Duke Institute for
Genome Sciences & Policy

Duke University
CIEMAS
101 Science Drive
Box 3382
Durham, NC 27708

T 919-684-4151
F 919-668-0795

www.genome.duke.edu

February 5, 2007

The Honorable Michael O. Leavitt
Secretary of Health and Human Services
The U.S. Department of Health and Human Services
200 Independence Avenue, S.W.
Room 434E
Washington, D.C. 20201

RE: Personalized Health Care RFI

Dear Secretary Leavitt:

We are pleased to provide a response to the Request for Information on Improving Health and Accelerating Personalized Health Care Through Health Information Technology and Genomic Information in Population- and Community-based Health Care Delivery Systems. We believe the field of personalized medicine has tremendous promise to prevent and treat disease resulting in improved health outcomes in a cost-efficient manner. However, many steps must be in place to enable delivery of these new technologies, namely the health information technology infrastructure and robust clinical evidence.

The Duke Institute for Genomic & Science Policy (IGSP) has been developing enabling strategies for personalized medicine's integration and adoption by healthcare systems. We believe our efforts may serve as a prototype for some aspects of this RFI. The IGSP consists of a multidisciplinary network of Centers and programs, including the Center for Applied Genomics & Technology, Center for Genome Ethics, Law & Policy, Center for Genomic Medicine, Center for Bioinformatics & Computational Biology, Center for Models of Human Disease, and Center for Population Genomics & Pharmacogenetics. Together, these centers form an integrated approach to advancing the Genome Revolution and addressing its implications for health and society.

In addition, Duke University was one of 12 academic health centers recently funded through NIH's Clinical and Translational Science Awards to transform how clinical and translational research is conducted, ultimately enabling researchers to provide new treatments more efficiently and quickly to patients. Under this award the Duke Translational Medicine Institute (DTMI) was established and organized around three integrated pillars: the Duke Translational Research Institute (DTRI), the Duke Clinical Research Institute (DCRI), and the Duke Center for Community Research. Among its many goals, the DTMI aims to transform how discoveries are translated into improved medical care and develop a community model to translate findings of research from bench to bedside to population. The IGSP has a central role in these translational efforts.

In the pages that follow we have provided comments on selected information points requested in the RFI based on our current efforts and expertise at Duke. We hope this information helps the committee gain a broader understanding of the efforts in this area and define future efforts needed to facilitate the translation of personalized medicine for the benefit of all Americans.

We look forward to the committee's efforts in this area and would be pleased to provide additional information upon request.

Sincerely,

Huntington F. Willard, PhD
Director, Institute for Genome Sciences & Policy
Vice Chancellor, Genome Sciences
Nanaline H. Duke Professor of Genome Sciences
E-Mail: hunt.willard@duke.edu

Geoffrey S. Ginsburg, MD, PhD
Director, Center for Genomic Medicine
Duke Institute for Genome Sciences & Policy
Professor of Medicine
Professor of Pathology
E-Mail: geoffrey.ginsburg@duke.edu

Susanne B. Haga, Ph.D.
Assistant Research Professor and IGSP Scholar
Duke Institute for Genome Sciences & Policy

Comments on Selected Information Points **Page**

I. Concepts on anticipated approaches for the use of EHR and population- and community-based health care system databases for longitudinal data collection in addressing: disease susceptibility, clinical course and outcomes, treatment response, evidenced-based clinical decision support, and optimal healthcare delivery systems 4

II. Anticipated applications of genomic-based clinical testing in medical decision-making, safety assessment, and risk management. 10

III. Establishment of biospecimen resources obtained from clinical medical services for application in research, clinical trials, health services planning, clinical effectiveness, and health outcomes evaluations 12

IV. Organizational or institutional practices to address ethical, legal, and social implications regarding the use of patient information, including genetic data, to support personalized health care 14

V. Feasibility and potential benefits for establishing linkages of institutional or organizational data resources with private and publicly available health databases 15

VI. Development of ontologies across different clinical data repositories that will facilitate the utility of the data for answering clinical research questions 17

VII. Models for linking clinical data repositories across disparate care providers 17

VIII. Examples of the use of disease registries to track specific diseases and response to drug therapies across different subpopulations 19

IX. Strategies for accumulating patient data necessary for research that may not be available through EHRs 20

X. Concepts or models on the potential use of clinical data and related resources for research applications 21

I. Concepts on anticipated approaches for the use of EHR and population- and community-based health care system databases for longitudinal data collection in addressing:

- **– Disease susceptibility**
- **– Clinical course and outcomes**
- **– Treatment response**
- **– Evidenced-based clinical decision support**
- **– Optimal healthcare delivery systems**

Duke Center for Community Research and Electronic Health Record. Duke was one of the first institutions to develop a computerized registry of patients in an interoperable fashion that produced clinical notes, database coded data for research and measured outcomes. Remarkably, this vision was enacted in 1967 under the leadership of Dr. Eugene Stead with the goal of creating a living "computerized textbook of medicine". To integrate isolated clusters of data, establish better communication among researchers, and improve the cooperation between researchers and community, the Biomedical Informatics Core is working to empower investigators to be leaders in the impending global effort to redefine human disease by using combined clinical and molecular profiling and to measure changes in health status across multiple dimensions as a routine function of biomedical informatics.

The Duke University Health System (DUHS) has begun the development of an Ambulatory Electronic Health Record (AEHR). DUHS and the Duke Faculty practice, the Private Diagnostic Clinic, have entered into an AEHR system co-development agreement with McKesson Corporation to produce an integrated Ambulatory EHR for the DUHS. The McKesson Ambulatory EMR system, referred to as Horizon Ambulatory Care, will complement existing technology and database structures rather than replace them, to the extent possible and appropriate. A massive effort, supported by in kind support from the health system will bring together research experts from the Duke Clinical Research Institute (DCRI) with health system operations leaders and community practitioners to define the codified data elements necessary to conduct quality longitudinal studies on our population and act on defined best practice clinical pathways which. At this time, most AEHR systems recognize the transactional operational efficiency gained by not demanding common nomenclature and data standards, but instead either allowing free text or providing templates that often are passively accepted in the notes. Through these efforts, we also hope to train the next generation of medical informaticians – clinicians with an understanding of clinical epidemiology, health services research, clinical practice operations and medical informatics.

In addition to the AEHR system, the Department of Psychiatry and Behavioral Sciences at DUHS has spent the last eight years developing and implementing the Clinical Research Information Systems (CRIS), a successful blending of both a mental health-specific EHR and electronic data capture (EDC) product into a single source documentation system. CRIS is unique in that it meets both regulatory requirements of HIPAA and FDA requirements outlined in the federal regulations 21CFR Part 11 (3-10) for clinical trials. CRIS is capable of functioning simultaneously as both an EHR and EDC system by using codified standard libraries of clinical data including diseases (DSM-IV, ICD-9), side effects, symptoms, and medication and

presenting these to the users as discrete data elements to capture all pertinent clinical patient information.

