20, 2010 public session in which Dr. McShane presented. All documents referenced here are publicly available, either from the IOM's public access file or from other sources, as indicated. In addition, the DPWG had prepared a list of specific questions to which Dr. McShane was asked to respond during the June 21, 2011 discussion.

Discussion of June 29, 2010 Slide Set:

The slide set referenced here was provided to IOM as a file called NCI_Meeting_Slides_29June2010-IOM.doc.

Dr. McShane indicated that NCI first became aware of the chemosensitivity predictors developed by the Potti/Nevins group in 2007, and staff were aware of the controversy surrounding errors in data and figures originally published with some of the key papers (slides 3-4). During this same time period, NCI CTEP received for review a protocol proposal for the study CALGB-30506. The original CALGB-30506 protocol proposed to use the Lung Metagene Score (LMS) recurrence predictor to guide use of chemotherapy for certain Stage I NSCLC patients. This predictor was described and reportedly validated in a paper by Potti et al 2006 NEJM. The CTEP reviewers did not have confidence in the validations reported in the NEJM paper due to some apparent methodologic errors in the analyses (strict separation of training and validation sets not maintained), and they were aware of data errors and changes in a subset of the data that had been obtained from one of the investigators funded under NCI's Director's Challenge Lung Study.

NCI reviewers requested another validation on a blinded data set held by NCI. Dr. Potti was sent microarray data from the blinded validation data set, and he sent back to NCI binary risk group predictions that he reported were produced from his LMS model. When NCI statisticians correlated the risk group indicators with the outcome data, the association went in the wrong direction, i.e., predicted favorable risk group had worse survival outcome than the predicted unfavorable risk group. NCI reviewers realized that once the Duke investigators had seen the initial (failed) results, the data could no longer be considered fully blinded. In particular, because the first validation attempt showed the predictor almost reaching statistical significance in the wrong direction, one could easily infer that a new predictor with most predictions reversed compared to the first predictor would show good performance.

NCI allowed Dr. Potti to perform several post hoc analyses to determine where the predictor might have gone wrong. Finally, one of these post hoc analyses showed promising performance of the predictor. The investigators explained that they had normalized the microarray data separately by the lab in which they were generated, and with this correction, the generated predictions seemed to correlate correctly with outcome. NCI noted in its response to the Duke investigators a puzzling aspect of the new results. The validation set contained approximately 50 samples from the NEJM publication and roughly 100 that were new to Potti et al. NCI statisticians kept track of what happened to each of those samples after the renormalization and found that about 70% of the predictions for the new set flipped while only 20-30% of the predictions for the previously used (unblinded) samples flipped. NCI asked the Duke investigators to confirm that no changes had occurred to the prediction model between the different sets of results. The Duke investigators asserted that no changes had been made to the model.

Due to NCI's concerns about the apparent sensitivity of the model to lab batch effects, but strong interest in studying the role of chemotherapy in stage I lung cancer patients, NCI approved the study but with a modified design. In the modified design (the only version ever approved by CTEP) all patients would be randomized and the LMS results would be kept blinded to patients and clinicians. This design ensured that no patient's therapy would be influenced by the LMS results (slides 4-5), so patients would not incur additional risk if the predictor turned out not to work.

Another issue that came to the attention of NCI was that some investigators from Duke University Institute for Genome Sciences & Policy (IGSP) were reporting in public that the CALGB-30506 trial was assigning patients to treatments based on results of the predictor, which is not consistent with the trial design described in the approved protocol. Jolly Graham and Potti published a paper (Current Oncology Reports, 2009; seeJolly-Graham&Potti_2009_CurrOncRep-IOM.pdf in the IOM public access file) that reported "A large National Cancer Institute–funded prospective phase 3 trial based on the lung metagene model is under way. In this study, patients who undergo resection of early-stage disease will receive further adjuvant chemotherapy if they are predicted to be at high risk of recurrence". NCI felt that this was a misrepresentation of the trial and Dr. Dansky-Ullmann and Dr. McShane wrote a letter to the editor of the journal (see Jolly-Graham&Potti_2009_CurrOncRep-Erratum-IOM.pdf in the IOM public access file) pointing out the incorrect description of the trial. From a google search for information about the CALGB-30506 trial, one can also obtain slides of talks presented by Dr. Ginsburg, Director of the Center for Genomic Medicine, that display incorrect information about the trial design and its CTEP approval status.

