Analytical and Bioanalytical Chemistry
https://doi.org/10.1007/s00216-026-06769-8

TRENDS

Redefining reference materials for the evolving metabolomics landscape

Tracey B. Schock[1] · Fabio Casu[1] · Ashley S. P. Boggs[1] · Amanda L. Bayless[1] · Hariharan Iyer[2] · Kristine Gierz[2] · W. Clay Davis[1]

Received: 24 June 2026 / Revised: 13 August 2026 / Accepted: 18 August 2026

Abstract

The evolution of metabolomics, a field aimed at comprehensively measuring the small organic molecule composition of a biological matrix, fundamentally reshaped the landscape of analytical measurement reliability. The 2011 release of Standard Reference Material (SRM) 1950 Metabolites in Frozen Human Plasma by the National Institute of Standards and Technology (NIST) marked the first reference material providing certification for 45 analytes to address the measurement precision of complex biological matrices. Assigning high-order values to many compounds is highly resource-intensive for a National Metrology Institute. Furthermore, certified values for individual metabolites cannot resolve the broad challenges associated with sample processing, analytical measurement, and data processing in untargeted workflows. The metabolomics community recognized the potential for SRM 1950 to serve as a benchmark material for method development and technical quality control (QC) rather than strictly for quantification. Recognizing this consumer-driven shift, NIST launched a novel reference material (RM) development strategy to provide accessible, fit-for-purpose materials evaluated by a new statistical framework for production homogeneity, the Coefficient of Disagreement. Here, we introduce the transition toward matrix-specific *QC Suites* featuring phenotypically distinct metabolite profiles, the first generation of which includes a human plasma suite, urine suite, liver suite, and fecal material. By prioritizing efficient production and embracing a community engagement model for consensus-based deep characterization, these suites offer a reliable tool for laboratory comparisons, instrument assessment, software development, and training. This new class of RMs underpins the next phase of measurement reliability, complementing traditional measurand certification while supporting the diverse needs of comprehensive metabolic profiling.

Keywords Metabolomics · Metrology · Reference materials · Coefficient of disagreement · Quality control

Introduction

The field of metabolomics seeks to provide a comprehensive snapshot of the small-molecule constituents within a biological system, offering a functional readout of cellular state and phenotype [1]. However, the inherent complexity of the metabolome characterized by a vast chemical diversity spanning several orders of magnitude in concentration and a wide array of physicochemical properties presents a formidable challenge for analytical chemistry [2]. Validating specific metabolites or complex metabolite profiles as reliable biomarkers requires more than findings from a single study; it necessitates the rigorous comparison of datasets generated across different laboratories and diverse analytical platforms, including liquid chromatography–mass spectrometry (LC–MS), gas chromatography–mass spectrometry (GC–MS), and nuclear magnetic resonance (NMR) spectroscopy. Without a common

This work is dedicated to Steve Wise, whose career of advancing measurement science is matched only by his passionate devotion to professional development and mentorship. His legacy of analytical rigor and dedication to standards development has shaped both the field of metrology and the careers of the scientists within it. We are especially grateful for his early willingness to champion and develop a metrology-based metabolomics program.

✉ Tracey B. Schock
tracey.schock@nist.gov

[1] Chemical Science Division, National Institute of Standards and Technology, 331 Ft. Johnson Road, Charleston, SC 29412, USA

[2] Statistical Engineering Division, National Institute of Standards and Technology, 100 Bureau Drive, Gaithersburg, MD 20899, USA

Published online: 02 September 2026

measurement scale to harmonize these disparate workflows, biological insights risk being confounded by technical variance, hindering the translation of metabolomic discoveries into clinical or industrial applications.

For decades, the National Institute of Standards and Technology (NIST) has served as the arbiter of measurement reliability, providing the *yardsticks* necessary to ground emerging scientific fields in metrological traceability. Under the leadership of Dr. Stephen A. Wise, NIST's Analytical Chemistry Division spearheaded the development of complex matrix Standard Reference Materials (SRMs) that moved from simple calibration solutions to embrace the intricacies of biological fluids and tissues. This culminated in the 2011 release of SRM 1950 Metabolites in Frozen Human Plasma, a landmark material developed with the National Institutes of Health (NIH) [3]. By providing certified values for nearly 50 analytes, SRM 1950 provided the nascent metabolomics community with its first true benchmark material.

