Executive Summary of the NIST-NCI Biospecimen Quality Assessment & Standards Development Workshop

October 21-22, 2010

Office of Biorepositories and Biospecimen Research
National Cancer Institute
National Institutes of Health
U.S. Department of Health and Human Services

Background

The US Congress has directed the National Institutes of Health (NIH), through recent health care reform legislation, to formulate a more aggressive strategy towards the translation of scientific discovery directly into the development of novel therapeutic and diagnostic capabilities. The National Cancer Institute (NCI) seeks to facilitate these efforts by targeting bottlenecks along the relevant development pipelines as it relates to treatment of cancer. Many of the bottlenecks that impede the pipelines for drug, biologic and device development are known, and include a lack of understanding and standardization surrounding the research biospecimens required to advance this research.

The National Cancer Institute Office of Biorepositories and Biospecimen Research (NCI/OBBR) hosted a workshop in coordination with the National Institute of Standards and Technology Materials Measurement Laboratory (NIST/MML) and the Food and Drug Administration (FDA) Office of Personalized Medicine to develop fit-for-purpose quality assessment tools for the human biospecimen samples used by the research community. Broad consensus by the scientific community has been repeatedly demonstrated that suggests the lack of sufficient quality human biospecimens is a critical limiting factor for translational research. The first step of this collaboration was to convene a workshop to gather recommendations from leaders in the scientific community.

The early vision for this collaboration would involve specimen procurement, some financial support, and scientific guidance from NCI, regulatory guidance from FDA, and the bulk of the investigation performed by NIST (with any available assistance that could be provided by hiring temporary research staff using funds from the NCI Cancer Human Biobank, caHUB, appropriation). Four specific laboratories were available to explore projects for the initial pilot period to pursue biospecimen quality assessment metrics for research biospecimens. The four participating labs can roughly attributed as those investigating DNA quality, RNA quality, serum-level protein quality, and formalin-fixed tissue quality.

It was decided that the following perspectives were required to gather the best recommendations for going forward:

- biomedical and biospecimen scientists
- both investigative and practicing pathologists and oncologists
- standards/metrology scientists
- commercial assay developers
- experimental design and research methods experts
- statisticians
- regulatory scientists

Specific individuals invited were those thought to be leaders in their respective fields based on personal recommendations and their publication records. Participants were for the most part restricted to domestic US individuals as it was thought that incorporating international regulatory issues would be beyond the scope of this pilot initiative.

Summary of workshop products and recommendations

Workshop participants developed a process roadmap with the following series of projects for NIST that would lead to the development of timely, fit-for-purpose, quality assessment metrics in the four focal areas:

- **RNA molecular quality assessment** – Deep sequencing of identified targets with known stressor variations in known tissues to derive a scorecard for RNA integrity
- **DNA molecular quality assessment** – Quantifier materials to address inter-lab variation and spike-in materials to address specimen mix-up and degradation
- **Formalin-fixed tissue quality assessment** – quantitative assessment of general phosphorylation, variability inherent to tissue staining, and optimal collections for specific biomarkers
- **Serum-level protein molecular quality assessment –** employ existing technologies for *in situ* storage conditions monitoring of storage tubes and a standardized multiplexed technology approach for serum quality assessment.

These critical assessment materials and methodologies were identified as being both important for the research community and achievable in a reasonable timeframe. The full report of workshop deliberations and presentations is available at the NCI/OBBR website (http://biospecimens.cancer.gov/resources/publications/workshop/psr.asp). The remainder of this summary provides a high-level description of the projects recommended by the working groups.