In early 2009, NCI CTEP first received the CALGB-30702 protocol proposal (slide 6). This protocol described a phase II study in advanced lung cancer in which six genomic chemosensitivity predictors would be used to optimize therapy. Dr. McShane explained that as part of the initial protocol review the NCI/CTEP reviewers asked for a precise description of the cell lines used to train the predictor and a list of the genes incorporated into each predictor. When the revised protocol containing that information was submitted for re-review, NCI identified many discrepancies. Little of what the investigators reported in the protocol matched with the supposed validation studies reported in the published articles. The pemetrexed predictor appeared reversed from the published paper. For other predictors the cell lines or

gene lists were discrepant. To support the use of the predictors in the CALGB trial, the PIs stated in the protocol that the predictors were already being used to guide therapy in trials at Duke. So the NCI began to search and look for trials and wondered whether other trial funders were aware of the potentially problematic discrepancies.

Dr. McShane sent a memo to the head of CTEP, Dr. Jeffrey Abrams, in September 2009, outlining her concerns about the trials and the potential for patient harm, particularly if any of the predictors were truly reversed. This led to NCI's contacting Dr. H. Kim Lyerly, who then immediately contacted Duke University officials responsible for oversight of the trials (slide 7). Duke officials put NCI in touch with the Duke IRB, which ultimately suspended the three trials named in the IOM committee's statement of task (NCT00509366, NCT00545948, NCT00636441).

After the three Duke trials were suspended, NCI made a request to CALGB to re-evaluate the pre-validation study results that had been produced as part of the protocol review and approval process for the CALGB-30506 LMS trial. A detailed report describing all aspects of NCI's re-review of the LMS predictor was provided by NCI to the IOM prior to the December 20, 2010 meeting and these were placed in the public access file (LMS_RE-EVALUATION_2-8-10-IOM.pdf, along with supplementary documents LMS-Appendices A-M-IOM.pdf, LMS-NCI response to Nevins 3-26-10-IOM.pdf, LMS-CTEP response from Nevins 3-8-10-IOM.pdf). Dr. McShane also presented the basic findings in her December 20, 2010 testimony.

Slides 8-17 describe NCI's evaluation of the cisplatin predictor that was reportedly developed and validated in the paper by Hsu et al (JCO 2007) and which was thought to be in use in the Duke trial NCT00509366 (TOP0602). In April 2010, NCI made its initial request to Dr. Potti for data and computer code to reproduce the results pertaining to the cisplatin and pemetrexed predictors presented in Hsu et al because it had identified an NCI grant awarded to Dr. Potti (R01 CA131049-01A1) in which these predictors were being studied. NCI also determined that some funds from the grant were being used to partially support the Duke trial NCT00509366, although NCI was not listed as a co-sponsor of the trial on ClinicalTrials.gov. NCI wanted to conduct its own review because NCI did not have access to the data and computer code provided to the external reviewers hired by the Duke IRB, and it did not feel its concerns had been fully addressed by the report produced by those external reviewers. As detailed on slides 10-17 (and in the documents Cis-Reanalysis_report_on_cisplatin_predictor_final_10June2010-IOM.pdf and Cis-Appendices-A-E-IOM.pdf supplied to the IOM in December 2010 and available from the public access file) NCI's analyses revealed substantial discrepancies in the training data and validation data sets for the cisplatin predictor when comparing the results presented in Hsu et al to the results produced by the data and computer code provided to NCI by Dr. Potti.

