PREDICTORS OF CHEMOTHERAPY RESPONSE: BACKGROUND INFORMATION

In relation to the topics and questions suggested by the IOM, the following document outlines some of the scientific and administrative responses around the attempted development of chemotherapy sensitivity predictors and is being provided by Duke Medicine administration to the IOM Committee in advance of its upcoming "Workshop on the Review of Omics-Based Tests for Predicting Patient Outcomes in Clinical Trials." The circumstances and events related to work conducted by Dr. Anil Potti, in collaboration with Dr. Joseph Nevins, at the Duke Institute for Genome Sciences & Policy that have led to the IOM Committee's review of this field are regrettable and have led to significant organizational self evaluation in an attempt to learn from the events of the past two years. Understanding that the committee is scheduled to receive additional information at future meetings, this document is limited in scope to some of the thought processes and actions relevant to the topics defined for this workshop. It is our hope and belief that the work of this committee will serve to provide much needed guidance for this emerging area of translational research and we are grateful to the IOM for taking on this project.

The IOM Committee's work is the result of a request by Duke University's Chancellor for Health Affairs, Dr. Victor J. Dzau, in consultation with Dr. Harold Varmus, Director of the National Cancer Institute, to Dr. Harvey Fineberg, President of the Institute of Medicine, on July 21, 2010, for a full and independent review of the issues around the genomic predictors of chemotherapy sensitivity, as well as to provide guidance on the appropriate scientific approaches to this rapidly evolving area of science and medicine.

As we now know, the foundational conclusions for the chemotherapy sensitivity predictors were compromised by corruption in validation datasets that invalidated those conclusions, and the clinical trials based on the science were stopped. The work of Drs. Keith Baggerly and Kevin Coombes was instrumental in uncovering the problems with these chemotherapy sensitivity predictors. However, the goal of developing universal standards for omics research remains and we look forward to the committee's recommendations.

Introduction to Genomic Studies in the Duke IGSP and Nevins' Laboratory

The genomic studies that resulted in the development of the predictors of chemotherapy drug response evolved within the context of an institution-wide effort involving faculty in the Schools of Medicine, Arts & Sciences and Engineering to apply genomic approaches and analyses to a wide variety of biological phenomena, disease areas and medical challenges. Among these was a program that made use of genomic analyses to address the complexity of cancer phenotypes and mechanisms of oncogenesis. Starting with a focus on the identification of critical gene regulatory proteins involved in the control of cell proliferation, and that were directly involved in oncogenesis, this work evolved to make use of genomic technologies including DNA microarray analysis as a powerful

means to evaluate the full spectrum of gene control mediated by these regulatory activities (Ishida, et al., 2001).

This was the genesis of a program, initiated in 2000, that brought together cancer biologists, informaticians and statisticians in a joint effort to define the critical questions and devise the experimental strategies and statistical methodologies to address these questions. The Computational and Applied Genomics Program (CAGP), a multidisciplinary research program developed by Dr. Joseph Nevins, then James B. Duke Professor of Genetics and Chair of the Department of Genetics and Dr. Mike West, Arts and Sciences Professor of Statistics and Director of the then Institute for Statistics and Decision Sciences (now Department of Statistical Science), focused on the creation, modeling, analysis and integration of multiple forms of data in basic and clinical biomedical studies. The core research in CAGP projects focused on basic genome science, complex modeling and statistical methods, and computational biology. CAGP projects began with both basic and applied research, with the applications arising from several related, multidisciplinary projects in specific biomedical areas. The applied contexts included projects involving molecular characterization studies, gene pathway studies and biological information integration in breast cancer and other cancers, cardiovascular studies, and neurological studies. Faculty, fellows and students involved in CAGP projects were based in various Duke departments including genetics, statistics, and medicine.

This integrated core group provided the expertise that led to a series of initial publications that focused on development of gene expression signatures predictive of breast cancer phenotypes (Huang, et al., 2003a; Pittman, et al., 2004; West, et al., 2001a). An additional area of work was focused on the development of expression signatures that predict activation of oncogenic signaling pathways (Huang, et al., 2003b), including the ability to link this activation with sensitivity to pathway-specific therapeutics (Bild, et al., 2006). Further work led to the development of expression signatures that could distinguish important cardiovascular disease states (Karra, et al., 2005; Seo, et al., 2004).

Aspects of the concepts and underlying methodology that emerged from Drs. Nevins, West, et al.'s research, particularly the focus in pathway activation signatures, have been extended in other laboratories. Lewis Chodosh and colleagues at the University of Pennsylvania have made use of similar methods - the software developed within the Integrative Cancer Biology Program (an NCI funded program led by Drs. Nevins and West) was applied to the analysis of mouse genetic models of cancer to develop signatures of pathway activity (Liu, et al., 2008). Joan Massague and colleagues at Sloan Kettering also made use of the Bild, et al., (2006) pathway signatures to identify a role for Src activity in late onset bone metastasis in breast cancer (Zhang, et al., 2009). Finally, investigators at Merck built upon the concepts described by Bild, et al., (2006) to link predicted pathway activity to the use of pathway-specific therapeutics (Loboda, et al., 2010).

Creation of the Duke Institute for Genome Sciences & Policy

In parallel with the emergence of the CAGP, the Duke University Institute for Genome Sciences & Policy (IGSP) was launched in early 2003, as an opportunity to bring together these diverse groups of investigators within the context of genomics research and to do so in an interdisciplinary environment. Dr. Huntington Willard was recruited as the Founding Director of IGSP and developed an organizational plan for the Institute's growth. His plan focused on developing a series of centers embedded within the IGSP that would provide the foundation for the diverse activities. The integrated program (CAGP) developed by Drs. Nevins and West became the basis for the development of the Center for Applied Genomics and Technology, one of these initial centers of the IGSP. Dr. Nevins became the Barbara Levine Professor of Breast Cancer Genomics and Director of the Center and initial recruitments included Dr. Sayan Mukherjee (Computational Biology), Dr. Uwe Ohler (Computational Biology), Dr. Jen-Tsen Chi (Molecular Genetics and Microbiology), and Dr. Phillip Febbo (Oncology).

In 2005, Dr. Geoffrey Ginsburg was recruited to the Duke IGSP as Director of the newly formed Center for Genomic Medicine. Previously Vice President of Molecular Medicine at Millennium Pharmaceuticals, Dr. Ginsburg arrived with the goal of translating scientific advances in genomics into clinical practice.

In 2006, the IGSP Center for Applied Genomics and Technology conducted a national search to identify a translational investigator who could develop a genomics-based research program within the IGSP alongside that of Dr. Febbo whose work was focused in the area of prostate cancer genomics. Two leading candidates emerged from this search: Dr. Sandeep Dave, who developed genomic characterizations of B cell lymphoma while a fellow in Dr. Louis Staudt's laboratory at the NCI; and Dr. Anil Potti, an oncology fellow working in Dr. Nevins' laboratory whose clinical focus was the application of genomic methods to lung cancer. The IGSP, in conjunction with the Department of Medicine, decided to make offers to both individuals, both accepted, and Dr. Potti established his own laboratory in the IGSP to continue his work on the application of genomic methods to lung cancer.

Dr. Potti began his fellowship studies at Duke in Hematology/Oncology in 2003, working with Dr. Thomas Ortel with a focus on work in hematology. He joined the Nevins' laboratory in 2004, as part of a collaborative effort to extend his work with Dr. Ortel in hematology using the previously described genomic technology and methodologies developed within the CAGP.