RNA Molecular Quality

The transcriptome of a tissue sample is a readily measured proxy for biological activity; gene expression profiling is part of routine assessment of cancer pathology samples. RNA degradation is a common challenge when measuring an expression profile from preserved tissue samples, as RNASEs are ubiquitous and active. RNA quality is commonly assessed with a proprietary heuristic measure, the "RNA Integrity Number," or RIN value. RIN is measured for total RNA, and is based on electrophoretic measurement of ribosomal RNA, degradation products of the ribosomal RNA, and the size population profile of mRNA. There is a wide range of experience with RIN, but little authoritative validation of it as a measure of the "soundness" or quality of a transcriptome. Development of a quantitative, objective measure of RNA integrity as an alternative to RIN is appropriate for high value samples, such as those to be deposited in a specimen bank, and might also serve to independently validate the use of RIN. Work to develop such a measure is consistent with programs underway in metrology for transcriptome measurements, and would leverage the availability of standards and techniques already in place.

A research path to develop a new RNA integrity measure was outlined in the breakout sessions addressing RNA integrity. The proposed measure would assess the integrity of the transcriptome with respect to chemical and biophysical degradation; effects on the transcriptome from perturbation (arising from patient treatment) would not be readily addressed in the proposed work.

The proposed work is to develop a multiple-gene "signature of intactness." Selecting a subset of genes is predicated on the hypothesis that a limited number of RNA degradation mechanisms

exist (on the order of 6-10), and that measuring the degree of degradation of genes that represent those mechanisms would permit a composite "intactness" score to be economically measured.

The research to test this hypothesis and develop the composite intactness assay would assess a variety of degradation mechanisms on the transcriptomes of a variety of tissue samples. The experiment design would include sufficient replication to permit a robust signature to be created. Various "stressors" of the transcriptome would be evaluated, including: storage temperature, time before stabilization, freeze-thaw cycles, age of sample, enzymatic degradation (RNases), chemical degradation, physical degradation (i.e. ultrasonic). Deep Ultra High Throughput sequencing of the RNA would be employed to directly measure the intactness of the RNA – both the messages, and the small RNA fractions.

The RNA intactness would be classified, initially using principal components analysis, and groups (i.e. genes subject to the same degradation pathways) would be identified. Representative genes from these groups would be selected to cover the dynamic range of relative abundance, and ultimately multiple qRT-PCR assays would be developed along the lengths of the RNA molecules, to provide a measure of the degree of intactness. These assay results would be aggregated into a composite intactness score that should establish the utility of the sample for further RNA analysis.

DNA Molecular Quality

The DNA quality integrity working group identified three primary projects, enlisted below. They are prioritized by increasing difficulty in achieving and the requirement for collaboration with investigators beyond the NIST. These were all identified as being of immediate and significant importance for the research community.

A quantitation standard reference material. This was identified as being a reasonably straightforward as it would involve the reissuance of the existing standard reference material (SRM) 2372. SRM 2372 is a Human DNA Quantitation Standard, which was developed for forensics laboratory applications. Coupled with robust fitness-for-purpose guidelines for using NIST products, it was felt that this would be achievable in a very short timeframe and would be of significant utility to the research community, especially for those engaged in team science projects.

A DNA fragmentation and spike-in material. It was felt that a longer-term project would be to modify and expand the Identifiler STR product from NIST to become a general tool for DNA quality assessment. This product could further inform the development a spike-in material that could be used as both an *in situ s*pecimen identifier as well as reporters for different kinds of degradation. It was thought that findings from the External RNA Control Consortium (ERCC) and materials toward that effort produced by NIST could potentially be leveraged for this end.

Robust DNA purification guidelines. It was thought that this project would require collaboration with other larger efforts across the research enterprise to understand which purification methods were most commonly used, especially those methods that involved the extraction and purification of multiple molecular types. This effort would involve assessing the impact of collection and storage variables.

Formalin-Fixed Tissue Quality

There currently exists no widely used metric for assessing the quality of FFPE tissue. One of the challenges in developing such a metric is the wide range of experimental techniques that are applied to FFPE material. Consequently, the FFPE Working Group reached a consensus that no universal standard or reference can be developed to address quality for all purposes. However, there was enthusiasm for having a metric to rate FFPE tissue that could then be applied to "fit-for-purpose" use. Using this metric, individual experimenters would determine the value that will predict the success of a particular assay using FFPE tissue. Within this framework, three possible strategies for the development of standards or reference materials by NIST were discussed, with general but not unanimous agreement that further pursuit of these strategies is worthwhile.

