Cross-Disciplinary Biomarkers Research: Lessons Learned by the CKD Biomarkers Consortium

Chi-yuan Hsu,* Shawn Ballard,† Daniel Batlle,‡ Joseph V. Bonventre,§ Erwin P. Böttinger,‖ Harold I. Feldman,†
Jon B. Klein,¶ Josef Coresh,** John H. Eckfeldt,†† Lesley A. Inker,‡‡ Paul L. Kimmel,§§ John W. Kusek,§§
Kathleen D. Liu,* Michael Mauer,†† Theodore E. Mifflin,† Mark E. Molitch,‡ Gary L. Nelsestuen,††
Casey M. Rebholz,** Brad H. Rovin,‖‖ Venkata S. Sabbisetti,§ Jennifer E. Van Eyk,**¶¶ Ramachandran S.
Vasan,*** Sushrut S. Waikar,§ Krista M. Whitehead,† and Robert G. Nelson††† for the CKD Biomarkers Consortium

Abstract

**Significant advances are needed to improve the diagnosis, prognosis, and management of persons with CKD.
Discovery of new biomarkers and improvements in currently available biomarkers for CKD hold great promise to
achieve these necessary advances. Interest in identification and evaluation of biomarkers for CKD has increased
substantially over the past decade. In 2009, the National Institute of Diabetes and Digestive and Kidney Diseases
established the CKD Biomarkers Consortium (http://www.ckdbiomarkersconsortium.org/), a multidisciplinary,
collaborative study group located at over a dozen academic medical centers. The main objective of the consortium
was to evaluate new biomarkers for purposes related to CKD in established prospective cohorts, including those
enriched for CKD. During the first 5 years of the consortium, many insights into collaborative biomarker research
were gained that may be useful to other investigators involved in biomarkers research. These lessons learned are
outlined in this Special Feature and include a wide range of issues related to biospecimen collection, storage,
and retrieval, and the internal and external quality assessment of laboratories that performed the assays. The
authors propose that investigations involving biomarker discovery and validation are greatly enhanced by
establishing and following explicit quality control metrics, including the use of blind replicate and proficiency
samples, by carefully considering the conditions under which specimens are collected, handled, and stored, and
by conducting pilot and feasibility studies when there are concerns about the condition of the specimens or the
accuracy or reproducibility of the assays.**

Clin J Am Soc Nephrol 10: 894–902, 2015. doi: 10.2215/CJN.11541114

*University of
California-San
Francisco, San
Francisco, California;
†University of
Pennsylvania Perelman
School of Medicine,
Philadelphia,
Pennsylvania;
‡Northwestern
University Feinberg
School of Medicine,
Chicago, Illinois;
§Brigham and Women's
Hospital, Boston,
Massachusetts; ‖Mount
Sinai Hospital, New
York, New York;
¶University of Louisville,
Louisville, Kentucky;
**Johns Hopkins
University, Baltimore,
Maryland; ††University of
Minnesota, Minneapolis,
Minnesota; ‡‡Tufts
Medical Center, Boston,
Massachusetts;
§§National Institutes of
Health National Institute
of Diabetes and
Digestive and Kidney
Diseases, Bethesda,
Maryland; ‖‖Ohio State
University, Columbus,
Ohio; ¶¶Cedars-Sinai
Medical Center, Los
Angeles, California;
***Boston University,
Boston, Massachusetts;
and †††National Institutes
of Health National
Institute of Diabetes and
Digestive and Kidney
Diseases, Phoenix,
Arizona

Correspondence:
Dr. Robert G. Nelson,
National Institutes of
Health National Institute
of Diabetes and
Digestive and Kidney
Diseases, 1550 E. Indian
School Road, Phoenix,
AZ 85014-4972. Email:
rnelson@nih.gov

Introduction

CKD imposes a substantial medical and public health
burden in the United States (1). Advances in screen-
ing, diagnosis, prognosis, and management of CKD
are needed to reduce the negative effect of this dis-
ease. Discovery of new biomarkers and improve-
ments in currently available biomarkers for CKD
hold great promise to achieve these necessary advances.
A widely accepted definition of a biomarker as "a
characteristic that is objectively measured and evalu-
ated as an indicator of normal biological processes,
pathogenic processes, or pharmacologic responses
to a therapeutic intervention" was put forth by the
National Institutes of Health Biomarkers Definitions
Working Group in 1998 (2).