Dr. Potti identified the potential utility these genomic strategies might provide in clinical oncology. In particular, he believed there was an opportunity to personalize the utilization of existing chemotherapies to achieve the best possible tumor response for the patient. While advances in the development of genomic tools offered significant promise for improved prognosis, particularly in breast cancer, the ability to identify the most appropriate therapy for an individual patient did not exist. His work in this area made use

of the same published statistical methodology developed by Dr. West in colaboration with Dr. Nevins and colleagues.

While a fellow in medical oncology, Dr. Potti's studies in the Nevins group culminated in 2006, with the publication of a study in Nature Medicine that described the development of gene expression signatures that could predict tumor sensitivity to various standard chemotherapies (*Nature Medicine*, 12:1294-1300 (2006)). This work made use of publicly-available data from a panel of cancer cell lines (the NCI-60 panel) to derive training sets for various chemotherapies involving cells resistant or sensitive to the drugs that then allowed the development of gene expression signatures that were purported to have validated in additional publically available datasets including several from human tumor samples. Although this work made use of methodology developed in the Nevins/ West collaboration, Dr. West was not actively involved in providing statistical expertise on the development of the chemotherapy sensitivity predictors. Dr. Potti carried out an additional study to derive gene expression signatures from lung tumor samples focused on prognosis in lung cancer that resulted in the development of a multi-gene predictor of lung cancer recurrence (the lung metagene score). This work was published in the New England Journal of Medicine in 2006.

Dr. Potti continued his work on genomic analysis of chemotherapy sensitivity, in collaboration with Dr. Nevins, et al., including extending the studies in the 2006 Nature Medicine paper to a validation opportunity in conjunction with a European Organization for Research and Treatment of Cancer (EORTC) breast cancer clinical trial. This validation work from the Potti laboratory was published in a Lancet Oncology paper in 2007. In addition, Dr. Potti and collaborators reported the development of a cisplatin chemotherapy response signature to direct the use of chemotherapy in a 2007 Journal of Clinical Oncology paper.

As will be described in more detail below, it is now recognized that the studies reported in Nature Medicine, Lancet Oncology, and the Journal of Clinical Oncology made use of corrupted validation data to assess performance, which invalidated their findings. In addition, attempts to reproduce results reported in the New England Journal of Medicine regarding the lung cancer recurrence predictor were not successful, because of discrepancies in clinical outcome information and high instability in the predictions due to stochastic components of the BinTree model. Therefore, each of these papers has been retracted. In evaluating these errors, it is clear that there were inadequate processes and data systems being used to assure data provenance and the ability of others to replicate the studies. Currently, Duke is evaluating the extent to which additional published studies may have made use of corrupted data or have a lack of reproducibility of findings. Duke administration's understanding of the need for more defined processes and data systems to be involved in transitioning scientific discovery to clinical trials has been the primary focus of our internal self-evaluation and immediate efforts to define institutional solutions moving forward.

Translation of genomic science to clinical studies

To translate genome-based discovery into clinical application requires high-quality clinical/genomic datasets for both discovery and validation of genomic biomarkers. The IGSP investigators recognized that prospective clinical trials would be required to adequately validate genomic biomarkers and demonstrate their utility in impacting clinical outcomes. Implementing prospective clinical trials using genomic biomarkers required a significant clinical-genomics infrastructure. Although there was already a robust clinical trials infrastructure in the Duke Clinical Research Institute (DCRI), the DCRI was not designed to implement genomic assays in oncology trials and its mission was focused on multi-site research. Therefore, the IGSP created an infrastructure and framework to launch the three clinical trials that made use of genomic signatures to guide chemotherapy selection.

The Duke Clinical Genomics Studies Unit (CGSU) was established in 2007, and fully developed over the course of the next two years, under the leadership of Dr. Geoffrey Ginsburg, Director of the IGSP Center for Genomic Medicine. The leaders of the CGSU intended from the outset, to develop standards for genome-based clinical studies aimed at biomarker discovery and validation and to develop an operational framework for implementing genomic studies involving the collection of clinical samples from human subjects and their relevant clinical data. A structure was set up within the CGSU to assess the scientific merit of new clinical study concepts aimed at validation of predictive genomic tests and to evaluate the technical and practical feasibility of the studies. Dr. Ginsburg's intent for the CGSU was to provide multidisciplinary rigor to these studies and to be an institutional resource for investigators at Duke.

The leaders of the CGSU recognized that implementing genomic assays in a clinical trial setting that may ultimately be translatable to routine clinical use if the trials were successful, required dedicated personnel and a clearly understood workflow. Many of the genomic procedures involved were novel and outside the context of a traditional clinical research study. The CGSU defined the operational standards to assure an unbroken chain of custody for tissue samples and data that included the following:

1. Sample acquisition, management and processing through a dedicated laboratory established in conjunction with Duke Clinical Molecular Diagnostics;
2. Precise and accurate genomic assays obtained from a genomics laboratory, Expression Analysis, Inc.;
3. Array data analysis through an automated system designed and controlled by a Duke faculty biostatistician; and
4. Dissemination of the results to key personnel through an automated system.

Dr. Ginsburg's intent for the CGSU was to conduct studies in the context of a dedicated genomic analysis laboratory, now a CAP-accredited and CLIA-certified environment that was created under the auspices of the Duke Clinical Molecular Diagnostics laboratory.

SOPs were developed for all aspects of the process for obtaining tissue samples, processing the samples, and quality control analysis.

A CGSU Management Team was instituted in 2007, consisting of the Director, the Clinical Director, the Director of Operations, the Clinical Genomics Study Manger, the Budget Director, and the Director of IGSP Administration. All aspects of clinical genomics research processes and standard operating procedures were defined by Dr. Ginsburg and the CGSU Management Team as part of this new unit. The Management Team met on a monthly basis to review all activities of the CGSU. This included ongoing review of the three genomic clinical trials.

The Clinical Genomics Study Manager, a Project Manager, and the Principal Investigator of each study comprised the Genomics Clinical Study Project Team and convened regular team meetings (at least bi-weekly) that consisted of genomics clinical research coordinators, sample processing personnel, data management personnel, and regulatory management. This project team initially designed, developed, and submitted the protocols and other necessary documents for review by the Cancer Center Protocol Review Committee (CCPRC) and then the IRB. Once approved by the CCPRC and the IRB, the project team was responsible for the execution of the study. The CGSU Management Team regularly reviewed the progress of the studies and any issues identified by the Principal Investigators.

Finally, with responsibilities similar to a Data, Safety, and Monitoring Board (DSMB) the DSMB 'Plus' was established in 2007, and its charter was defined in collaboration with the Duke Conflict of Interest Committee chaired by Dr. Ross McKinney, Jr., Director, Trent Center for Bioethics, Professor of Pediatric Infectious Diseases. The DSMB Plus was constituted to include investigators from outside of Duke who had the responsibility for monitoring study progress and reviewing each of the clinical studies for subject ascertainment and endpoints adjudication. The DSMB Plus included expertise in statistics as well as each of the oncology areas represented by the ongoing trials. The DSMB Plus reviewed the clinical studies at least yearly since 2008, and met at least twice per year with both an open and closed session. The 'plus' designation is to indicate that in addition to standard DSMB activities, this board was also charged with assessing conflict of interest issues in the design and operational features of the studies.

In summary, the leaders of the CGSU created the unit with the intent of meeting the demands of integrating genomic analysis into a complex clinical setting and particularly to evaluate the potential represented by the development of genomic signatures to guide chemotherapy use.