The application is scalable; it is presently deployed on over 3,000 geographically dispersed desktops to over 400 users throughout DUHS. While CRIS was designed with the specific intent of combining research and clinical care, its ultimate function is as a usable EHR. In addition to its successful integration into the practice of psychiatry at Duke, it has recently expanded to other medical centers, including department-wide at the University of Texas at Galveston and at the Medical College of Virginia (MCV) by psychiatrists in the family practice setting. The UT Galveston application is executed as a client-server application, while the MCV application is executed via the Internet under an application service provider model from a server located at Duke.

Our ability to extend the AEHR relies on our particularly strong network within Durham County. This includes the three hospitals of DUHS and the research buildings of the Duke Campus, Durham Regional Hospital, Lincoln Community Health Center, the Davis Ambulatory Surgical Center, numerous clinics of the Duke University Affiliated Physicians, multiple community health clinics, home health services, urgent Care centers and innumerable physicians' offices.

The Duke Health Enterprise (DHE) Network Infrastructure consists of the Core, Distribution, Servers Farm, and User Access layers. The Core is a dual path of multiple multi-layer gigabit switches, fully routed Internet Protocol (IP) only traffic. The Distribution layer is divided into multiple segmentations using Virtual Local Area Network (VLAN IEEE 802.1Q). The distributed domains use Hot Standby Router Protocol for fault tolerant IP routing. The connectivity from the Core to Distribution, to the Server Farms is gigabit in bandwidth using single mode or multi-mode fibers. There are three Server Farms, mesh configured as primary and secondary for failover, in three physically separated buildings. Servers in a server farm are fiber or copper connected to high-speed gigabit multi-layer switches The Users Access layer resides in a local communication room with fiber optic, gigabit uplink. Copper cables deliver the functionality of 10/100 Mbps switched ports to the users. Users also enjoy access to the network through an 802.11 wireless system. The DHE Network is protected by a dual perimeter firewall with gigabit connection to the outside.

To the end-user, this network infrastructure manifests as the ability to access a protected intranet throughout the health system. With the necessary rights, users and client applications, clinicians and researchers can access clinical resources distributed throughout the campus and health system. In addition to the wired infrastructure, Duke Health Technology Solutions (DHTS) has been expanding a wireless infrastructure in its major centers. Therefore, it is possible to access the same clinical information through wireless technology without any configuration change, using a laptop from any area of all the major health centers in the health system.

In the next year, DHTS will oversee the creation of a dedicated optical fiber network linking its major centers. The extraordinary increase in bandwidth and speed will make it possible to distribute data, even single databases, across geographically remote sites in the health system. There will be corresponding improvements in the ability to move large data files, such as

DICOM images and streaming video as well as improvements in the robustness and security of this network.

Evidence-based medicine. Evidence-based medicine aims to formally integrate clinical risk factors and patient characteristics in models to better inform the processes of individualized treatment decision-making and the development of policy decisions by health systems related to groups of patients. We aim to take this approach to a higher level by formally integrating multi-dimensional molecular data sets from gene expression patterns, metabolic profiles, single nucleotide polymorphisms (SNPs), and imaging information to develop detailed molecular signatures of disease processes. These data will be combined with traditional clinical factors to improve clinical prediction. In parallel, this information will be invaluable in the identification of genes, pathways, and systems biology that define the pathophysiological basis of disease and response to therapy. The result will be a series of tools and information that can be used by investigators, clinicians and administrators that will enable improved understanding and precision of outcome assessment, identification of potential new therapeutic targets, and comprehensive models that will underlie the next generation of systems to support personalized medicine, improve outcomes, and reduce the costs of healthcare delivery.

The following initiatives illustrate how Duke is approaching the challenges of longitudinal data collection using EHRs or other clinical and biological databases:

Kannapolis Longitudinal Study. This study will be the modern-day version of the "Framingham Heart Study" and will collect biological samples and medical information on a large segment of the Kannapolis population (in the Piedmont region of North Carolina) over a long time period (ideally 25 years or greater). The full resources of the Core Laboratory and Translational Medical Center, to be located at the Kannapolis Biotechnology Park, will be available for the study. The data collected will provide an invaluable resource for translational research efforts. Assuming the vast majority of the Kannapolis population participates, this will be an unprecedented opportunity to both provide important health information to this community as well as study perhaps even the rarest of diseases. The data collected will enable so-called "reverse translation" of knowledge from bedside to bench, enabling biomarker discovery, validation, and the correlation of imaging modalities with specific human disease. Specific cohorts of patients could be identified for recruitment into clinical trials or studies of lifestyles or interventional diets. Moreover, this study has the potential to fully engage the local community in becoming stakeholders in the biotechnology enterprise.

A second component of the population studies in Kannapolis is aimed specifically at redefining disease taxonomy based on its molecular architecture. Thus, while the longitudinal health outomes study is underway, specific cohorts of patients with diabetes, depression, vascular disease, cancer, obesity, pulmonary disease, and immune medicated diseases such as rheumatoid arthritis will be assembled for case control analysis. It is our hypothesis that with precision phenotyping of patients and matching on demographic characteristics that novel molecular descriptors of disease can be generated through 'pan-omic' analyses (genotypes, transcript, metabolite, and protein profiles) of blood and relevant tissue samples and that this can be accomplished with feasible numbers of individuals in each cohort (~400 cases and 400 controls).

The results of these analyses will create a new definition of disease and a new field of epidemiology that will lead to better tools to match patients to therapies – personalized medicine.

Establishing Pharmacogenetic Methods and Collaboration in Clinical Research. One area of critical importance to speeding translation and improving the interpretation of clinical trials and thereby improving the balance of benefit and risk in therapeutics, will be to identify genetic variants in the drug metabolism and transport systems that contribute to inter-individual variation in drug safety and/or efficacy. Variability in the efficacy and safety of drugs in relation to serum concentrations is a major problem in clinical medicine and drug development. A meta-analysis of 39 prospective studies from U.S. hospitals suggests that 6.7% of in-patients have serious adverse drug reactions (ADR) resulting in over 100,000 deaths per year in the U.S. Of equal relevance is the fact that most presently approved therapies are not effective in a significant proportion of patients for whom they are prescribed.

It is now evident that genetic variation in drug metabolizing enzymes, drug transporters, and drug targets can have a significant effect on the efficacy and toxicity of medications. The correlation between drug response and/or toxicity has been clearly established with the presence of specific inherited variations in genes involved in biotransformation such as the cytochrome P450 family, N-acetyltransferase and glucuronosyltransferase enzymes and drug transporters such as the multi-drug resistance genes. This has been complemented by the concurrent development of new technologies that provide reliable and cost effective means of identifying variant alleles in the human population. Analysis of the relationship between genotypes of single genes and drug response in humans has been termed “pharmacogenetics”.