In 2009, the National Institute of Diabetes and
Digestive and Kidney Diseases (NIDDK) announced an
open competition (http://grants.nih.gov/grants/guide/
rfa-files/RFA-DK-08-015.html) to support a multidisci-
plinary group of investigators to discover or validate
biomarkers for well defined CKD. The CKD Biomarkers
Consortium (http://www.ckdbiomarkersconsortium.
org/) benefited substantially from the synergy created
by bringing together experts in nephrology, clinical
chemistry, laboratory medicine, pathology, proteomics,

biostatistics, epidemiology, and informatics to work in
close collaboration. Because the consents from cohorts
contributing biosamples to the consortium permitted
use of coded samples and clinical data for this purpose,
the investigators established a global materials transfer
agreement across all institutions and cohorts and
launched a series of studies, involving >58,000 sam-
ples, to advance our understanding of the role of blood
and urine biomarkers in CKD. The scope and focus of
individual studies varied considerably and included
efforts to improve risk stratification for incident or pro-
gressive CKD using markers of inflammation, filtration,
fibrosis, and abnormal mineral metabolism. State-of-the
art proteomic technologies, including ELISA and mass
spectroscopy, were used to discover new biomarkers,
with plans for subsequent validation of promising tar-
gets using ELISA, multiple reaction monitoring (MRM),
and other approaches.

Given the multicenter nature of the consortium,
with numerous locations of stored biospecimens and
multiple laboratories performing assays (Table 1),
much initial effort was devoted to implementing pro-
cesses to monitor and ensure quality control (QC),
including maintenance of sample identity. During
the first 5 years of the CKD Biomarkers Consortium,

Table 1.   Studies providing samples and laboratories performing testing for the CKD Biomarkers Consortium

Studies Providing Samples	Laboratories Performing Testing
African-American Study of Kidney Disease and Hypertension (AASK)	Advanced Research and Diagnostic Laboratory, University of Minnesota
Atherosclerosis Risk in Communities Study (ARIC)	The Böttinger Laboratory at Mount Sinai
Cincinnati Children's Lupus Cohort	Brigham and Women's Hospital Kidney Biomarker Laboratory
Chronic Renal Insufficiency Cohort Study (CRIC)	Center for Mass Spectrometry and Proteomics, University of Minnesota
Joslin clinic cohort	Johns Hopkins University Bayview Proteomics Center and Johns Hopkins University NHLBI Proteomic Innovation Center on Heart Failure
Modification of Diet in Renal Disease Study (MDRD)	
Mount Sinai clinical care cohort	Ohio State University Glomerular Diseases Laboratory
National Health and Nutrition Examination Survey (NHANES)	Translational Core Laboratory, University of Pennsylvania Perelman School of Medicine
Natural History Study (NHS)	University College Dublin Core Laboratory
NIDDK American Indian Study	University of Louisville Clinical Proteomics Laboratory
Ohio SLE Study	
Renin Angiotensin System Study (RASS)	

NHLBI, National Heart, Lung, and Blood Institute; NIDDK, National Institute of Diabetes and Digestive and Kidney Diseases.

a number of lessons were learned that may be of interest to other investigators. These lessons learned span the domains of preanalytic processes, laboratory analytic procedures, and postanalytic data processing practices. They illustrate the value of utilizing rigorous informatics tools to limit sample identity errors; the critical influence of conditions under which samples are collected, processed, and stored; and the value of robust laboratory QC procedures.

QC Procedures Established by the Consortium

At the inception of the CKD Biomarkers Consortium, a QC committee was established to provide specification and oversight of minimum performance standards to be followed by participating investigators, cohorts, and laboratories within the consortium. Biospecimens used in the consortium came from 12 parent studies and were aliquoted, labeled, and shipped to nine laboratories for analyte measurements (Table 1). To facilitate the management of biospecimens and reduce the risk of identity errors, the consortium adopted a standard operating procedure (SOP) for sample labeling and processing that included the use of unique printed barcodes on all labels. The goal of this SOP was to ensure that samples were prepared and labeled accurately at their sites of origin, and were tracked centrally so that the assay results could be incorporated efficiently and accurately into a central database. A summary of these processes is shown in Table 2. We also developed SOPs for each analyte measured by the consortium (http://www.ckdbiomarkersconsortium.org/laboratory_SOPs.pdf) that provide quantitative data on assay range and reproducibility.