After receiving Dr. McShane's detailed report on the cisplatin predictor, Dr. Potti stated in an email sent to Dr. McShane on June 10, 2010: " . . . I am a bit dissappointed that you view not being able to EXACTLY reproduce the same figures in the original paper as a major point of concern, I am assuming, from the analysis you show, that you too were able to show in your hands that the predictor itself is valid, in predicting response to cisplatin in vitro and in vivo. As for why there are differences in the published

figures (from 4 years ago) to what you were able to generate (although valid in both instances) using the predictor and R code as is being used in the trial, there are several reasonable explanations that I am hoping an exchange between our scientific groups can discuss, clarify and come to a consensus on."

Several points about the cisplatin predictor analyses were highlighted in Dr. McShane's discussion with the DPWG. As shown on slide 11 there were discrepancies in the training data, and 45 genes were reported to have been included in the predictor generated from that training data. After correcting for an off-by-one error, there were still 4 important discrepancies in the gene list. (Each gene is represented by one or more probe sets, and typically "gene lists" are reported as lists of probe sets.) Two probe sets identified as being used in the predictor were not even on the array used in the study, and the two other discrepant probe sets had been used to explain in the paper the biological plausibility of the predictor. Dr. McShane reported that Dr. Potti had explained that the four discrepant probe sets (genes) had been produced by data handling errors. Specifically, Dr. Potti explained that the two probe sets that were not on the array had arisen as errors introduced when he was following a suggestion from one of the JCO reviewers to "map up" to the companion array (U133B, part of the U133 Plus 2 array set) which had not actually been used in the study. Dr. McShane explained that she had suggested to Dr. Potti that he ask JCO to send to NCI copies of the reviewer's report in which that suggestion was made. Dr. Potti said he sent JCO an email, but later reported to Dr. McShane that JCO declined to provide the reviews. Dr. Potti did not offer to send his copy of the reviews.

Slides 12-13 describe how NCI attempted to reproduce the ovarian cancer cell line validations published in Hsu et al. NCI statisticians reasoned that the drug sensitivity values for the validation cell lines should not have changed, even if the predictor had changed, so NCI statisticians matched up y-axis values from the plot in the paper to the y-axis values in the plot produced by the data and computer code provided to NCI by Dr. Potti. This exercise revealed that the validation data had changed from what was reported in the paper, and NCI also noticed that the ovarian cancer validation cell lines reported by name in Hsu et al did not completely match those reported to NCI by Potti (slide 12). Even though the p-values for the two plots were both less than 0.001 (slide 12), the underlying data were quite different and the predicted sensitivity probabilities had changed (slide 13). Dr. Potti suggested that these were just mistakes in the paper or perhaps were due to slight variations in the method of developing the predictor. It is of interest that the cell lines numbered #16, #17 and #18 by Dr. Potti in his documentation provided to NCI fell almost exactly on the regression line in the plot (slide 12, left side) that appeared in the Hsu article, but those cell lines were not mentioned by name in that article. Slides 14-17 document failure to confirm the lung cancer cell line validation results and failure to confirm the ovarian tumor validation results for the cisplatin sensitivity predictor.

Slides 19-23 present what NCI knows about the pemetrexed sensitivity predictor. In the TOP0703 trial, genomic predictors of response to vinorelbine and pemetrexed were used to determine which therapy early stage lung cancer patients would receive in combination with cisplatin. NCI did not have oversight or provide funding for that trial, but Dr. McShane explained that NCI did have interest in *development* of the pemetrexed predictor because Dr. Potti was receiving NCI funding through his R01 grant CA131049-01A1 to further develop and refine the pemetrexed predictor (grant aim #2b). In his submission of R01-

CA131049-01A1 in March 2008, Dr. Potti stated (see slide 21): "…we have only been able to validate the accuracy of the cisplatin predictor in independent patient samples…, not the pemetrexed predictor…it is probably a little bit premature to employ the pemetrexed predictor to stratify patients…" Nevertheless, accrual on TOP0703 had been opened in October 2007 and the pemetrexed predictor was in use to guide the choice of treatment (slide 21). Slide 22 indicates that the external reviewers hired by the IRB might have been confused about whether the pemetrexed predictor was being used in the Duke trials. The reviewers state that they "agree with Nevins and Potti that since the profile is not used in any of the clinical trials patients are not being endangered", but this would appear to be an incorrect statement because the pemetrexed predictor was being used in the trial NCT00545948 (TOP0703). Dr. McShane reported that during the June 29, 2010 meeting with the investigators she asked them about the readiness of the pemetrexed predictor for use in the TOP0703 trial. She said that the investigators acknowledged that the pemetrexed predictor was being used to direct treatment even though they had been quoted as stating this would be premature, but they argued no harm was done to patients because they were being assigned to standard therapies.