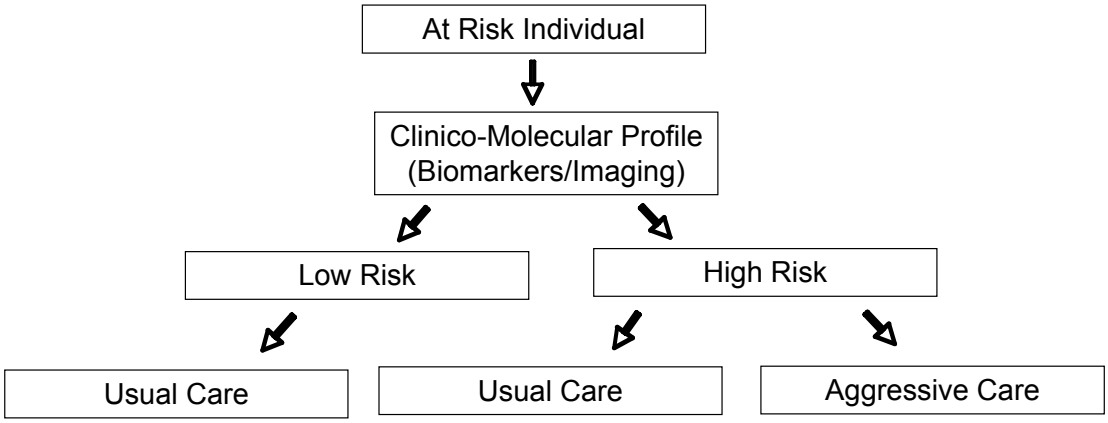
Understanding variation in drug response during the drug development process is of great interest to the Critical Path Initiative, which was originated by the FDA, but includes a consortium of industry, academia and multiple Federal agencies including the NIH. Recent publications by the FDA indicate the agency has a desire to complement the necessary empirical testing of clinical effects of drugs in broad populations with a more robust, mechanistic-based, hypothesis-driven process to understand ADRs and variability in drug efficacy. In 1999, over 150 clinical trials incorporated pharmacogenetic testing as part of the clinical protocol and as many as 80% of all trials collected DNA samples for potential genotyping.

The appropriate incorporation of pharmacogenetics into clinical research programs is critical for early phase studies and broader clinical trials. The goal of these initiatives is to identify genetic variants involved in biotransformation in order to understand how those genetic variants correlate with pharmacokinetic (PK) parameters that may ultimately relate to clinical efficacy or safety of a drug. These new methods will contribute to the development of novel therapeutics, perhaps combined with diagnostic testing to improve the balance of risk and benefit for individual prescriptions and health system policies.

Clinical Research Methods to Facilitate Delivery of Information to Clinicians for Clinical Decision Making. Many of the deliverables of scientific discovery programs will take the form of novel putative risk markers or biomarkers of disease or drug response. Of critical importance is to test both whether 1) the most promising associations discovered in smaller exploratory studies can be replicated in larger independent cohorts, and 2) whether these markers can be used

to explain variations in clinical outcomes and response to therapy. The former goal will result in the delivery of validated prognostic models for predicting individual patient outcomes that are substantially more powerful than currently available models. The latter goal will link these model predictions with variations in patient response to therapy.

In current clinical practice, it is often necessary to treat large numbers of patients with a given effective therapy in order to provide benefits for a few. This inefficiency, which is a direct result of our imprecise knowledge of the relationship between patient biology and mechanisms of therapeutic benefit, results in two particularly undesirable effects. First, many patients who are exposed to potentially harmful drugs and procedures in the name of modern therapeutics have only a small probability that they will actually be the ones to benefit. Second, routinely prescribing therapies that are actually ineffective in the majority of those who receive them results in a huge waste of resources (i.e., cost for no benefit). If therapies could be directed much more efficiently to those who will actually benefit, the resulting savings to the health care system would be enormous. Of perhaps equal importance is to suggest new relationships and models consisting of traditional and novel risk markers that can then feed back to the discovery projects for additional testing and exploration. The goal of these initiatives is to develop populations and frameworks for biomarker validation.



<u>GENErations Study</u>. The GENErations study aims to assess genetic variation of drug response in a community setting. It is being jointly conducted by IGSP and the Department of Community and Family Medicine. Unlike other pharmacogenetic studies which have carefully selected individuals in a clinical trial setting, this study is purposefully being conducted in the context of a busy family practice which we believe will be informative on not only the science of the genetics being studied, but the art (and science) of implementing DNA-based testing in an outpatient setting. Specifically, the study will determine the association of genetic variants known to affect drug response in commonly prescribed medications for hypertension, asthma, and hypercholesterolemia in patients newly diagnosed with these conditions. The study will also develop and test a clinical treatment algorithm based on these associations to determine the optimal therapy for an individual.

Prospective validation of biomarkers developed in discovery research against phenotypic measures of disease presence, severity, and outcomes. In general, this work will progress through 3 stages: 1) novel biomarkers or predictive models will be tested in isolation for association with the outcome of interest (univariable); 2) if a significant association is detected, the model will be retested, adding additional clinical factors (multivariable); 3) if the novel predictive model holds up, we will formulate a new model incorporating this factor as well as key clinical factors to optimize the model for prediction of patient outcome.

Several unique clinical cohorts have been established at Duke for this purpose including but not limited to the following:

1) Guilford (County) Genomic Medicine Initiative (GGMI). The Guilford County initiative is a Department of Defense-sponsored five year study to develop the models for clinical deployment of genomic medicine in the primary care setting. As part of this study several outpatient cohorts with variants in drug metabolizing enzymes, genes predisposing to vascular thrombosis (Factor V Leiden), and cancer (BRCA1/BRCA2) will be established and prospectively followed. The GGMI involves three partners: Moses Cone Health System (MCHS), University of North Carolina-Greensboro, and the Duke Center for Human Genetics with the goal of investigating and developing working models for incorporating medical genetics into patient care in an ethnically diverse community of patients, private-practice physicians (Guilford County) and a high-level, non-university medical center (MCHS).
2) The Duke Cardiovascular Disease Database contains detailed clinical, angiographic, therapeutic, and outcome data on over 180,000 patients. An ongoing effort to add genetic and other risk marker data will provide approximately 10,000 patients by 2005 with standard clinical and outcome data and banked blood for analysis. Follow-up remains 96% complete. This data set provides a low-power view of atherosclerosis phenotype (clinical cardiac events) as well as a high-power view of coronary atherosclerosis by angiography.
3) DCRI is coordinating five major multi-center cardiovascular clinical trials that have included prospective collection of banked blood for future analysis.
4) CATHGEN is a registry of all patients undergoing cardiac catheterization since 2000. Now numbering more than 6,000, these patients have undergone extensive cardiovascular phenotyping at the time of angiography and are followed prospectively yearly for subsequent major adverse cardiac events (MACE). Since 2000, DNA and serum has been collected and stored and since 2004, RNA has been extracted and stored. This registry has allowed the identification and validation of predictive candidate biomarkers for MACE with one and two years of angiography – a study that is the subject of a Program Project application submitted to the NHLBI in 2006.
5) GeneCard is a unique genetic population consisting of ~ 950 sibling pairs and extended families with coronary artery disease in males under age 50 and females under age 56. This NHLBI-funded study has identified 4 loci with LOD scores greater than 3 in multipoint linkage analysis for the following phenotypes: coronary artery disease, metabolic syndrome, acute coronary syndromes, and myocardial infarction. Linkage candidates have been identified and confirmed for two of these regions and fine mapping is underway for the other two.
6) Clinical Genomics Studies Unit (CGSU). The IGSP's CGSU is an enabler of gene- and genomics based clinical studies and trials. In 2007, the IGSP CGSU will initiate ten prospective oncology clinical studies using a microarray based predictor to assign patients to one therapeutic

