Trends in Analytical Chemistry 180 (2024) 117918

Contents lists available at ScienceDirect

Trends in Analytical Chemistry

journal homepage: www.elsevier.com/locate/trac

A reproducibility crisis for clinical metabolomics studies

Darcy Cochran [a,b], Mai Noureldein [a,b], Dominika Bezdeková [a], Aaron Schram [c], Réka Howard [c], Robert Powers [a,b,*]

[a] Department of Chemistry, University of Nebraska-Lincoln, Lincoln, NE, 68588-0304, USA
[b] Nebraska Center for Integrated Biomolecular Communication, University of Nebraska-Lincoln, Lincoln, NE, 68588-0304, USA
[c] Department of Statistics, University of Nebraska – Lincoln, Lincoln, NE, 68583-0963, USA

ARTICLE INFO

Keywords:
Meta-analysis
Cancer biomarkers
Clinical metabolomics
NMR
Mass spectrometry

ABSTRACT

Cancer is a leading cause of world-wide death and a major subject of clinical studies focused on the identification of new diagnostic tools. An in-depth meta-analysis of 244 clinical metabolomics studies of human serum samples highlights a reproducibility crisis. A total of 2,206 unique metabolites were reported as statistically significant across the 244 studies, but 72% (1,582) of these metabolites were identified by only one study. Further analysis shows a random disparate disagreement in reported directions of metabolite concentration changes when detected by multiple studies. Statistical models revealed that 1,867 of the 2,206 metabolites (85%) are simply statistical noise. Only 3–12% of these metabolites reach the threshold of statistical significance for a specific cancer type. Our findings demonstrate the absence of a detectable metabolic response to cancer and provide evidence of a serious need by the metabolomics community to establish widely accepted best practices to improve future outcomes.

1. Introduction

1.1. Cancer prevalence, cost, and screening

Cancer is a leading cause of death worldwide, where lung, colorectal and breast cancer have the highest mortality rates [1]. According to the International Agency for Research on Cancer and the World Health Organization, there were an estimated 20 million new cancer cases and 9.7 million cancer deaths worldwide in 2022 [2]. The overall cancer-related medical costs in the US for 2020 was estimated to be $208.9 billion, which includes the costs of cancer screening and diagnostics in addition to treatments and patient care [3]. Thus, cancer incurs a high financial and personal burden and, while progress has been made over the past decades, new treatments and diagnostic tools are still desperately needed [4].

Cancer is a diverse disease with over 100 different types that can appear anywhere in the body, which means that screening and diagnostics procedures are highly variable and dependent on the specific locations of the tumors. Additionally, early detection of cancer is the goal of any diagnostic tool since it significantly increases a positive outcome with cancer mortality rates decreasing by 33 % since 1991 due to improved cancer screening [3,5]. However, no screening method is 100 % effective, and many cancers lack routine tests especially in the absence of symptoms [6–8]. Other cancer screens have not been particularly successful at early detection and may lead to high false positive rates [9]. One means of improving the early detection of cancer would be the discovery of new, robust, and accurate molecular biomarkers [10,11].

1.2. Clinical metabolomics as a source for cancer biomarkers

Metabolomics has been especially impactful to issues of human health [12,13] because the metabolome changes rapidly in response to stressors like disease states. Thus, a common utilization of metabolomics is the search for molecular biomarkers from tissues [14], cell cultures [15], animal models [16–18], and biofluids [19–21] as a tool for disease diagnosis, prognosis, and precision medicine. Accordingly, clinical metabolomics has been applied multiple times to nearly every common human disease [22] from asthma [23] to zika virus infections [24]. This includes essentially every type of cancer [25]. The metabolome is particularly appealing source of biomarkers since metabolites are present in every biofluid that is easily accessible from a human patient, but, more importantly, metabolomics-derived biomarkers typically consist of a set of metabolites altered by the disease state instead of a single

* Corresponding author. University of Nebraska-Lincoln, Department of Chemistry, 722 Hamilton Hall, Lincoln, NE 68588-0304, USA.
E-mail address: rpowers3@unl.edu (R. Powers).

https://doi.org/10.1016/j.trac.2024.117918
Received 2 May 2024; Received in revised form 22 July 2024; Accepted 16 August 2024
Available online 19 August 2024

molecular entity. The accuracy and reproducibility of a diagnostic tool is expected to increase dramatically by the simultaneous monitoring of multiple molecular markers since it improves both sensitivity and selectivity. Thus, clinical metabolomics holds great promise to significantly and beneficially impact human health by improving both the diagnosis and treatment of human diseases [26–28].

