analytical chemistry

pubs.acs.org/ac Review

NMR: Unique Strengths That Enhance Modern Metabolomics Research

Arthur S. Edison,* Maxwell Colonna, Goncalo J. Gouveia, Nicole R. Holderman, Michael T. Judge, Xunan Shen, and Sicong Zhang

Cite This: https://dx.doi.org/10.1021/acs.analchem.0c04414 Read Online

ACCESS | Metrics & More | Article Recommendations

CONTENTS

Nuclear magnetic resonance (NMR) spectroscopy is an important analytical technique in metabolomics. Because it provides atomic-level detail of small molecules, NMR is well-known as an indispensable tool for unknown compound identification. In recent years, liquid chromatography–mass spectrometry (LC-MS) has become the dominant technology in metabolomics research because of its sensitivity and ability to record tens of thousands of features for each sample. The dominance of LC-MS has resulted in less visibility and use of NMR in publications, publicly deposited metabolomics data sets, and other less quantifiable ways such as fewer specific NMR metabolomics funding opportunities or fewer grant proposal reviewers with NMR expertise. We believe that this trend is largely because of a common misconception that NMR and LC-MS are largely overlapping technologies. In this review, we have focused on several unique strengths of NMR in metabolomics research and illustrate these strengths with examples from recent literature. We hope to show readers that NMR is indeed highly complementary to LC-MS and that it can significantly enhance the biological knowledge gained from metabolomics research.

Improvements in genomic sequencing technology have revolutionized science.[1] The initial human genome project[2] was a monumental effort estimated by the National Human Genome Research Institute (NHGRI) to cost $2.7 billion and require over a decade of effort by international teams of scientists. Today, the NHGRI estimates the cost of sequencing an entire human genome to be about $1000. As noted by many, metabolomics is more complicated than genomics from the perspective of analytical technology because of the large chemical diversity and dynamic nature of the metabolome.

Over the past two decades, biomedical and biology applications in metabolomics have steadily grown and show no signs of diminishing (Figure 1A). A large amount of that growth is from advances in mass spectrometry (MS),[3] especially LC-MS, which has become the dominant analytical platform used in metabolomics (Figure 1B). In addition to possessing extremely high sensitivity, the abilities of LC-MS are substantial, including measuring tens of thousands of features in one experiment, obtaining elemental formulas with

Special Issue: Fundamental and Applied Reviews in Analytical Chemistry 2021

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

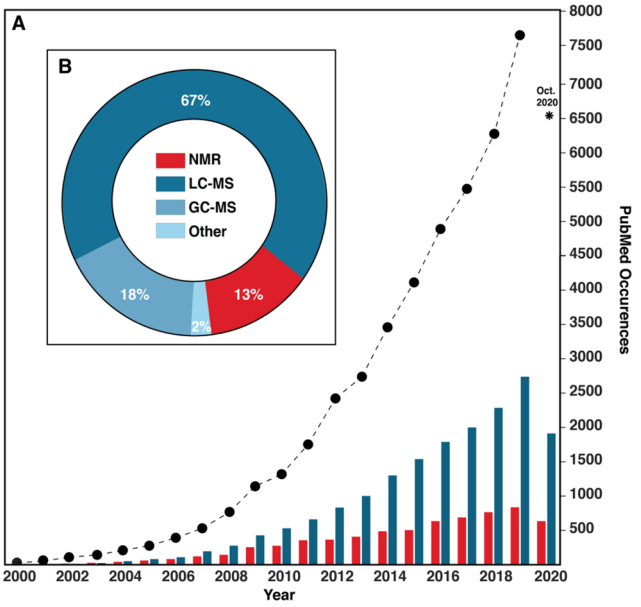

Figure 1. Overall trends. The bar chart in A shows PubMed search results for "metabolomics OR metabonomics" (black points connected by dashed lines), "metabolomics OR metabonomics AND mass spectrometry" (blue bars), and "metabolomics OR metabonomics AND NMR" (red bars). The inset pie chart B shows the current distribution of techniques used in studies deposited on the Metabolomics Workbench.[4] Both data were obtained Oct. 10, 2020.

high mass-resolution measurements, and matching metabolites to large and expanding MS databases with tandem mass spectrometry (MS/MS). These attributes, along with a highly competitive commercial landscape of MS vendors, make it an attractive choice for metabolomics studies.

What about the role of NMR in metabolomics?[5] Sensitivity is a regularly noted limitation of NMR, a spectroscopic technique with resonance frequencies that are several orders of magnitude lower than thermal energy at room temperature and standard field strengths available in most laboratories today. The Boltzmann distribution shows that only a few in 10^5 nuclear spins contribute to the signal (the majority cancel each other), but that low energy also allows for noninvasive magnetic resonance imaging (MRI), as well as nondestructive *in vivo* NMR. Further, many recent technical developments have greatly improved the sensitivity of NMR,[6] and state-of-the-art NMR probe technology can enable characterization of nanomole quantities of sample.[7] Thus, NMR is more sensitive than many people believe, but it is still a few orders of magnitude less sensitive than MS.

We routinely use both NMR and MS for metabolomics in our laboratory because they are largely complementary, especially for confident unknown compound identification.[8] While the trend in metabolomics research is clearly toward LC-MS, we have focused this review on applications that uniquely benefit from NMR spectroscopy. We are focusing on several strengths of NMR in metabolomics research (columns in Table 1), which will be organized in the review by examples of applications from the recent literature (rows in Table 1). We note that applications in biofluids from human samples are not emphasized here. These are clearly important areas of metabolomics that have numerous reviews already available. Many of the topics described below can be nicely applied to human biofluids, especially the use of fraction libraries for

Table 1. Strengths of NMR That Will Be Highlighted in Sections on Applications

	NMR strength				
Section	*In vivo* metabolism	Isotope tracing and flux	Ligand binding	Media analysis	Unknown compound ID
1) *In vivo* metabolomics	X	X		X	
2) Mammalian cell culture	X	X	X	X	
3) Metabolite-protein binding			X		
4) Model organisms	X	X		X	X
5) Marine and carbon cycling	X	X		X	X
6) Computation	X	X	X	X	X

improved ID and functional studies as well as virtually all of the topics covered under computation.

- *In vivo* NMR: MRI and magnetic resonance spectroscopy (MRS) are widespread and important in many clinical applications, but similar NMR technologies can be used to noninvasively study metabolism *in vivo*. These highly flexible methods can be applied to many different cultured systems and used to obtain data that would be difficult or impossible to collect with other techniques.
- Isotope tracing and flux: Unlike MS, which gets more complex when isotopic labels are employed, there are powerful methods in NMR to filter the signal during data acquisition to only see the isotopes and remove all other signals. This can be applied to many applications, including *in vivo* metabolomics, where isotopes can be used to investigate specific mechanisms in metabolic pathways.
- Binding studies: The dynamic nature of metabolism includes interactions between metabolites and macromolecules, especially proteins. These functional interactions are largely ignored in metabolomics research but have been routinely used in drug discovery. NMR can detect from very strong (K_d < nM) to very weak (K_d ~ μM-mM) binding affinities, and we present what we see as largely unexplored opportunities to use metabolite fraction libraries to study function.
- Media analysis: Because the sample does not touch the instrument, NMR is ideal for analysis of media sampled in real-time or in discrete time points. The quantitative and highly reproducible nature of NMR is ideal for characterizing nutrients consumed and metabolites released in cultures of microorganisms. This chemical "footprint" is simple to measure by NMR because the media needs no extraction or extensive preparation. More importantly, it can provide very useful information about the metabolic state of the culture.
- Unknown compound identification is one of the major strengths of NMR, especially when used with MS. We discuss a few examples of NMR for compound ID in genetic pathways of model organisms and marine natural products. Unknown ID and the use of fraction libraries

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

both serve to illustrate the close relationship between metabolomics and natural products discovery.[9]

- Computational methods: Whether data are collected by NMR or MS, all metabolomics studies generate a large amount of data that need to be processed and analyzed to extract biological knowledge. We end the review by summarizing recent developments in chemoinformatics and computational modeling that are especially important for NMR.

IN VIVO METABOLOMICS

Organisms inhabit different metabolic states depending on their developmental programming, internal resources, regulatory processes, and emergent dynamics (e.g., circadian clocks) and in response to changes in their environments. Single time-point measurements with little or no sample preparation allow for rapid snapshots of metabolism across multiple samples, conditions, or times. Continuous or real-time methods provide repeated measurements on the same sample over time scales with sufficient temporal resolution to observe a given dynamic phenomenon. NMR stands alone in its ability to noninvasively study the inner workings of metabolism within a wide range of living organisms due to its sensitivity, resolution, unbiased and quantitative nature, and ability to nondestructively measure complex samples with minimal modification.

Fingerprinting and the Utility of High-Resolution Magic Angle Spinning (HR-MAS) NMR Techniques. Metabolomics data are often considered the gold standard for phenotypic measurements. Thus the most basic use of *in vivo* metabolomics leverages the strengths of magnetic resonance to provide metabolic profiles (fingerprints) of samples to aid in *de novo* annotation, facilitate sample classification, and lend a scalable functional perspective[10,11] to metabolomics.

Sample preparation for *in vivo* NMR is typically quite simple; some cell suspensions can be measured directly in NMR tubes with only the addition of D_2O and a chemical shift reference.[12] However, high-resolution magic angle spinning (HR-MAS), which reduces effects of sample *inhomogeneity* (e.g., broad line shapes) for mixed-phase and solid biological samples, has gained popularity for *ex vivo* and *in vivo* metabolomics since its introduction for observing nonliving semisolids over 30 years ago.[13,14] In modern HR-MAS systems, a sample is simply placed with lock solvent and chemical shift reference in a small zirconia rotor (typically <80 μL). The rotor is then inserted into a specialized probe which tilts it to the "magic angle" (54.7° with respect to the applied magnetic field) and spins it pneumatically about its longitudinal axis.

One use of this technology is measuring regulatory and lifestyle changes, such as those which occur in the switch to pathogenicity.[15] HR-MAS provided the first *in vivo* metabolic profile of *Pseudomonas aeruginosa* to serve as a baseline for detection and identification of infections in clinical samples.[16] The lack of sample disruption in HR-MAS means clinical samples such as these can subsequently be preserved or used in downstream molecular analyses. This robust technique also provided classification of microalgae *in vivo* containing residual seawater, thus opening up metabolomics measurements for osmotically fragile marine cells in general[17] (discussed below). Additionally, it can link together pathways important to metabolic disease and aging and allows simultaneous observation of polar small molecules and lipids. Using *in vivo* NMR measurements of lipids and other metabolites in living *Drosophila melanogaster*, Righi et al. dissected relationships between injury in aging/immunodeficient flies and suggested that insulin signaling plays a role in both.[18] The group then found intramyocellular lipid components to be potential biomarkers for insulin resistance in mitochondrial flies, including unidentified peaks.[19] More recently, an exciting study by Sarou-Kanian et al. demonstrated that HR-MAS could both quantify and localize several metabolites (e.g., to reproductive organs) in living *D. melanogaster*.[20]

HR-MAS offers considerable advantages as a flexible *in vivo* metabolomics technique. However, samples can experience high acceleration while spinning.[21] While microorganisms such as yeast and bacteria tolerate and even grow in comparable hypergravity, these forces have some known effects on physiology.[22] Furthermore, these effects may be complex and difficult to distinguish from ordinary metabolic processes. Fortunately, two avenues of research are addressing this issue. First, new pulse sequences that allow for spinning speeds as low as 100 Hz have been demonstrated for intact *C. elegans* tissue and living freshwater shrimp,[23] greatly reducing the forces experienced by organisms during measurement. Second, an arrangement called high-resolution magic angle coil spinning (HR-MACS) uses a small resonator coil that is made in-house and placed inside the zirconia rotor. By reducing sample radius and spinning speeds, sub-μL yeast cell suspensions are subjected to far less acceleration with preserved data quality and improved mass-sensitivity.[24] We refer to a recent review of HR-μMAS approaches,[25] which show promise for further *in vivo* measurements of microorganisms.

Other challenges in HR-MAS based studies include water and macromolecule suppression, spinning sideband attenuation, and reduced flexibility in sample conditions. These, along with spinning speed, have been partially addressed by new 1D and 2D pulse sequences,[23,26] enabling application of HR-MAS to more sensitive specimens and extending the practical length of time-series experiments. Finally, probes with higher radio frequency (RF) power (e.g., Bruker composite multiphase (CMP) probes) have allowed comprehensive multiphase NMR measurements in freshwater shrimp which survive in the rotor for hours.[27] A hole can also be drilled in the rotor cap to allow for oxygenation with ambient air[21,27] or treatment with alternative gas mixtures.

Observing Fluxes Using Stable Isotope Labeling (SIL) Approaches. Stable isotope labeling (SIL) has historically allowed flux to be traced through the network of metabolism from a defined start point to multiple end points[28] where the label accumulates over time.[29] Most current SIL studies rely on ^{13}C,[30,31] which has a large chemical shift dispersion, ubiquity of incorporation in biomolecules, and relative low natural abundance that allows labels to be selectively observed without much background. A recent application of SIL produced a combined *in vivo* and *ex vivo* annotation of a ^{13}C-enriched water flea metabolome using several multidimensional ^{1}H and ^{13}C NMR experiments. This custom reference metabolome will be used in future studies for mapping results from real-time toxicity detection experiments.[32] The approach represents a promising general strategy of using reference data sets for future SIL studies in this nonmodel organism, underscoring the flexibility of *in vivo* NMR metabolomics. SIL approaches using NMR are reviewed in more detail below as well as in other recent reviews.[33,34] However, the utility of SIL approaches is

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

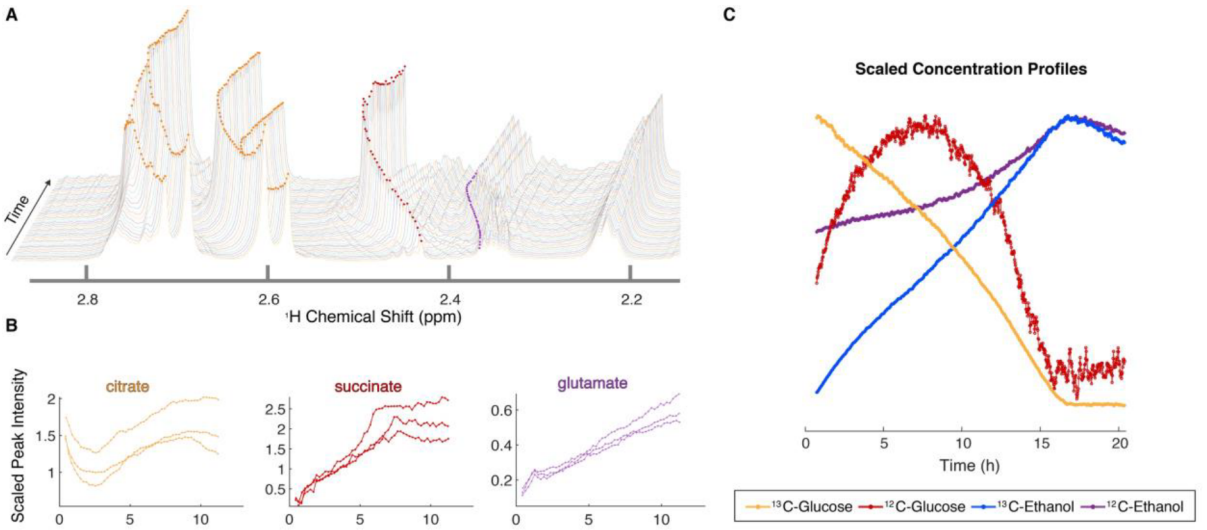

Figure 2. Continuous *in vivo* metabolism by NMR (CIVM-NMR) allows for global monitoring of known and unknown endo- and exometabolite pools in living cells or organisms with high temporal resolution. (A) Raw data can be traced using feature (peak/ridge) extraction algorithms, and (B) plots of combined peak intensity profiles for replicate samples show high reproducibility for metabolic trends. (C) Major fluxes can be observed when SIL is used to selectively monitor labeled derivatives, and differential dynamics in distinct pools of the same metabolite can be monitored in the same sample. Modified with permission from Judge, M. T.; Wu, Y.; Tayyari, F.; Hattori, A.; Glushka, J.; Ito, T.; Arnold, J.; Edison, A. S. Continuous in vivo Metabolism by NMR. *Front. Mol. Biosci.* **2019**, *6*, 26. doi: 10.3389/fmolb.2019.00026. (ref 21).

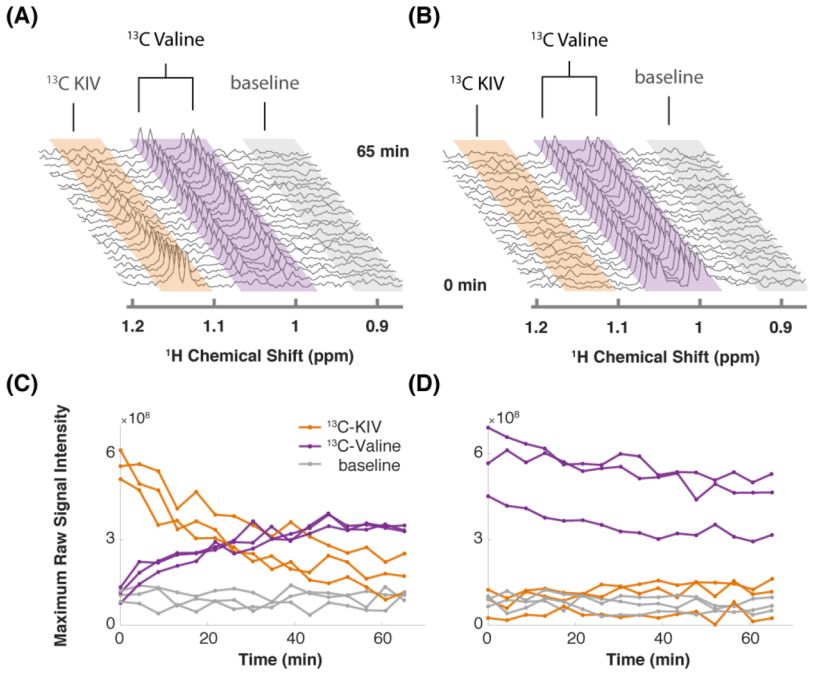

Figure 3. Targeted isotopic CIVM-NMR measurement of metabolic flux in human myeloid leukemia cells. (A) ^{13}C-labeled keto-isovalerate (KIV) was converted to valine. (B) ^{13}C-labeled valine was not converted to KIV, confirming unidirectional flux in ML cells. (C, D) Relative concentrations over time of ^{13}C-labeled KIV (orange) and ^{13}C-labeled valine (purple) compared to baseline noise (gray), obtained by taking the raw maximum spectral intensity within each region of the representative experiments in (A, B), respectively. Different lines show the data from 3 independent replicates of each experiment. Reprinted with permission from Judge, M. T.; Wu, Y.; Tayyari, F.; Hattori, A.; Glushka, J.; Ito, T.; Arnold, J.; Edison, A. S. Continuous in vivo Metabolism by NMR. *Front. Mol. Biosci.* **2019**, *6*, 26. doi: 10.3389/fmolb.2019.00026. (ref 21).

maximized when combined with real-time *in vivo* measurement (Figure 2).