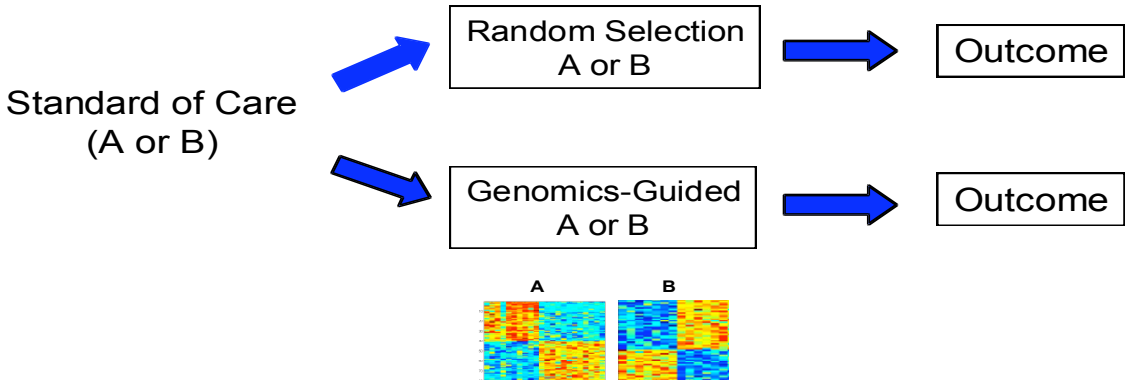
Development of the Clinical Trial Protocols

A major challenge for the translation of biomarkers from the discovery phase to clinical use is to establish the markers' clinical validity and utility. A primary motivation of designing and performing clinical genomic studies was the need to generate high-quality data that would provide an opportunity to evaluate the genomic predictors. While there

are various datasets that provide opportunities for validation of predictive signatures, they are often limited by small numbers of samples as well as quality issues with regard to the samples and clinical outcome data. The limited availability and quality of these 'samples of convenience' were viewed by the IGSP investigators as a major limitation and one that could only be addressed by the development of carefully controlled prospective studies that would allow the generation of high-quality data that could specifically assess the accuracy of potential predictive signatures to guide therapeutic decisions. Thus, the genomic signature guided trials were viewed by IGSP as opportunities not only to apply genomic strategies to important clinical questions, but also to generate high quality clinical and genomic data that would provide further opportunity to evaluate the performance of the signatures and other scientific advancements.

In developing the design of clinical utility studies, the IGSP investigators considered two basic trial designs. In one, patients would be treated with a given drug or drug options together with the collection of a tissue sample which would then be used to measure the genomic signature so as to evaluate at the end of the study the capacity of the signature to predict response. In this case, the signature would simply be a correlative science study, but the study would provide context in which to validate the performance of the signature.

The other design contemplated that the genomic test could be used prospectively to assess tumor responsiveness as a basis to improve on random choice of agents, but a patient enrolling in the trial would receive a widely used and accepted chemotherapy, irrespective of whether they were in the control arm or the experimental arm. This design was possible where multiple agents were considered acceptable therapies, thus providing a mechanism to simply guide the choice of otherwise equally acceptable treatments. This design also provided a mechanism to effectively validate the ability of the signature to perform. The design, as outlined in the figure below, became the framework for the breast cancer neoadjuvant treatment trial, initiated in April 2008, that evaluated the capacity of genomic signatures to guide the selection of either adriamycin or docetaxel.



The decision to use the genomic signatures to guide therapy was driven primarily by the IGSP investigators' belief that the design described above offered no more risk (above the risk of biopsy) to a patient than an observational study given the use of widely accepted drugs. What was seen as more challenging in putting these signatures to work in

a prospective trial was the development of the methodologies for transitioning from retrospective analyses involving large numbers of samples to methods that would allow for the assay of single samples, one enrollment at a time.

Methods to implement chemotherapy sensitivity signatures in clinical studies

In the IGSP investigators view, an important aspect of the methodology used to apply *in vitro* derived signatures to tumor samples is the recognition that the global expression profiles from the *in vitro* and *in vivo* contexts can differ substantially. For prospective clinical studies, this issue raises a major challenge since one must have the capacity to carry out an analysis on one patient sample at a time, which presents a unique challenge for carrying out the predictions, since the full data set will not be available for analysis.

Dr. William Barry, who joined Duke in April 2007, as a faculty member in the Department of Biostatistics and Bioinformatics and member of the IGSP, directed the effort of modifying existing processes so that the genomic signatures would be applied to the clinical studies in a verifiable manner. To this end, Dr. Barry outlined two previously identified mechanisms to integrate distinct forms of gene expression data (cells vs. human tumor data). The first strategy (termed a 'Type 1' analysis) merges cell line and tumor data sets and uses principal components from singular value decomposition (SVD) to partition the variation between and within data sets. Then, a Bayesian binary regression algorithm is used to generate predicted probabilities for the tumor sample. Importantly, no information regarding the clinical outcome/phenotype of tumor samples is contained in the merged data set, such that the SVD allows one to overcome differences in the expression profiles of tumor samples and cell line data without biasing a predictive model. This allows the genomic signature to perform in a robust manner when applied to the tumor data.

An alternative strategy, identified by the IGSP investigators, was to accommodate global differences in gene expression data between cell lines and human tumor data using normalization methods for correcting batch differences prior to decomposing the data by SVD (termed a 'Type 2' analysis). This was done by a simple linear discriminant analysis to standardize the mean and variance of expression data at the feature level. Principal components are built from a SVD of cell line data alone, and then the expression profiles of tumor samples are projected onto the normalized tumor data. In contrast to the two approaches, when principal components are built solely from the training set data (cell lines) and no methods are used to normalize the gene expression profiles with the tumor samples, the predicted probabilities are poorly distributed such that the model has no capacity to discriminate the responders from non-responders in tumor samples (described in a manuscript submitted to Clinical Cancer Research, now withdrawn due to corruption of the validation data).

Finally, IGSP investigators performed additional evaluations of the normalization procedures described above (Type 1 or Type 2) using completely independent tumor datasets that were not part of the validation data for the normalization. Under conditions where technical variability between reference and test datasets are minimized, the

performance was similar to the original results *in vitro* and thus, provided what was seen by the investigators as evidence that the validation of predictors would not be compromised by information leaking into the training data. This would later become important as questions were raised by Drs. Baggerly and Coombes, Professors of Bioinformatics and Computational Biology at MD Anderson.

This statistical methodology for normalization was a critical component of the implementation of the signatures in the clinical trials, providing a mechanism to use the signatures in prospective studies one sample at a time. Further, the development of the computational infrastructure for the prospective studies allowed for all components of the algorithm to be constructed in a way that ensured all results were completely reproducible. Prior to the implementation of the clinical trials, the predictive signatures were developed utilizing algorithms (named 'BinReg') programmed in Matlab, which had been created and maintained by Drs. West and Nevins for use on the local workstations of the IGSP investigators. The Matlab code involved stochastic simulation as a main computational tool and was not designed for routine applied use nor to create log files. When used in a trial, this could potentially have an effect of producing discrepant classifications for the same patient if the predictions were near the threshold used to categorize chemotherapy sensitivity (1-2%). Dr. Barry identified and addressed these issues by converting the methodology to R/Bioconductor environment such that analyses could be run on servers from primary data sources in a manner that allowed the stochastic components to be reproducible with a fixed numerical seed to the algorithm. This enabled each analysis run during the course of the prospective clinical trials to be identifiable, reproducible, and exchangeable across patients. This issue does not invalidate other uses of the code in discovery and retrospective validations.