1.3. Challenges and successes associated with novel biomarker discovery

The search for clinical biomarkers is not limited to metabolites as there are efforts being made across multiple other "-omics" fields like proteomics and genomics to find diagnostic biomarkers of human diseases [29]. However, it is quite difficult to transition any potential biomarker identified as significant in a research study to a validated diagnostic tool approved by the U.S. Food and Drug Administration (FDA), European Medicines Agency (EMA), or other government agencies for clinical usage [30]. The quality of a potential biomarker is typically evaluated through statistical metrics such as sensitivity, specificity, positive predictive value, and negative predictive value [31,32]. Drucker et al. (2013) report that the ratio of biomarker publications to biomarker patents is less than 6 %, demonstrating that very few of the annually discovered biomarkers (proteins, metabolites, genes, *etc.*) meet the necessary sensitivity and specificity criteria [30]. This lack of success can be attributed to errors in the study design that include improper sample collection and storage temperature, inadequate number of replicates or unreliable data collection and data analysis [33]. Metabolomics faces additional unique challenges in the reliable detection of biomarkers that includes: (i) chemical and enzymatic instability of metabolites, (ii) the metabolome being largely unknown or uncharacterized (*i.e.,* dark), (iii) incomplete coverage or detection of the metabolome by NMR and MS, and (iv) ambiguities in metabolite assignments.

While there are legitimate concerns about the discovery process, there is a strong history of successful molecular biomarkers that are being used in the clinic today. For example, breast cancer susceptibility genes BRCA1/2 and the cancer antigen 15–3 (CA 15–3) are biomarkers for breast cancer [34,35] and the prostate-specific antigen protein is a biomarker used in the detection of prostate cancer [34]. Blood testing of metabolites such as calcium, sodium, chloride, creatinine, glucose or cholesterol are routinely screened for kidney disease, diabetes, or cardiovascular disease [36]. A recent review by Qiu et al. (2023) highlighted other metabolites that are promising biomarkers for a variety of diseases such as traumatic brain injury, asthma, tuberculosis, cancers, and COVID-19 [37]. Overall, the road from biomarker discovery to clinical validation is a difficult, but worthwhile effort that has benefited numerous individuals.

1.4. Concerns about reproducibility in clinical metabolomics

Although metabolomics has the potential to transform cancer research, it is still a relatively new field that lacks community agreed-upon best practices for data collection and reporting criteria despite ongoing efforts by several groups and initiatives such as COordination of Standards in MetabOlomicS (COSMOS), Metabolomics Standards Initiative (MSI), Metabolomics Quality Assurance & Quality Control Consortium (mQACC), and Metabolomics Association of North America (MANA) [38–43]. As a result, there are substantial variations in protocols reported by clinical metabolomics studies regarding extraction, detection, and analysis methods, which can subsequently lead to inconsistent or contradictory outcomes. For example, different analytical methods (*i.e.,* NMR, MS, FT-IR, *etc.*) [44], LC columns (*i.e.,* HILIC, C18, IEC, *etc.*) [45], solvent extraction techniques (*i.e.,* aqueous, methanol, Folch, *etc.*) [46], and biomolecular removal protocols (*i.e.,* precipitation, filtration, intact) [47], among other factors, will lead to unique sets of detected metabolites. Simply, NMR measures the most abundant metabolites and MS detects metabolites that readily ionize. Differential physical, chemical, and structural properties of metabolites that includes solvent solubility, polarity, reactivity, molecular weight, oxidative and thermal stability, and biomolecular affinity will all uniquely impact the metabolites that remain following the sample preparation protocol. The selection of experimental and analysis method will similarly affect the precision and accuracy of the measured metabolite concentrations. These protocol decisions consists of the proper choice of internal standards (*i.e.,* blanks, isotopically labeled metabolite standards, *etc.*), quality control (QC) samples (*i.e.,* pooled case and control samples), feature selection method (*i.e.,* S/N, RSD, background removal, *etc.*), alignment and batch correction method (*i.e.,* regression models, normalization methods, *etc.*), statistical techniques (*i.e.,* PCA, Student's t-test, *etc.*), and sample randomization. Unfortunately, most clinical metabolomics studies exclude one or more of these essential protocols leading to erroneous results. Given this large diversity in study design choices, it is not surprising that replicate clinical metabolomics studies have reported discordant metabolites of interest with opposing metabolite directional changes partly due to the variable application of these experimental protocols.