Issues Associated with Sample Identity/Informatics
Errors in Sample Selection

We learned that errors can occur in selecting the appropriate type of stored blood samples for assay. In one set of consortium biosamples, six boxes of serum

samples were shipped to the testing laboratory for measurement of biomarkers related to mineral metabolism. In one box, the samples were stored in snap-cap Eppendorf tubes, whereas in the other boxes, the samples were stored in screw-capped cryovials with O-rings. Because the testing laboratory was blinded to the specimens' identity and prior history, technicians did not question this difference. Several months after completing the initial measurements, the testing laboratory was asked to measure some additional analytes including sodium and calcium. Concentrations of sodium were unphysiologically high in samples stored in the Eppendorf tubes (mean [SD]=177 [3.9] mmol/L) and not in the other samples (138 [2.9] mmol/L), and total calcium was lower (8.5 [0.52] mg/dl versus 9.2 [0.43] mg/dl). Knowing that the original cohort repository was storing both serum and citrate-anticoagulated plasma, the testing laboratory performed protein gel electrophoresis on several samples from the Eppendorf tubes, which revealed an obvious fibrinogen band, and also measured ionized calcium, which was undetectable. It was surmised that the samples in the Eppendorf tubes were in fact sodium citrate–anticoagulated plasma rather than serum as was requested. Fortunately, most analyte concentrations were not significantly different between the two sample types except for vitamin D, which was lower and needed statistical adjustment for sample type.

Errors in Sample Labeling

The importance of standardized procedures for labeling and QC reporting was demonstrated through the detection, investigation, and resolution of a processing error in one set of consortium biosamples. Urine samples from one cohort were aliquoted and labeled at the original cohort laboratory and shipped to two different laboratories for measurement of specific analytes in 2012. Personnel at the original cohort laboratory selected blind replicate samples at the time of aliquoting, choosing among eligible specimens that had

Table 2.   CKD Biomarkers Consortium QC procedures

Category of QC Procedures	Procedure
Historical performance measures	Each laboratory that performs assays provided the following information to the QC committee: Laboratory accreditation (*e.g.*, CLIA, CAP); for laboratories with no accreditation, information about the experience and qualifications of the supervisory staff and testing personnel was requested Instrument to be used for each analyte Experience with the particular instrument A description of measurement procedures on that instrument Westgard or other internal QC rules followed for acceptance of analytical batches Type and source of internal QC pools used for the measurement procedure Mean, SD, and percent coefficient of variation and recovery for the analyte
Sample preparation and handling	Each laboratory that prepared samples for assay provided the following information to the Consortium Coordinating Center or to the QC committee: List of specimens with sample identities and collection date for establishing a tracking database Procedures to be followed when preparing sample aliquots Procedures to be followed to randomize sample sequence for aliquoting, when feasible Procedures for selecting blind replicate samples. Laboratories were encouraged to provide replicates for 2% of the study samples or a minimum of 50 replicates per assay. Linkage information was maintained at the Consortium Coordinating Center and was not shared with the laboratories performing the assays
Prospective quality assurance	Each laboratory that performs assays provided the following information during and after completing each assay: Electronic listing of QC results and study data for the purpose of generating standardized internal and external QC reports Results of proficiency testing of QC pools of sufficient size to last the duration of consortium testing. These pools were collected from patients at various stages of CKD. Aliquots from these pools were added to the samples that were assayed to test for any drift in the assay

QC, quality control; CLIA, Clinical Laboratory Improvement Amendments; CAP, College of American Pathologists.

the greatest sample volume to avoid excess depletion of remaining specimens. Linkage of blind replicate sample identity barcodes to the parent sample identity was determined at the original cohort laboratory and transmitted to the Consortium Coordinating Center. This was in contrast with the process used for samples from other cohorts in which blind replicates were chosen in advance and labels were created by the Consortium Coordinating Center with the blind replicate identity barcodes already linked to the parent sample identity.

Upon completion of sample assays, the Consortium Coordinating Center received assay results from the two performance laboratories, each of which measured multiple analytes, including albumin, creatinine, liver-type fatty acid binding protein, transferrin, and uromodulin. QC reports generated by the Consortium Coordinating Center

indicated large discrepancies between the assay results of the blind replicates and the primary samples for multiple analytes measured in each laboratory, including analytes that generally have good reproducibility, such as urine creatinine and urine albumin. The fact that these discrepancies were seen across more than one laboratory performing assays and across multiple analytes implicated a labeling or identity management error, rather than a laboratory assay error.