Real-Time Metabolism Measurements. The gold standard for observing metabolism at work is a dynamic, real-time measurement. Because NMR excels at quickly and non-intrusively measuring metabolites in relatively small volumes, flow NMR systems were adapted long ago for monitoring cell cultures and conditioned media.[35] In these systems, a liquid sample (such as culture media or cell suspension) is pumped through the magnet bore into the probe in a closed or open loop and can be recycled or discarded after measurement. Alternatively, cells held within the probe can be continuously replenished with fresh media. Of particular interest, commercially available benchtop flow NMR systems extend the versatility of NMR for real-time monitoring of liquid reactions[36] while greatly reducing overhead and will be useful

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

for problems like media optimization and culture control.[37] Lastly, flow NMR is not limited to single-celled organisms as multicellular aquatic organisms were recently sustained in-probe with media flow to do real-time metabolic flux monitoring.[38]

For cell suspensions, simpler NMR setups can be used. Koczula et al. recently used solution ^{1}H NMR to collect time series spectra on agar-embedded chronic lymphoid leukemia cells, allowing them to observe real-time metabolome shifts without sedimentation.[39] Select peaks were traced using an in-house tool,[40] and pH was estimated by a known internal standard.[39] For more complex samples, such as the multicellular fungus *Neurospora crassa*, HR-MAS can be used for continuous *in vivo* monitoring of metabolism by NMR (CIVM-NMR). Judge et al. used CIVM-NMR to collect data (Figure 2, Panels A−B) with high temporal resolution (up to ∼30 s) under different oxygenation conditions without needing specialized flow NMR equipment and with minimal sample preparation.[21] More recently, *C. elegans* metabolism was monitored in real time using a custom NMR tube insert that separated worms from the D_2O lock solvent.[41] The approach used in this study offers the benefits of ease of application and high-quality real-time data without requiring an HR-MAS probe. In both this study and CIVM-NMR, features were extracted for both known and unknown peaks (Figure 2, Panels A−B).[21,41]

One particular strength of NMR is the unique ability to noninvasively conduct selective or parallel measurements of distinct pools of metabolites *in vivo*. Reed et al. recently used ^{13}C-edited NMR to track real-time incorporation of isotopically labeled glucose into multiple myeloma cells as well as define important considerations for SIL experiments.[42] Likewise, Judge et al. monitored fluxes derived from ^{13}C-labeled glucose in *Neurospora*, and protons on both labeled and unlabeled carbons were measured simultaneously to reveal unique dynamics for different glucose pools in the same organism (Figure 2, Panel C). In the same study, specific amino acid fluxes were measured in living cancer cells (Figure 3).[21]

Finally, while *in vivo* systems such as these are desirable, well-mixed cell cultures can be effectively sampled frequently with small volumes to yield very similar dynamic information. By injecting bacterial, yeast, or mammalian cell cultures into a mass spectrometer after in-line extraction, Link et al. were able to effectively track the response of bacterial cultures from starvation into feeding.[43] This study also highlights the need for modeling, as the high-density dynamic data generated from the technically impressive platform were used to fit kinetic models to interpret amino acid biosynthesis pathway dynamics.[43] Hyperpolarization NMR experiments, while challenging to implement and optimize, also produce data of similar temporal resolution for specific reactions.[34,44]

In Vivo Data Analysis. New approaches to data analysis are emerging for real-time metabolomics, as the ability to collect large amounts of real-time *in vivo* data currently outpaces our ability to extract and interpret information from it. In particular, advanced feature extraction tools and kinetic models trained on these new data are needed. Commercially available and proprietary software is limited for the analysis of extracted continuous NMR systems. MetaboLab from the Günther group[39] and in-house scripts from our own group[21] have been used for real-time *in vivo* feature tracing (Figure 2A) and extraction (Figure 2B). More recently, a computer vision-based tool called RTExtract has been developed for extracting peak intensities from NMR data with continuous changes in peak location and intensity, including those from CIVM-NMR or flow NMR experiments.[45] This tool greatly expands the potential of time-series *in vivo* NMR experiments by simplifying and expediting the feature extraction process as well as improving the capability to track overlapping peaks. This is exciting, as continuous spectral measurements on a system as quantitative and stable as NMR offer several advantages in solving deconvolution and alignment problems.[21,45] Ultimately, automated nonparametric feature extraction flexible enough for different measurement intervals and experimental formats is desirable. Improved approaches for spectral deconvolution are also needed, particularly for HR-MAS probes, which inherently have broader line widths that can mask coupling. Because real-time data are typically 1D, annotation is still a challenge. Faster 2D pulse sequences, which are more amenable to real-time measurements, would aid in annotation and resolution. Likewise, improved *de novo* annotation strategies could be paired with streamlined approaches for mapping peak annotations from extracted data to *in vivo* data to better leverage the benefits of both data types. Finally, integration of real-time data requires detailed kinetic models,[46] which do not make steady-state assumptions and are flexible with experimental format.

METABOLOMICS APPLICATIONS IN MAMMALIAN CELL CULTURES

Cell lines are advantageous for their relative accessibility, low cost, ease of manipulation, and experimental control. As the interest in cancer metabolism has increased, so has the application of metabolomics in cancer cell models. More recent developments in cell-based therapeutics have also created new opportunities for discovery by NMR. Here, we outline some recent work leveraging the unique capabilities of NMR to measure intra- and extracellular metabolites as well as specific metabolites and pathways implicated in disease or function in cultured mammalian cell models.

Targeted Metabolism in Mammalian Cells. There are several unique features of NMR that allow for targeted analysis of specific metabolite classes or pathways of interest to cancer metabolism with the use of cell lines. This is a particularly powerful capability when advancing from common profiling or screening studies to understanding mechanisms or changes in metabolic flux that produce observed changes.

For example, coenzyme A species, redox metabolites such as NAD+ and NADH, as well as energy molecules like ATP are particularly useful for understanding cancer metabolism *in vitro*. However, these classes of molecules are difficult to measure with MS-based techniques due to their highly labile nature and structural similarity. These are challenging for NMR as well due to their low concentrations and aforementioned structural similarity which can produce high spectral overlap. Recently, Nagana Gowda et al. have optimized extraction and sample preparation techniques that allow for preservation of these endogenous metabolites. A one-time addition of coenzyme standards to a reference sample was sufficient to identify and quantify signals from unique species using standard ^{1}H NMR across multiple samples.[47] Similarly, utilizing a combination of standard compound spiking and 2D correlation experiments, Nagana Gowda was also able to definitively identify and quantify redox coenzymes and adenosine phosphate species from extracted mammalian

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

cells.[48] These methods highlight how, when combined with robust chemical extraction methods, the reproducibility and stability of metabolite chemical shifts in an NMR spectrum can enable comprehensive profiling of coenzymes and energy molecules with a single experiment.

^{13}C-labeled substrates are useful for tracking the flux of carbon sources and understanding the unique metabolism of different cancer types. While using SIL is not unique to NMR, the ability to directly and selectively detect molecules and the positions of atoms containing tracers such as ^{13}C is unparalleled. NMR provides atom-specific information, making it ideal for isotopomer analysis. Lane et al.[49] provide an excellent example of applying both NMR and LC-MS for investigating cancer energetics using both ^{13}C- and ^{15}N-labeled glucose, glutamine, glycerol, and octanoate to profile differential nutrient utilization between breast cancer cell lines of different histological subtypes. By using ^{1}H-detected 1D ^{13}C-HSQC (heteronuclear single quantum correlation) experiments, they compared NMR J-couplings at different nuclei in metabolites labeled by different isotopic substrates and were able to deduce the specific pathways used to metabolize these carbon sources. Combined with information from 2D ^{1}H-TOCSY (total correlation spectroscopy) experiments, they were able to quantify relative amounts of ribose contained in nucleotides generated by oxidative vs nonoxidative pentose phosphate pathways, among other insights.[49] A review of the stable isotope resolved methods utilized by Lane was recently published by their group.[50] Winnike et al. also used ^{13}C-labeled glucose and glutamine to determine the relative flux of these nutrients in breast cancer cells using both directly detected ^{13}C 1D experiments as well as 2D ^{13}C-HSQC for metabolite annotation. ^{13}C detection provides greater chemical shift dispersion and less overlap than ^{1}H detected data.[30,51] Importantly, these studies and others have revealed that the typical classifications of breast cancer cells, such as proliferation rate or histological subtypes, do not necessarily predict metabolic pathway activity.

To probe metabolism even more specifically at the enzyme level, Hattori et al.[52] utilized a variety of 1D and 2D NMR experiments on cell extracts labeled with either valine or its ketoacid, keto-isovalerate (KIV). To determine the directionality of the transamination reaction, which is catalyzed by branched-chain amino acid aminotransferase 1 (BCAT1), they either added ^{13}C-valine with natural abundance KIV or ^{13}C-KIV with natural abundance valine. Similar experiments were done with ^{15}N to follow the amino group. These data showed that leukemia cells preferentially transaminate branched-chain keto acids to their respective branched-chain amino acids, uncovering a novel behavior of leukemia cells shown to enhance their malignancy.[52] Judge et al. were able to use CIVM-NMR to reproduce this result using the same matched pairs of substrates in live cells utilizing a continuous 1D ^{13}C-HSQC experiment to detect the protons that were connected to labeled carbons, giving a simple and direct real-time display of KIV turnover (Figure 3).[21]

While ^{1}H and ^{13}C atoms are the most commonly used nuclei for profiling metabolites, there are other nuclei that can be leveraged for targeted analysis of metabolism in cancer cells by NMR.

For instance, phosphocholine and phosphoethanolamine related molecules, which have been previously observed to be significant in cancer studies by NMR,[53,54] can be detected directly using ^{31}P NMR. Similar to ^{1}H, ^{31}P is an NMR active isotope and occurs at 100% natural abundance, which provides higher sensitivity than ^{13}C. There are fewer phosphorus resonances comprising a typical biological sample, resulting in a less crowded spectrum. Juranic and co-workers have developed useful NMR-based methods to characterize high-energy phosphometabolites like ATP.[55] Not only do they take advantage of ^{31}P NMR, but they also label samples with added $H_2{}^{18}O$. The addition of ^{18}O is indirectly detectable through isotope effects that manifest on the ^{31}P nuclei, allowing for elucidation of valuable functional information in perfused tissues. Shah et al. demonstrated the utility of ^{31}P NMR in capturing the dynamics of different phosphoethanolamine species across cancerous and nonmalignant cell lines, revealing differential dependence of cancer cells on phospholipid synthesis when biosynthetic genes were knocked out.[56] Veronesi et al.[57] used direct ^{19}F detection of fluorine-labeled substrates and their enzymatic products to monitor the activity of a specific enzyme in living cells. Similar to ^{31}P, ^{19}F is an NMR active isotope that occurs at 100% natural abundance, with essentially no background resonances in most biological samples. This allows tracking the fate of ^{19}F tracer molecules without signals from endogenous compounds. With their system, Veronesi et al. were able to quantify changes in fatty acid amide hydrolase activity upon treatment with inhibitors.[57] As the first example of this type of quantitative kinetic data obtained in intact cells, ^{19}F has applications for both cancer drug screening and targeted metabolism studies.

Media Analysis of Cancer Cell Cultures. The analysis of extracellular metabolites in culture media is important to the study of cancer cell metabolism *in vitro*. NMR is well suited for this due to the minimal sample preparation needed, which allows aliquots of culture media to be analyzed via NMR directly with the addition of a chemical shift reference in 5−10% D_2O. Complementing the intracellular metabolome with changes in the extracellular environment provides greater context for interpreting the results of metabolomics studies. In addition, the use of small diameter, low volume sample tubes with a high-sensitivity small volume probe enables sampling of media from the same culture repeatedly over time for time-course data.

Recently, Wojtowicz et al.[58] performed a time-course NMR analysis of media in the culture of breast cancer cells and compared the media profiles to NMR profiles of serum from breast cancer patients, to see if direct comparisons could reasonably be made between them when performing *in vitro* studies. Wojtowicz and colleagues collected media samples from breast cancer cell cultures 16 times over the course of 72 h, revealing dynamic, nonmonotonic changes in several detected metabolites. Lactate, alanine, glutamine, tyrosine, and glucose profiles in the culture media showed opposite trends of accumulation or depletion compared to changes in patient serum vs healthy controls, thus exhibiting some key limitations of direct comparisons of media with serum. However, the authors noted that for many metabolites, the interpretation of the results changed drastically depending on which time point was examined.[58] This again implicates the importance of collecting dyanmic measurements at biologically relevant resolution to contextualize and properly interpret metabolomics data.

Mahar and colleagues recently showed the utility of combining isotopomer analysis with conditioned media analysis by NMR to quantify the Warburg effect, a metabolic signature of cancer cells where most glucose is converted to

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

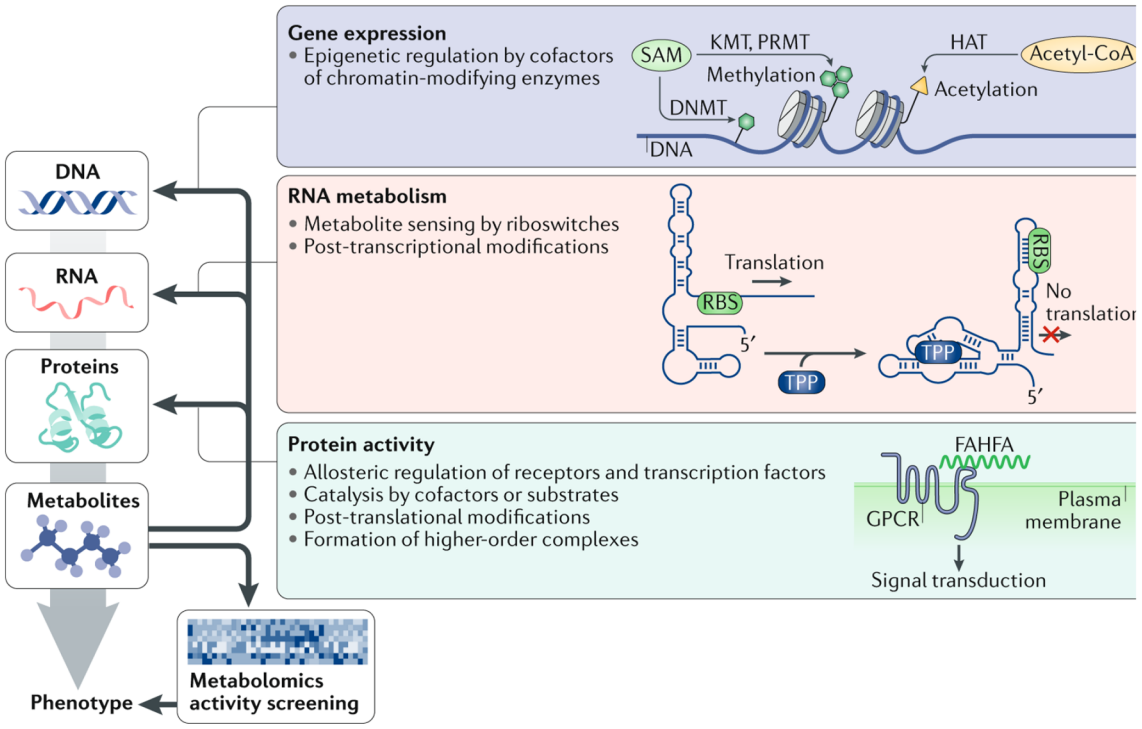

Figure 4. Metabolites as active participants of metabolism in all levels of the central dogma. Metabolites affect and regulate protein function in a variety of ways. They also influence transcription and RNA metabolism and regulate post-transcriptional modifications. At the uppermost level, metabolites as cofactors and cosubstrates are involved in epigenetic regulation. To uncover these relationships and interactions, novel strategies are needed to define metabolites function/activity and reveal their impact in the regulation of phenotypes. DNMT, DNA methyltransferase; FAHFA, fatty acid ester of hydroxyl fatty acid; GPCR, G protein- coupled receptor; HAT, histone acetyltransferase; KMT, lysine methyltransferase; PRMT, peptidyl-arginine methyltransferase; RBS, ribosome binding site; SAM, S-adenosylmethionine; TPP, thiamine pyrophosphate. Reprinted by permission from Macmillan Publishers Ltd.: NATURE REVIEWS MOLECULAR CELL BIOLOGY, Rinschen, M. M.; Ivanisevic, J.; Giera, M.; Siuzdak, G. *Nature Reviews Molecular Cell Biology* **2019**, *20*, 353−367. Copyright 2019. (ref 79).

lactate instead of pyruvate.[59] Both normal and cancer cell lines were cultured in media containing [$^2H_7$]glucose, and the media were sampled over the course of 5 h. Due to the isotopic effect of deuterium incorporation, they were able to use ^2H decoupled ^1H NMR to resolve and track the accumulation of lactate isotopomers excreted into the media over the culture period, after correcting for changes in T_1 relaxation, another isotopic effect. They showed not only that the cancer cell line consumed almost six-times the amount of glucose compared to normal liver cells, but also nearly all of the [$^2H_7$]glucose consumed by the cancer cells could be accounted for by glycolysis-produced lactate and monodeuterated water.[59] This study illustrates the power of meauring stable isotopes in conditioned media to derive definitive flux measurements through pathways enabled by the quanititative nature of NMR.