Our prior systematic review of pancreatic ductal adenocarcinoma (PDAC) papers found few metabolites were commonly reported across the 24 clinical metabolomics studies [48]. In fact, 87% of the 655 potential metabolite biomarkers for PDAC were reported by a single study. For the 16 most reported metabolites (*i.e.,* 5 to 11 studies), 10 of these metabolites were inconsistently identified as increasing or decreasing in PDAC patients. Unfortunately, other meta-analysis of clinical metabolomics studies uncovered similar inconsistencies and lack of biomarker reproducibility [49–52]. Herein, we present a systematic review and meta-analysis that expands upon our previous PDAC study to further explore the general reproducibility and consistency of proposed cancer metabolite biomarkers. Our meta-analysis of 244 clinical metabolomics studies of 19 homogenized cancer groups also provides an avenue to assess if a universal set of general cancer metabolites biomarkers exist and what the detection threshold for this panel may be. Importantly, we classified metabolomics and lipidomics studies separately and excluded solely lipidomics studies from our meta-analysis. Finally, our meta-analysis provides further evidence that community-wide standards and best practices are needed to ensure consistency across metabolomics studies to enable the harmonization of metabolomics data and results.

2. Discussion

2.1. Overview of the clinical metabolomics data set

An exhaustive search of scientific literature was conducted to find all clinical metabolomics studies from four major databases with an aim to identify diagnostic biomarkers for cancer (Fig. 1). Approximately 1,000 manuscripts were identified but after manually applying exclusion and inclusion criteria a total of 244 unique clinical metabolomics studies were identified. Notable inclusion criteria consisted of only human studies involving serum biofluids analyzed by NMR and/or MS to identify metabolites that differentiated cancerous from non-cancerous individuals. The 244 studies were then manually analyzed to extract a diversity of relevant data including details regarding the journal publication, the metabolomics and statistical methods, the cancer type and list of cancer dysregulated metabolites.

Considering metabolomics is a relatively new omics field, it was not surprising that most of the clinical metabolomics papers (82%) were published since 2015 (Fig. 2a). The oldest paper in the collection was from 2008. There were 43 unique cancer types as originally reported in the manuscripts, but after nomenclature homogenization that grouped similar cancer types together like laryngeal and nasopharyngeal cancer, the total number of cancer groups was reduced to 19.

(Fig. 2b). The lung (15%), colorectal (14%), hepatocellular (13%), and gastrointestinal (11%) cancer groups corresponded to over half of the clinical metabolomics studies comprising the data set. These cancer types are frequently ranked as the most commonly occurring cancers,