To determine which biosamples may have been affected by a labeling error, an additional aliquot from each parent sample was sent in 2013 to one of the laboratories for repeat measurement of urinary creatinine. The objective of repeat measurements was to compare the new results with historical values from the parent study assays performed before 2009, as well as with the consortium measurements made in 2012. With the exception of four outliers, the
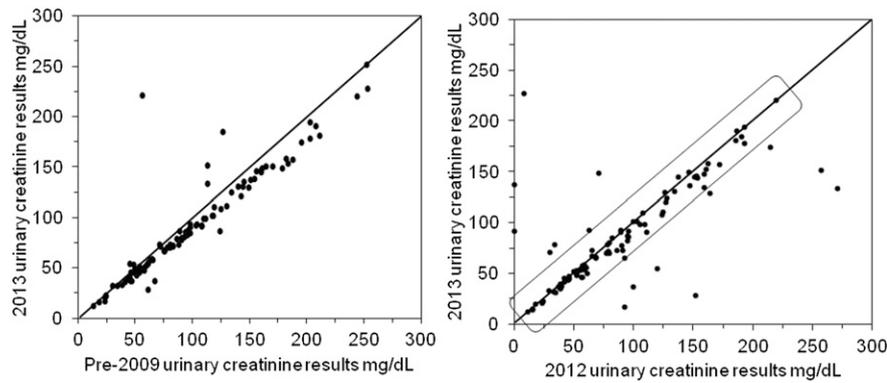
Figure 1. | Evaluation of sample labeling error. Comparison of urine creatinine concentration results in urine samples from one cohort assayed before 2009 and again in 2013 (left panel). Comparison of urine creatinine concentrations in samples from the same cohort from assays performed by a Biomarkers Consortium laboratory in 2013 to those from assays performed in the same laboratory in 2012 (right panel). Data points falling near the line of identity are shown in the rectangular box.

measurements performed in 2013 reliably reproduced the pre-2009 historical urine creatinine results (Figure 1, left panel). Next, investigators compared the measurements from 2013, excluding these outliers, with the assay results obtained in 2012 (Figure 1, right panel). When there was good correlation between the measurements, with values falling near the line of identity, the identities of the 2012 samples were deemed to be uncompromised and biomarker assays from those samples were used in subsequent analyses. Approximately 16% of samples were poorly correlated in either the pre-2009 versus 2013 or 2012 versus 2013 comparisons and were excluded from the final analyses.

Identification of labeling errors was made possible by the preparation and analysis of blind replicate samples, without which gross measurement errors would have gone unnoticed. Comparisons of new and historical results across laboratories salvaged a majority of the reliable data that would otherwise have been lost.

Errors in Sample Registration

We learned that errors can occur in data management when assay results are assigned to the wrong biospecimens. In one laboratory, samples from a single aliquot were subaliquoted into three wells so they could be assayed in triplicate. The

laboratory had changed the assay procedure specifically for CKD Biomarkers Consortium studies and this was the first use of the new procedure. Because of an incorrect template, the order in which parent samples were distributed in the plate was not captured properly by the software package, which assigned the order of sample results horizontally rather than vertically (Figure 2). Despite a reassuringly low coefficient of variation for triplicate measurements, an error in sample identity was only noted because of the inclusion of blind replicate samples. This error was readily detected in the first batch of samples assayed. Thus, blind replicate samples are critically important for assessing validity of measurements, especially when a laboratory protocol has changed, and they are a vital part of a comprehensive quality assurance strategy.

Issues Associated with Collection, Storage, and Retrieval of Banked Biospecimens

The majority of consortium studies included biosamples collected during prospective observational cohort studies and randomized controlled trials and stored in associated biorepositories. Such archival specimens may be particularly valuable for CKD biomarker studies given the extended
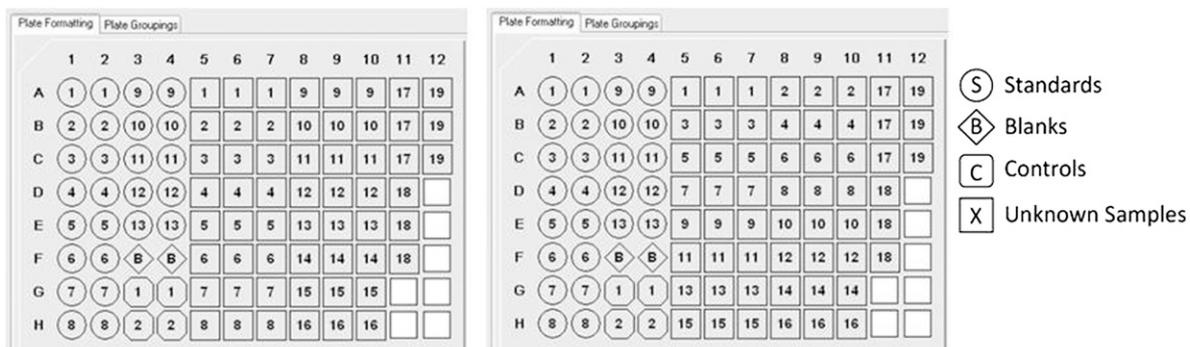