As is the case with any research study, it was recognized by the IGSP investigators from the beginning that there was significant uncertainty in being able to effectively bridge from the *in vitro* derived predictive chemotherapy signatures, and that those predictors might not be successful in predicting response in tumors. Therefore, the IGSP investigators designed the clinical trials to be therapeutically conservative, and only utilize widely used chemotherapies for lung cancer and breast cancer. Two genomic signature guided trials were initiated under the auspices of the CGSU within the IGSP in lung cancer in 2007, one focused on early stage disease and one on advanced stage disease. In both trial designs, the genomic signatures were used to assign therapy to one of two widely acceptable regimens. The early stage trial involved the use of adjuvant chemotherapy, each a widely used and acceptable regimen. For the late stage trial, the choice was between a cisplatin doublet or a pemetrexed doublet. The latter was an approved regimen for second line treatment of lung cancer and it was the view of the lung oncologists involved in the design of the study that this was an acceptable treatment option for first line use. Subsequent to the initiation of the two lung cancer trials, the results of other clinical studies began to demonstrate an inferiority for pemetrexed in lung cancer patients with squamous histology. This prompted discussions within the IGSP group, in consultation with the DSMB Plus, about whether it was appropriate to enroll patients with squamous disease into the trial who then might be treated with pemetrexed. As a result of these discussions, it was determined that the early stage trial would exclude

patients with squamous histology. For the advanced stage trial, an amendment was made that replaced pemetrexed-gemcitabine with docetaxel-gemcitabine for the treatment of cisplatin-resistant patients with squamous disease. These modifications were vetted with the DSMB Plus and revised statistical plans were developed to accommodate the protocol modification.

The initial design of these studies was vetted within the Duke oncology community in collaboration with scientists in the Duke IGSP. Indeed, a team of investigators from Duke Medical Oncology and the Duke Cancer Center considered the potential therapies and the trial designs to determine what would be the most appropriate course of study. The protocols that emerged from these discussions were then reviewed and approved by both the Cancer Protocols Committee of the Duke Cancer Center and the Duke Institutional Review Board. This was an extensive process that was intended to evaluate the range of potential issues surrounding the design of the studies, but did not address the fundamental issues of data provenance that provided the evidence for validation of the predictors.

Additionally, in a parallel review process, the breast cancer neoadjuvant study was submitted as a proposal to the Department of Defense Breast Cancer Program. It was peer reviewed, received a positive evaluation, and was funded to support the trial. Thus, all three trial designs and treatment arms were independently reviewed and approved by experts outside the IGSP.

It should be noted that these genomic predictors were not considered 'final products' ready for routine clinical use, but rather as promising research tools. The goal was to complete proof of principal phase II trials in a setting that would support appropriate care of patients, critically analyze the validity of the predictive signatures with internal and external biostatistical review, and design definitive randomized trials to test the hypothesis that signature directed therapy was superior to undirected therapy.

Questions arise regarding the chemotherapy signatures

In the fall of 2006 and continuing into 2007, Drs. Nevins and Potti received queries from Drs. Baggerly and Coombes regarding various aspects of their published work. Questions about published work, as well as requests for clarification or additional information, are a normal part of the scientific process. Drs. Nevins and Potti responded and provided clarifications over the span of several months. This communication was largely focused on Drs. Baggerly and Coombes' inability to reproduce the predictors when they used their own methods that were similar in nature to what was done by Drs. Nevins, Potti, et al. In the course of recreating the predictors, they identified multiple errors in the data that were presented, including mistakes in gene lists associated with genomic signatures.

Drs. Baggerly and Coombes published a letter to the editor of Nature Medicine in November 2007 (*Nature Medicine*, 13:1276-1277 (2007)), in which they described an inability to reproduce a series of aspects of the published work. This included the

selection of cell lines as resistant/sensitive, problems with gene lists and heat maps associated with signatures, and an inability to reproduce the basic finding of predicting drug response. In part, they attributed this failure to reproduce the primary result to the manner in which the software was implemented in the analysis of Drs. Nevins and Potti by not maintaining independence of training and validation data, as well as to problems with gene lists and validation datasets. Drs. Nevins and Potti responded to this letter with clarifications on the main issues, including the logic for selecting cell lines, acknowledging the problems with gene list errors, and responding to the criticism of maintaining separation of training and validation data. (*Nature Medicine*, 13:1277-1278 (2007)). Corrections were published regarding these issues (*Nature Medicine*, 13:1388 (2007)).

As indicated above, one of the issues raised in this initial communication from Drs. Baggerly and Coombes was a criticism of the manner in which the signatures were developed where samples from validation datasets were used as a source of information along with the training data for deriving principal components by singular value decomposition. It was thought at the time that Drs. Baggerly and Coombes were contending that this approach violated principles of validation such that information about the validation samples was 'leaking' into the model. Drs. Nevins and Potti held that there was nothing wrong in this approach as long as the clinical outcome information, which was the focus of the predictions, was not incorporated in the process, which they did not believe occurred. The main issue as described by Drs. Nevins and Potti was the need to accommodate the technical distinctions in the gene expression data derived from cell lines versus that derived from tumors. If this were not done, then they believed there was no ability to develop predictive models that could accurately predict.

Drs. Nevins and Potti noted that the need to accommodate the distinctions in cell line expression data and tumor data had also been recognized by others. In particular, Lee and Theodorescu at the University of Virginia had described an approach making use of cancer cell line drug sensitivity data, also from the NCI-60 panel, to predict response to chemotherapy in patient samples. They also recognized the need to incorporate aspects of the unique characteristics of the tumor data in building the model, in their case by identifying genes that showed evidence of co-expression in both the cell lines and tumors as the source of genes to be used in model building.

The issue of how the modeling was performed, recognizing the need to accommodate global distinctions in the nature of gene expression data derived from cell lines versus tumors, was presented by the Nevins and Potti group as the fundamental basis for the controversy with the MD Anderson scientists and shaped the manner in which they and ultimately the institution responded. There was considerable confidence within the Nevins/Potti group that the higher level methodological concepts were scientifically valid and appropriate to be developed into a translational framework. The signatures appeared to the Nevins/Potti group to be reproducible in further validation efforts, including using methods that directly addressed the criticism of the MD Anderson investigators. Specifically, the question of completely separating the validation and training datasets had been addressed in the work that developed the signatures for use in clinical trials.

This latter effort had not yet been published and was not known to the MD Anderson investigators. This was seen as critical evidence by Dr. Nevins that the methodologies were robust and that the questions being raised had been fully addressed.

Two further aspects of the work lent confidence to Dr. Nevins that the chemotherapy response signatures were robust – both of which he believed represented further validations beyond what had been described in the original work published in Nature Medicine. In one instance, signatures were validated in a dataset from an EORTC clinical trial, exhibiting very significant ability to distinguish responders from non-responders. Based on assurances by Dr. Potti that the validations were done blinded to the clinical outcome, and making use of the originally-defined signatures, this evidence was considered substantial in giving confidence in the methodology. However, it is now clear that the clinical data for this validation was provided to Dr. Potti from the beginning, raising questions about the extent to which the analysis was indeed blinded. Although Drs. Baggerly and Coombes claimed that in their hands the predictions did not work, it was the belief of Dr. Nevins that this failure was due to not accommodating the differences between cell line data and tumor data.