Analysis of marker expression. One proposal to create a reference standard involves the use of one or several markers that are commonly expressed in many types of cancers. HER2 was discussed as a representative example. HER2 levels will be analyzed after varying time to fixation and other collection parameters, and compared to the expression of a marker that is stable to degradation. The identity of this second marker will need to be determined. It was also agreed that two independent assays of HER2 expression will be required. The ratio of HER2 expression to the invariant marker will be a measure of tissue quality. This methodology requires the availability of specific antibodies and a method for quantitative analysis of immunohistochemical signal in FFPE material, all of which should currently exist. Additionally, a reference material will need to be developed (such as purified recombinant HER2 protein) that could be used as a standard in this assay.

Analysis of phosphoprotein expression. This proposal is similar to that involving analysis of markers such as HER2, but instead relies on the idea that the expression of some phosphoproteins will be dynamic, a key feature of any analyte used as an indicator of quality.

Analysis of general tissue quality using fluorescently-conjugated reagents. A more general approach to assessing FFPE tissue quality involves the use of chemical probes to visualize features of cellular integrity and molecular organization. Examples of features of interest include protein and organelle distribution, levels of reactive chemical groups, and lipid concentrations. To illustrate this concept, the following scenario considers the measurement of the ratio of free amines to thiols using fluorescently conjugated molecules with appropriate reactivities. In this scenario, several potential factors that are likely to affect tissue quality, include degree of formalin fixation, protein degradation, and protein denaturation, are expected to significantly alter the number of free amines in the sample while leaving the number of free thiols relatively constant. Quantitative measures of free amines normalized to free thiols will be compared in FFPE tissue collected after varying fixation time and other tissue collection parameters. Such comparisons will also be made after purposeful degradation of tissue quality using heat, protease, or nuclease treatment. The advantages of this approach include available reagents, and the relative ease of assay performance and quantitation. These features are likely to result in reduced variability when relative to immunohistochemical methods. However, it is unknown if amines will vary in a predictable manner or thiol levels will remain constant as tissue quality is altered.

For any of these quality assessment assays to provide meaningful information, they must be able to guide the selection of tissue for a given assay or measurement; ideally by providing a numerical cutoff above which that assay can be reliable used. Therefore, tissue of varying quality, as determined by any of the assays that are developed, will be interrogated for expression of biomarkers using a range of common techniques (FISH, immunohistochemistry, in situ hybridization, etc.).

Note: The NCI Biospecimen Research Network (BRN) program is currently funding Dr. David Rimm (Yale) to investigate analysis of marker and phosphoprotein expression as a measure of quality. NIST can support this effort by developing a methodology or reference material to allow standardization and normalization of this assay in different laboratories. Specific project plans will be developed in coordination with Dr. Rimm (Yale) and Dr. Helen Moore (Program Director of BRN).

Serum-Level Protein Molecular Quality

Serum is an important biological sample in clinical medicine because it reflects the overall proteomic profile of an individual and is readily accessible. It is also generally accepted that new diagnostic (or prognostic) tests will be based on detection of disease markers in blood or other body fluids. Serum makes up approximately 55% of blood by volume. To isolate serum from raw plasma, fibrinogen and other clotting factors are removed. Serum is composed of approximately 92% water by weight, 7% protein and 1% other solutes. Moreover, approximately 95% of the protein content is represented by fewer than 25 proteins, which would seem to indicate that the protein component is relatively simple. However, the dynamic range of serum has been estimated to span some 12 orders of magnitude, containing hundreds to thousands of proteins, and indicating that the remaining 5% of the protein fraction contributes a great deal of complexity. It is within this latter fraction that most discovery assays are focused.

Mass spectrometry is the method of choice for sensitive identification of protein analytes in complex biological samples, including serum. While these instruments provide sensitive identification and quantitative capabilities, they are not immune to analytical and pre-analytical variability such as that caused by sample degradation.