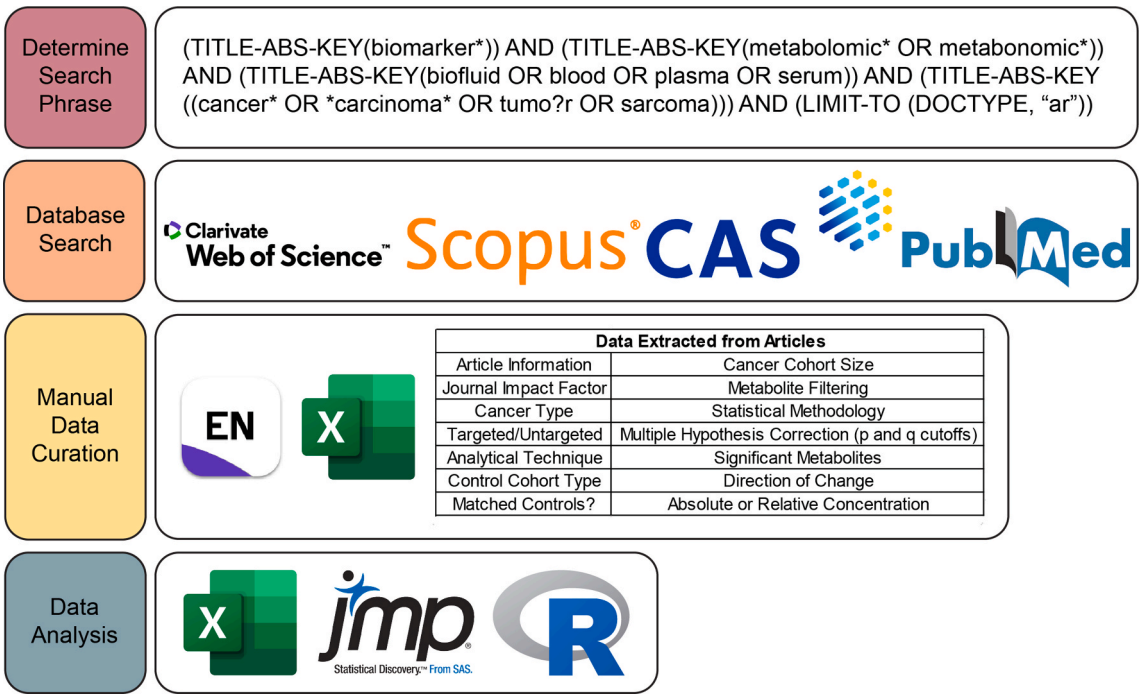


Fig. 1. Data Curation Workflow. Top to bottom stepwise diagram depicting the experimental workflow. Boolean search terms (Table S1) were applied to each database to acquire articles of interest (Table S2). Data extracted from the articles are listed in Tables S4–S12.