A unique twist on the analysis of conditioned media is to use it to simulate an *in vivo* environment for cancer cells. Luis et al. used conditioned media from cultured adipocytes to expose MCF-7 cells to an environment simulating an *in vivo* obesity phenotype. ^1H NMR analysis of the conditioned media before and after culturing MCF-7 cells showed an "inversion" of the Warburg effect as evidenced by increased glycolytic intermediates in the culture medium, along with increased proliferation and invasiveness characteristics of the cells.[60] The straightforward sampling and preparation of media for NMR analysis enabled the authors to conclude that the conditioned media metabolite profiles may be relevant to *in vivo* disease and have provided more evidence for how and why breast cancer prognoses are worse in obese women.

While the methods and approaches for interrogating mammalian (cancer) cell metabolism *in vitro* continue to become more advanced and comprehensive, the need for more physiologically relevant culture models amenable to these techniques is apparent. Ultimately, the application of these techniques to more complex models such as tissue-on-a-chip, coculture, or organoid systems will be able to provide metabolic discoveries with more relevance to disease and physiology.

Mammalian Cell-Based Manufacturing. Aside from basic biology research, another significant application of mammalian cell culture is in industrial production of biotherapeutics. While the industry of cell expressed biologics continues to grow and diversify,[61] a new generation of cells-as-therapies is also on the horizon. Here we will describe recent uses of NMR for mammalian cell-based biologics manufacturing in existing industrial processes and the future of NMR as a potential key technology for manufacturing of cell therapies.

The use of NMR, metabolomics, and other techniques for mammalian cell production of biologics manufacturing is not new,[36,62,63] but they have recently gained popularity for their utility in cell culture engineering and prediction of product quality. Recent work by Ali et al. and Zürcher et al. has highlighted the use of metabolomics data collected during cell culture bioprocesses to predict and improve the quality of products produced in cells.[64,65] However, there are some unique applications of NMR technology to improving cell-based bioprocesses. Brinson and Marino recently demonstrated the use of 2D NMR spectroscopy to provide a high-

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

order structural fingerprint of cell-expressed protein products.[66] Since three-dimensional structure is critical to protein function, these NMR signatures can be used as metrics to assess product quality. This application is an example of the utility and versatility of structural information that is provided by NMR in cell manufacturing.

In addition, Blondeel et al. used time-course ^{1}H NMR metabolomics of culture media in cells expressing recombinant proteins in order to identify metabolites that were rapidly consumed, thus limiting cell density and production. By supplementing additional nutrients observed by NMR to be rapidly depleted, they were able to achieve a nearly 75% increase in cell density.[67] As mentioned elsewhere, the recurring measurement of metabolites in culture media is uniquely suited to NMR due to the simplicity of sample preparation and inherently quantitative nature of measurements.

As an evolution from cell expression-based production of biologics, cell therapies (such as CAR-T cells, mesenchymal stromal cells, and stem cell therapies) are poised to become a significant sector of biopharmaceutical manufacturing.[68−70] Since cell therapies are even more complicated than the expression of biologics, cell products require increasingly comprehensive characterization and optimization for manufacture. Recent studies have already demonstrated the utility of metabolomics measurements to characterize and improve growth of cell therapy products.[71−74] As with cell expression systems described previously, optimized, rationally designed media formulations are also needed for effective cell therapy manufacturing and can be aided by NMR metabolomics in similar ways.[75] ^{1}H NMR was recently used by Agostini et al. to characterize and predict the quality of platelet-derived media supplements for industrial cell manufacturing, leveraging the ease of sample preparation and throughput of media analysis by NMR.[76] Continuous flow NMR techniques for monitoring mammalian cell culture systems have existed for decades[35] but have not yet become a central technology in cell manufacturing. However, NMR could fill the need for a nonintrusive, high information content, online monitoring technique that can be leveraged to predict and improve product quality.

METABOLITE−PROTEIN INTERACTIONS

Metabolites play an active role in metabolism, contributing to regulation at all levels of the metabolic central dogma and serving as direct modulators of biological processes and phenotypes (Figure 4).[77−79] However, current experimental strategies to understand how metabolites interact with macromolecules are still limited. Novel experimental approaches are slowly emerging, and most rely on MS-based indirect measurements.[77,79−84] NMR spectroscopy has a long history of contributions to protein−ligand research[85] and has been widely used in the pharmaceutical industry for lead drug discovery.[86] A recent review by Becker et al.[85] gives an overview of the large repertoire of techniques available. NMR is capable of measuring weak- or tight-binding interactions. Weak binding generally translates to fast exchange, which can be measured by ligand-detected NMR experiments like saturation-transfer difference and WaterLOGSY. These techniques have been used in drug discovery to screen natural products and other libraries against drug targets.[81,82,85] They can detect equilibrium dissociation constants (K_d) greater than μM. This approach was also used by Waller and co-workers, who were able to show that a novel Fe/S enzyme called YgfZ from *E. coli* (but found in all domains of life) was able to bind to a folate ligand. This binding was predicted based on mutation analysis, but the binding was too weak to be detected by other methods.[87]

Despite the fact that ligand-binding studies can be relatively high-throughput by using the same automation technology already in place for NMR metabolomics (autosamplers, racks of 96 tubes, autoshimming/tuning, etc.), not many applications have taken advantage of this powerful technology to functionally characterize physiologically relevant metabolites with protein interactions in a biochemical model.[83] The rationale and methodology for utilizing this systematic approach was well laid out by Nikolaev and colleagues.[82] The authors used magnetization transfer NMR experiments to suppress protein resonances and identify changes in metabolite features that corresponded to metabolite-protein binding. Four well-characterized bacterial and mammalian proteins were assayed against a panel of 33 synthetic metabolite mixtures, detecting all of the well-known interactions and uncovering novel metabolite−protein interactions previously undescribed.[82] An exciting future prospect of functional metabolomics is the extension of NMR metabolomics workflows to incorporate these ligand-detected NMR experiments. The output from metabolomics studies are physiologically relevant, metabolite rich, and matrix free extracts that reach an end point after analysis because NMR is nondestructive and nonintrusive. However, spectral overlap of chemically similar compounds in these extracts can lead to ambiguous assignment of potential protein−metabolite interactions.[85]

Whiley et al.[88] have developed a protocol that incorporates semipreparative fractionation as a complement to metabolomics studies. Parallels can be drawn to natural products chemistry. Routinely used chromatographic fractionation and fraction concentration simplify and concentrate complex extracts for downstream analysis.[80,89,90] The authors created a library of elutants (fractions) collected at distinct time intervals from a chromatography system. The fractions can be concentrated to bridge the sensitivity gap between NMR and LC-MS, as well as to reduce spectral overlap, thus improving the process of comprehensive metabolite annotation. The authors demonstrate that the combination of NMR and MS could be used to identify several unknown metabolites in human urine. Furthermore, this method paves the way to create valuable libraries for ligand-detected NMR experiments. NMR metabolomics extracts or reference materials can be similarly transformed into libraries that circumvent the need for synthetic standards and become screening substrates for previously undiscovered metabolite interactions with protein(s) of interest. These methods can serve as building blocks for metabolomics-based protein interaction studies, with the promise of increased throughput not only for metabolite identification but also to better define their respective biological function.

MODEL MICROORGANISMS AND METABOLISM

Investigators have used model microorganisms for the characterization of metabolism throughout the history of biochemistry research.[91,92] Biochemistry experiments have long relied on their short generation times, easy manipulation, and fast response to systematic perturbations of their metabolism. Microbiologists have developed powerful tools to manipulate micro-organisms, and these tools are useful in metabolomics studies.[91,93] These model organisms are

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

particularly well suited for studying functional interactions, as described above (Figure 4).[77,94]

Classical Biochemistry and Metabolomics. Classical genetics and biochemistry experiments define testable models of detailed biochemical reactions.[95] However, classical experiments are unable to characterize the overall metabolic network. Metabolomics can provide a global view of the metabolic network but often falls short in defining function or mechanism.[77,91] Here, we outline examples of experimental approaches that highlight the complementary integration of metabolomics and classical genetics/biochemistry with the unique capabilities of NMR.

The use of NMR metabolomics with classical biochemistry and genetics is illustrated by the RidA paradigm of enamine stress in the model organism *Salmonella enterica*. Borchert and co-workers used NMR metabolomics and transcriptomics to better understand how the absence of RidA (a protein responsible for the hydrolysis of enamine/imine species) causes detrimental accumulation of the enamine 2-aminoacrylate (2AA).[96] The omics analyses revealed that the *ridA* mutation caused global metabolic changes in *S. enterica.* In particular, folate and branched-chain amino acid metabolism pathways were disrupted.[96] Additional work on the *ridA* biochemical model capitalized on these findings in a nutrient supplementation experiment. NMR measurements of chemically defined spent minimal media, spent minimal media supplemented with isoleucine, and the resulting cell pellets demonstrated that key metabolite changes between *ridA* and wild-type strains were due to the IlvA-dependent generation of 2AA. Further media supplementations circumvented the downstream pathway damage to GlyA, largely restoring those same endogenous and exogenous metabolites to wild-type levels.[97] This classical genetic perturbation and metabolome rescue approach, and the agreement between media analysis and bacterial endometabolome, disentangled the complex metabolic relationships between enzymes and further expanded the interacting edges of the known *ridA* biochemical model to previously unexplored pathways.[96,97]

Genetics and Metabolomics. Genetically modifying model organisms to target specific genes and protein complexes is a well-established approach. This introduces disruptions that facilitate detailed mechanistic insights otherwise difficult to reach.[98] Without *a priori* knowledge, investigators can use NMR to quantify known and unexpected metabolite changes in disrupted pathways and in pathways that were not previously thought to be connected. Furthermore, these metabolites can be known or novel, and both can be quantified with or without the need for chemical standards.[97,99−101] This versatility has been illustrated by Marshall et al.[99] by inactivating *alr* (an alanine racemase) to determine its effects on the cellular metabolism of *Mycobacterium smegmatis.* NMR profiles resolved a much-debated controversy over the role of *alr* in D-alanine biosynthesis. The authors used SIL by supplementing cultures with ^{13}C D-alanine and recording 2D ^{13}C-HSQC experiments. They were able to quantify 38 metabolites that were produced from the ^{13}C D-alanine carbon source, including several that changed significantly with an *alr* mutant. These data allowed them to uncover an alternate biosynthetic pathway and to conclude that inhibitors of the Alr protein would not be bacteriocidal.[99]

In addition to gene knockout and knockdown methods, overexpression studies of target proteins (or recombinant proteins) can use metabolomics as a powerful tool to probe into these overactive pathways. Modern NMR spectrometers have an extremely large dynamic range, and it is possible to achieve over 2 million to 1 in dynamic range, depending on many factors, including shimming and water suppression to minimize radiation damping (personal communication, Dr. Clemens Anklin, Bruker Biospin). This large dynamic range allows for reliable quantitative measurements of metabolites over a wide range of concentrations. In a study using genetically edited yeast to overexpress the human oncoprotein NSD3, Rona and co-workers used NMR to quantify small, but statistically significant, differences in valine, glutamate, and phosphocholine between NSD3 and the structurally similar overexpressed yeast Pdp3.[102] No significant differences were observed for aspartate and arginine, both with low concentrations, a finding that is consistent in light of the structural similarity of the two proteins. Yet, together with alanine, these metabolites were significantly different from the *Saccharomyces cerevisiae* wild-type strain and consistent with cancer phenotypes. This suggested that despite the near identical metabolic profile of these modified organisms, phosphocholine, valine, and glutamate indicate distinct pathways that ultimately recruit different downstream signaling complexes.[102]

Microbial protein production systems rely heavily on overexpression methods to produce large quantities of proteins of interest.[103] These can be hard to optimize or troubleshoot when yields are lacking or no protein is produced. Chae et al.[103] hypothesized that NMR metabolomics paired with isotopic labeling could identify metabolite profiles that characterize optimal protein production conditions. They collected 2D NMR on 71 *Escherichia coli* cultures, each overexpressing a different gene, and were able to identify 17 metabolites that reflected the optimal conditions for protein expression. These metabolites, with changes between groups ranging from an order of magnitude to as low as 0.2 fold, were used to reduce the number of trial-and-error iterations necessary to achieve high protein yields and set indicators for external stresses that could modulate metabolism to an optimal protein production environment.[91,102,103]

Role of *Caenorhabditis elegans* in Metabolomics. *C. elegans* has been a powerful model organism for nearly half a century.[104] Initially the focus of genetics and developmental biology, *C. elegans* has become an important organism in metabolomics research. There are two main advantages of *C. elegans*: a large number of available mutations of known pathways[105] and the ability to culture large amounts of material with the option of isotopic labeling. Clendinen and co-workers developed a method to uniformly label *C. elegans* with ^{13}C and performed 2D NMR INADEQUATE (Incredible Natural Abundance DoublE QUAntum Transfer Experiment) experiments on these samples to obtain covalent fragment information for all abundant metabolites.[106] They were able to match some of these fragments to an *in silico* database of INADEQUATE spectra that was created from assigned ^{13}C 1D NMR spectra in the Biological Magnetic Resonance Bank (BMRB) database. One drawback to uniform ^{13}C labeling is the large directly bonded ^{13}C couplings ($^{1}J_{CC}$ ∼ 30−75 Hz), which cause overlap and reduce sensitivity. Geier et al. showed that a constant-time 2D ^{13}C-HSQC experiment—which decouples ^{13}C from itself in the indirect ^{13}C dimension—can greatly improve the spectral quality in uniformly labeled ^{13}C-labeled *C. elegans.*[107] Using this approach, they were able to more effectively match their data to spectral databases,

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

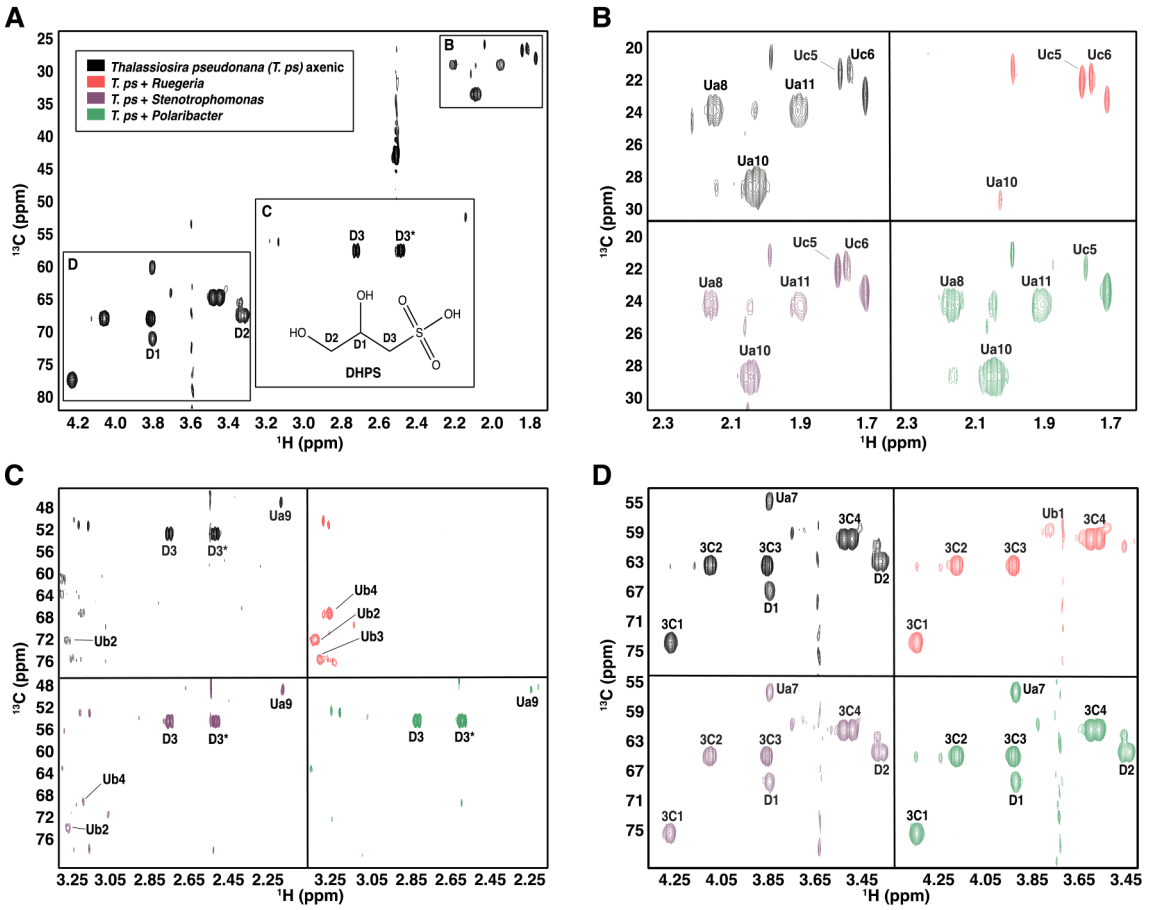

Figure 5. ^{13}C-HSQC spectra reveal differential metabolite uptake and secretion in axenic phytoplankton *Thalassiosira pseudonana* (*T. ps*) and three marine bacteria cocultures with *T. ps*. (A) Full representative spectra from *T. ps* with insets B, C, D, and E. Panel colors indicate coculture identity (black—axenic *T. ps*; red—*T. ps* + *Ruegeria*; purple—*T. ps* + *Stenotophomonas*; green—*T. ps* + *Polaribacter*). The structure of DHPS is shown with numbered peaks (D1, D2, and D3/D3*; * indicates a C−H bond in DHPS that lies within T1 noise). (B−E) Detailed NMR signals of 18 features in axenic and coculture spent media. Spectra from one of 3 representative replicate samples are shown. Figure adapted with permission from ref 118 under a Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

including the very useful COLMAR database (complex mixtures by magnetic resonance).[108]

Research in *C. elegans* also nicely illustrates the complementary roles of NMR and MS in unknown compound identification. Since the 1970s, geneticists knew about a pheromone that causes *C. elegans* to develop into the dauer stage when conditions are harsh and food scarce.[109] Nearly 30 years later, the chemical identity of the dauer[110] and related mating pheromones was identified using 2D NMR and LC-MS.[111] The Schroeder group then developed an NMR-based approach called differential analysis by 2D NMR spectroscopy (DANS) that compared 2D COSY NMR data between the reference strain N2 and a genetic mutant *daf-22* that led to the identification and *de novo* annotation of three known and four previously undescribed pheromones from the same family called ascarosides.[112] Using the structural information obtained by NMR, the Schroeder lab developed an LC-MS/MS assay highly selective for ascarosides that monitored for the neutral loss of 73 amu.[113] This assay was only possible after the initial structural characterization by 2D NMR, leading to the discovery of hundreds of additional ascarosides and related signaling metabolites.[114] Similarly, Shou et al. used both NMR and LC-MS to discover a novel class of small endogenous peptides, nemamides, found to promote *C. elegans* survival during starvation-induced larval arrest.[115]

MARINE ENVIRONMENTS AND CARBON CYCLING

Phytoplankton perform half of all photosynthesis on Earth, generating 50% of the oxygen and contributing nearly half of the turnover in the global carbon cycle. A quarter of the carbon fixed by phytoplankton becomes part of the metabolite pool of small molecules known as dissolved organic matter (DOM). DOM is an assemblage of organic metabolites that play vital roles in global carbon, nitrogen, sulfur, and phosphorus cycles; metabolism of nutrients and xenobiotics; and marine biodiversity.[116] Crucially, DOM is scavenged by mutualistic marine bacteria in loose association with phytoplankton and is altered *via* bacterial catabolism into numerous metabolites as part of the microbial loop. This loop cycles transformed nutrients back into all trophic layers of the marine food web and is a crucial step in elemental cycling. Characterizing the identity of metabolites released by phytoplankton and transformed by bacteria is crucial to understanding the dynamism of reactivity, metabolic turnover, and chemical transformation essential to supporting life.