strategy vs another. CGSU studies will measure clinical and health economic outcomes in each of these investigations.
7) Specialized Programs of Research Excellence (SPORE) in Breast Cancer. The Breast Cancer Research Program of the Duke Comprehensive Cancer Center is one of only 10 programs in the country to have a SPORE grant from the National Cancer Institute (NCI). The NCI established the highly competitive SPORE program in 1992 to foster innovative research and bring new findings from the laboratory to the clinic. The SPORE grant funds four research projects and an ongoing developmental research project within the Duke Breast Cancer Research Program: hypoxia and chemoresistance in breast cancer; T-helper responses to HER2/neu in breast cancer patients; genetic modifiers of BRCA1 and BRCA2; application of pharmacogenomics to treatment of breast cancer; and identification of Epigenetic Events in Early mammary Carcinoma. As part of the SPORE, a tissue specimen biorepository was developed (see below).

II. Anticipated applications of genomic-based clinical testing in medical decision-making, safety assessment, and risk management.

Development of a Personalized Medical Information System. The future of disease treatment and prevention relies on a complex model of therapeutic interactions between patients and providers called "personalized medicine." Detailed knowledge of a person's genes, proteins, morphological and psychological characteristics, and preferences will allow providers to weigh alternatives and arrive at rational decisions that maximize health outcomes. Through the use of novel genomic technologies, we aim to advance study of the relationships between the characteristics of individual patients and clinical outcomes with selected diseases, their interactions with the environment, and the development of targeted diagnostic and therapeutic strategies. For example, pharmacogenetic research methods may specifically be applied in a clinical trial, analysis of hyper- or hypo-responders to the drug, or individuals with unexpected responses may feed discovery activities to identify novel drug mechanisms. Similarly, the results of such clinical investigations will inform the design of population studies of the drug.

Before genomic, biological, and mathematical data can be translated into a usable form that can assist and influence decisions among health care deliverers, an improved understanding of how to best apply scientific knowledge in conjunction with the preferences of patients and the skills of their doctors is needed. To accomplish this, we are developing a Personalized Medical Information System (PMIS) that will allow each patient to interact with his or her physician and other health care providers armed with individual information and predictions. Prototypes of the information system will be populated with clinical data and predictive models presently at the DCRI and IGSP. The architecture of the PMIS will allow rapid integration of additional data as it becomes available. In order to develop these tools in effective ways, we will evaluate preferences and capabilities of patients, providers, and administrators to deal with probabilistic information. In addition, we will evaluate the economic and financial factors affecting the ability to deliver the right therapy to the right patient.

The PMIS will formulate individualized predictions for baseline risk and for effects of key preventive and therapeutic interventions. The system will integrate the predictive models validated in large cohorts with clinical trials and databases managed by the DCRI and its global

collaborators. We have created software that selects an appropriate predictive model for a given problem based on the patient data provided by the user and have implemented it in the cardiology patient care system. This is a primordial "model-management system," in which the user is seamlessly directed to the model best equipped to answer it. We will expand this concept by incorporating existing models and algorithms into an organized, catalogued system (the PMIS) that can intelligently interact with users with varying levels of sophistication. The model-management system makes families of sophisticated statistical models easier to use. It also will create an infrastructure for continued updates and enhancements to the disease-state models as new evidence emerges. By acting as a virtual "reference librarian," directing users to appropriate resources, we can project the power of evidence-based medicine into everyday clinical practice. Additional pilot approaches are being taken in the Kannapolis and Guilford County programs (described later). Eventually, we expect PMIS to be integrated into the AEHR system to provide real-time decision support in practice.

Development of a disease-based model where a specific device to predict and identify the appropriate, personalized weight loss program for individual obese patients. The Sarah W. Steadman Nutrition and Metabolism Center at Duke University, led by Dr. Chris Newgard, has developed a mass spectrometry-based metabolomics laboratory located within this center at Duke University. The laboratory currently conducts targeted GC/MS and MS/MS-based measurements of over 100 metabolites of known chemical identity in 4 classes (free fatty acids, acylcirnitines, organic acids and amino acids). The center also has the ability to perform unbiased profiling of up to 2,200 metabolites per sample via LC/MS. Additionally, the laboratory performs multiplex platform assays for approximately 20 hormones associated with energy balance (e.g., leptin, gherlin, adiponectin) and 20 inflammatory cytokines.

A preliminary pilot study evaluated the metabolomic profiles in obese subjects and subsequently evaluated the efficacy of a variety of weight loss interventions including a rice diet, the Atkins diet, a low calorie diet, a high carbohydrate diet and others. The researchers have identified certain particular pre-weight-loss intervention metabolomic profiles that are predictive of weight loss for each of these interventions with an acceptable degree of accuracy. Given these preliminary findings, the Steadman Center, in collaboration with the protein engineering group from the Departments of Biochemistry and Pharmacology and Cancer Biology, DUMC, the molecular biosensors group in the Fitzpatrick Center for Photonics and Communication Systems in the Pratt School of Engineering and an external partner, has developed the conceptual idea that these particular metabolite profiles could be developed into an office-based testing device, similar to a home blood glucose monitoring device, that could then predict in any individual obese patient the most appropriate and effective weight loss intervention for that patient with that particular profile.

This concept of a translational program that could potentially lead to more effective weight loss programs for obese patients epitomizes the bench to bedside process and the possibility of personalizing therapy for this significant and growing medical problem. Further development of the program and the device will require significant resources to determine applicability both in subsequent early and later stage clinical trials. If successful, evaluation will be needed in the final stage of translation to test efficacy and acceptance in the community setting. Taken further, this process of applying core metabolomic, proteomic and genomic technologies in conjunction

with the engineering of such devices could be applied more widely to transform other disease states, such as the detection of individuals at risk for early onset cardiovascular disease.

Development of a systematic approach to understanding preferences and probabilistic decision making strategies of patients, physicians and administrators. While much of clinical science is based on mechanistic and quantitative understanding, medical decision-making requires an interpretation from the framework of the individual. Improved quantitative prognostic information is useful only if the information is incorporated into decisions made by physicians and patients at a case by case level. However, in practice, physicians and patients often disregard important, powerful evidence, resulting in the failure to provide effective therapies. For example, despite definitive evidence that patients with atherosclerosis should be treated with aspirin and a "statin" at a minimum, doctors and health systems often fail to prescribe them, and patients frequently stop taking the medicines over time. National Medicare data show that only 25% of patients started on a statin remain on the drug six months later.