Since its release, NIST SRM 1950 quickly became the most widely adopted reference material for metabolomics and lipidomics. However, with the rapid growth of high-throughput, untargeted profiling, the limitations of traditional reference material models have become apparent. The vast chemical diversity within biological matrices surpasses the practical reach of the resource-intensive process of individual compound certification. Modern untargeted workflows demand a breadth of coverage that includes unforeseen analytes, making a compound-by-compound approach insufficient for the evolving needs of the field. Additionally, it became increasingly clear that a single material cannot meet all clinical and analytical needs, prompting community requests for RMs representing specific populations (e.g., male and female cohorts) or disease states. The current fundamental challenge is detecting and reproducing fold-changes or differential signatures defining biomarkers between phenotypically distinct groups. This has inspired a paradigm shift at NIST toward fit-for-purpose Reference Materials (RMs), alongside the introduction of the Coefficient of Disagreement (COD) as a new statistical tool for assessing production-level homogeneity and variability in metabolomics profiles across an RM lot. These materials are intended to provide the community with accessible tools [4] needed for validating accuracy and precision of metabolomics measurements, for evaluating comparability and harmonization across studies and laboratories [5], for creating benchmarking datasets for software development [6], and for underpinning training and proficiency testing.

Fit-for-purpose QC Suites

NIST's new RM development strategy, the Metabolomics Quality Assurance and Quality Control Materials (MetQual) program, focuses on providing cost-friendly, matrix-matched materials designed specifically for untargeted and differential analysis. Unlike single-point reference materials, these "QC Suites" include phenotypically distinct metabolite profiles to mimic the variations encountered in clinical and research settings. Serving as versatile QA/QC tools for the metabolomics community, these suites support routine measurements, analytical method suitability, and analyst training. Furthermore, they enable researchers to evaluate data comparability across different laboratories and analytical approaches, while simultaneously providing a robust benchmark for assessing the confidence of metabolite annotation and identification. By providing these varied suites and qualitative characterizations with lists of confidently annotated metabolites, NIST is laying the foundational groundwork to help researchers evaluate whether observed phenotypic differences in the metabolome reflect true biological variation or arise from variability in sample processing, measurement techniques, or data handling.

In this Trends article, QC Suite donor characteristics are detailed in Table 1. Additional context on both the composition and the philosophy of each suite can be found in the NIST Reference Material Information Sheets (RMIS) that accompany each RM (https://shop.nist.gov); these documents provide comprehensive information on donor pool demographics, collection procedures, processing, and distribution format, and serve as the primary technical record for each material, including data lists of confident metabolite annotations and identifications (also found in Supplementary Information). Furthermore, the work of Lippa et al. [4] outlines definitions, design considerations, and intended uses of reference materials in untargeted metabolomics ("Fit-for-purpose QC suites"), including sections describing guiding principles that informed the QC suite compositions.

RM 8231 frozen human plasma Suite

Blood matrices are the most widely used biofluid in clinical trials and metabolomics due to their accessibility and reflective nature of systemic physiology. Providing a reference material in a matching matrix ensures that the quality control steps more closely mimic the clinical samples being analyzed, thereby helping plasma suiteto account for matrix-dependent effects during extraction or ionization. As an alternative to SRM 1950, RM 8231 is a three-part suite designed specifically for untargeted, differential metabolomics.

The suite includes unadulterated pooled human plasma from distinct cohorts (Table 1):

- Diabetic Plasma (Pool 1): pooled from donors (6 male, 5 female) with glucose > 126 mg/dL and low/normal triglycerides (< 150 mg/dL) (Table S1)

Table 1 Donor characteristics for each metabolomics QC Suite. SD denotes standard deviation

RM 8231	Diabetic Plasma (Part 1)	Hypertriglyceridemic Plasma (Part 2)	African-American Plasma (Part 3)
Age (years)	34 to 68	31 to 72	20 to 25
Sex (Male/Female)	55/45	100/0	50/50
Race	55% White 36% African American or Black 9% Asian	55% White 36% African American or Black 9% Hispanic	100 % African American or Black
Fasted	Yes	Yes	Yes
Anticoagulant	Lithium heparin	Lithium heparin	K2 EDTA
Glucose (mg/dL)	143	109	N/A
Triglycerides (mg/dL)	124	367	N/A