Marine Exometabolome. The high salt content required for culturing marine microorganisms or for analyzing natural

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

populations in seawater interferes with both MS and NMR measurements.[117] Additionally, the concentration of DOM is in the nanomolar to micromolar range, so it must be concentrated for NMR, but the salt is also concentrated. This step is typically followed by membrane-based or solid phase extraction (SPE) to remove salts from the solution, with some notable exceptions, but in this situation SPE does not effectively work for smaller metabolites.

A challenge of detecting LMW (low molecular weight, < 1 kDa) metabolites from seawater media is that the smallest and most polar metabolites behave like salt. They are often difficult to separate from the high-salt media background, particularly at such low concentrations. Ferrer-Gonzalez et al. developed a novel method to extract such LMW metabolites from the coculture exudates of bacteria *Ruegeria pomeroyi* and phytoplankton *Thalassiosira pseudonana*, two marine microorganisms commonly shown in ecological association.[118] *R. pomeroyi*—*T. pseudonana* cocultures demonstrate bacterial upregulation of several genes linked to the transport and catabolism of dihydroxypropanesulfonate (DHPS), a small metabolite rich in carbon and sulfur.[119] *T. pseudonana* releases large quantities of DHPS, a concentrated source of DOM, which *R. pomeroyi* is able to consume as a sole carbon source. Three amplified genes (*hpsK, hpsL, and hpsM*) were hypothesized to form a tripartite ATP-independent periplasmic (TRAP) transporter for import of DHPS into the cytoplasm.[119] To verify that DHPS was significant to *R. pomeroyi* carbon exchange, methods were developed for extracting and analyzing ^{13}C-labeled spent media from *R. pomeroyi*—*T. pseudonana* cocultures via NMR using lyophilization, reconstitution in DMSO-d6, and analysis in a 1.7 mm cryogenic probe at 800 MHz. The smaller diameter probe enhances mass sensitivity and minimizes loss due to salt. The group developed an *R. pomeroyi* transporter mutant (Δ*hpsKLM*) in which the putative transporter genes were knocked out to determine whether DHPS was significant to carbon exchange in an *R. pomeroyi*—*T. pseudonana* coculture. ^{13}C-HSQC was used to analyze extracted spent media from axenic *T. pseudonana, R. pomeroyi*—*T. pseudonana*, and Δ*hpsKLM*—*T. pseudonana* coculture pairs. ^{13}C-Labeled DHPS was absent in wild-type samples, indicating the metabolite was taken up by *R. pomeroyi*. ^{13}C-Labeled DHPS was present in both the axenic *T. pseudonana* and the Δ*hpsKLM* coculture import mutant of DHPS. The extraction method was later applied to expand knowledge of metabolic interactions between ocean microbes across a range of bacterial taxa commonly shown in ecological association with phytoplankton. Spent media from axenic *T. pseudonana* and coculture pairs of *T. pseudonana* with marine bacteria *R. pomeroyi, Stenotrophomonas sp.*, and *Polaribacter sp.* were investigated via 2D ^{13}C-HSQC. The 2D helped by reducing overlap and also took advantage of the uniform ^{13}C-labeling of the cocultures. These analyses of extracted spent media demonstrated that metabolic interactions between phytoplankton and marine bacteria generate niche-specific fingerprints that differentiate cocultures, revealing differential patterns of uptake and enhancement of certain DOM species between coculture groups versus axenic *T. pseudonana* (Figure 5).[118]

Despite the challenges with low concentrations, direct ^{1}H NMR spectra of DOM in natural, unaltered samples are possible. Efficient water suppression is critical, so Lam et al. combined well-known sequences to take advantage of presaturation, water dephasing, and peak selectivity into an innovative sequence: SPR-W5-Watergate. SPR (shaped pre-irradiation) offers a greater magnitude of control over the saturation region versus standard sequences. W5-Watergate uses a "gradient−180*−gradient" combination where the initial gradient pulse dephases every signal, 180* indicates a selective 180-degree pulse on H_2O resonance, and the second gradient further dephases the water but refocuses the rest of the signals. In combination, these sequences provide clean baselines and may retain some exchangeable amide protons, although at the expense of signal loss within 0.5−1 ppm of the water peak.[117]

Marine Endometabolome. NMR and UHPLC-MS (ultrahigh performance liquid chromatography−mass spectrometry) have also been combined in ecological metabolomics research, providing clues about allelopathy, a type of competition in which species release inhibitory or lethal chemical compounds to impair competitor species.[120] *Karenia brevis* is a dinoflagellate associated with toxic red-tide algal blooms. Two diatom competitors—*Asterionellopsis glacialis* and *Thalassiosira pseudonana*—were exposed to allelopathic compounds from *K. brevis* and evaluated for physiological, proteomic, and metabolomic impact. *T. pseudonana* exposed to *K. brevis* had a statistically significant 85% reduction in growth rate and population decline, whereas *A. glacialis* had a nonsignificant 35% growth reduction. Partial least-squares-discriminate analysis (PLS-DA) was performed to determine which compounds deviated from the control in response to *K. brevis*. *A. glacialis* had 6−9% variation in metabolites and few protein changes due to *K. brevis*, indicating resistance to the allelopathic compounds likely evolved from exposure in nature. Conversely, *T. pseudonana* metabolism was dramatically impacted by exposure to the red-tide dinoflagellate, likely due to sparse prior exposure in nature. A large number of proteins were affected, and metabolites indicative of changes in energy metabolism and cell stress were discovered. UHPLC-MS analysis revealed decreases in metabolites associated with cell wall structure, oxidative stress, and carbon metabolism. ^{13}C-HSQC NMR analysis revealed complementary data including decreases in metabolites associated with osmoregulation, amino acid metabolism, carbon metabolism, glycolysis, photorespiration, and pyrimidine metabolism.[120] This nicely illustrates how NMR and MS can be used together to obtain a more complete description of the metabolome. The lipidome of *A. glacialis* and *T. pseudonana* exposed to *K. brevis* was subsequently interrogated using the same analytical platforms. ^{13}C-HSQC revealed that 80 lipids in *T. pseudonana* were significantly altered in response to *K. brevis* exposure, whereas just six were significantly altered in *A. glacialis*.[121] The authors thus concluded that *K. brevis* allelopathy disrupts cell membrane lipid metabolism, increases permeabilization of the cell wall, and decreases photosynthetic efficiency in competitors, particularly those without a robust response due to prior natural selective pressure.

Research into the intracellular metabolome of marine microorganisms has enhanced understanding of how bacteria transform metabolites for the microbial loop as well. Using stable-isotope-guided NMR to augment metabolite signals, intracellular metabolites of a natural coastal bacterial community following 48 h incubation with $^{13}C_6$-glucose were examined via ^{13}C-HSQC.[122] Twenty-two compounds were assigned based on correlation signals determined from HSQC-TOCSY and ^{13}C-decoupled 2D-JRES (two-dimensional ^{1}H−^{1}H *J*-resolved spectroscopy) and included amino acids,

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

dipeptides, carboxylic acids, nucleosides, nucleobases, carbohydrates, and amino alcohols. This chemical repertoire suggests that $^{13}C_6$-glucose entered and was transformed via the glycolysis and TCA (tricarboxylic acid) cycles, providing a glimpse into the way DOM is transformed within the microbial loop.[122] The range of different NMR experiments is one of its major strengths. In this case, the 2D J-RES data were used as a way to profile samples without interference of $^1H-^1H$ J-couplings, and both ^{13}C-HSQC and HSQC-TOCSY experiments allowed for database matching and compound ID.

Marine Natural Product (NP) Discovery. As noted above, metabolomics and natural products discovery are similar, especially in unknown compound identification.[9] An important area of marine metabolomics has been the use of NMR for NP discovery.[123−125] While structural and biological characterization of small marine compounds is a challenge due to low metabolite concentrations, development of the capillary and microcryoprobes has enabled identification of many of these metabolites—even at the nanomole scale.[124] The mollusk *Hexabranchus sanguineus* consumes a diet of sponges from which it isolates trisoxazole macrolides, powerful antifungal compounds with cytotoxicity. Samples were collected from the mollusk in the late 1980s, and major trisoxazole structures were determined with the available NMR tools, but minor components remained structurally elusive for almost 20 years. Using a custom 1 mm 600 MHz microcryoprobe[126] with ^1H NMR, COSY, ^{13}C-HSQC, and ^{13}C-HMBC (heteronuclear multiple bond correlation), Dalisay et al. were able to determine the structure of these minor metabolites through overlapping NMR correlations afforded by the different 2D experiments[127]

Molinski and co-workers have conducted numerous studies utilizing the microcryoprobe approach to elucidate the structure of marine NPs from sponges and ascidians, colloquially known as sea squirts.[128−130] Isolates from the sea sponge *Trikentrion flabelliforme* have been shown to possess indole structures known as trikentrin. Previous research demonstrated antibacterial and cytotoxic properties in various configurations of trikentrin, so COSY, NOESY (nuclear Overhauser effect spectroscopy), ^{13}C-HSQC, and ^{13}C-HMBC spectra were collected to further elucidate the metabolic contents of *T. flabelliforme* fractions from over two decades ago.[128] Later, bromotryptosine residues from the sponge *Aplysina lacunose* were investigated due to interest generated by prior research into associated antibacterial, anti-inflammatory, anticancer, and antifungal properties. Structures were elucidated through similar 2D NMR experiments with a microcryoprobe.[129] Finally, the Molinski group also investigated the ascidian *Didemnum mole* for natural products effective against cell proliferation and infection, and using NMR, they elucidated the structure of two cyclic hexapeptides which warrant further pharmacological investigation.[130]

The diversity and large number of NPs yet to be explored is a challenge. The structural elucidation process of these complex molecules limits the throughput of bioactive metabolite discovery. Zhang et al.[131] utilized the nearly standardized NMR structural elucidation process and highly discriminating HSQC experiments to create an AI-based dereplication and analysis tool that can rapidly associate newly isolated NPs with their known analogues. This tool was used by Reher et al.,[132] in combination with a MS molecular networking tool to elucidate a new chimeric swinholide-like macrolide, symplocolide A, as well as the annotation of swinholide A, samholides A−I, and several new derivatives from a filamentous marine cyanobacterium. Approaches like this are possible because NMR chemical shifts are highly reproducible and provide a fingerprint of the molecule. Furthermore, chemical shifts can be accurately calculated using *ab initio* methods,[133] and recent advances have greatly improved the throughput of such calculations, making them more practical for metabolomics and AI applications.[134] These approaches are promising advancements in structural identification of novel metabolites, increasing throughput and efficiency in both natural products research and metabolomics.[132]

CHEMOINFORMATICS AND COMPUTATIONAL MODELING

All of the applications highlighted above require considerable data analysis, and in this final section we highlight some current trends in NMR metabolomics analysis. Some of these methods, especially machine learning, also apply to MS metabolomics and other omics fields. This section is not meant to be a comprehensive review of chemoinformatics or modeling but rather of approaches that we think are especially promising for NMR applications.

Spectral Processing. The physical and chemical properties of a sample can affect the chemical shifts of some NMR peaks. This makes it hard to compare peaks across spectra, so alignment and/or division of NMR spectra into smaller regions (binning) is usually applied to manage this problem.[135] However, results from this step are not always optimal, especially in complicated samples. Takis et al. used modeling strategies for this problem by considering the chemical shift of a signal as the function of a mixture's total chemical composition, pH, and temperature.[136] They built a model including sample pH, temperature, concentrations of 11 ions, and chemical shifts and concentrations of 40 abundant metabolites to estimate the chemical shifts of these metabolites on ∼4000 artificial urine samples. The algorithm begins by matching five navigating signals and then exports estimations of chemical shifts and concentrations of the targeted metabolites and ions. The algorithm demonstrated high predictive accuracy in real urine samples. It also deconvoluted overlapped peaks, thus improving annotation and quantification.

As another alternative to the "align and/or bin" strategy, "speaq 2.0" used wavelets to extract features from raw spectra.[137] In this method, Mexican hat wavelets were used for peak picking because they are robust to baseline distortions. Picked peaks were grouped for signals from the same nuclei across different spectra. In contrast to using peak integrals for quantifications, wavelet coefficients were used here to represent the abundance of picked peaks for later analysis. This method showed tolerance to small chemical shift variations and could effectively extract features from the simulated and published data sets.

Due to the abundance of signals, peaks are often overlapped in a 1D ^1H metabolomics spectrum. 2D NMR experiments could better separate overlapped signals, but due to long acquisition times, they are usually used only for peak annotation or on a small sample set. With the development of fast 2D NMR experiments,[138,139] 2D spectra have the potential to be used for relative quantification. Therefore, tools for quantifying 2D peaks have recently been improved. Two-dimensional spectra can be vectorized[140] or projected on one

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

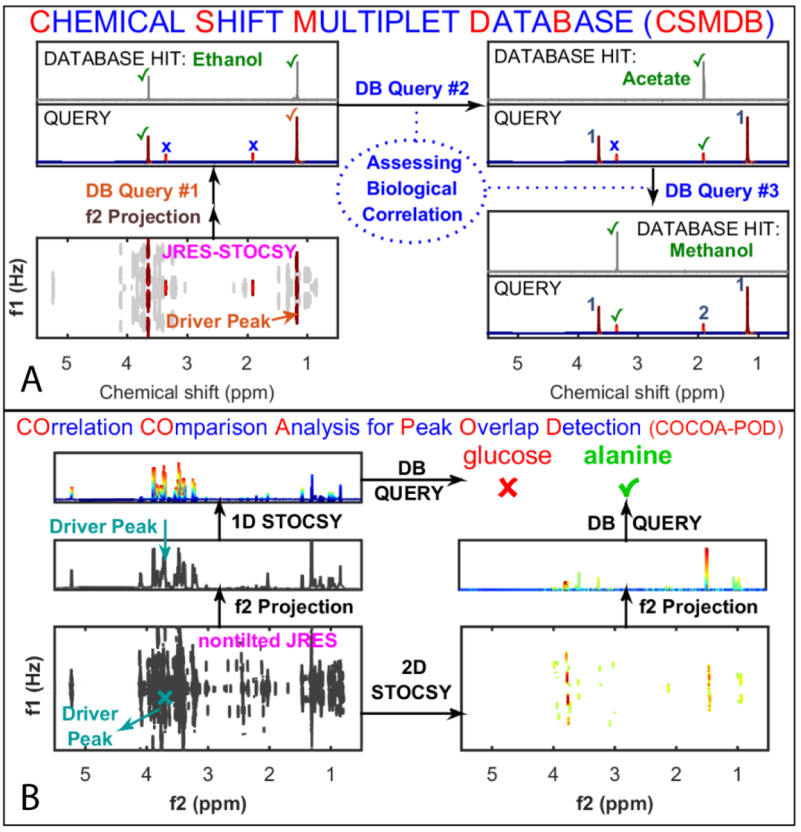

Figure 6. Workflows of using projection of STOCSY traces from tilted and symmetrized JRES spectra (A) and nontilted JRES spectra (B) for database query. (A) Example of consecutively querying chemical shift multiplet database with p-(JRES-STOCSY) on driver peak at 1.181 ppm on tilted and symmetrized JRES spectra. (B) Example of comparing results of querying database with (p-ntJRES)-STOCSY and p-(ntJRES-STOCSY) on the same driver peak at 3.783 ppm. (A) Adapted with permission from Charris-Molina, A.; Riquelme, G.; Burdisso, P.; Hoijemberg, P. A. *J. Proteome Res.* **2020**, *19* (8), 2977−2988 (ref 148). Copyright 2020 American Chemical Society. (B) Adapted with permission from Charris-Molina, A.; Riquelme, G.; Burdisso, P.; Hoijemberg, P. A. *J. Proteome Res.* **2019**, *18* (5), 2241−2253 (ref 149). Copyright 2019 American Chemical Society.

dimension[141] to suit both 1D spectra processing methods and statistical analysis methods. For example, the projection of JRES spectra on the chemical shift dimension (pJRES) can be binned by JBA (pJRES Binning Algorithm).[141] JBA extends the concept of statistical recoupling of variables[142] by using the collinearity of adjacent points to help define bin boundaries and is able to retain small signals more efficiently.

Two-dimensional peaks can also be binned[143,144] or line-shape fitted[145] directly for quantification and matrix size reduction. While nonuniform binning can better quantify peaks than uniform binning, the nonuniform binning algorithms are underdeveloped for 2D spectra. The binning step in HATS-PR (Hierarchical Alignment of Two-dimensional Spectra-Pattern Recognition) can adjust bins by combining multiplets and extending uniform bins to the next bin or to the maximum user defined length.[146] A more flexible multidimensional binning algorithm, Generalized Adaptive Intelligent (GAI) binning, was recently proposed.[143] It extended adaptive intelligent binning from one-dimensional to multidimensional data so that 2D spectra could be binned with flexible bin sizes automatically.