We know relatively little about 1) how physicians incorporate aggregate clinical statistics with their prior beliefs about a patient's prognosis, or 2) how patients use aggregate clinical statistics to arrive at a judgment about their confidence in their own individual outcome. These issues are critical for understanding how important risk data might shape patient and physician preferences. A medical decision-making expert group will provide input into the development and use of models, including assessment of the cognitive and behavioral implications of alternative approaches to presenting quantitative support for choices to patients and providers.

III. Establishment of biospecimen resources obtained from clinical medical services for application in research, clinical trials, health services planning, clinical effectiveness, and health outcomes evaluations

The Duke-IGSP Biospecimen Repository. The IGSP has established a centralized and accessible biological resource for Duke investigators and research programs. This resource will support *the provision of a standardized, HIPPA and IRB compliant process for storage and processing of biological specimens and ascertainment of clinical data with adjudicated phenotypes.* The Duke-IGSP Biospecimen Repository is a standardized, regulatory-compliant and genetic privacy protected repository for fluid and tissue biospecimens, and allows Duke Investigators to reduce costs, and eliminate redundancies and significant risks associated with past bio-banking practices. The Biospecimen Repository serves a significant role as a "knowledge-based collection facility." While historically repositories have focused solely on the cataloging of samples, our goal is to increae the value of specimens by associating clinical phenotypes at the point of storage. The Duke-IGSP biorepository is responsible for accessing, warehousing, tracking, and dispensing samples from human cohorts for Duke investigators.

Operationally, the Duke-IGSP Biospecimen Repository is divided into two units located at Duke University Medical Center campus and the Research Triangle Institute (RTI).

(1) *The Duke 'Portal' Facility* operates as the on-campus gateway to and from the long-term storage facility (described below), accommodates samples that are expected to be used immediately, and serves as a buffer to meet periods of high demand for sample retrieval.
(2) The Long-term Storage Facility is located at and operated by RTI and serves as a dedicated and scalable physical repository for the receipt, accessioning, long-term storage, and distribution of biospecimens. The RTI facility receives specimens passing through the Duke 'portal' facility (via courier), as well as direct receipt and accessioning of specimens from remote sites in the US and internationally, for long-term storage. RTI is known for its operational excellence and commitment to execution, and has considerable experience with GLP/GMP practices that are directly relevant to specimen collections and software validation required for the Biorepository.

Both the Duke Portal Facility and the RTI Storage Facility have been established in accordance with industry and NIH-established best practices including facility security to limit access to specimens to appropriate biorepository personnel only, power and freezer redundancy, 24/7 freezer monitoring and alarms, appropriate environmental control, SOPs, and trained and experienced staff.

Repository Informatics. The foundation of any repository is the informatics platform. The Biorepository has utilized institutional resources to develop a robust and flexible software platform, Medical Assistant on the World Wide Web (MAW3®), which provides comprehensive biospecimen and data banking services and inventory management capabilities. The web-based MAW3® system, in conjunction with SOPs, provide the program management, protocols, training, inventory management tools and software components necessary to maintain a comprehensive and efficient banking program within the appropriate bioethics framework. All project investigators will have access to sample information via web access to MAW3. MAW3 supports the integrated collection, processing, annotation, storage, and distribution of biospecimens and data through a secure, web-based interface that is deployable at remote sites for data entry and query. Barcode-based chain-of-custody and inventory management functionalities track the precise location and status of every specimen at all times. The MAW3® system is able to define roles-based access privileges for repository staff and users, and to store electronic signatures that meet 21CFR Part11 requirements including a full audit trail of changes made to the database and provision of security measures necessary to ensure data privacy and integrity (HIPAA-compliant). MAW3® has been approved for use in tissue banking by the College of American Pathologists and is funded to serve as an adopter system for caTISSUE Core biorepository software platform standard of the NCI's **ca**ncer **B**iomedical **I**nformatics **G**rid (caBIG™; https://cabig.nci.nih.gov/) project. The infrastructure and tools created by caBIG™ also have broad utility outside the cancer community. MAW3® has been utilized with an excess tissue collection protocol over the past five years to build a tissue repository in support of multiple tissue banking programs, and has collected and annotated in excess of 25,000 diseased and normal tissue samples from over 4,900 consenting events for use by Duke and external researchers. This combination of characteristics is essential to fully support scientific opportunities and advance translational research programs in a diverse and decentralized research environment.

Clinical Data / Annotation. MAW3® has been developed to allow standardized (minimal common data elements), yet flexible (custom data elements) sample annotation, from basic

providing the ability to query across complex clinical data and sample and inventory data, to identify available specimens meeting specific criteria.

Enabling research. The integrated logistical platform supporting the Duke-IGSP Biospecimen Repository facilitates the integration of clinical research programs and registries with specimen processing capabilities and molecular profiling technologies in support genomic medicine and personalized medicine programs at Duke. Investigators have the opportunity to design projects, recruit and consent participants, and collect, process, and store specimens and associated data, with full chain-of-custody tracking. Furthermore, investigators can query across large sets of demographic and clinical data to identify and access the appropriate high quality specimens meeting specific criteria, and direct molecular analyses utilizing a suite of integrated platform molecular analysis technologies, thereby driving translational research programs.