RM 8232	Female Non-Smoker Urine (Pool 1)	Female Smoker Urine (Pool 2)	Male Non-Smoker Urine (Pool 3)	Male Smoker Urine (Pool 4)
Age (years)	33 to 63	28 to 47	51 to 58	45 to 51
Sex (Male/Female)	0/100	0/100	100/0	100/0
Gravimetric Density (g/mL)	1.01068 ± 0.00108 SD	1.01614 ± 0.00229 SD	1.01826 ± 0.00352 SD	1.02499 ± 0.00056 SD

RM 8462	Normal Liver	Congested Liver	Fatty Liver
Age (years)	14 to 25	46 to 52	43 to 50
Sex (Male/Female)	100/0	100/0	50/50
BMI (average)	25.5	27.5	35.3
Liver Histology Notes	Liver within normal limits, mild congestion	Moderate vascular congestion, normal mild congestion	Mild fatty change

RM 8048	Omnivore Diet Stool	Vegetarian Diet Stool
Age (years)	unknown	unknown
Sex (Male/Female)	unknown	unknown
BMI (average)	unknown	unknown
Diet Logs and Dietary Profiles	NIST SP260-252 [9] Section 2 and Appendix A	NIST SP260-252 [9] Section 2 and Appendix A

- Hypertriglyceridemic Plasma (Poolt 2): pooled from donors (11 males) with triglycerides > 300 mg/dL and glucose < 100 mg/dL (Table S2)
- African-American Plasma (Pool 3): pooled from healthy donors (8 male, 8 female) between 20 and 25 years of age (Table S3)

Early analyses using untargeted ^{1}H NMR metabolomics (Fig. 1) identified varying levels of glucose, lactate, acetate, betaine, glycine, EDTA (collection anticoagulant), and hippurate across the donor pools while lipidomics identified lauric acid-containing lipids in the hypertriglyceridemic pool and oxidized fatty acid-containing phospholipids in the African-American pool [7], thereby confirming the unique cohort profiles of the phenotypically different groups.

RM 8232 frozen human urine Suite

While plasma represents a tightly regulated systemic snapshot of the body's circulating molecules, urine offers a different but complementary view of the human metabolome. Urine is a biofluid excreted from the kidneys and the product of the body's primary filtration and waste clearance system. Urine is heavily enriched with specific classes of compounds (i.e., waste and by-products of metabolism from foods, drugs, or environmental contaminants, microbiome metabolism, excess electrolytes, etc.) that are cleared quickly from the blood or present there only in trace amounts. This RM suite provides matrix-matched materials for urinary profiling allowing for the precise monitoring of exposure profiles and the evaluation of excretory system integrity. Such data are vital for healthcare providers in diagnosing metabolic imbalances and assessing the physiological impact of external stressors. To ensure distinct metabolic signatures suitable for differential analysis, the suite comprises pooled human urine from four phenotypically distinct cohorts (Table 1):

- Female non-smoker (Pool 1) (Tables S4 and S8)
- Female smoker (Pool 2) (Tables S5 and S9)
- Male non-smoker (Pool 3) (Tables S6 and S10)
- Male smoker (Pool 4) (Tables S7 and S11)

^{1}H NMR spectra revealed distinct pool-level metabolic profiles (Fig. 1), with the detection of several metabolites. For example, signals assigned to ibuprofen and its metabolites, 2-hydroxyibuprofen and carboxyibuprofen, were observed in the female pools, whereas glutarate, 3-hydroxyphenylacetate, and N-acetylcysteine–acetaminophen were among the features differentiating the male non-smoker, female non-smoker, and female smoker pools, respectively. These observations reflect NMR-detectable differences in