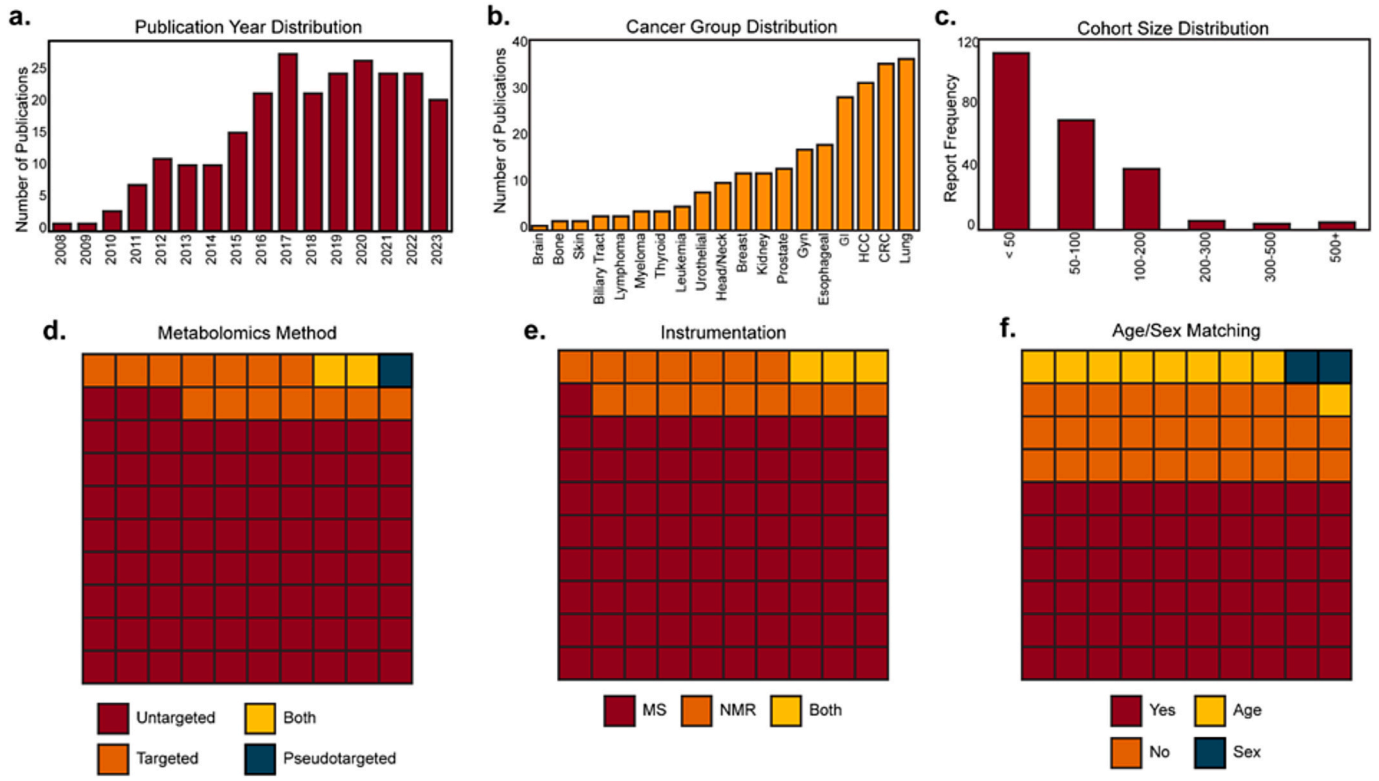


Fig. 2. Cohort Characteristics. a.-c. Bar chart depicting the publication year, cancer group, and cohort size distribution of the 244 studies. **d.** Waffle chart showing the percentage of metabolomics methods utilized in the 244 studies: untargeted (red), targeted (orange), targeted and untargeted (yellow), pseudotargeted (blue). **e.** Waffle chart showing the percentage of instrumentation methods utilized in the 244 studies: MS (red), NMR (orange), MS and NMR (yellow). **f.** Waffle chart showing the percentage of 244 studies that utilized age and sex matched cohorts: Both age and sex matched (yes, red), neither age nor sex matched (no, orange), only age matched (age, yellow), only sex matched (sex, blue). Abbreviations: Gyn – Gynecological, GI – Gastrointestinal, HCC – Hepatocellular Carcinoma, CRC – Colorectal Cancer, MS – Mass Spectrometry, NMR – Nuclear Magnetic Resonance. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

which likely increased the availability and size of cohorts and funding opportunities. However, only 5% of the 244 studies examined breast cancer despite it being the fifth leading cause of global deaths. Worldwide, over 684,000 women died from breast cancer in 2020. The suspiciously low number of breast cancer metabolomics studies may be attributed to the fact that research into diseases that primarily affect women is disproportionately underfunded [53].

Sample size and statistical power should be a primary consideration when organizing a clinical study [54]. Almost half (47.7%) of the included studies had <50 samples in their cohort and only 6.2% had more than 200 samples (Fig. 2c). To take conclusions from a small population and extrapolate the results to a larger population, the study must have sufficient statistical power and, subsequently, enough replicates to reach a meaningful outcome [55]. A recent clinical lipidomics study suggests a minimum cohort size should consist of more than 100 samples and 100 controls to attain statistically valid results [56]. Similarly, a large-scale study of 1,200 patients suggested that several hundred samples would be needed to be representative of the population of 1,200 [57]. Furthermore, the analytical method must include suitable quantitative internal standards and appropriate QC samples, and ideally the conclusions need to be verified by a second, independent laboratory to achieve the successful identification of reliable metabolite biomarkers. Unfortunately, and as clearly identified by our meta-analysis, these analytical methods and protocols are rarely adhered to in a clinical metabolomics study. Instead, accessibility, shipping, location, storage, throughput, manpower, practicality and cost concerns are often cited as being detrimental to a study's ability to acquire and analyze several hundred samples. Consequently, these studies may not have enough statistical power to reliably identify metabolic biomarkers. Unsurprisingly, there are currently no widely adopted standards regarding a minimum sample size for metabolomics clinical studies.

The clinical metabolomics data set consisted mostly of untargeted metabolomics studies that used MS (68%) as the analytical method for metabolite detection (Fig. 2d). Untargeted metabolomics has the advantage of being discovery driven and hypothesis generating, which allows for the elucidation of novel biomarkers. Targeted metabolomics studies comprised a smaller proportion (14%) of the clinical metabolomics data set, but often allowed for better absolute quantitation since calibration curves and isotopically labeled standards were implemented into the workflow for the metabolites of interest. A point of potential concern was the observation that only 2% of the clinical studies combined untargeted and targeted methods, which could be used to confirm and validate the potential metabolite biomarkers and provide an absolute quantification of the cancer-induced metabolite concentration changes.

In similar proportions, mass spectrometry (81%) combined with gas or liquid chromatography was the popular choice of analytical method due to its higher sensitivity and broader coverage of the metabolome (Fig. 2e). Nonetheless, NMR (16%) was still commonly used in these clinical metabolomics studies, where investigator experience and expertise are likely factors in the choice of analytical method. Notably and despite the inherent complementarity of NMR and MS, only 3% of the studies used both NMR and MS.