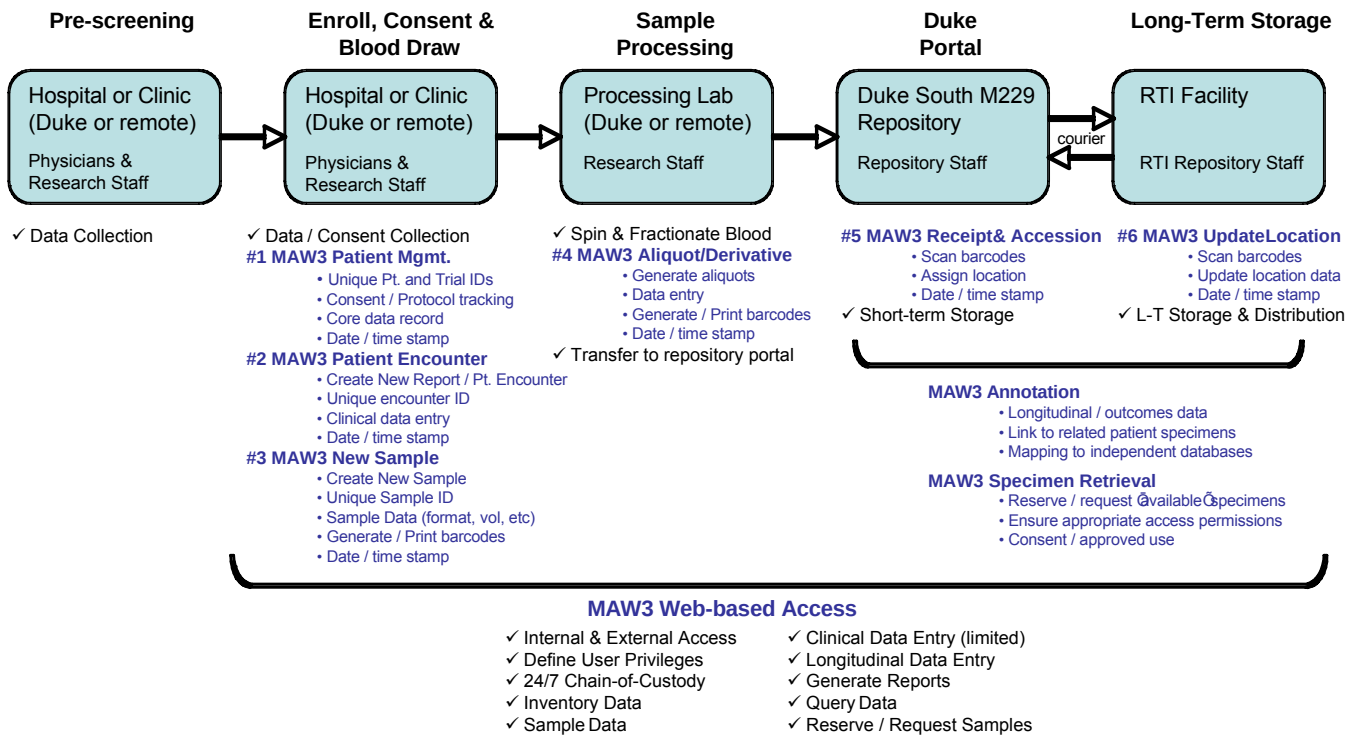


Figure 2. Biospecimen and Data Process Flow. This diagram represents a typical process flow for biological specimen (blood product) procurement, processing, annotation, banking, and request/retrieval. MAW3 data entry steps are indicated in blue.

IV. Organizational or institutional practices to address ethical, legal, and social implications regarding the use of patient information, including genetic data, to support personalized health care.

Research ethics consultation. A number of complex ethical issues can arise in the design and conduct of clinical and translational genomic research. Duke University is one of the few

universities (and perhaps only) with a full-time position dedicated to research ethics consultation. Housed in the IGSP's Center for Genome Ethics, Law & Policy, the Associate Director for Research Ethics Consultation coordinates a multidisciplinary team to help anticipate ethical issues and identify appropriate policy options so that investigators can plan their research and dissemination of results accordingly. The level of assistance is tailored to the need at hand and can take a number of forms, including:

- Serving as a consultative resource on specific issues, to researchers and/or others involved in the research enterprise (*e.g.*, the IRB)
- Participating in an ongoing advisory capacity on the team that is planning and carrying out the research
- Designing, conducting, and/or collaborating on:
  - Empirical studies to examine ethical and policy issues
  - Projects to inform and/or develop policies, guidelines, and tools related to ethical and policy issues

Biorepository informed consent. Dr. Laura Beskow (IGSP) developed an informed consent template for the Duke-IGSP Biorepository based on federal regulations, best practice guidelines for biorepositories, guidance from the Duke IRB, and sample forms from other academic institutions and government agencies. A meeting was held with key Duke investigators in November 2006 and the template revised based on their input. Further efforts are underway to gather input from the Duke IRB and to develop an empirical project to obtain feedback from individuals generally representative of those who may one day be asked to contribute to a specimen.

Improving Communication between Researchers and Community to create an interdisciplinary definition of human disease. Peers, policy-makers, and the general public need to be aware of the published scientific information so that it can be adequately evaluated and used to change health care patterns when appropriate. In order to increase the public visibility of research publications achieved by Duke researchers, we have developed a Web application called Duke Knowledge. Duke Knowledge is a knowledge management system that collects information about the progress of scientific studies and other dissemination tools (web seminars, public presentations, pamphlets, internet sites, etc) with a major focus on publications of impact that have just been accepted in peer-reviewed journals and makes that information available to the central news office for Duke Medicine, called Duke News. Within Duke News, Duke Knowledge matches the topic of scientific articles with the interest of journalists all over the U.S. and the world. These matches are provided to journalists in Duke News as suggestions, empowering them to spread the information through most adequate media routes. All information is released with a documented embargo, therefore ensuring that journalists are

bounded by the regulations of the scientific journal publishing the manuscript. Ultimately, Duke Knowledge provides more exposure of the research findings to the policy-makers and the general public, thus increasing the value of the information to the public.

V. Feasibility and potential benefits for establishing linkages of institutional or organizational data resources with private and publicly available health databases

Data Importation and Integration. While we envision great value for the integration of internally derived data, it is also necessary to address the need for the extension of that integration to external data sets. While there has been some work in the informatics community to attempt to provide linkages of identical data types (e.g., Life Science Identifiers) or ontologies of biological objects (GO consortium), these tools are not presently capable for the widespread integration of biological data. We plan to use technologies developed by firms such as IBM as a building block upon which to integrate data sets of interests. By using tools that already exist, we have the advantage of not having to recreate a costly wheel. Additionally, we can put into place a system by which our tools are not only freely available to the Duke community, but to the scientific community in general.

We aim to develop tools that will enable analysis across data types and integrate with public repositories. The system functions will include:

- standards-based data repository
- study management tools
- tools for uploading and downloading data and metadata
- ontology-enabled query tools
- ontology-management tools
- data navigation tool
- statistical analysis and visualization tools
- data merge and integration tool
- biological pathways tool.

The foundation of the system will be a repository of data containing but not limited to the following data types:

- clinical
- microarray
- protein chip
- pathology
- flow cytometry
- mass spectrometry proteomic
- genotype (including SNP).
- medical imaging

The result of this system will be an enabling technology for investigators to assemble and access the needed data into one functional unit further analysis. Ultimately, these data will form a component of the much larger Common Data Repository (described below).

Development of Investigator-Accessible High Dimensional Datasets. The integration of internal and external data will allow for the creation of an informatics environment which will act as a discovery "toolbox." The pulling together of data along with the appropriate relationships between each data type can be integrated with analytic tools to enable cutting-edge research. These analytical tools (e.g., a microarray visualization and analysis tools) will contain both internally developed as well as commercially and publicly available software. Our approach is to bring together the "best of breed" tools which allow the user community to analyze their data.

Where possible, the system will be built in an open source model, with the ability of other users to "plug in" their own applications or their licensed versions of commercial applications. Thus, if a group wishes to use Spotfire for expression analysis, our system will have an interface into which Spotfire can be attached, but the user of the system will not be required to have a Spotfire license unless that it a local tool they wish to use.

VI. Development of ontologies across different clinical data repositories that will facilitate the utility of the data for answering clinical research questions

Merging Clinical and Biological Information. In order to successfully analyze and interpret relevant but potentially disparate datasets, investigators must extract and integrate data from multiple repositories, often in many different formats. Frequently, these researchers do not have access to satisfactory tools with which to manage and analyze their own data, let alone share and integrate data for meta-analysis. Clinical data and mechanistic (basic biological) data share the trait that both are internally and externally heterogeneous. The same clinical trial may have several variables that refer to the same measurement. This internal heterogeneity is complicated when one attempts to match data across different clinical trials. Mechanistic data have less heterogeneity within the same study, but externally there are many complications in ensuring that a biological object (e.g., a gene) is the exact object referred to in other datasets.