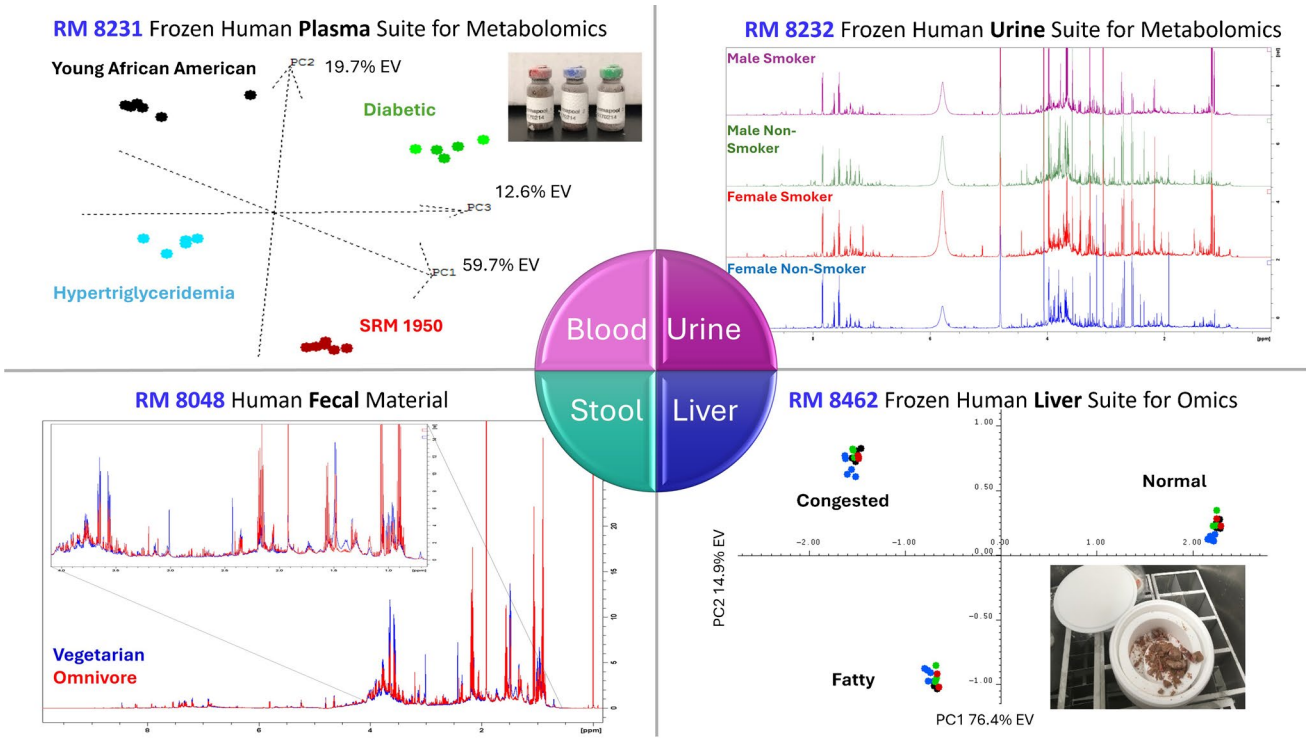

Fig. 1 NIST Metabolomic QC Suites, clockwise from left to right: plasma (RM 8231), urine (RM 8232), liver (RM 8462), and stool (RM 8048). Phenotypically distinct donor pools for each material are conveyed with ^{1}H NMR spectra or principal components analysis (PCA) of ^{1}H NMR spectra. PCA for RM 8462 represents tissue extracts prepared from lyophilized samples (blue) and frozen, wet tissue (black, expert analyst; red, intermediate analyst; green, novice analyst) for each donor pool during an analyst training exercise

the pooled materials; signals not observed in a given pool may be present below the NMR detection limit. This urine suite approach allows researchers to validate methods and software designed to detect differences driven by environmental exposure and lifestyle factors.

RM 8462 human liver Suite

While biofluids like plasma and urine offer an excellent systemic overview of an organism, analyzing solid tissue introduces a completely different set of biochemical and analytical challenges. NIST prioritized the development of tissue-based materials to support invasive diagnostic research. The analytical challenges posed by tissue-specific matrix effects, cellular heterogeneity, massive concentrations of lipids, and complex sample preparation demand rigorous quality control to harmonize data across measurement platforms. Derived from human liver specimens from the EPA Human Liver Biobanking Project [8], archived cryogenically at the NIST Biorepository, RM 8462 supports differential proteomics and metabolomics. The suite categorizes materials into three health states based on pathological data collected at the time of autopsy and body mass index (BMI) (Table 1):

- Normal liver: BMI < 27.5 (Tables S12, S15, S18)
- Congested liver: identified by pathologist visual inspections (Tables S13, S16, S19)
- Fatty liver: BMI average of 35.3 (Tables S14, S17, S20)

Preliminary NMR (Fig. 1) and LC–MS data demonstrate that these materials produce measurable and reproducible differences, with principal component analysis (PCA) scores showing distinct separation between the three health states. By providing matrix-matched tissue RMs, NIST reduces the analytical uncertainty inherent in solid-organ profiling, helping ensure that metabolic signatures are a true reflection of cellular pathology rather than artifacts arising from incomplete homogenization, non-representative sampling, or suboptimal extraction of complex tissues.