After completion of the review of the June 29, 2010 slide set, the DPWG engaged in further discussion with Dr. McShane.

Further Discussion with Dr. McShane:

In response to questions about what review the trial protocols had undergone and who had monitoring oversight for the trials, Dr. McShane stated that she had no direct knowledge of Duke's system beyond what had been reported by Duke to the IOM in the document "Duke Historical Perspective 3 29 11.doc". She pointed the committee to the section in that document entitled "Translation of genomic science into clinical trials." This section described the development of the Duke Clinical Genomics Studies Unit (CGSU) as a separate entity from the existing Duke Clinical Research Institute (DCRI) that already had a robust clinical trials infrastructure in place. It is indicated that once the IRB provided approval, the project team was responsible for the execution of the study, with the CGSU Management Team regularly reviewing study progress and any issues identified by the PIs. Oversight was also to be provided by a group called a "DSMB-plus". Dr. McShane stated her personal impression that even if a clinical trial protocol is embedded in a grant application that is reviewed by an NIH study section, the review of the trial protocol by the study section is not generally at the same detailed level as would be expected by the review body for the institution that is sponsoring and responsible for oversight of the conduct of the trial.

Concerning who was overseeing the IRB-led investigation of the trials, Dr. McShane reminded the DPWG that NCI had no part in the IRB-led review beyond recommending and contacting potential reviewers, as reported in her previous testimony. She stated she was aware that the reviewers had an in person meeting with Drs. Potti, Nevins, and Barry and that Dr. Potti had been involved in preparation of the data supplied to the reviewers. She could not determine who was responsible for overseeing the review. Although Dr. Kornbluth had been NCI's main contact during the time the IRB review was taking place, Dr. McShane expressed her opinion that Dr. Kornbluth did not seem to have detailed knowledge

of what had been examined by the external statistical reviewers. For example, when asked by Dr. McShane if the reviewers had verified all sensitivity predictions for all patients enrolled on the trials using the predictors reportedly used in those trials, Dr. Kornbluth was uncertain and sought the answer from Dr. Potti. Dr. McShane also asked whether the reviewers had been able to verify the results in the published papers because the reviewers stated in their report that they were unable to identify a place where the statistical methods were described in sufficient detail to independently replicate the findings of the papers. NCI made clear its view that verifying and correcting the published record was important in addition to evaluating the validity of the predictors used in the trials, if the two sets of predictors were in fact different. Duke officials appear to have considered what they believed were the trial versions of the predictors separately from the material in the published papers. Dr. McShane suggested that the DPWG might query Duke officials about who was overseeing the IRB-led review and what specific instructions had been provided to the reviewers beyond what is stated in the introductory paragraph of the reviewer's report (Review of Genomic Predictors from Duke-IOM.pdf).

In further discussions about several of the papers that had been retracted, Dr. McShane expressed her opinion that it might be valuable for the committee to speak with some individuals who were coauthors or who were acknowledged as having provided data management or statistical analysis assistance. She felt it might also be valuable for the committee to speak with Dr. Barry who was hired to assist IGSP in April 2007. Although he was not a coauthor on any of the retracted papers it was he who, according to Dr. McShane's understanding, had been chiefly responsible for writing R code for implementation of the predictors in the clinical trials. Dr. McShane expressed her view that the committee might gain valuable insights from speaking to Dr. Barry about how he received the data and earlier versions of computer code, and what instructions and processes he followed to develop the clinical trial versions of the predictor code.