We are in the initial stages of creating a meta-thesaurus across clinical trials to map those data to external vocabularies or ontologies such as SnowMed. The result will be the ability for any researcher to determine unambiguously the data gathered for each variable collected during a clinical trial. This process will also result in the creation of system that would allow for quick meta-study analysis across multiple studies. As an example, a researcher interested in the effects of analgesics on men over the age of 50 would be able to query the meta-data warehouse to determine what information is available on this subset of patients within multiple studies.

As a result of the relatively inexpensive and rapid genotyping, we expect the volume of biospecimens associated with patients to increase substantially. In particular, this will result in the accumulation of numerous genotypes associated with each patient per study. These "patient samples" will be electronically linked to each patient in a study. For ongoing studies, links to the Laboratory Information Management Systems (LIMS) of the repositories will be established which will keep the samples and patients in sync. Ultimately, as data become attached to the samples (e.g., pathology data, genotypes), these will be linked to the patient records. The ultimate goal is to create not only a warehouse of valuable clinical and mechanistic data, but also the technologies for their collection and curation. Where possible, collaborations with IT firms may be used to more quickly develop robust tools; however, any resulting software from these activities will be open source.

VII. Models for linking clinical data repositories across disparate care providers

Common Data Repository. The Duke clinical informatics environment already includes a variety of capabilities, including the Common Data Repository (CDR), a comprehensive DB2-based clinical application, and the Decision Support Repository, an Oracle-based decision support system. The CDR contains a unified inpatient and outpatient longitudinal record spanning over 40 million reports on over 1 million patients. The reports include General Labs (over 120 million coded test results) and Microbiology (over 450,000 coded isolates), as well as semi-structured text reports for Radiology, Pathology, other specialties, and Medical Records. The CDR currently receives approximately 50,000 ambulatory documents per month (including but not limited to clinic notes and results and expanded to include digital electrocardiographic and magnetic resonance images, as well as data feeds from Clinical Systems at Durham Regional Hospital (Siemens) and Duke Health Raleigh Hospital (Meditech). The CDR is accessible via a web front end called eBrowser.

The biomedical computing initiatives at Duke leverage existing computing resources across various areas of Duke, including clinical, translational, and community-based research. Advances in genomics, proteomics, metabolomics, and clinical patient data (such as the CDR) are expanding the complex streams of information on which physicians and biomedical researchers can base decisions. These resources will be integrated using grid computing technologies in two ways: sharing of computing capacity and linking file systems and data. The resulting integration facility will be accessible through open web services interfaces so that clinical, translational, and community-based data. Ultimately, all Duke researchers will be able to access the integrated datasets as well as aggregate those datasets with others for local analysis.

Decision Support Repository - A Data Repository Infrastructure. The Decision Support Repository (DSR) is an integrated enterprise data warehouse. Architecturally, the DSR is a custom-built relational database using dimensional modeling and batch extraction, transformation, and loading routines that extracts both inpatient and outpatient data from multiple institutional source systems ranging from clinical to financial. Information in DSR is arranged by subject area: demographic, labs, vital, charges, budgets, diagnosis, procedures, staffing, etc. The scope of the data domain is comprised of five overlapping areas. DSR is currently used to generate and support over 3,000 operational and decision support reports and numerous research studies. The underlying database management system is Oracle and supports any ODBC-compliant reporting tools and is HIPAA-compliant.

DSR follows the traditional data warehouse development methodology and utilizes dimensional data model into star schemas. Corporate data, including clinical and research, are physically integrated into the data warehouse on nightly or weekly basis from the operational systems. Data are scrubbed, cleansed and transformed according to the specific business rules prior to loading into the data warehouse to ensure data consistency and accuracy and allow patients to be matched across disparate transactional systems. A newly generated surrogate key is used to identify patients and is also used to support de-identified queries. Master patient identification mapping table can be maintained in the data warehouse that allows the research databases, such as microarray data, to remain de-identified. End user queries can be performed either against the enterprise data warehouse directly or via custom-built data marts designed for a specific department or research study.

Data security and integrity at the DSR are maintained at several levels:

a) Role-based Access: User accounts are categorized into pre-defined groups with matching database profiles. These security profiles define the subject area/data marts that any given user is authorized to access.
b) De-identified Data: For IRB-approved research projects, DSR data can be presented to the researchers as a one-time data feed or via direct access through pre-defined data marts built according to the study protocol. For data mining on possible studies prior to IRB submission, researchers can query the entire data warehouse of clinical information via de-identified filter. For studies that meet the IRB exempt protocol, DSR supports the HIPAA "limited data set" requirements and can provide custom-built research database for a specific study.

VIII. Examples of the use of disease registries to track specific diseases and response to drug therapies across different subpopulations

Establishment of registry of healthy individuals with known variants in relevant drug metabolizing enzymes (DMEs) and drug targets for use in Phase I studies. Pre-clinical drug metabolism and pharmacokinetic (DMPK) studies usually establish the metabolic fate and enzymatic degradation and transformation process of drugs well in advance of first-in-human studies. As part of the *post-hoc* analysis, genetic variation should be studied to determine if any association exists with pharmacokinetic outliers. However, this strategy has one potentially fatal flaw – most phase I studies involve a relatively small number of subjects (single and multiple ascending doses often enroll >50 individuals and in some cases, >20 subjects are deemed sufficient for safety testing). Thus, for DME genetic variants of <5% frequency in the population, the likelihood that any one individual will have the mutation and is included in the study cohort for a Phase I study is extremely low. However, if the over-/under-metabolism of a drug leads to an adverse event, when larger Phase II or III studies are conducted, the identification of these individuals may be critically important for the overall success of the drug to reach the market without unexpected toxicity. Moreover, if investigators were to insist that subjects with relevant variants be proactively recruited into an early Phase I study (a strategy rarely if every contemplated), a prospective screen of hundreds of individuals would need to be undertaken to find both individuals with the variant(s) and who are also willing to participate in an investigational study. Genetic variation targeting of a drug will also be of importance to assess the genetic influence on pharmacodynamics (PD).

We will establish a local group of genotyped "healthy volunteers" who have known DME variants and who would be willing to participate in early phase I or II clinical investigations. Moreover, a mechanism will be established to 'replenish' this cohort on a yearly basis to assure a minimal pool of individuals would be available for study. To accomplish this goal, we will take advantage of the National Institute of Environmental Health Sciences "Environmental Polymorphisms Registry (EPR)" established at both Duke University and University of North Carolina, Chapel Hill. Established in 2004, one of the major goals of the registry is to collect DNA from 20,000 individuals in the Durham-Chapel Hill area. This registry will serve as a valuable resource for studies seeking to understand how genetic variation might be associated with human responses to environmental exposures such as diet and drugs. The registry will be

maintained for 25 years and will provide a mechanism for re-contacting subjects with DNA variations of interest to participate in research protocols. To date, 3,000 EPR DNA samples have been banked and an assessment of DME variants in this population is underway.