RM 8048 human fecal material

As research into the human microbiome explodes, stool has transitioned from a discarded waste product to a goldmine of biological information with fecal metabolomics revealing the functional aspect of the microbial community. To support this field, NIST with support from the Institute for the Advancement of Food and Nutrition Sciences (IAFNS) developed RM 8048 [9]. From an analytical chemistry

perspective, stool is difficult to standardize. It is inherently non-homogeneous, containing varying amounts of undigested dietary fiber, water, shed intestinal epithelial cells, and dense bacterial pellets. A single, homogeneous pool acts as a stable benchmark for this community. The NIST material consists of homogenized stool from multiple donors resuspended in water at 100 mg/mL. It is divided into two dietary parts (Table 1):

- Omnivore donors: pooled from five self-reported healthy donors (Tables S21, S23, S25)
- Vegetarian donors: pooled from three self-reported healthy donors (Tables S22, S24, S26)

RM 8048 underwent extensive multi-omic characterization including metagenomics and metabolomics by LC–MS and ^{1}H NMR [9]. The inclusion of two distinct dietary cohorts supports development and validation of methods designed to detect differences. ^{1}H NMR data differences between omnivore and vegetarian donors are represented in Fig. 1.

For the microbiome industry to achieve widespread commercial success, it must move beyond fragmented discovery toward analytical standardization. NIST reference materials provide the "metrological anchor" necessary to harmonize measurements, fostering the customer trust and data comparability required to scale microbiome-based technologies into global markets.

Quantifying reliability: the coefficient of disagreement

The transition toward complex matrix RMs for untargeted metabolomics requires a corresponding shift in statistical methodologies for assessing material production consistency. Traditional methods for evaluating homogeneity and stability, such as analysis of variance (ANOVA) or trend analysis, often reduce high-dimensional data to binary "homogeneous/heterogeneous" decisions and only focus on statistical significance, which are frequently insufficient for determining whether a material is fit for its intended purpose in untargeted workflows [10]. To address this, NIST has introduced the Coefficient of Disagreement (COD) as an intuitive framework that characterizes the range of variability an end-user can reasonably expect when comparing two randomly selected samples from a material production [11].

For a pair of vials, v_i and v_j, COD is calculated across K measured attributes, such as metabolite peaks, spectral buckets, or taxa. First, a pairwise discrepancy, feature-to-feature discrepancy ($FFD_{ij}(k)$) is calculated for each attribute k. The specific FFD is selected to match the data type and intended interpretation. For example, for non-negative peak-height data, the relative discrepancy can be expressed as

$$FFD_{ij}(k) = \frac{max\{y_{ik}, y_{jk}\}}{min\{y_{ik}, y_{jk}\}} - 1,$$

where y_{ik} and y_{jk} are the measurements for attribute k in the two vials. This metric is directly interpretable as the proportional difference between the two measurements. For presence/absence data, such as selected NMR spectral-bin analyses, a Jaccard distance can instead quantify disagreement in detected features. For LC–MS and other data that span several orders of magnitude, discrepancies may be calculated after log transformation to emphasize relative rather than absolute differences.

The COD for a vial pair is then defined as the $(1-\beta)$ quantile of all attribute-level discrepancies:

$$COD_{\beta}(v_i, v_j) = Q_{1-\beta}\{FFD_{ij}(k) : k = 1, \ldots, K\}.$$

The trimming parameter β is selected according to the application and is reported with the result. Choosing a small nonzero β limits the influence of a small number of extreme or undefined discrepancies while retaining a conservative, interpretable assessment of agreement across the large majority of measured attributes. Thus, COD does not ask merely whether two samples differ statistically; it describes how much they can differ for the specified proportion of attributes.