Figure 2. | Data templates on two Luminex instruments. The left panel shows sample order assigned horizontally and the right panel shows sample order assigned vertically.
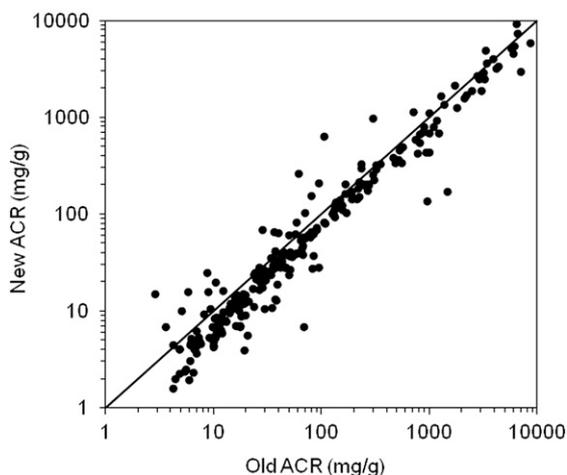
Figure 3. | Scatter plot of urine ACR measured at the time of sample collection (old ACR) and after long-term frozen storage (new ACR). The line of identity is shown. ACR, albumin/creatinine ratio.

duration of follow-up available for analysis of biomarkers in relation to clinically relevant end points, such as a doubling of serum creatinine concentration, the need for RRT, or death. The laboratory storage conditions for these precious older samples, however, may not be ideal by current laboratory practice standards for analytes that are more prone to degradation. Suboptimal storage conditions could potentially influence assay results and, therefore, the findings of biomarker studies (3).

Effect of Sample Collection and Storage Conditions

For some of the studies used by the consortium, urine samples were originally shipped on cold packs (4°C) to a central location where laboratory measurements were made and the residual samples were stored at −70°C. By contrast, samples obtained from other studies were processed under more standardized conditions and frozen at −70°C within 2 hours of sample collection.

Such information on biosample storage conditions provided an appropriate context when interpreting results from pilot studies that were designed specifically to probe the integrity of these biospecimens. In one example, the consortium had access to several cohorts of CKD patients spanning from the late 1980s through the 2010s. Before embarking on thousands of biomarker measurements across multiple cohorts, we performed a pilot study in which we measured urinary biomarkers of interest in 100 samples (50 individuals who experienced CKD progression, and 50 who did not) in each of the three large cohort studies where eligible samples were available. We observed that the proportion of samples with undetectable biomarker levels among the cohorts in which sample freezing was delayed was higher relative to the cohorts in which urine specimens were frozen promptly, even after adjustment for other potential factors that may be associated with low biomarker levels, such as GFR. Because we could not determine with certainty if the observed differences in nondetectable biomarker values were causally related to differences in the handling of urine specimens at the time of sample collection (versus true biologic differences), we chose a conservative strategy of using only laboratory specimens from the cohorts with standardized biosample processing protocols in which the specimens were promptly frozen.

Biomarker Degradation during Long-Term Storage

Some of the blood and urine biosamples considered by consortium investigators were collected for different purposes decades before the formation of the consortium. An extended storage period may be associated with sample degradation over time. The general contemporary consensus is that biosamples stored at −20°C for long periods of time are not recommended for use in scientific studies of biomarkers, given the known degradation of common analytes over time (4–6). An important reason samples stored
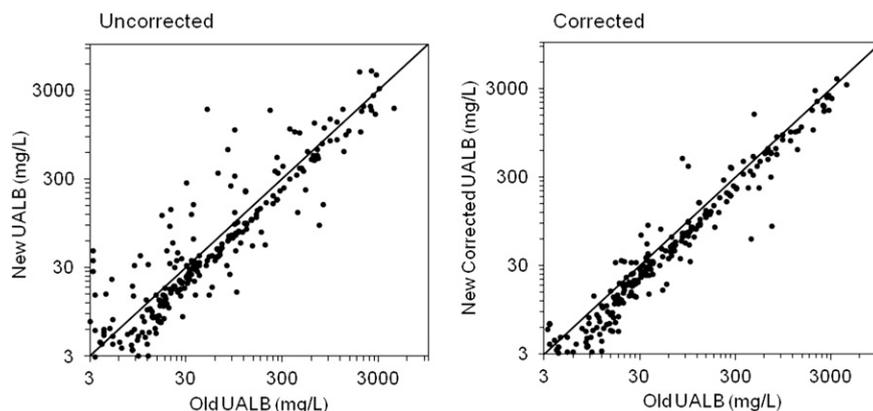