Additionally, in the course of developing methods for the clinical trials, a validation study was performed using a dataset from MD Anderson of 133 breast cancer patients treated with TFAC chemotherapy. Although a portion of this data had been used for validation in the original Nature Medicine publication, the majority of the samples were new. Based on the analysis with this new data that provided validation of the original predictor using the methods developed for the trial, this appeared to provide further compelling evidence for the robust performance of the predictor as originally described. Taken together, these measures appeared to provide further validation and confidence to Dr. Nevins that many of the issues raised by Drs. Baggerly and Coombes, while certainly identifying errors in data presentation, were not identifying fundamental flaws in the predictive signatures. However, it is also now clear that the clinical data for this validation set was corrupted in a manner that improved the apparent performance of the chemotherapy response signature (discussed in more detail later).

In 2008, Drs. Baggerly and Coombes wrote a second letter to Nature Medicine citing deficiencies in the Nevins/Potti reply and raising further issues. The editor of Nature Medicine asked for a response to this letter, which was provided by Drs. Nevins and Potti on behalf of the research collaborators. Ultimately, Nature Medicine accepted the response and decided not to publish the correspondence, but only a correction of one analysis based on our investigators' recognition of duplication of samples in a validation dataset (*Nature Medicine,* 14:889 (2008)).

Drs. Baggerly and Coombes wrote additional letters of concern regarding data that had been published in a 2007 JCO paper and a 2007 Lancet Oncology paper. In each case, Duke leadership understood Drs. Nevins and Potti to have responded with either corrections or a response to the concerns. The debate and dialogue on these issues was seen as a disagreement between the two groups on the application of the methodology, and despite these multiple elements of correspondence, it was not presented or

recognized as a criticism or implication of underlying data corruption. The Nevins/Potti group remained confident in the results because of various other assessments in which what they saw as the fundamental point of disagreement (use of validation data to build principal components) was addressed through other means (for instance, using an independent dataset not part of the validations as the source of data for building principal components as described above).

In September 2009, the situation changed when Drs. Baggerly and Coombes published an article in Annals of Applied Statistics, re-stating some of the issues previously raised in the Nature Medicine correspondence, including those issues that Nature Medicine chose not to publish. They further stated that although they believed the signatures developed by the Duke investigators to predict response to chemotherapy did not work, in the event that they were incorrect and the predictors were functional, there was a further problem in that the pemetrexed signature in the published work appeared to be reversed. Most importantly, they expressed concern that this reversal might be putting patients at risk by directing therapy in a manner opposite to that intended.

On September 22, 2009, Dr. Jeff Abrams, Associate Director of Cancer Therapy Evaluation Program (CTEP) of NCI contacted Dr. Kim Lyerly, Director of the Duke Comprehensive Cancer Center, and subsequently Dr. John Falletta, the Senior Chair of the Duke IRB, to express CTEP's concerns regarding the genomic predictors developed by Drs. Potti and Nevins. Dr. Abrams referenced the Annals of Applied Statistics article, which came to NCI's attention during the course of NCI's review of a protocol submitted for CTEP review. He noted that the Baggerly/Coombes paper documented numerous instances of data errors and inconsistencies in a strikingly large number of publications originating from the research of Drs. Nevins and Potti over a period of several years. Further, Dr. Abrams indicated that NCI had been able to confirm some of the discrepancies comparing information in published papers to information about these predictors that had been submitted to NCI in review settings. Dr. Abrams told Dr. Lyerly that given the large number of issues raised, it would be prudent to more fully investigate the reliability and credibility of the series of genomic predictors published and promoted by this team of investigators.

Dr Lyerly notified Dr. Sally Kornbluth, Vice Dean for Research in the Duke School of Medicine, having been appointed to that position in 2009. Dr. Kornbluth, a basic scientist whose background is in Molecular Cancer Biology, in turn informed senior leadership within Duke Medicine. After discussing the issues with Drs. Nevins and Potti at length and reviewing the journal exchanges, Dr. Kornbluth reported to senior leadership that this appeared to be a matter of legitimate scientific disagreement appropriately being debated in the scientific journals. On September 30, a conference call took place between Drs. Kornbluth and Lyerly, and Dr. Abrams, Dr. Lisa McShane, and four other NCI staff to ensure all parties understood the concerns. Drs. Abrams and McShane stated that they had come across inconsistencies and errors in the Nevins/Potti data in the course of reviewing a then current CALGB trial application to CTEP that they believed could have implications for the on-going clinical trials. It was also discussed how Duke, as an institution, would address these concerns.

One of the issues was that of patient safety in the ongoing clinical trials based on the possibility that a signature for pemetrexed sensitivity was reversed with respect to the training samples that were used for distinguishing sensitive and resistant cell lines. The suggestion was that if the signature were reversed, patients in the clinical trial could be placed at risk by being treated in the opposite manner than what was intended. The clinical trial referred to in this statement was one in which the pemetrexed signature was actually not used to guide therapy (NCT00509366). In this trial, patients predicted to be resistant to cisplatin based on the signature were treated with pemetrexed. Nevertheless, the pemetrexed signature was being used to guide treatment in a second lung cancer trial (NCT00545948) and thus, a reversal would be a problem. Based on the concern for patient safety and the apparent need for a deeper investigation into the issues surrounding the genomic predictor methodology, Duke administration asked the principal investigators on the three clinical trials to voluntarily close their trials to enrollment while concerns about the underlying methodology could be addressed. Evaluation of the predictor model for pemetrexed found that the signature was not reversed relative to the cell-line identities. Further, the gene expression data applied in the *in vitro* derived models were confirmed to match publicly available sources (from NCI-60 data from the NCI DPT site for pemetrexed). Thus, there were no irregularities identified in the provenance of the cell line data, or application of the chemotherapy predictive models in the adjuvant chemotherapy trial.

Regarding Drs. Baggerly and Coombes suggestion that the pemetrexed sensitivity signature was reversed, it was believed that their conclusion was based on the use of a published heat map for the pemetrexed signature and then working back to identify cell lines that could produce that heat map. Then, when they examined these cell lines they found that there was a reversal of the labels for sensitive and resistant. This possible reversal of sensitivity/resistant labels could have been the result of a reversal of the training set as suggested by Drs. Baggerly and Coombes, which would have indeed negated the predictor. However, as it turned out, the published heat map had been reversed in the course of the journal preparing the figure for publication (a correction was published in JCO, Jun 1, 2805 (2010)). The identification and acknowledgement of this error in the spring of 2010, was also viewed by Duke leadership as further validation of its external reviewers' conclusions as will be discussed in further detail later.

In response to the patient safety concerns and issues raised by CTEP, a plan was developed for an independent review of the Duke methodology. This plan was vetted by Dr. Nancy Andrews, Dean of the School of Medicine and Dr. Victor Dzau. At Duke's request and in order to ensure an independent review, Drs. Abrams and McShane provided the names of several statisticians whom they considered to have the appropriate expertise to review the validity of the science.

Because her husband collaborates closely with Dr. Nevins, Dean Andrews recused herself from further involvement in the Duke investigation of the science. Additionally, because the issues being raised involved not only scientific issues, but also clinical care, Dr. Michael Cuffe, Vice Dean for Medical Affairs was brought in to work directly with

Dr. Kornbluth, Dr. John Harrelson (Duke IRB Chair standing in for Dr. Falletta) and Dr. Dzau on these issues. This leadership group was advised by others as well.

In the process of putting the clinical trials on hold in early October, Duke also immediately initiated discussions about whether patients already enrolled in the trials should continue with their protocol prescribed chemotherapy regimens. Dr. Lyerly wrote to several cancer center directors asking them to assist the IRB on this issue. Dr. Harrelson then contacted these directors describing the three trials, the treatment arms, and the situation that led to the trials being closed to enrollment, and inquired whether it would be standard practice to complete the treatment phase for the individual patients already in the midst of treatment within the trial. The recommendations were to keep patients on their assigned chemotherapy regimens.