Extracting Information for Peak Annotations. Grouping signals from the same metabolites for database matching can improve the accuracy of annotation. A spectrum from a pure compound can be directly queried and matched in the database. However, for mixtures, peaks from the same metabolite need to be found before querying. Because signals from the same compound should be highly linearly correlated with each other, Statistical TOtal Correlation SpectroscopY (STOCSY) uses Pearson correlation coefficients to gather signals from the same metabolite and builds a pseudospectrum for database query.[147] However, STOCSY performance may be compromised where peaks overlap. JRES can efficiently reduce overlap by separating chemical shifts and multiplicities to two dimensions, but JRES databases are limited and difficult for peak matching.[148] The Hoijemberg group introduced two strategies to circumvent this challenge by querying peaks from the projection of JRES for 1D databases (Figure 6).[148,149] The first strategy (Figure 6A) uses projection of STOCSY traces from tilted and symmetrized JRES (p-(JRES-STOCSY)) as pseudospectra.[148] Because projection of tilted and symmetrized JRES (pJRES) spectra differ from 1D spectra in terms of multiplicity, pJRES spectra cannot be matched directly to 1D databases. Therefore, they built a library (Chemical Shift Multiplet Database) using curated pJRES spectra and their traces on the J-coupling dimension obtained from the Birmingham Metabolite Library.[150] They also built a tool for querying this database. This tool allows for repeated use and includes correlated but small peaks in the query list to avoid false-negative matching. Correlated but unmatched peaks can also be queried against this database for unravelling biological associations. The second strategy (Figure 6B) uses projection of STOCSY results on nontilted JRES spectra (p-(ntJRES-STOCSY))[149] instead of STOCSY on p-ntJRES ((p-ntJRES)-

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

STOCSY)[151] to mimic 1D spectra. Therefore, the deconvoluting power of JRES is preserved.

Although including overlapped peaks in a query list may introduce irrelevant metabolites in the query result, simply discarding the overlapped peak also runs the risk of increasing the false-negative rate. POD-CAST (Peak Overlap Detection by Clustering Analysis and Sorting of Traces) took overlapping information from the clustering of all STOCSY traces, which came from each peak used as a driver peak, to complement the peak list for database query.[152]

These computational techniques could be employed with other spectroscopical or physical separating techniques to enhance the efficiency of peak annotation. In a recent approach to compound identification in NMR metabolomics, the Nicholson group proposed a system to sequentially use computational and experimental annotation strategies for a metabolomics study.[153] This system was shown to efficiently reduce manual input and improve annotation accuracy and thus is expected to be generally utilized in the future.

Database coverage also limits NMR peak annotation. With the development of computational biology, structural information can now be predicted from chemical shifts by machine learning at the motif,[154] molecular,[155] and compound family[131] level. In addition, the 1D spectrum from one metabolite can be simulated under any magnetic field strength according to its spin parametrized system matrix in the GISSMO (Guided Ideographic Spin System Model Optimization) library.[156] These approaches are not only helpful for peak annotations but they also enable the enhancement of current databases with additional putative reference spectra.

Workflows. Owing to the diversity of tools available for NMR spectral analysis, researchers usually assemble their own workflows to record the functions and parameters they use. There are several general tools which serve as pipelines that include widely used steps and methods, such as NMRProcFlow,[157] ASICS R,[158] and Chenomx (Chenomx, Edmonton, Canada). Recently published pipelines include AlpsNMR[159] and PepsNMR[160] for preprocessing, Lipspin[161] for lipids profiling, InterSpin[162] for low-resolution NMR (such as benchtop NMR and solid state NMR), and SigMa[163] for complex spectra processing. For complex spectra, processing regions with different methods then combining them can sometimes provide a better result than processing the entire spectra with one method.[164] SigMa divides a 1D NMR spectrum into three categories: signature signals (SS), signals of unknown spin systems (SUS), and bins (BINS) which are too complicated to align and annotate. SS and SUS are further aligned and quantified by line-shape fitting, but the signals in BINS are integrated directly. By doing this, more information can be extracted from the spectra.

With the rapid development of tools and workflows, the need for workflow management is drawing attention.[165] Verhoeven et al.[166] recently wrote a review on KNIME[167] and Galaxy[168] workflow management platforms, where users can assemble, automatically run, and record their workflows. Workflow4Metabolomics (W4M)[169] is another workflow management infrastructure based on Galaxy. Beyond just building and running workflows, this platform is also designed to be a workflow repository. Usually, workflows are deposited with data in data repositories, such as MetaboLights[170] and Metabolomics Workbench,[4] but with this platform, workflows can be cited and shared directly. PhenoMeNal[171] is a recently built cloud-based metabolomics analysis e-infrastructure. While it incorporates W4M, it provides a greater variety of established tools than W4M. PhenoMeNal enables calculations to be run on the cloud and therefore makes analysis both time- and resource-efficient. While current pipeline tools aim to provide automatic solutions for analysis, an advanced user may tend to have more flexibility in choosing methods for each step. Therefore, a central language-independent tool repository for metabolomics research would be invaluable for users to learn and explore functions.

Feature Selection. Though modeling often uses many features, only a few may be related to outcomes in NMR metabolomics. To speed up the learning process as well as increase model performance, feature selection is often performed before the model training process. Feature selection uses specific strategies to select features that contribute most highly to the prediction variable. Many feature selection methods have been developed—from traditional univariate selection methods to modern machine learning algorithms.[172] While a discussion of different feature selection algorithms is beyond the scope of this review, we will focus on two main hurdles of NMR metabolomics data—class-imbalance and nonlinearity.

LASSO (least absolute shrinkage and selection operator) is an L1 norm regularization technique that allows the coefficient value of less important variables to be 0, significantly eliminating feature quantity. It has, therefore, been used as a feature selection method.[173] When dealing with large-scale, class-imbalanced metabolomics data, Fu et al. found that sparse regularization can stabilize training results.[174] They proposed that minimizing the degree of overlap between imbalanced classes can make the data set more separable. They used the portion of minority cases incorrectly classified by K-nearest neighbors (KNN) as the indicator of overlap degree between imbalanced classes.

Two major Lasso-based feature selection strategies have been developed—Minimizing Overlapping Selection under No-Sampling (MOSNS) and Minimizing Overlapping Selection under SMOTE (MOSS). SMOTE is a technique that oversamples the minority group with replaced values.[175] Four different metabolomics data sets, which included one simulation data set, were tested by these algorithms integrated with a support vector machine (SVM). They found that MOSS achieved higher classification performance on all four data sets, indicating that rebalancing processing is beneficial in increasing the true positive rate, whereas MOSNS performed well in identifying important features.[176] Along with LASSO-regularization, both algorithms can effectively alleviate class imbalance effects, thus outperforming other algorithms in their study.

The imbalanced class is not the only concern of NMR metabolomics data analysis. The inherent nonlinearity characteristic of metabolites also makes feature selection problematic. Although LASSO can significantly reduce feature quantity for downstream analysis, it does not handle the nonlinearity of metabolomics data since it assumes that all variables have linear relationships. To address this issue, Yamada et al.[177] developed the Hilbert-Schmidt independence criterion (HSIC) Lasso nonlinear feature selection method. HSIC Lasso extracts predictor variables that are independent of each other and are evaluated by nonparametric HSIC dependency score statistics.[178] Takahashi et al. used this novel HSIC Lasso-based prediction model in a study to predict depressive symptoms using metabolic data from 897 plasma samples.[179] They compared the results with state-of-the-art

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

prediction models including, for example, Lasso, SVM, Random Forest (RF), and Partial Least Squares (PLS) and found that the HSIC Lasso-based feature selection integrated with SVM yields the best predictive power. Feature frequency was used to select key metabolites as predictors of depressive symptoms. No strong dependency was observed among the selected metabolites. Therefore, the HSIC Lasso-based feature selection strategy can handle nonlinearity as well as avoid redundancies in variable selection.

One shortcoming of NMR metabolomics studies is limited sample sizes, with some studies containing 10 or fewer replicates in each group. Since the Takahashi et al. study used a relatively large sample size, the effectiveness of HSIC Lasso-based feature selection on small sample sizes needs to be tested. The small sample size makes feature selection and dimensionality reduction extremely important and difficult to perform. A common strategy is to combine univariate analysis with multivariate analysis to perform feature selection.[180,181] To date, no perfect method can handle the issue of small sample sizes.[182,183]

Machine Learning (ML) Algorithm Comparison. With the development of high throughput NMR and MS data collection in metabolomics, ML algorithms have been widely used in this field for multivariate metabolite analyses in order to diagnose disease,[184] predict risk,[185] and reveal underlying biological mechanisms between human health and disease.[186] Figure 7 lists popular machine learning algorithms in the NMR

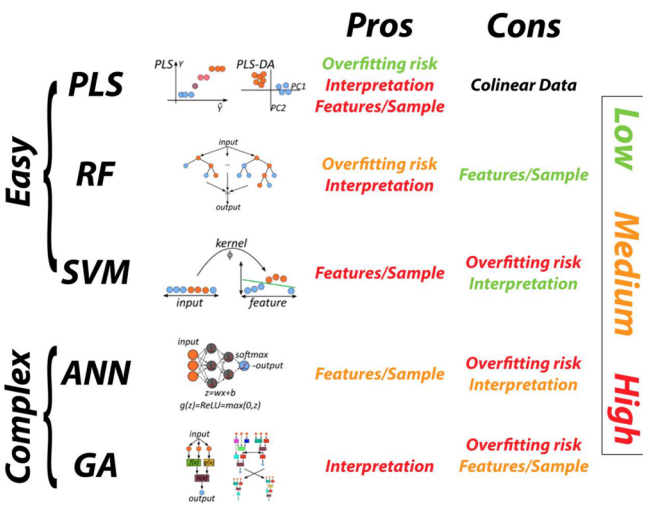

Figure 7. Advantages and disadvantages of popular ML algorithms. The terms on the left ("Easy" and "Complex") refer to implementation. For more detailed information, please see the reference by Liebal and co-workers. Adapted with permission from Liebal, U. W.; Phan, A. N. T.; Sudhakar, M.; Raman, K.; Blank, L. M. *Metabolites* **2020**, *10* (6), 243. (ref 187). Creative Commons Attribution 4.0 International License (https://creativecommons.org/licenses/by/4.0/).

metabolomics field.[187] PLS is a regression method that projects features into linear structures to maximize the explained variance of data sets. PLS-DA is an extension of the PLS algorithm to classify binary classes. RF is a decision tree-based ensemble learning method. By constructing multiple decision trees and combining trees using a majority voting rank, it can be used for classification. SVM outputs a map of sorted data with the clearest margins to separate the two groups. Sometimes, a kernel method can be used in SVM to transform the margin from linear to nonlinear. Artificial neural networks (ANNs) collect the connected units, which allow signal travel between each layer and modelling of a biological brain.[188] As the overview of deep learning (DL) in metabolomics is beyond the scope of this review, please refer to Sen et al.[189] for a detailed description. Genetic algorithms are stochastic methods for function optimization based on biological evolution.[190] Each method has pros and cons, which make choosing proper ML algorithms a challenge.

The Broadhurst group has tested eight different linear and nonlinear ML approaches for their performance in binary classification on 10 clinical data sets from metabolomics studies.[191−201] These 10 data sets were acquired either from NMR or MS, data set size varied from 59 to 968, and the number of metabolite variables of each data set varied from 29 to 689. The eight ML approaches are partial least-squares regression (PLSR), principal component regression (PCR), principal component logistic regression (PCLR), linear kernel support vector machines (SVM-Lin), radial basis function kernel support vector machines (SVM-RBF), RF, linear ANNs, and nonlinear ANNs. The results showed that overall the linear ML algorithms and nonlinear algorithms achieved similar prediction performance in the binary classification metabolomics scenario. RF performed poorly with a small number of samples when the number of variables was large.[202] It also emphasized that if applying Occam's razor principle, the PLS-DA remains the first choice in binary classification problems.[203]

Similarly, Powers' group also evaluated five different ML algorithms (Orthogonal PLS-DA, PLS, SVM, RF, and principal components−linear discriminant analysis (PC-LDA)) in binary classification NMR metabolomics studies.[204] They simulated a 50-metabolite-NMR data set to mimic human urine sample data with known within-group variances, between-group variances, and precisely defined group separation. Thirty-three out of these 50 metabolites were commercially available and used to collect experimental NMR spectra. Overall, equivalent performance was achieved from the five ML algorithms when analyzing high-quality data sets with low noise, small within-group variance, and large between-group variance. When the group separation contributor was limited to one single variable, OPLS-DA and PC-LDA outperformed other models.

From these papers, a simple conclusion can be drawn: PLS-DA and OPLS-DA are currently the gold standard for binary classification. However, things are different in multiclass classification cases. Multiclass classification trains a system to discriminate different classes for various unknown objects. With the increase of class numbers, the complexities of the model also increase, making it more complicated than in a binary classification.[205] Therefore, the strategy of multiclass classification analysis should also be thoroughly addressed. The common strategy to handle multiclass classification is to transform the problem into a binary case, which is called binary decomposition.[206] One-against-all (OAA) and one-against-one (OAO) are two common binary decomposition strategies. OAA divides each class and all other classes into two groups, transforming the K-class classification problem into K parallel binary classifications,[207] while OAO generates binary classification between each pair of classes and total K(K-1)/2 parallel binary classifications are conducted.[208] A study of multiclass discrimination in untargeted metabolomics has been conducted by Trainor et al. to evaluate the performance of six

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

different classifiers.[209] In this case, the OAO approach was used in multiclass classification. The simulated data were incorporated with realistic blockwise correlation and partial correlation of structures to mimic the correlations and metabolite clustering in untargeted metabolomics studies and then tested with six ML approaches including PLS-DA, Sparse PLS-DA, RF, SVM, ANN, KNN, and naïve Bayes. The results showed that SVM and RF outperformed other models when the studies incorporated non-normal error distributions, unbalanced phenotype allocation, outliers, missing values, and dimension reduction. When training with the three class data sets, SVM and RF also perform better than other algorithms. Therefore, knowing the intricate pattern of the data set, considering the statistical power of analysis, and choosing an algorithm accordingly are important and essential when applying ML algorithms in metabolomics studies.

Integration of NMR with Other Data. The quantitative and reproducible qualities of NMR spectroscopy make it ideal to use in data integration. Because of the complementary nature of NMR and MS, combining them in a study is advantageous. They can be combined in tandem or in parallel for structure elucidation or for a better metabolite coverage.[111,115,210] The Brüschweiler lab developed an approach to integrate NMR and high-resolution MS data called SUMMIT MS/NMR (Structure of Unknown Metabolomic Mixture components by MS/NMR).[211] With high-resolution MS data, it is possible to obtain an accurate molecular formula for an unknown MS feature. NMR chemical shifts are calculated for every possible structure, and the corresponding NMR data are searched for the best match. This conceptually simple approach can become quite complicated with larger values of m/z, which can lead to a large number of structures. It also depends on accurate calculations of NMR chemical shifts, which as noted above are now quite accurate with high-level theory.[134] The SUMMIT approach would nicely complement the metabolite fraction libraries described above.[88]

With recent improvements in quantification accuracy and the development of compatible sample preparation protocols for MS and NMR techniques, interplatform correlation is reinforced.[210,212,213] For example, Clendinen et al. used NMR and LC-MS to look for potential biomarkers of prostate cancer recurrence.[213] They measured polar extractions of human serum samples by both NMR and hydrophilic interaction liquid chromatography (HILIC)-MS and nonpolar extractions by reversed phase liquid chromatography (RPLC)-MS. Each platform detected some unique analytes. The authors used correlations between signals across platforms to confirm peak annotations. Also, features from the same metabolites with low or negative interplatform correlation might indicate unreliable quantifications for either platform; therefore, those features were excluded from statistical analysis. For multivariate statistical analysis, NMR and MS data were concatenated and feature selected. As a result, a set of 20 metabolites (3 from NMR) were reported to be potential biomarkers. In addition, correlations between signals from metabolites measured from different platforms were observed, which supplemented metabolic crosstalk information. Together, these results revealed the strength of interplatform correlation on improving peak annotation confidence and relative quantification and unravelling biological relationships between metabolites. Moreover, Nagana Gowda and co-workers showed that interplatform correlation could make absolute quantification in MS samples easier when NMR quantification from the same sample is used as reference.[214]

In many cases, processed NMR and MS data are statistically analyzed separately and integrated on a pathway level. Combining data matrices before multivariate analysis is relatively rare partly due to the matrix size issue.[215] For most cases, the number of variables is much larger than the number of samples, which is not favorable to most statistical analyses. Concatenating matrices will further amplify these differences. The concept of penalized multiblock analysis accompanied with feature selection is suitable for this situation.[215,216] It also manages the imbalance of signal scales between platforms. Deng et al. reported efficient classification of sample groups in integrated LC-MS and NMR data with feature selection for multiblock PLS-DA.[217] They showed that simply concatenating matrices did not produce better performance than did a single matrix, but integrated matrices with feature selection did outperform a single matrix with feature selection. The availability of these statistical tools should be recognized for a more thorough usage of information in data integration.

NMR-based metabolomics can also be integrated with other omics, such as genomics, transcriptomics, proteomics, or microbiomics, in order to develop a more comprehensive understanding of biological systems.[218] Sheikh et al. recently introduced metabolomics to glycomics studies in *C. elegans*.[164] Besides glycomics data, the authors also integrated metabolomics data with worm population distribution to analyze the relationship between metabolites, glycans, and size as a proxy for development. In this study, synchronized worm samples were measured for their metabolome by NMR, glycome by LC-MS/MS, and population distribution by large-particle flow cytometry. Different sized worms showed distinct patterns of glycan and metabolite levels. A correlation network between the three data matrices also showed associations between metabolites, glycans, and worm size. Furthermore, NMR-measured metabolites provided a substrate-level detail of glycan modification and glycosylation. For example, the authors observed that phosphocholine was positively correlated to some developmental-stage-specific N-glycans. This result suggested those glycans may be potential substrates for phosphocholine modifications. The correlation between UDP-*N*-acetylglucosamine (UDP-GlcNAc), O-glycans, and worm sizes indicated possible changes in O-glycan utilization with worm growth. Therefore, together with glycan-level changes, metabolomics results shed light on glycan dynamics during worm development.