The flexibility of this framework permits assessment across the diverse data types represented by the QC Suites. For example, in RM 8048 Human Fecal Material, 600 MHz NMR analyses using relative peak-height discrepancies and $\beta = 0.01$ produced approximate 95% upper confidence bounds for COD of 0.1069 and 0.1323 for the omnivore and vegetarian materials, respectively (Figure S3). Complementary 700 MHz NMR analyses based on the presence or absence of spectral features yielded approximate 95% upper confidence limits of 0.1435 for the omnivore material (Figure S4) and 0.1632 for the vegetarian material (Figure S5). These platform-specific results provide complementary evidence that the materials are sufficiently consistent for their intended metabolomics applications.

Because COD values are calculated from overlapping vial pairs, the resulting pairwise observations are not independent. NIST therefore uses statistical methods appropriate to the sampling design, including Ove Frank U-statistic estimators, to estimate population-level COD means and variances and to establish approximate upper confidence bounds. Full platform-specific FFD definitions, COD calculations, statistical procedures, and suite-level results are provided in the Supplementary Information.

Qualitative characterization and nominal properties

The utility of the next-generation QC Suites extends beyond providing physical samples; it encompasses the rigorous assignment of qualitative, or nominal, properties. While traditional SRMs focus on quantitative measurands, modern untargeted metabolomics is equally dependent on the accurate identification, or annotation, of chemical species. Nominal properties, as defined by the international standard ISO 33406, are properties of a phenomenon, body, or substance that have no magnitude, such as the identity of chemical species within each matrix or the phenotypic label of a biological cohort [12]. For the untargeted metabolomics community, where the identification bottleneck remains a significant hurdle, providing RMs with well-documented qualitative identities is essential for software benchmarking and the validation of annotation pipelines. To establish these properties, NIST employs a suite of complementary analytical techniques, including liquid chromatography-tandem mass spectrometry (LC–MS/MS) and nuclear magnetic resonance (NMR) spectroscopy (1D and 2D) for metabolite profiling, alongside next-generation sequencing (NGS) to characterize community composition specifically for materials like the Human Fecal Material (RM 8048).

For mass spectrometry-based characterization, NIST has developed a metrological workflow adapted from the "Two or More Independent Reference Measurement Procedures" mode outlined in NIST Special Publication 260-136 [10]. This rigorous protocol for nominal property assignment is built upon a consensus formula requiring two or more independent sample preparation methods, two or more analytical measurement platforms, and three distinct search algorithms for spectral matching. In the characterization of the Frozen Human Plasma Suite (RM 8231), this involved parallel extractions using methanol and ethanol protein precipitation, with data collected on both Q-Exactive and Fusion Lumos Orbitrap mass spectrometers. The resulting spectra were searched using three distinct algorithms, HighChem-HighRes, HighChemDP, and NIST, within the Compound Discoverer environment against both the NIST20 mass spectral library and 2019 mzVault library, providing access to large, curated collections of experimental MS/MS spectra from authentic standards at different experimental conditions.

Within this consensus model, highly confident identifications are defined only as those metabolites matched by all three algorithms across both instrumentation platforms and extraction methods. This high-stringency approach, which filters for matches and requires consensus across 48 datasets per sample, yields a reproducible core metabolome for each pool. However, we explicitly recognize that library matching alone, no matter how consistent across platforms or laboratories, does not eliminate the risk of misannotation, particularly in the context of sparse libraries and the combinatorial complexity of metabolite structures. Accordingly, these identifications are reported as putative annotations (Metabolomics Standards Initiative Level 2 [13]), unless supported by authentic standards in the same matrix (MSI Level 1), and should not be interpreted as absolute ground truth. Rather, they provide a transparent, consensus-based reference against which new deconvolution algorithms, library-search strategies, and data-processing tools can be evaluated. As community datasets accumulate, including cases where MSI Level 1 identifications are established across multiple laboratories, it may become appropriate to consider additional reporting categories for consensus-validated metabolites, but such developments will require broad community discussion. This qualitative foundation is nonetheless a necessary prerequisite before consistency and variability can be quantified through more advanced assessments and, where warranted, before future certified value assignments are considered.

Outlook: community engagement and consensus characterization

The transition toward qualitative characterization also reflects a broader metrological shift in the life sciences. While quantitative certification requires metrological traceability to the International System of Units (SI), qualitative properties are often established against high-confidence spectral databases or through consensus among expert laboratories. As detailed in NIST SP 260-136 [10], the assignment of these properties can involve results from multiple independent methods or interlaboratory comparisons to ensure robustness. In departure from internal certification models, NIST is embracing a community engagement model for these materials. While NIST performs foundational homogeneity and stability assessments, deep characterization is increasingly driven by consensus-based characterization through international interlaboratory studies.