Dr. McShane's responses to questions asked by DPWG

Question 1:
Have genomic signatures either developed or validated on Potti's data been used to determine patient assignments in trials other than Duke's previously identified three trials (the two in lung cancer and one in neo-adjuvant breast cancer)? Are any other investigators at any other universities or cancer centers involved in the design and conduct of trials using genomic signatures linked to Potti? If so, who sponsored these trials and are these sponsors fully informed about possible integrity issues in these trials? Have NCI and other sponsors conducted appropriately comprehensive investigations into any suspected integrity issues?

McShane response:
The way that NCI originally identified the three Duke trials was by a search of the public website ClinicalTrials.gov. NCI staff searched on "genomic lung Duke", "genomic guided" and "genomic directed" to try to identify the genomically guided trials referenced in the CALGB-30702 protocol proposal that NCI/CTEP was reviewing in September 2009. Dr. McShane demonstrated these searches for the DPWG. Dr. McShane indicated that she was personally unaware of another source of information available to NCI that would easily identify trials for which NCI was not a sponsor or was not responsible for oversight.

She explained that while NCI could identify any trial being run by the cooperative therapy groups it funds, NCI would not have easy access to information about trials that might be independently funded through a cancer center or university except for what could be found by searching ClinicalTrials.gov. An indirect method to find some trials might be for NCI to search its grants database for any information suggesting a link to a clinical trial, even if the trial was not being funded by the grant. Dr. McShane indicated that, in her opinion, this indirect search is tedious and often not fruitful unless one has a specific grant number or some information about the trial one is looking for, for example a PI's name, institution, a drug name, etc. A publicly available resource for obtaining basic information about grants is now available through the website projectreporter.nih.gov. Dr. McShane also suggested that any grants acknowledged in the published papers from the Potti/Nevins group could be searched by grant number using this database.

Dr. McShane explained her understanding that although NCI does not routinely reveal information about grants or trials beyond what is in the publicly available databases, individuals or organizations can put in special requests for more information through the Freedom of Information Act (FOIA) or other legal mechanism. Dr. McShane suggested to the DPWG members that they develop a list of trials and grants for which they are interested in additional information, and submit the list to NCI. NCI would then determine what records could be released through FOIA or other appropriate mechanism. Further, she pointed out that the release of information by NCI in December 2010 concerning Duke trials and grants was made possible, in part, by confidentiality releases signed by Duke and three of its investigators.

Question 2:

We understand Duke has conducted a review of the integrity of the three clinical trials (the two in lung cancer and one in neo-adjuvant breast cancer). What is known from that review and other insights regarding the quality of the design and conduct of those trials, beyond the irregularities in the genomic signature data? Were proper clinical trials procedures engaged to ensure the integrity of the safety and efficacy data in these three trials? Have clinical research coordinators, data managers, statisticians and others with detailed insights about these issues and involved with these trials recently, or any time since initiation of the trials, been queried?

McShane response:

Dr. McShane responded that it is publicly known that an FDA audit had been conducted at Duke in January 2011, and it is possible that the results of that audit could shed some light on the answers to these questions. She also expressed her opinion that Duke officials would likely have been informed of any aberrations detected in the audit. Dr. McShane indicated that she had not personally been involved in investigating the conduct of the three trials. Responding to a direct question asking whether she was personally aware of any problems with the conduct of the trials, Dr. McShane stated that she was aware of allegations that good practice procedures for clinical trials (e.g., data maintained in locked clinical databases with controls on who could access the data) might not have been followed in at least one of the three Duke trials that had been investigated. She indicated that due to the sensitive nature of the allegations she received, she did not feel she could comment any further or provide a copy of the allegations at this time. Dr. McShane suggested that the DPWG ask Duke officials about the allegations, and she mentioned that they might contact Dr. Robert Califf, specifically. She expressed her personal

view that Dr. Califf appears to have deep insight into the issues and a thorough appreciation for good clinical trials practice. When asked if she could provide information about the allegations if Duke officials would not or could not provide the information, Dr. McShane responded that such a decision would have to be made by NCI .