Figure 4. | Evaluation of sample desiccation. Scatter plots of urine albumin concentration measured at the time of sample collection (Old UALB) and after long-term frozen storage (New UALB) before (left panel) and after (right panel) correcting for water loss in the sample during storage. Water loss was estimated by computing the ratio of urine creatinine concentration measured at the time of sample collection and the concentration measured after long-term storage. The urine albumin concentration measured after storage was then multiplied by this ratio. The lines of identity are shown. UALB, urine albumin.
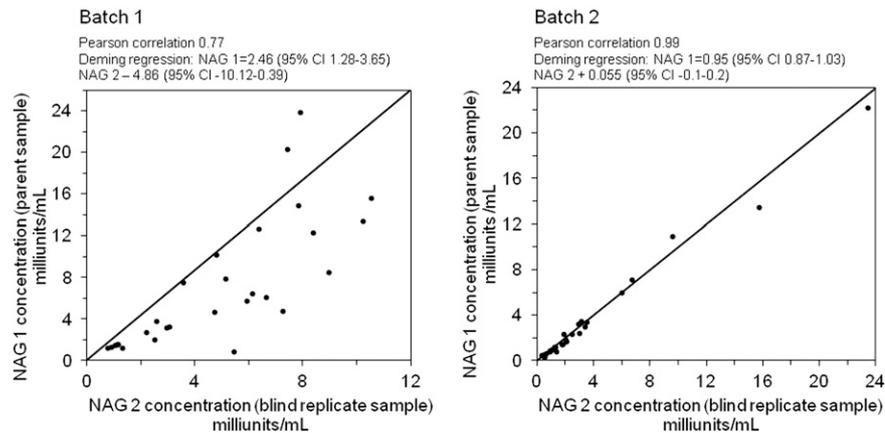
Figure 5. | **Evaluation of incomplete mixing.** Scatter plots of urine NAG concentration measured in the parent sample (NAG 1, vertical axis) or blind replicate sample (NAG 2, horizontal axis) in samples aliquoted in two large batches (batch 1, left panel; batch 2, right panel). The Pearson correlation and Deming regression for each batch are shown. The lines of identity are shown. 95% CI, 95% confidence interval; NAG, *N*-acetyl-β-D-glucosaminidase.

at −20°C are suboptimal is that freeze-thaw cycles can still occur at −20°C. These freeze-thaw cycles can lead to a progressive decline in measured protein concentrations, particularly when antibody-based laboratory assay methods are used, because the degraded proteins might not be detected by antibodies.

CKD Biomarkers Consortium investigators observed that when urinary albumin concentrations were measured by mass spectrometry using a peptide-level MRM assay method, the results were consistent with the antibody-based albumin measurement performed at sample collection even in samples stored for many years at −20°C. Overall, our findings suggest that although protein analytes in specimens stored for up to 30 years at −20°C may not be optimal for assay by antibody-based methods, they may be amenable to assay protocols that measure the analytes by mass spectrometry (*e.g.*, MRM). Target peptides for MRM analysis should avoid oxidation-sensitive amino acids detected in top-down analyses, and oxidation-resistant creatinine is accurately replicated after long-term storage at −20°C. If this approach is validated for other analytes, it may greatly enhance the usability of older existing sample collections not optimally stored at −70°C or below for future clinical research.

Desiccation of Stored Samples

We learned that even with appropriate storage of bio-samples at −70°C or lower temperatures, there can be unanticipated problems in assaying biomarkers after long-term storage. Samples from one cohort were stored at −70°C and sent to a consortium laboratory for ELISA testing using a Luminex platform. During a pilot study, some samples (even after centrifugation) were sticky (*i.e.*, there was bead aggregation on a Luminex platform) and had a high coefficient of variation in duplicate measurements. Bead aggregation may alter analyte values by clogging the needle on the assay instrument, or possibly by carry-over of the beads from one sample into another sample in the plate. Further investigation of the sticky samples included determination

of linearity of the assay after diluting samples at 1:4, 1:8, and 1:10 concentrations. At these dilutions, results were far more consistent and had reduced bead aggregation.

Aliquots from the same samples were also sent to another laboratory for measurement of urine creatinine and urine albumin concentrations. These measurements were repeats of measurements made on the original samples collected 10–20 years earlier. We noted a lack of concordance between consortium measurements and the original measurements of both analytes, with systematically higher concentrations in the recent measurements compared with the original values. Yet when expressed as a ratio (*i.e.*, the albumin/creatinine ratio), consortium measurements were comparable to the same ratio in the original measurements (Figure 3).