Dr. Harrelson officially replied to the NCI on October 6, 2009, informing them that the IRB would initiate the following actions:

1. To ensure the safety of subjects in these trials, we will conduct a review that will utilize independent external experts to evaluate the advisability of allowing currently enrolled subjects to remain on study.
2. We will request the relevant data safety monitoring boards submit a current review of the studies to the IRB within the week.
3. We will be reviewing the scientific underpinnings of these trials with the use of external experts, to assess, analyze and comment on the specific issues cited by Baggerly and Coombes in their Annals of Applied Statistics manuscript.
4. We will inform the sponsors of the studies of the actions we are taking.

Of the experts suggested by NCI, Duke's IRB approached three with one agreeing to serve on a review panel. Therefore, that reviewer was asked for the name of another qualified potential reviewer, which was provided and submitted by Duke to Dr. McShane for vetting. She expressed her confidence in, and approval of, this second reviewer based on the reviewer's bioinformatic expertise. This second reviewer was invited to participate and accepted.

Dr. Daniel George, Associate Professor of Oncology and Director of the Duke Clinical Trials Shared Resource, also initiated actions to have the trials re-examined by the DSMB Plus, and the protocols re-reviewed by the Duke Cancer Protocol Review Committee. Neither body identified any patient safety concerns.

By mid-October, the external review was beginning under the peer review auspices of the Duke IRB. The IRB peer review process was utilized specifically to provide anonymity to the reviewers to enhance their independence to conduct the review and provide their candid assessment of the Nevins/Potti science and the objections of Drs. Baggerly and Coombes based entirely on review of the data. Dr. Harrelson sent a letter to the external reviewers asking them to address the issues that Duke understood, at that time, to be the primary questions:

1. Have the methodology errors originally communicated by the MD Anderson Cancer Center researchers, Baggerly and Coombes, been adequately addressed by the Duke researchers?
2. Do the methods as originally developed and as applied in the context of these trials remain valid?

In engaging the reviewers, Duke's intent was that the reviewers would have unfettered access to all of the data, software, analyses and that they could request any other information needed from Drs. Nevins and Potti. In addition, Drs. Nevins and Potti were asked to prepare extensive material for the reviewers, including all information needed to address the concerns of Drs. Baggerly and Coombes. Despite the assumption of Duke leadership that the reviewers would have complete and unfettered data access, we now know that unknown to Dr. Nevins, the sample labels associated with the gene expression data for the cisplatin signature had been changed from the tumor ID to a simple numbering system (Sample 1 to Sample 59) for the review and that the reviewers were unknowingly testing methodology using corrupt datasets that could only inaccurately yield positive predictors.

The Duke IRB's engagement of the external experts to perform an independent evaluation of its investigators' research was an unusual step for the institution. At that point in time, while there were mounting concerns about the data, there were no allegations of research misconduct and the researchers remained adamant in their conviction that their work was reliable and accurate. Therefore, Duke leadership was committed to enabling a rigorous and comprehensive independent review while also ensuring that it was as fair and balanced as possible to the investigators, one of whom, Dr. Nevins, is an esteemed Duke researcher who has served at Duke since 1987, in numerous leadership positions both at Duke and in professional associations. Duke administration - based on the lack of proof of flawed science, along with the investigators' insistence of the quality of the data - wanted to avoid any premature judgment or condemnation of this work in the absence of an objective review.

On November 9, 2009, Dr. Kornbluth received an email from Dr. Baggerly describing his analysis of supplementary data posted to a Duke web page that was in relation to the previously published JCO paper regarding the cisplatin and pemetrexed signatures. He stated that since Drs. Potti and Nevins had said that they were preparing a paper clarifying their methods, and since this JCO paper was a subject of dispute, he assumed that there would be corrections regarding previously identified clerical errors. Dr. Baggerly went on to document two primary issues – that the pemetrexed signature was reversed, and that many, if not all, of a collection of ovarian cancer samples were incorrectly labeled. This Duke web page had been used for the sharing of data by the investigators involved in developing the new manuscript, as well as preparing for the review. Dr. Kornbluth forwarded Dr. Baggerly's communication directly to Dr. Harrelson at the IRB so that it would be entered into the peer review process (with the assumption that this information would be provided to the external reviewers along with any other

material prepared by Drs. Nevins and Potti). Dr. Kornbluth also thanked Dr. Baggerly and informed him that the material he sent would be provided to the IRB.

Dr. Harrelson then forwarded the material from Dr. Baggerly to Drs. Nevins and Potti noting that he was not sure if this represented old information or whether there were any additional claims that they wanted to address in the response document being prepared for the external reviewers. At this point, Dr. Nevins expressed his strong objection to Dr. Kornbluth and others on the leadership team, believing that this was an improper intrusion by Dr. Baggerly into an independent review process commissioned by the Duke IRB, that the pemetrexed issue was not new and had already been addressed, and that the issues relating to sample mislabeling had to do with data being assembled for the review involving work in the new manuscript in which there were changes in predictors for implementation of the trial. Based on his commitment to fairness to faculty, Dr. Nevins' conviction and arguments, and in recognition of his research stature, Dr. Dzau concurred that the reviewers should examine the data independent of the Dr. Baggerly email. It was believed that the conclusions of a thorough and objective review of all of the data would speak for itself. There was no other discussion on this point and Dr. Baggerly's communication was not further disseminated.

The external reviewers proceeded expeditiously, running the data provided by Drs. Nevins and Potti using their own methodology, similar to the Duke methodology, and guided by the parameters supplied by Drs. Nevins and Potti. The reviewers were provided with what Duke leadership believed, and were assured, was all of the data used to develop and validate the predictors. Importantly, it was believed that there was no restriction on the reviewers access to source data, either in Duke databases or publicly available datasets, and there was no limitation on their ability to verify the integrity of data used in the development of predictors, as well as the validation of predictors, although that was not specifically within their charge. However, as noted above, this was not the case and by its very nature, the corruption in the underlying datasets was not easily discernable as will be discussed further.

On December 22, 2009, the external reviewers provided a written report.

> "We were given two charges by Dr. Harrelson and the Duke IRB. The first was 'Have the methodology errors originally communicated by the M.D. Anderson Cancer Center researchers, Baggerly and Coombes, been adequately addressed by the Duke researchers?' and the second 'Do the methods as originally developed and as applied in the context of these trials remain valid?' We reviewed the responses provided, read the accumulated literature, reviewed the R code, attempted to replicate the methods proposed, and conducted a two-hour in person meeting with Drs. Potti, Nevins and Barry on December 16, 2009. For the first charge we think that the Duke investigators have, with a few caveats expanded upon in the next two paragraphs, responded sufficiently to the comments of Baggerly and Coombes. For the second charge we were able to show with an independent analysis that the approaches used in the Duke clinical

> predictors are viable and likely to succeed. In summary we believe the predictors are scientifically valid and with a few additions can be fully responsive to the comments of Drs. Baggerly and Coombes."

Based on the external reviewers' report, which was shared with the principal investigators of the three clinical trials, Duke informed the NCI on January 7, 2010, of the principal investigators' intent to reopen the clinical trials.

On January 8, Dr. Nevins informed the leadership team that he had been previously contacted by Dr. Richard Shilsky of the Cancer and Leukemia Group B (CALGB). Dr. Shilsky informed Dr. Nevins that he had been contacted by Drs. McShane and Abrams who expressed interest in re-evaluating the work that formed the basis for the CALGB 30506 lung cancer prognosis trial (a Phase III trial focused on early stage lung cancer that used the Drs. Nevins and Potti recurrence predictor). This was the first that the leadership team was aware of this further challenge to the work.