We will establish a similar registry in the Durham community to evaluate specifically Drug-Gene Interactions (DGI). The DGI registry will complement the EPR and would specifically consent individuals for DNA banking and analysis of genetic variants relevant to drug metabolism or drug targets. Individuals consenting to this study would understand that they might be approached to volunteer for phase I studies under a new informed consent. Several benefits can be gained from study of individuals collected in these registries: 1) defining the frequency of relevant DME and PD variants in the general population and across ethnic groups, 2) enabling immediate ascertainment of subjects for study (vs. prospective screening), 3) allowing for appropriate strategic design of early stage dosing studies to establish definitive genetic and PK relationships, and finally 4) informing design of phase II and phase III studies that validate the results in phase I and potentially exclude certain genetic populations from later stage studies due to abnormal metabolism of the drug(s)and increased risk of adverse reactions. Taken together, this strategy should have a significant impact on the successful design of, recruitment for, and execution of clinical drug development studies.

IX. Strategies for accumulating patient data necessary for research that may not be available through EHRs

Data Standards and Data Collection – Integrating data clusters to redefine human disease through the use of combined clinical and molecular profiling. In a recent administrative survey conducted among research members of the Duke Medicine community, the lack of dynamic Web-based applications for data collection has been identified as one of the primary sources of concern. As a result, several Web applications have been developed to meet the needs of researchers.

The first application, called DADOS-prospective, is an open-source, HIPAA-compliant Web-based data collection tool. DADOS-prospective allows for centralized creation of research forms that can then be made available to multiple clinics and research through the Web. Centralizing data collection also facilitates the transmission of regulatory documentation (IRB protocol and consent forms), allowing for a regulatory compliance among all participating sites. Among its many features, DADOS-prospective provides researchers with the ability to track any changes to the data collection form once the initial data has been entered. These changes are all displayed in a single audit trail table. Audit trails tables can then be compared to the data collection forms themselves, data in a spreadsheet format, and original source documentation converted to electronic files in PDF. The combination of multiple sources of data into a single location thus provides researchers with the ability to oversee quality for data collection even for remote sites in multicenter research.

Another feature of DADOS-prospective is its ability to use previously existing data collection forms as a template for the creation of new forms. This feature allows for time saving in that researchers frequently use pre-existing forms as their new studies, thus expediting the initial

phases of the study. Finally, all raw data associated with each study can be immediately retrieved by the primary investigator and designated research coordinators at any point in time. This feature allows for real-time control of data quality as well as oversight of accrual levels at the individual site level. By providing a central tool for the collection of clinical, translational and community-based research data, a new opportunity is created to use common standards for data collections throughout the multiple clinics and laboratories within CTSA. Combined with our experience with national data collection standards such as CDISC and HL7, this common tool and data standardization will empower investigators to redefine human disease through the use of combined clinical and molecular profiling.

As two companion tools to the DADOS-prospective tool, we have also compiled an outcomes scale repository and a tool that allows for determination of the total number of patients with a specific combination of diagnostic and procedure codes in a service. The outcomes scale repository currently contains references for over 250 different self-reported outcome scales, some of them containing links to the scales in full-text with copyright authorization by authors and journals. By compiling the literature on self-reported evaluations on a single location, this repository allows for easy retrieval of the most appropriate measures to evaluate patient outcomes. The second tool allows for retrieval in real-time of the total number of patients with a certain combination of diagnostic and procedure codes. For example, if a researcher wants to know the total number of patients with lung cancer undergoing a lobectomy, the appropriate diagnostic and procedure codes can be retrieved in less than a minute and the total number of patients with that combination. An additional spreadsheet can then be downloaded demonstrating how the procedure codes have been distributed among the different diagnosis codes. The system, which currently is available on a limited basis, does not contain patient identifiers in compliance with HIPAA regulations. In combination with the scales repository, this tool allows for better design of prospective studies, increasing the precision of researchers' estimates of the total number of patients who may be eligible for a given clinical or translational study depending on variations in the inclusion and exclusion criteria.

With the rapid advances in “-omics” technologies, biospecimen banks need to provide stringent and integrated logistics for the collection, tracking, and inventory management of large numbers of biological specimens, track derivative materials (RNA, DNA, protein) that are used for molecular analysis, and annotate specimens with accumulating clinical and experimental data so that they can be used for epidemiological and correlative studies. Given the importance and complexity of the proposed human biological specimen collections associated with translational programs, it is imperative to develop the appropriate processes and infrastructure to handle these specimens in an accurate and secure manner.

X. Concepts or models on the potential use of clinical data and related resources for research applications

Duke Clinical Research Unit (DCRU). Equipped with state of the art electronic data capture systems, the DCRU will allow for integration of laboratory and clinical data, thus ensuring a timely retrieval of variables for clinical and translational studies. The Core Laboratory will manage samples from all parts of the campus and integrate this information with output from Kannapolis and Singapore to produce standardized and regulatory compliant electronic data.

The DCRU laboratory also provides state of the art facilities for sample storage equipped with an informatics system that facilitates sample retrieval and linkage with associated research and clinical data. To streamline processes and decrease the number of inaccuracies, research participant charts, identification wrist bands, and laboratory samples will be bar-coded. Bar coding systems also allow for precise time tracking of each of the events related to a research encounter, thus ensuring that our research processes are constantly evaluated for process optimization. Electronic systems will red flag patients who might be eligible for participation in research studies based on their diagnoses and comorbidities, thus embracing the vision that all patients receiving care at the Duke University Health System should be offered the opportunity to be in a research study. Currently employing a web-based electronic data capture system, data will be collected through tablet computers using wireless connections, thus ensuring that investigators can track the whole medical encounter while making appropriate research notes. All systems are compliant with HL7/CDISC, and caBIG data standards. Finally, an indirect interface with Duke University's SAP accounting system, linked by the site management system developed in the NIH CTN Best Practices Roadmap contract will allow the merging of research and cost data, thus creating opportunities for associated cost-effectiveness studies.

Cancer Biomedical Informatics Grid (CaBIG) Activities at Duke. The Duke Comprehensive Cancer Center is a funded adopter of a clinical database management system known as the Cancer Centralized Clinical Database (C3D), with local implementation at Duke. C3D is the Clinical Trials Database and a key component of NCI's C3DS suite. Oracle Clinical serves as the foundation of the C3D by supporting clinical trial definition, data capture, multiple site reporting, data definition and usage standardization. Trial definition is based on curated CDEs from an existing library of adaptable template Case Report Forms. Oracle Clinical's Remote Data Capture (RDC) provides a user-friendly interface that allows local and remote data entry and electronically confirms source data verification (electronic signature). This 21 CFR Part 11-compliant system provides an optimal environment for sharing clinical data among cancer research units. Given that the DCRI is expert in the Clintrial System, this gives Duke researchers access to the two other Part 11 compliant systems in current use worldwide.