Combining these observations (sample stickiness on a Luminex platform and higher concentrations of both urinary albumin and creatinine), we inferred sample desiccation as the likely source. Specifically, to explain why some of the samples had substantially higher concentrations after long-term storage than they did at the time of collection (Figure 4, left panel), we estimated the water loss that occurred during storage from the ratio of urine creatinine concentration measured at the time of sample collection to the concentration measured after storage. After correcting for water loss by this method, the urine albumin concentrations of the samples measured after long-term storage were equivalent to those measured at the time of collection (Figure 4, right panel). Accordingly, all urinary biomarker concentrations reported in this cohort were normalized to the urine creatinine concentration of the stored specimen to account for any sample desiccation that may have occurred.

Further investigation confirmed that some of the samples were stored using tubes that lacked a rubber seal and thus may not have been consistently air tight. Specifically, instead of cryovials with neoprene or silicon gaskets that provide a tight seal between the cap and the tube, non-gasketed tubes were used and some of the seals had failed. We suspect that some samples desiccated slowly over time, leading to higher viscosity and higher concentrations of

Table 3.   Key lessons and recommendations

Lesson	Recommendation
Incorrect biologic specimens may be selected for assay (*e.g.,* plasma instead of serum)	Barcode stored biologic specimens to simplify tracking and characterization
Aliquots may be labeled incorrectly	Employ standardized procedures for labeling and reporting Include blind replicate samples
Assay results may be reported incorrectly even with automated reporting systems	Standardize assay reporting procedures Pilot test assay reporting procedures that have changed Include blind replicate samples
Proper collection, processing, and storage of biologic specimens are essential	Steps in collection, processing, and storage of biologic specimens must be clearly specified For many biologic specimens, storage at $-70°C$ is preferable to $-20°C$ to prevent degradation or other forms of artificial protein modification Proper frozen storage requires cryovial tubes with screw caps and silicon gaskets to prevent sample desiccation When storing samples for future research, consider preparing multiple small aliquots at initial collection to prevent multiple freeze-thaw cycles and to reduce the likelihood of aliquoting errors related to inappropriate mixing of large sample volumes Perform pilot assays on subsets of samples when there are any concerns about the suitability of the available biospecimens
Perform critical assays in experienced laboratories	Develop standard operating procedures for every assay that include details of the laboratory's internal QC practices Develop external QC procedures to supplement these internal practices. These procedures should include: Blind replicate samples Proficiency samples from QC pools
Routinely examine QC data	Journals should ask for quality control data as part of standard reporting practice

both urinary albumin and creatinine, but preserved values for the urine albumin/creatinine ratio.

This experience demonstrates three important lessons: (*1*) appropriate storage containers for biosamples are important, in addition to storage temperature; (*2*) well designed QC laboratory measurements are able to identify preanalytic errors in storage or processing; and (*3*) urine samples from archived samples should be assayed for creatinine routinely, even when analyte/biomarker values are being evaluated without normalization to creatinine, to test for the possibility of desiccation. Urinary creatinine, as a stable analyte over time, will be expected to change only in so far as the assay standard might have drifted or intentionally changed. If its concentration increases over time, it is possible that water has been lost from the samples.

Incomplete Mixing during Aliquoting of Stored Urine

Measurement of analytes in blind replicate samples from a commercial repository demonstrated unacceptable reproducibility. The discrepancy was evident across two analytes measured by the laboratory. There were many possible explanations for this result. However, when we split the results by aliquot batch, it became clear that the poor reproducibility occurred with only one of the two batches assayed and reported at the same time, suggesting that the problem was not one of assay reproducibility (Figure 5). In addition, the correlation of blind replicate and parent assay results improved when the biomarker concentrations were normalized to urinary creatinine concentrations, suggesting that the issue was more likely related to sample handling than labeling *per se* (Figure 5). The performance laboratory also confirmed that the labeling process had followed the SOP.

Further testing of different aliquots from the same sample clearly identified variability in biomarker values as a result of insufficient mixing at the aliquoting step. The original parent aliquot contained 9 ml of urine. When multiple daughter aliquots were prepared, incomplete mixing resulted in lack of homogeneity in the concentration of the analytes in different daughter tubes.

We ultimately determined that a change in operator at the repository resulted in a change in how the samples were mixed before aliquoting. This finding was confirmed by thawing multiple daughter aliquots from the same parent sample and showing variation in concentrations of urinary sodium, albumin, and creatinine, analytes in which the assays are extremely robust and well established. Without detailed SOPs and blind replicate sample analysis, it is unlikely that we would have identified the problem at this level or used the samples.