With the development of statistical methods, data integration is becoming more flexible. For example, Le Moyec et al. used an unsupervised multiblock model to analyze NMR-measured metabolites and biochemical assay results, which contained heterogeneous analytes such as specific lipid levels, protein levels, and enzymatic activities, for understanding equine energy metabolism during horse racing.[219] Furthermore, data integration can in turn help NMR peak annotation with knowledge from other omics. Wang et al. built a network with NMR-measured metabolite levels, microbiome gene abundance in rumen fluid samples, and compound knowledge in the KEGG database.[220] NMR features were associated with genes through linear and nonlinear correlations. Those genes were mapped in the KEGG database for connected compounds through reaction knowledge. In this way, NMR features were connected to compound names, thus helping to extrapolate peak identity.

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

The developers of Metabomatching also proposed the idea to annotate metabolites with their associated genetic traits (e.g., SNP).[221] In a mGWAS experimental set, they gathered peaks highly associated with one SNP to generate a pseudospectrum for database query. This approach was tested to work on some known metabolites. While this work may not yet fully replace routine annotation, for example by 2D NMR, it can provide some idea of unknown peaks, as metabolites that generate such signals would be associated with enzymes coded by the genes. However, this kind of annotation technique is limited by the genetic diversity of samples and requires a specific experimental design. It is thus more suitable for an integrated study rather than an unaccompanied metabolomics study.

Statistical methods are developing quickly for multiomics studies,[216,222] but methods of integrating metabolomics with other omics data are limited. Using methods developed on other omics-integrated approaches for metabolomics-involved integration is a promising avenue for future research.

CONCLUSION AND FUTURE PERSPECTIVE

We have highlighted several strengths of NMR and shown how these can add considerable value to a metabolomics study. In some examples, such as monitoring *in vivo* metabolism in real-time and in protein-metabolite binding studies, NMR provides information that would be inaccessible with any other technology. In other cases, such as culture media analysis and quantitative analysis of genetic mutants, other technologies such as LC-MS could be used, but NMR is far simpler, quantitative, and reproducible and thus has advantages that arguably outweigh its lower sensitivity. As illustrated by discoveries in natural products chemistry, true unknown compound identification in metabolomics almost always needs NMR, especially in the assignment of stereochemistry.

It is counterproductive to think of NMR and LC-MS or other MS technology to be competitive technologies. Our laboratory regularly uses both, and properly designed studies can benefit greatly from their complementary data. Computational tools exist that allow integration of NMR and MS, though there is considerable room for development in integrative approaches.

Technology advances drive applications, which in turn, expose areas ripe for new technology. New superconducting materials are fueling growth in ultrahigh magnetic fields that exceed 1 GHz. These new NMR systems will have greater resolution and sensitivity and will also be ideal for detecting nuclei other than ^{1}H. NMR probe technology continues to increase sensitivity and reduce volumes of sample. Technologies such as dynamic nuclear polarization (DNP), which are beyond the scope of this review, have the promise of improving NMR sensitivity by orders of magnitude.[6] And the use of NMR spectroscopy, in combination with MRI and some of the other emerging technologies such as DNP, have the potential to provide even a greater bridge between metabolomics and function.

AUTHOR INFORMATION

Corresponding Author

Arthur S. Edison − *Departments of Biochemistry & Molecular Biology, Genetics, and Institute of Bioinformatics, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, USA;* orcid.org/0000-0002-5686-2350; Email: aedison@uga.edu

Authors

Maxwell Colonna − *Departments of Biochemistry & Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, USA*

Goncalo J. Gouveia − *Departments of Biochemistry & Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, USA*

Nicole R. Holderman − *Departments of Biochemistry & Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, USA*

Michael T. Judge − *Genetics, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, USA*

Xunan Shen − *Institute of Bioinformatics, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, USA*

Sicong Zhang − *Departments of Biochemistry & Molecular Biology, Complex Carbohydrate Research Center, University of Georgia, Athens, Georgia 30605, USA*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.analchem.0c04414

Notes

The authors declare no competing financial interest.

Biographies

Arthur S. Edison is a Georgia Research Alliance Eminent Scholar and Professor of Genetics and Biochemistry and a member of the Complex Carbohydrate Research Center and Institute of Bioinformatics at the University of Georgia. He received his Ph.D. in biophysics from the University of Wisconsin—Madison, where he developed and applied NMR experimental and theoretical methods for protein structural studies under the supervision of John Markley and Frank Weinhold. He joined the faculty at the University of Florida and the National High Magnetic Field Laboratory in 1996. He advanced from Assistant to Full Professor in the UF Department of Biochemistry & Molecular Biology. Professor Edison was the founding PI and Director of the NIH-funded Southeast Center for Integrated Metabolomics, and his research focuses on the role of small molecules in biology and disease. In 2015, Edison moved to the University of Georgia, where he directs the CCRC NMR facility, which supports research in both metabolomics and structural biology. Edison's research group collaborates on several metabolomics projects including NMR technology, unknown compound ID, and applications from microbes to humans.

Maxwell B. Colonna is a Ph.D. candidate in the Department of Biochemistry and Molecular Biology at the University of Georgia. He received his undergraduate education from the University of Texas at Austin, completing a B.S. in Biology with a focus on Cell and Molecular Biology in 2011. After graduating, he worked as a laboratory instructor at UT Austin and as an adult education instructor at Austin Community College. In 2016 Max entered the Integrated Life Sciences program at the University of Georgia, ultimately joining the laboratory of Art Edison. His doctoral research focuses on improving cancer immunotherapy manufacturing and understanding interactions between cancer and immune cell metabolism through NMR metabolomics.

Gonçalo J. Gouveia is a Ph.D. candidate in the Department of Biochemistry and Molecular Biology at the University of Georgia. He received his B.Sc. (HONS) degree from the University of Wales in Forensic Science with a focus in Analytical Chemistry. After graduating, he was accepted into a Forensic Science M.Sc. program at the London South Bank University. He became a forensic scientist

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

working for the Forensic Science Services and later for LGC Forensics. He started his scientific career as a Drugs Analyst and progressed to Forensic Toxicology Reporting Officer. He joined the Biochemistry and Molecular Biology Ph.D. program in 2018 at the University of Georgia. Gonçalo is currently conducting his graduate research under the supervision of Art Edison. His doctoral research focuses on metabolomics, metabolomics QA/QC systems, and metabolite identification methods.

Nicole R. Holderman is a Ph.D. candidate in the Department of Biochemistry and Molecular Biology at the University of Georgia. She received her B.S. in Public Health and B.A. in Psychology from Indiana University in 2012. In 2016, she earned her M.S. in Life Science Research from Lincoln Memorial University in Harrogate, Tennessee, while working as a Research Assistant at the University of Alabama at Birmingham. Nicole became a graduate student in the Integrated Life Sciences program at the University of Georgia in 2017 where she later joined the lab of Art Edison. Her research centers on the metabolomics of marine microorganisms and the role their metabolic interactions play in the global carbon cycle.

Michael T. Judge is a Ph.D. candidate in the Department of Genetics at the University of Georgia. He received his B.Sc. in Cell/Molecular Biology from Appalachian State University in Boone, North Carolina, in 2015. While there, he wrote a thesis on paralog divergence in *Arabidopsis* under Annkatrin Rose. He also worked with Anna Stepanova and José Alonso at North Carolina State University in 2013 and Jonathan Arnold at the University of Georgia in 2014 as part of two NSF REU programs. Michael returned to the University of Georgia for his Ph.D. and joined the Art Edison laboratory. His research focuses on the use of NMR for real-time *in vivo* metabolomics of microorganisms.

Xunan Shen is a Ph.D. candidate in the Institute of Bioinformatics at the University of Georgia. She received her B.Sc. in Pharmaceutical Science from Nanjing University of Chinese Medicine in 2015. She then received her M.S. in Pharmacology and Toxicology from the University of Kansas in 2017. Xunan joined the Ph.D. program at the University of Georgia and is working under the direction of Art Edison to characterize stem cells using metabolomics techniques.

Sicong Zhang is a Ph.D. candidate in the Department of Biochemistry and Molecular Biology at the University of Georgia. She received her B.Sc. in Biological Sciences from China Agricultural University in 2016. She then entered the Integrated Life Sciences program at the University of Georgia and joined the Art Edison lab. Sicong's research is focused on developing and applying computational tools for NMR spectral analysis for metabolomics studies.

■ ACKNOWLEDGMENTS

We thank all current and former members of the Edison lab, especially John Glushka, Mario Uchimiya, and Fariba Tayyari for NMR support; Yue Wu, Rahil Taujale, and Olatomiwa Bifarin for ideas in computation; Amanda Shaver and Brianna Garcia for *C. elegans* applications; and Karen Howard for help in editing this review. Jonathan Arnold, Facundo Fernandez, Tobias Kind, and Andrew Borchert all provided helpful discussions. This was supported by the following grants: NIH U2CES030167-03, NIH R01GM120151-04, NIH R01HD087306, NSF 1648035, NSF 1713746, the Gordon and Betty Moore Foundation, and the Georgia Research Alliance.

■ REFERENCES

(1) Heather, J. M.; Chain, B. *Genomics* **2016**, *107*, 1−8.
(2) Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; Funke, R.; Gage, D.; Harris, K.; Heaford, A.; Howland, J.; Kann, L.; Lehoczky, J.; LeVine, R.; McEwan, P.; McKernan, K.; et al. *Nature* **2001**, *409*, 860−921.
(3) May, J. C.; McLean, J. A. *Annu. Rev. Anal. Chem.* **2016**, *9*, 387−409.
(4) Sud, M.; Fahy, E.; Cotter, D.; Azam, K.; Vadivelu, I.; Burant, C.; Edison, A.; Fiehn, O.; Higashi, R.; Nair, K. S.; Sumner, S.; Subramaniam, S. *Nucleic Acids Res.* **2016**, *44*, D463−470.
(5) Markley, J. L.; Bruschweiler, R.; Edison, A. S.; Eghbalnia, H. R.; Powers, R.; Raftery, D.; Wishart, D. S. *Curr. Opin. Biotechnol.* **2017**, *43*, 34−40.
(6) Ardenkjaer-Larsen, J. H.; Boebinger, G. S.; Comment, A.; Duckett, S.; Edison, A. S.; Engelke, F.; Griesinger, C.; Griffin, R. G.; Hilty, C.; Maeda, H.; Parigi, G.; Prisner, T.; Ravera, E.; van Bentum, J.; Vega, S.; Webb, A.; Luchinat, C.; Schwalbe, H.; Frydman, L. *Angew. Chem., Int. Ed.* **2015**, *54*, 9162−9185.
(7) Dalisay, D. S.; Rogers, E. W.; Edison, A. S.; Molinski, T. F. *J. Nat. Prod.* **2009**, *72*, 732−738.
(8) Monge, M. E.; Dodds, J. N.; Baker, E. S.; Edison, A. S.; Fernandez, F. M. *Annu. Rev. Anal. Chem.* **2019**, *12*, 177−199.
(9) Robinette, S. L.; Bruschweiler, R.; Schroeder, F. C.; Edison, A. S. *Acc. Chem. Res.* **2012**, *45*, 288−297.
(10) Fuhrer, T.; Zampieri, M.; Sévin, D. C.; Sauer, U.; Zamboni, N. *Mol. Syst. Biol.* **2017**, *13*, 907.
(11) Mülleder, M.; Calvani, E.; Alam, M. T.; Wang, R. K.; Eckerstorfer, F.; Zelezniak, A.; Ralser, M. *Cell* **2016**, *167*, 553−565. (e512).
(12) Himmelreich, U.; Somorjai, R. L.; Dolenko, B.; Lee, O. C.; Daniel, H.-M.; Murray, R.; Mountford, C. E.; Sorrell, T. C. *Appl. Environ. Microbiol.* **2003**, *69*, 4566−4574.
(13) Li, W. *Analyst* **2006**, *131*, 777−781.
(14) Maricq, M. M.; Waugh, J. *J. Chem. Phys.* **1979**, *70*, 3300−3316.
(15) Righi, V.; Constantinou, C.; Kesarwani, M.; Rahme, L. G.; Tzika, A. A. *Biomed. Rep.* **2013**, *1*, 707−712.
(16) Faure, E.; Kwong, K.; Nguyen, D. *Front. Immunol.* **2018**, *9*, 2416−2416.
(17) Chauton, M. S.; Optun, O. I.; Bathen, T. F.; Volent, Z.; Gribbestad, I. S.; Johnsen, G. *Mar. Ecol.: Prog. Ser.* **2003**, *256*, 57−62.
(18) Righi, V.; Apidianakis, Y.; Mintzopoulos, D.; Astrakas, L.; Rahme, L. G.; Tzika, A. A. *Int. J. Mol. Med.* **2010**, *26*, 175−184.
(19) Righi, V.; Apidianakis, Y.; Psychogios, N.; Rahme, L. G.; Tompkins, R. G.; Tzika, A. A. *Int. J. Mol. Med.* **2014**, *34*, 327−333.
(20) Sarou-Kanian, V.; Joudiou, N.; Louat, F.; Yon, M.; Szeremeta, F.; Meme, S.; Massiot, D.; Decoville, M.; Fayon, F.; Beloeil, J. C. *Sci. Rep.* **2015**, *5*, 9872.
(21) Judge, M. T.; Wu, Y.; Tayyari, F.; Hattori, A.; Glushka, J.; Ito, T.; Arnold, J.; Edison, A. S. *Front Mol. Biosci* **2019**, *6*, 1−16.
(22) Deguchi, S.; Shimoshige, H.; Tsudome, M.; Mukai, S.-a.; Corkery, R. W.; Ito, S.; Horikoshi, K. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, *108*, 7997.
(23) Mobarhan, Y. L.; Struppe, J.; Fortier-McGill, B.; Simpson, A. J. *Anal. Bioanal. Chem.* **2017**, *409*, 5043−5055.
(24) Wong, A.; Boutin, C.; Aguiar, P. M. *Front. Chem.* **2014**, *2*, 38−38.
(25) Lucas-Torres, C.; Wong, A. *Metabolites* **2019**, *9*, 29.
(26) Bastawrous, M.; Jenne, A.; Tabatabaei Anaraki, M.; Simpson, A. J. *Metabolites* **2018**, *8*, 35
(27) Mobarhan, Y. L.; Fortier-McGill, B.; Soong, R.; Maas, W. E.; Fey, M.; Monette, M.; Stronks, H. J.; Schmidt, S.; Heumann, H.; Norwood, W.; Simpson, A. J. *Chem. Sci.* **2016**, *7*, 4856−4866.
(28) Eakin, R.; Morgan, L.; Gregg, C.; Matwiyoff, N. *FEBS Lett.* **1972**, *28*, 259−264.
(29) Ugurbil, K.; Brown, T.; Den Hollander, J.; Glynn, P.; Shulman, R. *Proc. Natl. Acad. Sci. U. S. A.* **1978**, *75*, 3742−3746.
(30) Clendinen, C. S.; Stupp, G. S.; Ajredini, R.; Lee-McMullen, B.; Beecher, C.; Edison, A. S. *Front. Plant Sci.* **2015**, *6*, 611.