A robust and reliable clinical study necessitates an appropriate study design, which includes, among other considerations, age, and sex matched cohorts. In this regard, it is encouraging to report that ~72% of the clinical metabolomics data set reported utilizing age and sex matched cohorts. Specifically, 60% of the manuscripts reported both age and sex matched cohorts with 8.6% reporting only age matched cohorts and 2.8% reporting only sex matched cohorts (Fig. 2f). However, 28.6% and 11.4% of the manuscripts can be considered as reporting an improperly or incompletely designed clinical trial. Of course, practical considerations and unavoidable limitations may negatively impact the final cohort composition that are out of investigator control, but it still raises serious concerns about bias, and the reliability and applicability of the study's outcomes. A 1:1 matching between controls and cases is a commonly accepted cohort design where adding more controls may only increase statistical power up to a 4:1 ratio [58]. Major confounding factors such as age and sex should always be matched to avoid or minimize bias [59]. Simply, it has been well-documented that human diseases manifest differently according to the sex and age of the individual [60]. Thus, age and sex matched cohorts should be the standard practice for all metabolomics studies. This combination of inadequate and diverse designs of clinical metabolomics studies will likely negatively impact the reproducibility, reliability, and accuracy of the cancer biomarkers identified from the metabolomics data set.

2.2. Variability of statistical techniques used in clinical metabolomics studies

The type of control group chosen is important for biomarker discovery. Conversely, changing the control group could significantly affect the number and type of metabolites identified as disease-dependent, and dictate the specific utility of these disease biomarkers. For example, the choice of control group would determine if biomarkers were useful for diagnosing cancer (*i.e.,* healthy controls), identifying the cancer stage (*i.e.,* stage 1 cancer patients), or for precision medicine (*i.e.,* cancer patient prior to initiating treatment). Accordingly, 88% of studies comprising the clinical metabolomics data set used healthy individuals as a control group, 54% of the studies used individuals with a related disease, and 43% of the studies used both healthy individuals and patients with a related disease (Fig. 3a). Using a related disease as a control may be beneficial to metabolite biomarker discovery given the potential of narrowing and focusing the outcomes to the specifics of the cancer type being investigated. In effect, common responses to any disease, like an immune response, may be canceled out and the remaining dysregulated metabolites would presumably be a direct result of the cancer type. However, it is still possible to miss metabolites of interest that may vary moderately between the related disease and cancer. Additionally, choosing the correct related disease can be challenging. Is it best to choose a benign tumor, an inflammatory disease, or an earlier stage of cancer? Despite these potential issues, we believe the benefit of adding manuscripts that used a related disease as a control and maximizing the number of replicate studies negated any other concerns. Nevertheless, the choice and diversity of control groups used in the metabolomics data may negatively impact the reproducibility and the reliable application of any proposed cancer biomarker.

Multiple hypothesis or false discovery rate (FDR) correction is another key factor that directly determines the number of metabolites identified as statistically dysregulated by cancer. Accordingly, all metabolomics studies need to apply an FDR or equivalent protocol because errors propagate exponentially as each additional metabolite of statistical significance is added to a *set* (eqn. (1)):

$$p = 1 - (1 - \alpha)^m \tag{1}$$

where p is the p-value, m is the number of metabolites and α is the significance level , usually 0.05 or less. Again, as the number of potential metabolite biomarkers increases the likelihood of falsely rejecting the null hypothesis (*i.e.,* false positives) increases [61]. Troublingly, only 45% of the clinical metabolomics data set employed any type of multiple hypothesis correction method (Fig. 3b). This is a modest improvement over the 34% of the NMR metabolomics studies published in 2020 that used FDR, but it is still a serious concern [62]. Of the 110 multiple hypothesis corrected studies, 41% utilized the Benjamini-Hochberg FDR correction method, where 15.5% of studies used the Bonferroni method (Table S9). Surprisingly, while other papers (24%) mentioned the use of a multiple hypothesis correction, the specific test employed was not reported. Despite the common omission of an FDR correction, 86% of studies did report the application of a p-value <0.05 as a threshold for statistical significance (Table S10). While this p-value is a popular choice for statistical significance, our prior meta-analysis of PDAC