Issues at the Analytic Stage
Outliers

There are many analytic issues related to biomarker studies that are beyond the scope of this article. Nevertheless, when analyzing data from blind replicates, or comparing contemporary measurements with prior measurements of an analyte from the same parent aliquot, notable outliers (e.g., off by an order of magnitude) are occasionally observed (on the order of 1%). Presence of these occasional notable outliers may not indicate that there are serious QC problems. The reasons for their existence are often not clear, although some due diligence is necessary to be sure the discrepancy is not because one of the values is close to the lower limit of detection of the assay or there has been mislabeling of the samples or other identity problems.

Recommendations

The past decade has seen a substantial increase in interest in biomarkers in the field of nephrology. Few reports however, have addressed important methodologic issues in the conduct of such studies. We believe that enhancing methodologic rigor is a fundamental requisite to further research in this field. Key lessons learned in the first 5 years of conducting large-scale consortium-driven biomarker research are summarized in Table 3. In designing future biomarker studies, investigators should consider the following:

(1) *Establishing QC procedures and standard operating protocols*: Assembling standardized operating protocols and continuously monitoring the preanalytic, analytic, and postanalytic components of biomarker measurement is critical to ensure the integrity and highest quality of the resulting scientific findings.
(2) *Assessing reproducibility*: Blind replicates are invaluable and should be considered essential in the design of any biomarker study.
(3) *Ensuring optimal sample storage conditions*: Little is known about the effects of long-term storage on most analytes. Clinically useful biomarkers must be able to withstand deviations in conditions that occur during routine clinical chemistry laboratory handling. Nevertheless, during discovery and early validation, great attention must be paid to the conditions under which biosamples are collected, handled, and stored, because these conditions may affect assay results of putative biomarkers (3). For most purposes, samples stored at −70°C are preferred to those stored at −20°C. Storage of multiple aliquots will help to ensure that adequate samples are available that have not been exposed to multiple freeze-thaw cycles.
(4) *Conducting pilot and feasibility studies*: Conducting such studies on a smaller subset of samples as initial steps can be very helpful to recognize problems with sample identification, effects of thawing and subaliquoting, and shipping. These studies may also be informative regarding the integrity of samples (e.g., proportion with undetectable biomarkers) and how the biomarker assays perform in the hands of the actual laboratory doing the testing (beyond what is described by the manufacturer, which may overstate assay performance; e.g., percent coefficient of variation of duplicate measurements).

In summary, we provide an overview of key lessons we have learned in the initial phase of the CKD Biomarkers Consortium, with the intent of facilitating ongoing and future biomarker science in renal medicine.

Disclosures

C.H. received donation of biomarker reagents/measurements for research studies from Abbott. J.V.B. is coinventor on KIM-1 patents, which are licensed by Partners Healthcare, and has received royalty income from Partners Healthcare and grant funding from Novo Nordisk as well as income for consulting on biomarkers for Pharmaco-Kinesis, Abbvie, and Medivation. J.H.E. is a paid consultant for Gentian. K.D.L. received donation of reagents by Abbott

References
1. US Renal Data System: *USRDS 2013 Annual Data Report: Atlas of Chronic Kidney Disease and End-Stage Renal Disease in the United States*, Bethesda, MD, National Institutes of Health National Institute of Diabetes and Digestive and Kidney Diseases, 2013
2. Biomarkers Definitions Working Group: Biomarkers and surrogate endpoints: Preferred definitions and conceptual framework. *Clin Pharmacol Ther* 69: 89–95, 2001
3. Star R, Kimmel P, Bonventre J: Best practices for sample storage: A report from the Workshop on Urine Biospecimen Handling, 2010. Available at: http://archives.niddk.nih.gov/urine/Best_Practices_for_Sample_Storage.pdf. Accessed November 21, 2014
4. Manley SE, Burton ME, Fisher KE, Cull CA, Turner RC: Decreases in albumin/creatinine and N-acetylglucosaminidase/creatinine ratios in urine samples stored at -20 degrees C. *Clin Chem* 38: 2294–2299, 1992
5. Parekh RS, Kao WH, Meoni LA, Ipp E, Kimmel PL, La Page J, Fondran C, Knowler WC, Klag MJ; Family Investigation of Nephropathy and Diabetes Research Group: Reliability of urinary albumin, total protein, and creatinine assays after prolonged storage: The Family Investigation of Nephropathy and Diabetes. *Clin J Am Soc Nephrol* 2: 1156–1162, 2007
6. Schultz CJ, Dalton RN, Turner C, Neil HA, Dunger DB; The Oxford Regional Prospective Study Group: Freezing method affects the concentration and variability of urine proteins and the interpretation of data on microalbuminuria. *Diabet Med* 17: 7–14, 2000