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

(31) Edison, A. S.; Le Guennec, A.; Delaglio, F.; Kupče, Ē. In *NMR-Based Metabolomics: Methods and Protocols*; Gowda, G. A. N., Raftery, D., Eds.; Springer New York: New York, NY, 2019; pp 69−95.
(32) Anaraki, M. T.; Lysak, D. H.; Soong, R.; Simpson, M. J.; Spraul, M.; Bermel, W.; Heumann, H.; Gundy, M.; Boenisch, H.; Simpson, A. J. *Analyst* **2020**, *145*, 5787.
(33) Jang, C.; Chen, L.; Rabinowitz, J. D. *Cell* **2018**, *173*, 822−837.
(34) Kumar, A.; Kuhn, L. T.; Balbach, J. *Int. J. Mol. Sci.* **2019**, *20*, 378.
(35) Gonzalez-Mendez, R.; Wemmer, D.; Hahn, G.; Wade-Jardetzky, N.; Jardetzky, O. *Biochim. Biophys. Acta, Mol. Cell Res.* **1982**, *720*, 274−280.
(36) Bradley, S. A.; Ouyang, A.; Purdie, J.; Smitka, T. A.; Wang, T.; Kaerner, A. *J. Am. Chem. Soc.* **2010**, *132*, 9531−9533.
(37) Friebel, A.; von Harbou, E.; Münnemann, K.; Hasse, H. *Ind. Eng. Chem. Res.* **2019**, *58*, 18125−18133.
(38) Anaraki, M. T.; Lane, D.; Bastawrous, M.; Jenne, A.; Simpson, A. J. *Methods Mol. Biol.* **2019**, *2037*, 395−409.
(39) Koczula, K. M.; Ludwig, C.; Hayden, R.; Cronin, L.; Pratt, G.; Parry, H.; Tennant, D.; Drayson, M.; Bunce, C. M.; Khanim, F. L.; Gunther, U. L. *Leukemia* **2016**, *30*, 65−73.
(40) Ludwig, C.; Günther, U. L. *BMC Bioinf.* **2011**, *12*, 1−6.
(41) Nguyen, T. T. M.; An, Y. J.; Cha, J. W.; Ko, Y.-J.; Lee, H.; Chung, C. H.; Jeon, S.-M.; Lee, J.; Park, S. *Anal. Chem.* **2020**, *92*, 7382−7387.
(42) Reed, M. A. C.; Roberts, J.; Gierth, P.; Kupče, E.; Günther, U. L. *ChemBioChem* **2019**, *20*, 2207−2211.
(43) Link, H.; Fuhrer, T.; Gerosa, L.; Zamboni, N.; Sauer, U. *Nat. Methods* **2015**, *12*, 1091−1097.
(44) Brindle, K. M. *J. Am. Chem. Soc.* **2015**, *137*, 6418−6427.
(45) Wu, Y.; Judge, M. T.; Arnold, J.; Bhandarkar, S. M.; Edison, A. S. *Bioinformatics* **2020**, DOI: 10.1093/bioinformatics/btaa631.
(46) Link, H.; Christodoulou, D.; Sauer, U. *Curr. Opin. Biotechnol.* **2014**, *29*, 8−14.
(47) Nagana Gowda, G. A.; Abell, L.; Tian, R. *Anal. Chem.* **2019**, *91*, 2464−2471.
(48) Nagana Gowda, G. A. *Metabolites* **2018**, *8*, 32.
(49) Lane, A. N.; Tan, J.; Wang, Y.; Yan, J.; Higashi, R. M.; Fan, T. W. *Metab. Eng.* **2017**, *43*, 125.
(50) Lin, P.; Lane, A. N.; Fan, T. W. M. *Methods Mol. Biol.* **2019**, *2037*, 151−168.
(51) Winnike, J. H.; Stewart, D. A.; Pathmasiri, W. W.; McRitchie, S. L.; Sumner, S. J. *Int. J. Breast Cancer* **2018**, *2018*, 1−12.
(52) Hattori, A.; Tsunoda, M.; Konuma, T.; Kobayashi, M.; Nagy, T.; Glushka, J.; Tayyari, F.; McSkimming, D.; Kannan, N.; Tojo, A.; Edison, A. S.; Ito, T. *Nature* **2017**, *545*, 500−504.
(53) Verma, A.; Kumar, I.; Verma, N.; Aggarwal, P.; Ojha, R. *BBA clinical* **2016**, *5*, 170−178.
(54) Podo, F.; de Certaines, J. D. *Anticancer Res.* **1996**, *16*, 1305−1315.
(55) Juranic, N.; Nemutlu, E.; Zhang, S.; Dzeja, P.; Terzic, A.; Macura, S. *J. Biomol. NMR* **2011**, *50*, 237−245.
(56) Shah, T.; Krishnamachary, B.; Wildes, F.; Wijnen, J. P.; Glunde, K.; Bhujwalla, Z. M. *NMR Biomed.* **2018**, *31*, e3936.
(57) Veronesi, M.; Giacomina, F.; Romeo, E.; Castellani, B.; Ottonello, G.; Lambruschini, C.; Garau, G.; Scarpelli, R.; Bandiera, T.; Piomelli, D.; Dalvit, C. *Anal. Biochem.* **2016**, *495*, 52−59.
(58) Wojtowicz, W.; Wróbel, A.; Pyziak, K.; Tarkowski, R.; Balcerzak, A.; Bębenek, M.; Młynarz, P. *Metabolites* **2020**, *10*, 173.
(59) Mahar, R.; Donabedian, P. L.; Merritt, M. E. *Sci. Rep.* **2020**, *10*, 8885.
(60) Luis, C.; Duarte, F.; Faria, I.; Jarak, I.; Oliveira, P. F.; Alves, M. G.; Soares, R.; Fernandes, R. *Life Sci.* **2019**, *223*, 38−46.
(61) Sawyer, D.; Sanderson, K.; Lu, R.; Daszkowski, T.; Clark, E.; Mcduff, P.; Astrom, J.; Heffernan, C.; Duffy, L.; Poole, S. *Biomanufacturing technology roadmap* **2017**, 1−48.
(62) Chrysanthopoulos, P. K.; Goudar, C. T.; Klapa, M. I. *Metab. Eng.* **2010**, *12*, 212−222.
(63) Lewis, A. M.; Abu-Absi, N. R.; Borys, M. C.; Li, Z. J. *Biotechnol. Bioeng.* **2016**, *113*, 26−38.
(64) Ali, A. S.; Chen, R.; Raju, R.; Kshirsagar, R.; Gilbert, A.; Zang, L.; Karger, B. L.; Ivanov, A. R. *Biotechnol. J.* **2020**, *15*, e1900565.
(65) Zürcher, P.; Sokolov, M.; Brühlmann, D.; Ducommun, R.; Stettler, M.; Souquet, J.; Jordan, M.; Broly, H.; Morbidelli, M.; Butté, A. *Biotechnol. Prog.* **2020**, *36*, e3012.
(66) Brinson, R. G.; Marino, J. P. *J. Magn. Reson.* **2019**, *307*, 106581.
(67) Blondeel, E. J. M.; Ho, R.; Schulze, S.; Sokolenko, S.; Guillemette, S. R.; Slivac, I.; Durocher, Y.; Guillemette, J. G.; McConkey, B. J.; Chang, D.; Aucoin, M. G. *J. Biotechnol.* **2016**, *234*, 127−138.
(68) Feigal, E. G.; DeWitt, N. D.; Cantilena, C.; Peck, C.; Stroncek, D. *Nat. Immunol.* **2019**, *20*, 955−962.
(69) Roddie, C.; O'Reilly, M.; Dias Alves Pinto, J.; Vispute, K.; Lowdell, M. *Cytotherapy* **2019**, *21*, 327−340.
(70) Arjmand, B.; Sarvari, M.; Alavi-Moghadam, S.; Payab, M.; Goodarzi, P.; Gilany, K.; Mehrdad, N.; Larijani, B. *Front. Endocrinol.* **2020**, *11*, DOI: 10.3389/fendo.2020.00430
(71) Doron, G.; Klontzas, M. E.; Mantalaris, A.; Guldberg, R. E.; Temenoff, J. S. *Biotechnol. Bioeng.* **2020**, *117*, 1761−1778.
(72) Wang, D.; Zhao, L.; Zheng, H.; Dong, M.; Pan, L.; Zhang, X.; Zhang, H.; Gao, H. *Mol. Neurobiol.* **2018**, *55*, 1112−1122.
(73) Lussey-Lepoutre, C.; Hollinshead, K. E. R.; Ludwig, C.; Menara, M.; Morin, A.; Castro-Vega, L.-J.; Parker, S. J.; Janin, M.; Martinelli, C.; Ottolenghi, C.; Metallo, C.; Gimenez-Roqueplo, A.-P.; Favier, J.; Tennant, D. A. *Nat. Commun.* **2015**, *6*, 8784.
(74) Wilkins, J.; Sakrikar, D.; Petterson, X.-M.; Lanza, I. R.; Trushina, E. *Metabolomics* **2019**, *15*, 83.
(75) Ghassemi, S.; Martinez-Becerra, F.; Master, A.; Richman, S. A.; Heo, D.; Leferovich, J.; Tu, Y.; Garcia-Canaveras, J. C.; Ayari, A.; Lu, Y.; Wang, A.; Rabinowitz, J. D.; Milone, M. C.; O'Connor, R. *Cytotherapy* **2020**, *22*, S133.
(76) Agostini, F.; Ruzza, M.; Corpillo, D.; Biondi, L.; Acquadro, E.; Canepa, B.; Viale, A.; Battiston, M.; Serra, F.; Aime, S.; Mazzucato, M. *PLoS One* **2018**, *13*, e0203048.
(77) Johnson, C. H.; Ivanisevic, J.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 451−459.
(78) Guijas, C.; Montenegro-Burke, J. R.; Warth, B.; Spilker, M. E.; Siuzdak, G. *Nat. Biotechnol.* **2018**, *36*, 316−320.
(79) Rinschen, M. M.; Ivanisevic, J.; Giera, M.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2019**, *20*, 353−367.
(80) de Carvalho, L. P.; Zhao, H.; Dickinson, C. E.; Arango, N. M.; Lima, C. D.; Fischer, S. M.; Ouerfelli, O.; Nathan, C.; Rhee, K. Y. *Chem. Biol.* **2010**, *17*, 323−332.
(81) Nikolaev, Y.; Ripin, N.; Soste, M.; Picotti, P.; Iber, D.; Allain, F. H. *Nat. Methods* **2019**, *16*, 743−749.
(82) Nikolaev, Y. V.; Kochanowski, K.; Link, H.; Sauer, U.; Allain, F. H. *Biochemistry* **2016**, *55*, 2590−2600.
(83) Diether, M.; Nikolaev, Y.; Allain, F. H.; Sauer, U. *Mol. Syst. Biol.* **2019**, *15*, e9008.
(84) Tsukidate, T.; Li, Q.; Hang, H. C. *Curr. Opin. Chem. Biol.* **2020**, *54*, 19−27.
(85) Becker, W.; Bhattiprolu, K. C.; Gubensak, N.; Zangger, K. *ChemPhysChem* **2018**, *19*, 895−906.
(86) Lepre, C. A.; Moore, J. M.; Peng, J. W. *Chem. Rev.* **2004**, *104*, 3641−3676.
(87) Waller, J. C.; Alvarez, S.; Naponelli, V.; Lara-Nunez, A.; Blaby, I. K.; Da Silva, V.; Ziemak, M. J.; Vickers, T. J.; Beverley, S. M.; Edison, A. S.; Rocca, J. R.; Gregory, J. F., 3rd; de Crecy-Lagard, V.; Hanson, A. D. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 10412−10417.
(88) Whiley, L.; Chekmeneva, E.; Berry, D. J.; Jimenez, B.; Yuen, A. H. Y.; Salam, A.; Hussain, H.; Witt, M.; Takats, Z.; Nicholson, J.; Lewis, M. R. *Anal. Chem.* **2019**, *91*, 8873−8882.
(89) Weller, M. G. *Sensors* **2012**, *12*, 9181−9209.
(90) Lim, A. S. T.; Vincent, I. M.; Barrett, M. P.; Gilbert, I. H. *ACS Omega* **2019**, *4*, 19199−19215.
(91) Downs, D. M.; Bazurto, J. V.; Gupta, A.; Fonseca, L. L.; Voit, E. O. *AIMS Microbiol* **2018**, *4*, 289−303.

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

(92) Buchner, E. *Nobel Lecture* **1907**.
(93) Reed, L. K.; Baer, C. F.; Edison, A. S. *Curr. Opin. Chem. Biol.* **2017**, *36*, 7−14.
(94) Prosser, G. A.; Larrouy-Maumus, G.; de Carvalho, L. P. *EMBO Rep.* **2014**, *15*, 657−669.
(95) Beadle, G. W.; Tatum, E. L. *Proc. Natl. Acad. Sci. U. S. A.* **1941**, *27*, 499−506.
(96) Borchert, A. J.; Walejko, J. M.; Guennec, A. L.; Ernst, D. C.; Edison, A. S.; Downs, D. M. *Metabolites* **2020**, *10*, 12
(97) Borchert, A. J.; Gouveia, G. J.; Edison, A. S.; Downs, D. M. *mSystems* **2020**, *5*, DOI: 10.1128/mSystems.00843-19
(98) Prelich, G. *Genetics* **2012**, *190*, 841−854.
(99) Marshall, D. D.; Halouska, S.; Zinniel, D. K.; Fenton, R. J.; Kenealy, K.; Chahal, H. K.; Rathnaiah, G.; Barletta, R. G.; Powers, R. *J. Proteome Res.* **2017**, *16*, 1270−1279.
(100) Mashruwala, A. A.; Eilers, B. J.; Fuchs, A. L.; Norambuena, J.; Earle, C. A.; van de Guchte, A.; Tripet, B. P.; Copie, V.; Boyd, J. M. *J. Bacteriol.* **2019**, *201*, DOI: 10.1128/JB.00188-19
(101) Marshall, D. D.; Sadykov, M. R.; Thomas, V. C.; Bayles, K. W.; Powers, R. *J. Proteome Res.* **2016**, *15*, 1205−1212.
(102) Rona, G. B.; Almeida, N. P.; Santos, G. C., Jr; Fidalgo, T. K.; Almeida, F. C.; Eleutherio, E. C.; Pinheiro, A. S. *J. Cell. Biochem.* **2019**, *120*, 5377−5385.
(103) Chae, Y. K.; Kim, S. H.; Markley, J. L. *PLoS One* **2017**, *12*, e0177233.
(104) Brenner, S. *Genetics* **1974**, *77*, 71−94.
(105) Stiernagle, T. *WormBook: the online review of C. elegans biology* **2006**, 1−11.
(106) Clendinen, C. S.; Pasquel, C.; Ajredini, R.; Edison, A. S. *Anal. Chem.* **2015**, *87*, 5698−5706.
(107) Geier, F. M.; Leroi, A. M.; Bundy, J. G. *Front Mol. Biosci* **2019**, *6*, 27.
(108) Bingol, K.; Li, D. W.; Bruschweiler-Li, L.; Cabrera, O. A.; Megraw, T.; Zhang, F.; Bruschweiler, R. *ACS Chem. Biol.* **2015**, *10*, 452−459.
(109) Riddle, D. L. *Genetics* **1977**, *86*, S51−S52.
(110) Jeong, P. Y.; Jung, M.; Yim, Y. H.; Kim, H.; Park, M.; Hong, E.; Lee, W.; Kim, Y. H.; Kim, K.; Paik, Y. K. *Nature* **2005**, *433*, 541−545.
(111) Srinivasan, J.; Kaplan, F.; Ajredini, R.; Zachariah, C.; Alborn, H. T.; Teal, P. E.; Malik, R. U.; Edison, A. S.; Sternberg, P. W.; Schroeder, F. C. *Nature* **2008**, *454*, 1115−1118.
(112) Pungaliya, C.; Srinivasan, J.; Fox, B. W.; Malik, R. U.; Ludewig, A. H.; Sternberg, P. W.; Schroeder, F. C. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 7708−7713.
(113) von Reuss, S. H.; Bose, N.; Srinivasan, J.; Yim, J. J.; Judkins, J. C.; Sternberg, P. W.; Schroeder, F. C. *J. Am. Chem. Soc.* **2012**, *134*, 1817−1824.
(114) Artyukhin, A. B.; Zhang, Y. K.; Akagi, A. E.; Panda, O.; Sternberg, P. W.; Schroeder, F. C. *J. Am. Chem. Soc.* **2018**, *140*, 2841−2852.
(115) Shou, Q.; Feng, L.; Long, Y.; Han, J.; Nunnery, J. K.; Powell, D. H.; Butcher, R. A. *Nat. Chem. Biol.* **2016**, *12*, 770−772.
(116) Hedges, J. I. *Why Dissolved Organics Matter*; Academic Press: San Diego, CA, 2002.
(117) Lam, B.; Simpson, A. J. *Analyst* **2008**, *133*, 263−269.
(118) Ferrer-Gonzalez, F. X.; Widner, B.; Holderman, N. R.; Glushka, J.; Edison, A. S.; Kujawinski, E. B.; Moran, M. A. *ISME J.* **2020**, DOI: 10.1038/s41396-020-00811-y.
(119) Durham, B. P.; Sharma, S.; Luo, H.; Smith, C. B.; Amin, S. A.; Bender, S. J.; Dearth, S. P.; Van Mooy, B. A. S.; Campagna, S. R.; Kujawinski, E. B.; Armbrust, E. V.; Moran, M. A. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112*, 453−457.
(120) Poulson-Ellestad, K. L.; Jones, C. M.; Roy, J.; Viant, M. R.; Fernandez, F. M.; Kubanek, J.; Nunn, B. L. *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 9009−9014.
(121) Poulin, R. X.; Hogan, S.; Poulson-Ellestad, K. L.; Brown, E.; Fernández, F. M.; Kubanek, J. *Sci. Rep.* **2018**, *8*, 9572.
(122) Uchimiya, M.; Tsuboi, Y.; Ito, K.; Date, Y.; Kikuchi, J. *Metabolites* **2017**, *7*, 52
(123) Chaugule, S. R.; Indap, M.; Chiplunkar, S. *Front. Mar. Sci.* **2017**, *4*, DOI: 10.3389/fmars.2017.00384
(124) Molinski, T. F. *Nat. Prod. Rep.* **2010**, *27*, 321−329.
(125) Blunt, J. W.; Copp, B. R.; Keyzers, R. A.; Munro, M. H. G.; Prinsep, M. R. *Nat. Prod. Rep.* **2015**, *32*, 116−211.
(126) Brey, W. W.; Edison, A. S.; Nast, R. E.; Rocca, J. R.; Saha, S.; Withers, R. S. *J. Magn. Reson.* **2006**, *179*, 290−293.
(127) Dalisay, D. S.; Rogers, E. W.; Edison, A. S.; Molinski, T. F. *J. Nat. Prod.* **2009**, *72*, 732−738.
(128) Salib, M. N.; Molinski, T. F. *J. Org. Chem.* **2018**, *83*, 1278−1286.
(129) Salib, M. N.; Jamison, M. T.; Molinski, T. F. *J. Nat. Prod.* **2020**, *83*, 1532−1540.
(130) Salib, M. N.; Molinski, T. F. *J. Org. Chem.* **2017**, *82*, 10181−10187.
(131) Zhang, C.; Idelbayev, Y.; Roberts, N.; Tao, Y.; Nannapaneni, Y.; Duggan, B. M.; Min, J.; Lin, E. C.; Gerwick, E. C.; Cottrell, G. W.; Gerwick, W. H. *Sci. Rep.* **2017**, *7*, 14243.
(132) Reher, R.; Kim, H. W.; Zhang, C.; Mao, H. H.; Wang, M.; Nothias, L. F.; Caraballo-Rodriguez, A. M.; Glukhov, E.; Teke, B.; Leao, T.; Alexander, K. L.; Duggan, B. M.; Van Everbroeck, E. L.; Dorrestein, P. C.; Cottrell, G. W.; Gerwick, W. H. *J. Am. Chem. Soc.* **2020**, *142*, 4114−4120.
(133) Wang, B.; Dossey, A. T.; Walse, S. S.; Edison, A. S.; Merz, K. M., Jr. *J. Nat. Prod.* **2009**, *72*, 709−713.
(134) Das, S.; Edison, A. S.; Merz, K. M., Jr. *Anal. Chem.* **2020**, *92*, 10412−10419.
(135) Vu, T. N.; Laukens, K. *Metabolites* **2013**, *3*, 259−276.
(136) Takis, P. G.; Schäfer, H.; Spraul, M.; Luchinat, C. *Nat. Commun.* **2017**, *8*, 1662.
(137) Beirnaert, C.; Meysman, P.; Vu, T. N.; Hermans, N.; Apers, S.; Pieters, L.; Covaci, A.; Laukens, K. *PLoS Comput. Biol.* **2018**, *14*, e1006018.
(138) Marchand, J.; Martineau, E.; Guitton, Y.; Dervilly-Pinel, G.; Giraudeau, P. *Curr. Opin. Biotechnol.* **2017**, *43*, 49−55.
(139) Ludwig, C.; Viant, M. R. *Phytochem. Anal.* **2010**, *21*, 22−32.
(140) Tabatabaei Anaraki, M.; Bermel, W.; Dutta Majumdar, R.; Soong, R.; Simpson, M.; Monnette, M.; Simpson, A. J. *Metabolites* **2019**, *9*, 16
(141) Rodriguez-Martinez, A.; Ayala, R.; Posma, J. M.; Harvey, N.; Jimenez, B.; Sonomura, K.; Sato, T. A.; Matsuda, F.; Zalloua, P.; Gauguier, D.; Nicholson, J. K.; Dumas, M. E. *Bioinformatics* **2019**, *35*, 1916−1922.
(142) Blaise, B. J.; Shintu, L.; Elena, B.; Emsley, L.; Dumas, M. E.; Toulhoat, P. *Anal. Chem.* **2009**, *81*, 6242−6251.
(143) Worley, B.; Powers, R. *Chemom. Intell. Lab. Syst.* **2015**, *146*, 42−46.
(144) Puig-Castellvi, F.; Perez, Y.; Pina, B.; Tauler, R.; Alfonso, I. *Chem. Commun. (Cambridge, U. K.)* **2018**, *54*, 3090−3093.
(145) Niklasson, M.; Otten, R.; Ahlner, A.; Andresen, C.; Schlagnitweit, J.; Petzold, K.; Lundström, P. *J. Biomol. NMR* **2017**, *69*, 93−99.
(146) Robinette, S. L.; Ajredini, R.; Rasheed, H.; Zeinomar, A.; Schroeder, F. C.; Dossey, A. T.; Edison, A. S. *Anal. Chem.* **2011**, *83*, 1649−1657.
(147) Cloarec, O.; Dumas, M. E.; Craig, A.; Barton, R. H.; Trygg, J.; Hudson, J.; Blancher, C.; Gauguier, D.; Lindon, J. C.; Holmes, E.; Nicholson, J. *Anal. Chem.* **2005**, *77*, 1282−1289.
(148) Charris-Molina, A.; Riquelme, G.; Burdisso, P.; Hoijemberg, P. A. *J. Proteome Res.* **2020**, *19* (8), 2977−2988.
(149) Charris-Molina, A.; Riquelme, G.; Burdisso, P.; Hoijemberg, P. A. *J. Proteome Res.* **2019**, *18*, 2241−2253.
(150) Ludwig, C.; Easton, J. M.; Lodi, A.; Tiziani, S.; Manzoor, S. E.; Southam, A. D.; Byrne, J. J.; Bishop, L. M.; He, S.; Arvanitis, T. N.; Günther, U. L.; Viant, M. R. *Metabolomics* **2012**, *8*, 8−18.