This model is anchored in the success of the de facto anchor in metabolomics QA/QC, SRM 1950. A 2017 NIST lipidomics interlaboratory study set the stage for community-driven characterization by engaging 31 laboratories to establish consensus concentration estimates for lipids in SRM 1950 [14]. While the values are not metrologically traceable, they provided a critical benchmark for a community operating in an environment largely deplete of internal standards. The agreement between these

community-derived consensus medians and NIST-certified values demonstrated that aggregated interlaboratory data could provide reliable estimates suitable for benchmarking and harmonization.

Building on this blueprint, the Metabolomics Quality Assurance and Quality Control Consortium (mQACC) (www.mqacc.org) Reference and Test Materials Working Group led a crowdsourcing effort to expand the utility of SRM 1950 to the wider metabolomics community. The project is compiling both quantitative and qualitative data for SRM 1950 from 25 community submissions and 23 peer-reviewed publications, encompassing measurements from LC–MS, GC–MS, direct injection (DI-MS), capillary electrophoresis (CE-MS), and NMR platforms under diverse analytical conditions. Consensus median concentrations will be reported along with qualitative identifications.

The success of the SRM 1950 projects now informs the characterization of the next-generation QC Suites. For RM 8231, NIST distributed samples to five external entities, including the International Lipidomics Society (ILS), Biocrates®, and academic centers like the University of Florida and the University of Lausanne. This collaboration utilizes varied platforms to generate quantitative consensus values for approximately 1000 analytes per plasma pool. Through this collaborative model, NIST provides the foundational physical benchmarks while the global community contributes the essential "omic" depth. This rich collection of consensus data will facilitate the identification of high-priority analytes for future SI-traceable certification, ensuring that metrological reliability remains at the forefront of an increasingly complex and rapidly shifting analytical landscape.

Conclusion

By transitioning from measurand certification of single-point materials to qualitative and consensus concentrations for phenotypically diverse QC Suites, NIST continues to honor the foundational philosophy of measurement excellence led by Dr. Steve Wise while adapting to the needs of modern biology. By prioritizing usability, commutability, and cost-effectiveness, these fit-for-purpose QC Suites provide the stable benchmarks necessary to underpin measurement reliability and accelerate the advancement of metabolomics.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00216-026-06769-8.

Acknowledgements We would like to express our gratitude to our NIST colleagues Katrice Lippa, Christina Jones, Debra Ellisor, Stephanie Servetas, Samuel Forry, Monique Hunter, Jennifer Dootz, Aaron Urbas, Sandra Da Silva, Kirsten Parratt, Scott Jackson, and Nancy Lin for their contributions to the development and analysis of these reference material suites.

Author contribution Material preparation, data collection, and analysis were performed by Tracey Schock, Amanda Bayless, Fabio Casu, Hariharan Iyer, Kristine Gierz, and Clay Davis. The first draft of the manuscript was written by Tracey Schock, editorialized by Ashley Boggs, and all authors provided feedback on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding The authors of this work are or were US Government employees during the development and analysis of these NIST reference materials. Financial support for the development of RM 8048 was provided by the Institute for the Advancement of Food and Nutrition Sciences (IAFNS, formerly International Life Sciences Institute North America (ILSI NA)).

Data availability RM 8231 Frozen Human Plasma Suite for Metabolomics, RM 8232 Frozen Human Urine Suite for Metabolomics, RM 8462 Frozen Human Liver Suite for Omics: https://doi.org/10.18434/mds2-3894.

RM 8048 Human Fecal Material: https://doi.org/10.18434/mds2-3633.

Declarations

Conflict of interest The authors declare no competing interests.

NIST disclaimer Certain commercial equipment, instruments, or materials are identified in this paper in order to specify the experimental procedure adequately. Such identification is not intended to imply recommendation or endorsement by the National Institute of Standards and Technology, nor is it intended to imply that the materials or equipment identified are necessarily the best available for the purpose.