Dr. Kornbluth received an email from Dr. McShane on January 20, to update her on NCI's analyses pertaining to CALGB-30506 (lung metagene trial). Dr. McShane stated that she remained unable to reproduce the results obtained by the Duke investigators two years earlier and would be discussing the matter further with Drs. Nevins and Potti to give them the opportunity to clarify how they obtained their original results. After multiple exchanges involving CTEP and the CALGB leadership, the results still could not be reproduced. Eventually, it was decided that the trial would proceed, but that the genomic analysis would no longer be a co-primary aim of the study.

In early April 2010, Dr. Potti was contacted by Dr. William Timmer, who was the CTEP program director for Dr. Potti's R01 grant CA131049-01A1. The letter stated that since the NCI was not involved in the original external review of the research and because the advanced lung trial overlapped aims of his award, CTEP wanted to review the data and code associated with the cisplatin and pemetrexed sensitivity predictors. Dr. Jeff Abrams, Associate Director of CTEP, and Dr. Roy Wu, Chief of the CTEP Clinical Grants and Contracts Branch, were copied on the email, and Dr. McShane's email was provided so that any questions or clarifications about the items being requested could be directed to her. The letter issued a request for clarifications, as well as provision of supporting data and software, to attempt to reproduce the work that had been published in JCO in 2007, and used as supporting data for Dr. Potti's R01. This work focused on the development of the gene expression signatures predicting responses to cisplatin and pemetrexed. In response to the request, Dr. Potti was to provide clarifications and explanations as well as the software and data that would allow Dr. McShane to reproduce the work. However, even with the additional details, she was unable to reproduce key results. Including a result demonstrating validation of the cisplatin signature that used a panel of ovarian cancer cell lines treated with cisplatin. Dr. Potti provided only some data and no computer code or explanations for the pemetrexed sensitivity predictor, so Dr. McShane was not able to examine the basis for that predictor.

Following up on Dr. McShane's inability to reproduce the expected results, the leadership team decided it would be helpful for Dr. Potti to speak directly with Dr. McShane to determine how best to resolve these outstanding issues. After several communications between Dr. Potti and the NCI regarding the questions about the cisplatin results in the grant, it was determined that a meeting would be beneficial to address the issues and concerns. Dr. Kornbluth contacted Drs. Abrams and McShane and it was agreed that a meeting would be productive. At the time, Dr. McShane also spoke with Dr. Kornbluth to clarify the nature of her concerns, which included possible biased selection of cell lines and/or tumor data for analysis, i.e. cherry picking of data. Even at this time, the Duke leadership team believed it likely that if its investigators provided complete detail of the development of the methodology to Drs. Abrams and McShane that the work would be validated.

On June 29, 2010, the Duke group (Drs. Kornbluth, Willard, Barry, Nevins and Potti) travelled to Bethesda to meet with the CTEP group. At the meeting at NCI, a number of issues were brought into focus for Duke leadership. It was clear that CTEP would continue to have concerns until they could be provided with the exact code and data from the JCO paper such that they could run the analysis and get the same output as reported in JCO and as presented in the R01 grant. Dr. Potti had led Dr. Kornbluth to believe that the BinReg code (the original code developed by Mike West, used by the laboratories of Drs. Nevins and Potti during the time frame of this research) necessary for Dr. McShane to recreate the results had been provided, but in actuality, Dr. Potti provided a version of the code that was not what was originally used in the JCO paper, and did not produce the figures therein. At the conclusion of the meeting, Dr. Barry agreed to attempt to reproduce the original JCO analysis and providing Dr. McShane with the original code and necessary data transformation instructions, so that she could reliably obtain the published results.

These analyses were begun soon after the group returned to Duke and were well underway when, on July 15, 2010, inconsistencies in Dr. Potti's curriculum vitae came to light and he was immediately placed on administrative leave while these issues were investigated. Due to the nature of the curriculum vitae issues, i.e. alleged falsification of credentials that were shared immediately with the principal investigators of the three clinical trials, the principal investigators and Duke administration agreed to voluntarily close the trials to enrollment without delay. In the following week, NCI received a letter from a group of the country's leading biostatisticians expressing their concerns related to this situation and the safety of patients in the clinical trials.

Dr. Barry, who as previously noted was not a recipient of Dr. Baggerly's November 2009 communication, continued to evaluate the data sources and analysis methods of the work described in the 2007 JCO paper with a focus on the ovarian cancer cell line data since this was the primary focus of the NCI review. Dr. Barry located a primary source to the experimental data that was inconsistent with a table provided for the NCI review that appeared to incorrectly match drug sensitivity results for individual cell lines. This was provided to Dr. Nevins for confirmation on August 5, 2010, at which point Dr. Nevins identified further discrepancies in the actual measures of drug sensitivity derived from the

experimental data. Further, Dr. Barry determined that these mismatches had the net result of improving the prediction of drug sensitivity with the cisplatin signature. In light of what appeared to be a non-random mislabeling of samples in this cell line dataset that resulted in a positive effect on predictions, Dr. Nevins was prompted by the previous November 2009 communication from Dr. Baggerly concerning mislabeling of ovarian tumor samples, to further examine the ovarian tumor data. This analysis revealed that incorrect labeling of the samples also appeared to be non-random and yielded robust predictions of drug response while predictions with correct clinical annotation did not give accurate predictions.

Further analyses revealed corruption of multiple datasets compiled by Dr. Potti that had been used as sources of validation of the various chemotherapy sensitivity signatures. These included data derived not only from Duke sources, but also publicly available data. As an example, a dataset of 133 samples from a neoadjuvant breast cancer study at MD Anderson involving patients treated with the combined regimen TFAC was used for validation of an adriamycin signature. The clinical annotation that was assumed to be used by Dr. Potti included 34 responders and 99 non-responders, the same distribution as reported by MD Anderson. However, a detailed comparison of the two datasets revealed that the response information was reversed for 24 cases with 12 labeled incorrectly in each direction. In this case, the corrupted data yielded positive validation results whereas the accurate data did not provide evidence for validation. Similar findings of corruption of data in key validation datasets were observed in other instances.

As a result, three publications were retracted, a manuscript describing the methods for implementing signatures in the clinical trials that was under review was removed from further consideration, and other publications are currently being analyzed. Dr. Potti issued his resignation statement on November 19, 2010, and a statement of responsibility for the problems with the work. A research misconduct investigation is in progress.

In light of the findings concerning the lack of reproducibility of chemotherapy response predictors, and considering the concerns raised by the NCI with respect to the lung metagene score recurrence predictor, a re-evaluation of the data described in the 2006 NEJM paper was also undertaken. This analysis failed to reproduce the key original findings regarding prediction of recurrence within two independent validation datasets. As a result, the authors of this study have retracted that paper.

There are many lessons to be learned from this experience, but the immediate lessons that Duke and the IGSP have learned are that all data and methods for clinical research must be assessed at multiple levels and that quantitative expertise is needed for complex analyses; furthermore, for translation to clinical trials these analyses must be done using systems that maintain independence between the data generation and the analysis and enable replication of the results, along with documentation of all changes to data and analyses. The IGSP is committed to ensuring full publication of data and methods going forward and would note the Gatza 2010 publication as an example of that commitment.