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

(151) Johnson, C. H.; Athersuch, T. J.; Wilson, I. D.; Iddon, L.; Meng, X.; Stachulski, A. V.; Lindon, J. C.; Nicholson, J. K. *Anal. Chem.* **2008**, *80*, 4886−4895.
(152) Hoijemberg, P. A.; Pelczer, I. *J. Proteome Res.* **2018**, *17*, 392−401.
(153) Garcia-Perez, I.; Posma, J. M.; Serrano-Contreras, J. I.; Boulange, C. L.; Chan, Q.; Frost, G.; Stamler, J.; Elliott, P.; Lindon, J. C.; Holmes, E.; Nicholson, J. K. *Nat. Protoc.* **2020**, *15*, 2538−2567.
(154) Wang, C.; Zhang, B.; Timari, I.; Somogyi, A.; Li, D. W.; Adcox, H. E.; Gunn, J. S.; Bruschweiler-Li, L.; Bruschweiler, R. *Anal. Chem.* **2019**, *91*, 15686−15693.
(155) Kang, S.; Kwon, Y.; Lee, D.; Choi, Y.-S. *J. Chem. Inf. Model.* **2020**, *60*, 3765.
(156) Dashti, H.; Wedell, J. R.; Westler, W. M.; Tonelli, M.; Aceti, D.; Amarasinghe, G. K.; Markley, J. L.; Eghbalnia, H. R. *Anal. Chem.* **2018**, *90*, 10646−10649.
(157) Jacob, D.; Deborde, C.; Lefebvre, M.; Maucourt, M.; Moing, A. *Metabolomics* **2017**, *13*, 36.
(158) Lefort, G.; Liaubet, L.; Canlet, C.; Tardivel, P.; Pere, M. C.; Quesnel, H.; Paris, A.; Iannuccelli, N.; Vialaneix, N.; Servien, R. *Bioinformatics* **2019**, *35*, 4356−4363.
(159) Madrid-Gambin, F.; Oller-Moreno, S.; Fernandez, L.; Bartova, S.; Giner, M. P.; Joyce, C.; Ferraro, F.; Montoliu, I.; Moco, S.; Marco, S. *Bioinformatics* **2020**, *36*, 2943−2945.
(160) Martin, M.; Legat, B.; Leenders, J.; Vanwinsberghe, J.; Rousseau, R.; Boulanger, B.; Eilers, P. H. C.; De Tullio, P.; Govaerts, B. *Anal. Chim. Acta* **2018**, *1019*, 1−13.
(161) Barrilero, R.; Gil, M.; Amigo, N.; Dias, C. B.; Wood, L. G.; Garg, M. L.; Ribalta, J.; Heras, M.; Vinaixa, M.; Correig, X. *Anal. Chem.* **2018**, *90*, 2031−2040.
(162) Yamada, S.; Ito, K.; Kurotani, A.; Yamada, Y.; Chikayama, E.; Kikuchi, J. *ACS Omega* **2019**, *4*, 3361−3369.
(163) Khakimov, B.; Mobaraki, N.; Trimigno, A.; Aru, V.; Engelsen, S. B. *Anal. Chim. Acta* **2020**, *1108*, 142−151.
(164) Sheikh, M. O.; Tayyari, F.; Zhang, S.; Judge, M. T.; Weatherly, D. B.; Ponce, F. V.; Wells, L.; Edison, A. S. *Frontiers in molecular biosciences* **2019**, *6*, 49−49.
(165) O'Shea, K.; Misra, B. B. *Metabolomics* **2020**, *16*, 36.
(166) Verhoeven, A.; Giera, M.; Mayboroda, O. A. *Analyst* **2020**, *145*, 3801−3808.
(167) Fillbrunn, A.; Dietz, C.; Pfeuffer, J.; Rahn, R.; Landrum, G. A.; Berthold, M. R. *J. Biotechnol.* **2017**, *261*, 149−156.
(168) Afgan, E.; Baker, D.; Batut, B.; van den Beek, M.; Bouvier, D.; Cech, M.; Chilton, J.; Clements, D.; Coraor, N.; Gruning, B. A.; Guerler, A.; Hillman-Jackson, J.; Hiltemann, S.; Jalili, V.; Rasche, H.; Soranzo, N.; Goecks, J.; Taylor, J.; Nekrutenko, A.; Blankenberg, D. *Nucleic Acids Res.* **2018**, *46*, W537−W544.
(169) Guitton, Y.; Tremblay-Franco, M.; Le Corguille, G.; Martin, J. F.; Petera, M.; Roger-Mele, P.; Delabriere, A.; Goulitquer, S.; Monsoor, M.; Duperier, C.; Canlet, C.; Servien, R.; Tardivel, P.; Caron, C.; Giacomoni, F.; Thevenot, E. A. *Int. J. Biochem. Cell Biol.* **2017**, *93*, 89−101.
(170) Haug, K.; Cochrane, K.; Nainala, V. C.; Williams, M.; Chang, J.; Jayaseelan, K. V.; O'Donovan, C. *Nucleic Acids Res.* **2020**, *48*, D440−D444.
(171) Peters, K.; Bradbury, J.; Bergmann, S.; Capuccini, M.; Cascante, M.; de Atauri, P.; Ebbels, T. M. D.; Foguet, C.; Glen, R.; Gonzalez-Beltran, A.; Gunther, U. L.; Handakas, E.; Hankemeier, T.; Haug, K.; Herman, S.; Holub, P.; Izzo, M.; Jacob, D.; Johnson, D.; Jourdan, F., et al. *GigaScience* **2019**, *8*, DOI: 10.1093/gigascience/giy149
(172) Li, J.; Cheng, K.; Wang, S.; Morstatter, F.; Trevino, R. P.; Tang, J.; Liu, H. *ACM Computing Surveys (CSUR)* **2018**, *50*, 1−45.
(173) Tibshirani, R. *Journal of the Royal Statistical Society: Series B (Methodological)* **1996**, *58*, 267−288.
(174) Fu, G. H.; Yi, L. Z.; Pan, J. *J. Chemom.* **2019**, *33*, e3177.
(175) Chawla, N. V.; Bowyer, K. W.; Hall, L. O.; Kegelmeyer, W. P. *Journal of artificial intelligence research* **2002**, *16*, 321−357.
(176) Fu, G.-H.; Wu, Y.-J.; Zong, M.-J.; Yi, L.-Z. *Chemom. Intell. Lab. Syst.* **2020**, *196*, 103906.
(177) Yamada, M.; Jitkrittum, W.; Sigal, L.; Xing, E. P.; Sugiyama, M. *Neural computation* **2014**, *26*, 185−207.
(178) Gretton, A.; Bousquet, O.; Smola, A.; Schölkopf, B. In *International conference on algorithmic learning theory*; Springer, 2005; pp 63−77.
(179) Takahashi, Y.; Ueki, M.; Yamada, M.; Tamiya, G.; Motoike, I. N.; Saigusa, D.; Sakurai, M.; Nagami, F.; Ogishima, S.; Koshiba, S. *Transl. Psychiatry* **2020**, *10*, 1−12.
(180) Walejko, J. M.; Chelliah, A.; Keller-Wood, M.; Wasserfall, C.; Atkinson, M.; Gregg, A.; Edison, A. S. *Metabolites* **2020**, *10*, 350.
(181) Walejko, J. M.; Antolic, A.; Koelmel, J. P.; Garrett, T. J.; Edison, A. S.; Keller-Wood, M. *American Journal of Physiology-Endocrinology and Metabolism* **2019**, *316*, E546−E556.
(182) Athreya, A. P.; Banerjee, S. S.; Neavin, D.; Kaddurah-Daouk, R.; Rush, A. J.; Frye, M. A.; Wang, L.; Weinshilboum, R. M.; Bobo, W. V.; Iyer, R. K. *2017 IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB)*; IEEE, 2017; pp 1−9.
(183) Kageyama, Y.; Kasahara, T.; Nakamura, T.; Hattori, K.; Deguchi, Y.; Tani, M.; Kuroda, K.; Yoshida, S.; Goto, Y.-i.; Inoue, K. *Int. J. Neuropsychopharmacol.* **2018**, *21*, 207−215.
(184) Abbiss, H.; Maker, G. L.; Trengove, R. D. *Metabolites* **2019**, *9*, 34.
(185) Merino, J.; Udler, M. S.; Leong, A.; Meigs, J. B. *Curr. Diabetes Rep.* **2017**, *17*, 135.
(186) Johnson, C. H.; Ivanisevic, J.; Siuzdak, G. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 451−459.
(187) Liebal, U. W.; Phan, A. N.; Sudhakar, M.; Raman, K.; Blank, L. M. *Metabolites* **2020**, *10*, 243.
(188) Min, S.; Lee, B.; Yoon, S. *Briefings Bioinf.* **2017**, *18*, 851−869.
(189) Sen, P.; Lamichhane, S.; Mathema, V. B.; McGlinchey, A.; Dickens, A. M.; Khoomrung, S.; Orešič, M. *Briefings Bioinf.* **2020**, DOI: 10.1093/bib/bbaa204.
(190) Gilbert, R. J.; Goodacre, R.; Woodward, A. M.; Kell, D. B. *Anal. Chem.* **1997**, *69*, 4381−4389.
(191) Ganna, A.; Salihovic, S.; Sundström, J.; Broeckling, C. D.; Hedman, Å. K.; Magnusson, P. K.; Pedersen, N. L.; Larsson, A.; Siegbahn, A.; Zilmer, M. *PLoS Genet.* **2014**, *10*, e1004801.
(192) Ganna, A.; Fall, T.; Salihovic, S.; Lee, W.; Broeckling, C. D.; Kumar, J.; Hägg, S.; Stenemo, M.; Magnusson, P. K.; Prenni, J. E. *Metabolomics* **2016**, *12*, 4.
(193) Hilvo, M.; Gade, S.; Hyötyläinen, T.; Nekljudova, V.; Seppänen-Laakso, T.; Sysi-Aho, M.; Untch, M.; Huober, J.; von Minckwitz, G.; Denkert, C. *Int. J. Cancer* **2014**, *134*, 1725−1733.
(194) Stevens, V. L.; Wang, Y.; Carter, B. D.; Gaudet, M. M.; Gapstur, S. M. *Metabolomics* **2018**, *14*, 97.
(195) Armstrong, C. W.; McGregor, N. R.; Lewis, D. P.; Butt, H. L.; Gooley, P. R. *Metabolomics* **2015**, *11*, 1626−1639.
(196) Thévenot, E. A.; Roux, A.; Xu, Y.; Ezan, E.; Junot, C. *J. Proteome Res.* **2015**, *14*, 3322−3335.
(197) Zheng, X.; Huang, F.; Zhao, A.; Lei, S.; Zhang, Y.; Xie, G.; Chen, T.; Qu, C.; Rajani, C.; Dong, B. *BMC Biol.* **2017**, *15*, 1−15.
(198) Fahrmann, J. F.; Kim, K.; DeFelice, B. C.; Taylor, S. L.; Gandara, D. R.; Yoneda, K. Y.; Cooke, D. T.; Fiehn, O.; Kelly, K.; Miyamoto, S. *Cancer Epidemiol., Biomarkers Prev.* **2015**, *24*, 1716−1723.
(199) Sakanaka, A.; Kuboniwa, M.; Hashino, E.; Bamba, T.; Fukusaki, E.; Amano, A. *Sci. Rep.* **2017**, *7*, 42818.
(200) Franzosa, E. A.; Sirota-Madi, A.; Avila-Pacheco, J.; Fornelos, N.; Haiser, H. J.; Reinker, S.; Vatanen, T.; Hall, A. B.; Mallick, H.; McIver, L. J. *Nat. Microbiol.* **2019**, *4*, 293−305.
(201) Chan, A. W.; Mercier, P.; Schiller, D.; Bailey, R.; Robbins, S.; Eurich, D. T.; Sawyer, M. B.; Broadhurst, D. *Br. J. Cancer* **2016**, *114*, 59−62.
(202) Friedman, J.; Hastie, T.; Tibshirani, R. *The elements of statistical learning*; Springer series in statistics; New York, 2001; Vol. *1*.

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX

(203) Mendez, K. M.; Reinke, S. N.; Broadhurst, D. I. *Metabolomics* **2019**, *15*, 150.
(204) Vu, T.; Siemek, P.; Bhinderwala, F.; Xu, Y.; Powers, R. *J. Proteome Res.* **2019**, *18*, 3282−3294.
(205) Fu, H.-Y.; Yin, Q.-B.; Xu, L.; Goodarzi, M.; Yang, T.-M.; Li, G.-F.; She, Y.-B. *Chemom. Intell. Lab. Syst.* **2016**, *157*, 43−49.
(206) Elkano, M.; Galar, M.; Sanz, J.; Bustince, H. *Inf. Sci.* **2016**, *332*, 94−114.
(207) Clark, P.; Boswell, R. *European Working Session on Learning*; Springer, 1991; pp 151−163.
(208) Brunner, C.; Fischer, A.; Luig, K.; Thies, T. *J. Machine Learning Res.* **2012**, *13*, 2279−2292.
(209) Trainor, P. J.; DeFilippis, A. P.; Rai, S. N. *Metabolites* **2017**, *7*, 30.
(210) Marshall, D. D.; Powers, R. *Prog. Nucl. Magn. Reson. Spectrosc.* **2017**, *100*, 1−16.
(211) Bingol, K.; Bruschweiler-Li, L.; Yu, C.; Somogyi, A.; Zhang, F.; Bruschweiler, R. *Anal. Chem.* **2015**, *87*, 3864−3870.
(212) Li, X.; Luo, H.; Huang, T.; Xu, L.; Shi, X.; Hu, K. *Anal. Bioanal. Chem.* **2019**, *411*, 1301−1309.
(213) Clendinen, C. S.; Gaul, D. A.; Monge, M. E.; Arnold, R. S.; Edison, A. S.; Petros, J. A.; Fernandez, F. M. *J. Proteome Res.* **2019**, *18*, 1316−1327.
(214) Nagana Gowda, G. A.; Djukovic, D.; Bettcher, L. F.; Gu, H.; Raftery, D. *Anal. Chem.* **2018**, *90*, 2001−2009.
(215) Boccard, J.; Rudaz, S. *J. Chemom.* **2014**, *28*, 1−9.
(216) Csala, A.; Zwinderman, A. H. In *Computational Biology*; Husi, H., Ed.: Brisbane (AU), 2019.
(217) Deng, L.; Gu, H.; Zhu, J.; Nagana Gowda, G. A.; Djukovic, D.; Chiorean, E. G.; Raftery, D. *Anal. Chem.* **2016**, *88*, 7975−7983.
(218) Pinu, F. R.; Beale, D. J.; Paten, A. M.; Kouremenos, K.; Swarup, S.; Schirra, H. J.; Wishart, D. *Metabolites* **2019**, *9*, 76
(219) Le Moyec, L.; Robert, C.; Triba, M. N.; Bouchemal, N.; Mach, N.; Riviere, J.; Zalachas-Rebours, E.; Barrey, E. *Front Mol. Biosci* **2019**, *6*, 45.
(220) Wang, M.; Wang, H.; Zheng, H.; Dewhurst, R.; Roehe, R. *IEEE Trans Nanobioscience* **2020**, *19*, 518−526.
(221) Rueedi, R.; Mallol, R.; Raffler, J.; Lamparter, D.; Friedrich, N.; Vollenweider, P.; Waeber, G.; Kastenmüller, G.; Kutalik, Z.; Bergmann, S. *PLoS Comput. Biol.* **2017**, *13*, e1005839−e1005839.
(222) Huang, S.; Chaudhary, K.; Garmire, L. X. *Front. Genet.* **2017**, *8*, 84.

https://dx.doi.org/10.1021/acs.analchem.0c04414
Anal. Chem. XXXX, XXX, XXX−XXX