NIST proposes a two-phase study aimed at identifying a panel of protein degradation markers, and qualifying those markers given a biologically diverse sample set. While our initial investigations will focus on identifying peptide markers, metabolites that respond to treatment conditions will also be sought. The goals of this proposal are to *(1) identify new markers in serum that reproducibly correlate with protein degradation, and (2) determine the sensitivity and specificity of these markers, and existing markers (e.g., bradykinin), in a population of samples where biological variability is a significant factor.*

Phase 1 - Discovery and measurement reproducibility. NIST proposes a series of discovery experiments aimed at identifying a panel of protein degradation markers that can be reproducibly measured. Part of this phase will include an iterative process, during which methods and metrics are refined. In Phase 1, a small number of serum and/or plasma samples (3-5) will be subjected to conditions known to affect sample integrity (see below). All samples will be procured from the caHUB repository which are collected under highly-controlled conditions. The experimental assumptions are as follows:

1. Samples obtained from caHUB are 'best case scenario' and therefore represent degradation point zero with respect to pre-storage protein degradation.
2. If the identification of degradation markers and patterns of decay are not reproducible in a small number of samples, they will not be reproducible in a large number of samples. We therefore limit screening sample size to <5.

Common sample storage and handling (e.g., increased temperatures, freeze/thaw cycles, etc.) as well as processing conditions (e.g., no digestion, partial digestion, mass filtration, etc.) will be varied on specimens collected from selected individuals (3-5 individuals, as indicated above) to generate a matrix of conditions. Once samples have been received at NIST, they will be processed according to the experimental design matrix. All discovery-based measurements will be made using LTQ or Orbitrap mass spectrometers following short LC separations. Quantitative validation, including that of bradykinin (see below), will be performed using multiple reaction monitoring methods on triple quadrupole mass spectrometers. It is expected that Phase 1 will generate many hundreds of data files.

Data analysis will include the following:

1. QC analysis according to methods published by the participating NIST divisions for proteomics samples
2. Ion current-based quantitative analysis for differential testing (discovery)
3. Development of new metrics for monitoring protein degradation
4. Statistical ranking of storage, handling and processing variables

Each sample will be run with enough technical replication to ensure adequate statistical power.

Phase 2 - Validation in the face of biological variability. While the logical progression of this proposal lends itself to using only the degradation markers identified in Phase I in validation studies, we propose adding at least three additional markers to these follow-up studies: (1) human serum albumin (HSA), (2) bradykinin, and (3) prostate specific antigen (PSA). For the past two years, NIST has been performing extensive experimentation on the digestion properties of HSA and, through these studies, has amassed a large knowledgebase of experimental analysis of this protein that indicate that some HSA peptides are differentially susceptible to analytical sample treatments. We propose targeting some of these HSA peptides in validation studies. Additionally, Becton Dickinson (BD) (http://www.bd.com/proteomics/pdfs/ASMS06-Yi_11x17_prnt.pdf) has presented data indicating that bradykinin is specifically affected by a host protease cascade, and that it might therefore serve as a sensitive marker for protein degradation. PSA was also proposed by the NIST advisory working group as a marker to consider because it exhibits differential stability based on state (free, bound or truncated.) Together, with markers discovered in Phase 1, the following experimental approach will be taken:

1. Procure ~50 serum samples from caHUB covering a broad demographical distribution
2. Design MRM assays to measure peptide or metabolite targets
3. Apply degradation conditions from Phase I that cause significant changes
4. Apply sample processing conditions that maximize sample integrity from Phase I

5. Perform multiplexed MRM assays
6. Analyze data and calculate the value of each degradation marker across the sample cohort

At the end of Phases I and II, the following expectations should be met:

1. A good understanding of the storage, handling and processing variables affecting protein degradation will be established
2. A good understanding of the difficulties in discovering protein degradation markers (or lack thereof) in serum will be established
3. A good understanding of the relative impact of biological variability on protein stability testing in serum will be established
4. If Phase I successfully produces a qualified panel of degradation markers, next steps will include examining ways to engineer a high-throughput assay for protein degradation in serum