Sustained statistical collaboration is critical to assure proper management of these complex datasets for translation to clinical utility, as illustrated by the efforts of Dr. Barry to re-evaluate prior work without clear primary sources for the data, and records of the precise use of statistical methodologies and programs. The fundamental methods of managing data and validating statistical algorithms are not something basic scientists are generally familiar with, thus statisticians need to take an active role in participating in basic science research, both in terms of teaching research methods and in improving the design of studies.

Part of what contributed to the inability of Duke's investigators to publish complete data and methods, even after the external reviewers recommended just such action, was an inability to fully track down the exact data and analyses underlying the early genomic predictor publications. Therefore, the implementation and utilization of systems that provide the ability to track and record each step in these types of complex projects is critical. There are often a very large number of variables that go into the model building process and during the exploratory phase, literally hundreds of variations might be evaluated. The importance of the ability to record this trail of exploration cannot be underestimated. Thus, systems that can track this automatically, providing records of what was done and an ability to trace all steps in the analysis is enormously valuable. Examples of such automated systems include SWEAVE as utilized at MD Anderson, by many faculty members in the Duke Biostatistics and Bioinformatics Department, and other institutions and a system developed during this past year at Duke (QUADRA) that has been implemented in recent studies.

In response to the foregoing series of events, Dr. Dzau charged a group of Duke's leading researchers to evaluate the processes and procedures at Duke for ensuring the highest quality of work for genomic science destined for clinical trials. The work of this group, the Translational Medicine Quality Framework (TMQF) committee, has been provided to you in a separate document. Implementation of the TMQF recommendations is underway at Duke, but we look forward to learning much more from the results and recommendations of this committee.

Finally, as stated in the first paragraph of this document, this situation represents a regrettable event that has had a broad impact. At this point, it is imperative that we, the IOM committee, and the research community broadly learn everything possible from this experience in order to ensure that sophisticated and redundant processes are created together to optimally guide and safeguard this important area of translational research. Again, we appreciate the work of this committee and look forward to your ultimate recommendations and assessment.

References

Black, E. P., Hallstrom, T., Dressman, H. K., West, M., and Nevins, J. R. (2005). Distinctions in the specificity of E2F function revealed by gene expression signatures. Proc. Nat'l. Acad. Sci. USA 102, 15948-15953.

Bild, A., Yao, G., Chang, J.T., Wang, Q., Potti, A., Chasse, D., Joshi, M.-B., Harpole, D., Lancaster, J.M., Berchuck, A., Olson, J. A., Marks, J. R., Dressman, H. K., West, M. and Nevins, J. R. (2006). Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature *439*, 353-357.

Carvalho, C., Chang, J., Lucas, J., Nevins, J.R., Wang, Q., and West, M. (2008). High-dimensional sparse factor modelling: applications in gene expression genomics. J Am Stat Assoc *103*, 1438-1456.

Chang, J.T., Carvalho, C., Mori, S., Bild, A., Gatza, M., Wang, Q., Lucase, J.E., Potti, A., Febbo, P., West, M., and Nevins, J. R. (2009). A genomic strategy to elucidate modules of oncogenic pathway signaling networks. Molecular Cell *34*, 104-114.

Gatza, M.L., Lucas, J.E., Barry, W.T., Kim, J.-W., Wang, Q., Crawford, M., Datto, M., Kelley, M., Mathey-Prevot, B., Potti, A., and Nevins, J. R. (2010). A pathway-based classification of human breast cancer. Proc Nat'l Acad Sci *107*, 6994-6999.

Gatza, M. L., Lucas, J. E., Berchuck, A., Nevins, J. R., and Cheng, Q. (2011). Linking pathway activity to genome alterations in a pathway-defined classification of ovarian cancer. Manuscript submitted.

Hallstrom, T.C., Mori, S., and Nevins, J.R. (2008). An E2F1-dependent gene expression program that determines the balance between proliferation and cell death. Cancer Cell *13*, 11-22.

Huang, E., Cheng, S.H., Dressman, H., Pittman, J., Tsou, M.H., Horng, C.F., Bild, A., Iversen, E.S., Liao, M., Chen, C.M., West, M., Nevins, J. R., and Huang, A. T. (2003a). Gene expression predictors of breast cancer outcomes. Lancet *361*, 1590-1596.

Huang, E., Ishida, S., Pittman, J., Dressman, H., Bild, A., Kloos, M., D'Amico, M., Pestell, R.G., West, M., and Nevins, J.R. (2003b). Gene expression phenotypic models that predict the activity of oncogenic pathways. Nat Genet *34*, 226-230.

Ishida, S., Huang, E., Zuzan, H., Spang, R., Leone, G., West, M., and Nevins, J.R. (2001). Role for E2F in the control of both DNA replication and mitotic functions as revealed from DNA microarray analysis. Mol Cell Biol *21*, 4684-4699.

Karra, R., Vemullapalli, S., Dong, C., Herderick, E.E., Song, X., Slosek, K., Nevins, J.R., West, M., Goldschmidt-Clermont, P.J., and Seo, D. (2005). Molecular evidence for arterial repair in atherosclerosis. Proc Natl Acad Sci U S A *102*, 16789-16794.

Liu, Z., Wang, M., Alvarez, J.V., Bonney, M.E., Chen, C.-c., D'Cruz, C., Pan, T.-C., Tadesse, M.G., and Chodosh, L.A. (2008). Singular value decomposition-based regression identifies activation of endogenous signaling pathway *in vivo*. Genome Biology *9*, R180.181 - R180.111.

Loboda, A., Nebozhyn, M., Klinghoffer, R., Frazier, J., Chastain, M., Arthur, W., Roberts, B.E., Zhang, T., Chenard, M., Haines, B., Andersen, J., Nagashima, K., Paweletz, C., Lynch, B., Feldman, I., Dai, H., Huang, P., and Watters, J. (2010). A gene expression signature of RAS pathway dependence predicts response to P13K and RAS pathway inhibitors and expands the population of RAS pathway activated tumors. BMC Medical Genomics *3*, 1775-8794.

Pittman, J., Huang, E., Dressman, H., Horng, C.-F., Cheng, S.-H., Tsou, M.-H., Chen, C.-M., Bild, A., Iversen, E.S., Huang, A.T., Nevins, J. R., and West, M. (2004). Integrated modeling of clinical and gene expression information for personalized prediction of disease outcomes. Proc Nat'l Acad Sci *101*, 8431-8436.

Seo, D., Wang, T., Dressman, H., Herderick, E.E., Iversen, E.S., Dong, C., Vata, K., Milano, C.A., Rigat, F., Pittman, J., Nevins, J. R., West, M., and Goldschmidt-Clermont, P. J. (2004). Gene expression phenotypes of atherosclerosis. Arterioscler Thomb Vasc Biol *24*, 1922-1927.

Spang, R., Zuzan, H., West, M., Nevins, J.R., Blanchette, C., and Marks, J. (2002). Prediction and uncertainty in the analysis of gene expression profiles. In Silico Biol *2*, 369-381.

West, M., Blanchette, C., Dressman, H., Huang, E., Ishida, S., Spang, R., Zuzan, H., Olson, J.A., Jr., Marks, J.R., and Nevins, J.R. (2001a). Predicting the clinical status of human breast cancer by using gene expression profiles. Proc Natl Acad Sci USA *98*, 11462-11467.

Zhang, X.H., Wang, Q., Gerald, W., Hudis, C.A., Norton, L., Smid, M., Foekens, J.A., and Massague, J. (2009). Latent bone metastasis in breast cancer tied to Src-dependent survival signals. Cancer Cell *16*, 67-78.