References

1. Patti GJ, Yanes O, Siuzdak G. Metabolomics: the apogee of the omics trilogy. Nat Rev Mol Cell Biol. 2012;13:263–9. https://doi.org/10.1038/nrm3314.
2. Alseekh S, Aharoni A, Brotman Y, Contrepois K, D'Auria J, Ewald J, et al. Mass spectrometry-based metabolomics: a guide for annotation, quantification and best reporting practices. Nat Methods. 2021;18:747–56. https://doi.org/10.1038/s41592-021-01197-1.
3. Phinney KW, Ballihaut G, Bedner M, Benford BS, Camara JE, Christopher SJ, et al. Development of a standard reference material for metabolomics research. Anal Chem. 2013;85:11732–8. https://doi.org/10.1021/ac402689t.

4. Lippa KA, Aristizabal-Henao JJ, Beger RD, Bowden JA, Broeckling C, Beecher C, et al. Reference materials for MS-based untargeted metabolomics and lipidomics: a review by the metabolomics quality assurance and quality control consortium (mQACC). Metabolomics. 2022;18:24. https://doi.org/10.1007/s11306-021-01848-6.
5. Broadhurst D, Goodacre R, Reinke SN, Kuligowski J, Wilson ID, Lewis MR, et al. Guidelines and considerations for the use of system suitability and quality control samples in mass spectrometry assays applied in untargeted clinical metabolomic studies. Metabolomics. 2018;14:72. https://doi.org/10.1007/s11306-018-1367-3.
6. Riquelme G, Zabalegui N, Marchi P, Jones CM, Monge ME. A Python-based pipeline for preprocessing LC–MS data for untargeted metabolomics workflows. Metabolites. 2020;10:416. https://doi.org/10.3390/metabo10100416.
7. Aristizabal-Henao JJ, Jones CM, Lippa KA, Bowden JA. Nontargeted lipidomics of novel human plasma reference materials: hypertriglyceridemic, diabetic, and African-American. Anal Bioanal Chem. 2020;412:7373–80. https://doi.org/10.1007/s00216-020-02910-3.
8. Gills T, Rook H, Durst, R, Goldstein G. The national environmental specimen bank research program for sampling, storage, and analysis. 1979. EPA-600/1–79–017. https://doi.org/10.6028/NBS.IR.79-1704
9. Servetas S, Forry S, Hunter M, Dootz J, Schock T, Davis W, Bayless A, Casu F, Ellisor D, Urbas A, Da Silva S, Parratt K, Gierz K, Iyer H, Jackson S, Lin N. Characterization of human fecal material (RM 8048) from omnivore and vegetarian donors. Nat Inst Stand Technol Gaithersburg, MD. 2026. NIST SP 260–252. https://doi.org/10.6028/NIST.SP.260-252.
10. Beauchamp CR, Camara JE, Carney J, Choquette SJ, Cole KD, DeRose PC, Duewer DL, Epstein MS, Kline MC, Lippa KA, Lucon E, Molloy J, Nelson MA, Phinney KW, Polakoski M, Possolo A, Sander LC, Schiel JE, Sharpless KE, Toman B, Winchester MR, Windover D. Metrological tools for the reference materials and reference instruments of the NIST Material Measurement Laboratory. National Institute of Standards and Technology (U.S.), Gaithersburg, MD. 2021. NIST SP 260–136–2021. https://doi.org/10.6028/NIST.SP.260-136-2021.
11. Gierz K. The gut check: on stability and homogeneity methods for reference materials. J Stat Theory Pract. 2025;19:79. https://doi.org/10.1007/s42519-025-00490-y.
12. Trapmann S, Ellison SLR, Davis WC, Possolo A, Zheng J, Botha A. Approaches for the production of reference materials with qualitative properties—the new International Standard ISO 33406. Accredit Qual Assur. 2024;29:269–76. https://doi.org/10.1007/s00769-024-01602-0.
13. Sumner LW, Amberg A, Barrett D, Beale MH, Beger R, Daykin CA, et al. Proposed minimum reporting standards for chemical analysis. Metabolomics. 2007;3:211–21. https://doi.org/10.1007/s11306-007-0082-2.
14. Bowden JA, Heckert A, Ulmer CZ, Jones CM, Koelmel JP, Abdullah L, et al. Harmonizing lipidomics: NIST interlaboratory comparison exercise for lipidomics using SRM 1950–metabolites in frozen human plasma[S]. J Lipid Res. 2017;58:2275–88. https://doi.org/10.1194/jlr.M079012.