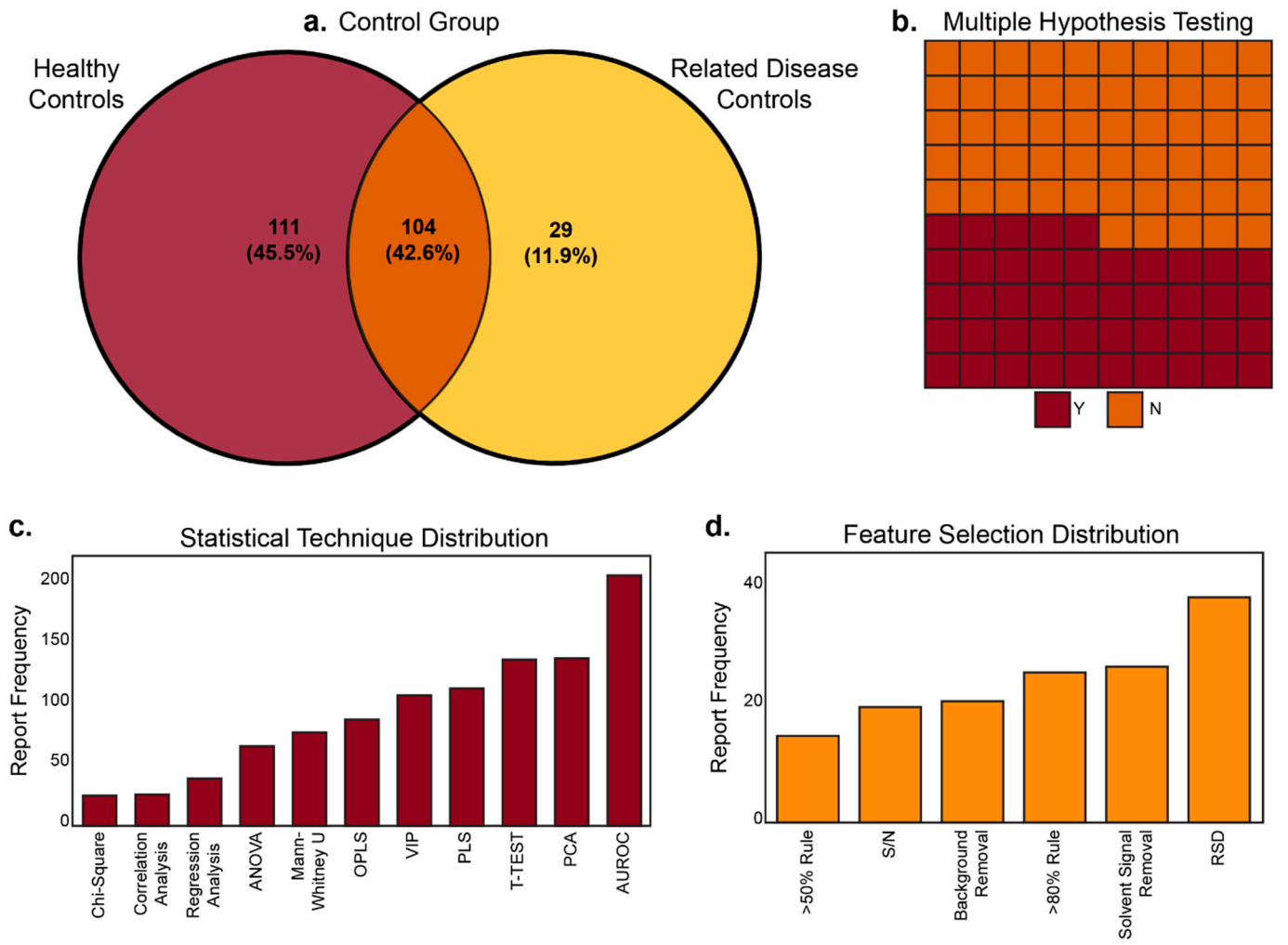


Fig. 3. Cohort Characteristics (cont.). a. Venn diagram showing the number and percentage of the 244 studies that used either a healthy control, related disease control, or both. **b.** Waffle chart showing which percentage of studies implemented any form of multiple hypothesis testing: Yes (red), No (orange). The full distribution and type of multiple hypothesis testing correction method used can be found in Supplementary Table S9 c-d. Bar chart depicting the number of times that each statistical technique or feature selection method was used across the 244 studies. The full distributions can be found in Supplementary Tables S11–S12. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

metabolite biomarkers suggested an aggressive choice of *p*-value may be partly responsible for the low reproducibility of clinical metabolomics studies. Unfortunately, only 7.4% of the studies in our cancer metabolomics data set utilized a more conservative choice of *p*-value <0.01, 0.005 or 0.001. More concerning was the observation that 3.7% of papers used much higher *p*-value thresholds of <0.1, <0.15 or <0.25, where 2 %of studies did not even report a *p*-value. The common omission of an FDR correction and the range of p-value choices is expected to contribute to a potentially low reproducibility in the cancer biomarkers identified from the clinical metabolomics data set.