Question 3:

How independent were the statisticians who were engaged by the Potti/Nevins group for the design and conduct of the trials? Specifically, were the statisticians faculty members in a statistics or biostatistics department, and to what extent did these statisticians report directly to Nevins or Potti as his/her superior or have salary or other support directly controlled by Potti or Nevins?

McShane response:

Dr. McShane responded that she does not have in depth knowledge of the administrative or oversight structure for statisticians at Duke, and cannot confirm who pays the salaries or provides other support for the statisticians involved. She noted that the background information provided by Duke to the IOM committee does indicate that Dr. William Barry was hired in April 2007 as a faculty member in the Department of Biostatistics and Bioinformatics and member of the IGSP.

Question 4:

How independent were the bioinformaticians, computer scientists, and data managers involved in preparing or overseeing computer code or databases for the trials? Did they report directly to Nevins or Potti as their superior or have salary or other support directly controlled by Potti or Nevins? Were there any changes in personnel that could have adversely affected the data integrity or computer code in use in the clinical trials?

McShane response:

Dr. McShane indicated she was aware that there had been some turnover in bioinformaticians and data managers involved with the trials, but she did not know who their supervisors were or who controlled the salaries or other support for the individuals who had left those positions. She suggested that it might be useful for the DPWG to interview the bioinformaticians, computer scientists, and data managers involved in preparing or overseeing computer code or databases for the trials. Further, she recommended that if the DPWG decided to interview any of these individuals, they might want to include among their interviewees individuals involved since the time immediately prior to initiation of the trials, regardless of whether those individuals were still employed in those same positions throughout the time during which the trials were in progress or were under investigation.

Question 5:

Have all coauthors of the four papers retracted to date been interviewed to understand the roles and responsibilities of each of the members of these research teams?

McShane response:

Dr. McShane has not been involved personally in interviewing any of the coauthors, but in her opinion, conducting interviews with the coauthors might provide some valuable insights for the DPWG.

Question 6:

Is our understanding correct that Potti was the PI of trials that were evaluating genomic signatures for which he held or had applied for a patent, or for which he might have had a financial interest in companies providing services directly related to the genomic predictors under investigation in the trials? If so, this seems to be a serious conflict of interest. What measures were employed to properly address this? Also, were potential conflicts disclosed fully and in a timely fashion to any federal agencies involved in funding these trials or related research projects?

McShane response:

Anil Potti was PI on the trial TOP0602 (NCT00509366) until October 6, 2009 when the trial was suspended (see publicly available information in the ClinicalTrials.gov history) and the Duke IRB investigation was begun. According to publicly available information on the website http://www.faqs.org/patents/inventor/anil-7/, Anil Potti was listed on at least five patent applications, all involving either predicting responsiveness to cancer therapies or predicting lung cancer tumor recurrence. Public information indicates that these patent applications were published during the period 4/23/09 to 11/4/10. Disclosures on published papers and other publicly available documents indicate interests in the companies Expression Analysis and CancerGuide DX by several Duke investigators who were coauthors on papers published by the Potti/Nevins group.

Speaking in general terms and not specifically about Duke, Dr. McShane explained her personal view that institutions generally are responsible for having in place management plans for investigators' financial conflicts of interest. These management plans, as they relate to clinical trials, typically include stipulations such as disclosures on trial consent forms and barring an investigator from being a PI on a trial in which he or she has a perceived conflict of interest in the outcome of the trial. Dr. McShane noted that conflicts could arise and resolve over time, so that any management plan might change over time to adapt to the changing landscape of potential financial conflicts. Thus, one might need to look at all versions of IRB-approved clinical trial protocols and consent forms to understand how the disclosures and background information might evolve over time.

Dr. McShane explained that in her experience with NIH-funded grants, it appears that investigators are expected to disclose to their institutions their potential financial conflicts of interest, and it is expected that the institution has in place a COI management plan. Dr. McShane's impression of the usual procedure is that institutions report to NIH when there has been any change in the COI status and the institutions provide assurance that they have a management plan in place. In her experience, NCI usually does not receive further details of COIs or management plans in the ordinary course unless it specifically asks.