Metabolomics data sets were usually analyzed with a combination of univariate and multivariate statistical techniques. Common multivariate techniques included principal component analysis (PCA), orthogonal projection to latent structures (OPLS) or partial least squares (PLS). Multivariate techniques should be validated by using a permutation test, CV-ANOVA, or ideally a cross-validation technique that involves dividing the data into training and validation sets. Unfortunately, and as we previously observed, only 30–40% of studies properly validated their statistical models [62].

Unsurprisingly, t-tests, Mann-Whitney U tests, and ANOVAs were frequently used for a univariate statistical analysis (Fig. 3c–Table S11). Notably, multivariate statistical methods combined with an area under the receiver operating characteristic curve (AUROC) analysis [63,64] was the overwhelming choice in 89% of the studies. In general, investigators used multiple parametric ROC curves based on a set of metabolites that individually passed common minimal parameters like VIP (>1) and fold change (>2). Typically, the multiple parametric ROC curves were iteratively optimized to identify the optimal set of metabolites that yielded the best overall predictive outcome. While AUROC is a valuable approach to assess how well a metabolomics model may predict a cancer diagnosis, community standards or best practices have not been established or widely adopted for its application. For example, what measurables and parameter settings (*i.e.,* VIP >1, FC > 2, p-value <0.05, *etc.*) should be used to include or exclude a metabolite from a ROC curve analysis? What is the maximum number of metabolites (*i.e.,* 5, 10, 25, *etc.*) that should be used in a ROC curve analysis? The perceived accuracy of any multiparameter fit nearly always appears to improve with the number of added variables, but this is also likely to lead to an over-fit and unreliable model. Also, the ROC model should be cross validated by separating the data set into a test and validation set, which requires a significantly larger cohort than commonly available (Fig. 2c). The variability in the use and application of ROC curves across the clinical metabolomics data set may contribute to a high variability in the identification of potential cancer biomarkers.

2.3. Various feature selection methods used in clinical metabolomics studies

After the data has been acquired, the feature selection and filtering protocol used to curate and analyze the metabolomics data can be as deeply impactful to the outcomes of a biomarker study as the choice of statistical methods. Multiple literature reviews are available that detail the options, merits, and limitations of feature selection and filtering techniques [65–68], but a community consensus regarding best practices has not been established or widely adopted. Instead, a diversity of protocols is routinely employed by metabolomics investigators based on their experience, data structure, and other relevant concerns. The feature selection and filtering techniques most reported in our clinical metabolomics data set are shown in Fig. 3d. The complete distribution is listed in Table S12. Concerningly, many papers (41%) did not list specific feature selection parameters, or the methodology description was too vague to categorize. Removing known solvent signals, background, and peaks under the limit of detection was a routinely used method of data cleaning and simplification. The most common choice for feature selection was a minimum threshold for an individual metabolite being present within a group. A minimum threshold approach was reported in 63 or 26% of the studies (Fig. 3d). With this approach, the metabolomics field routinely employs the "80% rule", which excludes any metabolite that is present in less than 80% of the samples within a group. The 80% rule is rather arbitrary and many of the publications in the clinical metabolomics data set used alternative cutoffs that ranged from 20% to 100%. The second most common choice for a feature selection method was the percent relative standard deviation (RSD) cutoff, which was reported in 39 or 16% of the studies. A threshold RSD value of <30% was typically used to exclude metabolites by studies utilizing this method. As with other study design decisions, the diversity and lack of details regarding feature selection and filtering protocols will affect the metabolites identified as potential cancer biomarkers leading to reproducibility and accuracy concerns.

2.4. Low reproducibility of metabolites reported as potential cancer biomarkers

A total of 2,206 unique metabolites were reported as statistically significantly changing in the serum of cancer patients across the 244 studies comprising our clinical metabolomics data set. Only 624 (28%) out of the 2,206 metabolites were reported by more than one study (Fig. 4a). Fig. 4b shows the number of manuscripts reporting each of these metabolites, where 480 (77%) out of 624 were reported by only 2 to 5 studies. The top five reported metabolites were glutamine and glutamic acid (59 studies), alanine (49 studies), and lactic acid and tyrosine (45 studies). The top 50 most highly reported metabolites were only detected by 6–24% of the 244 studies. Even more concerningly, 1,582 (72%) out of the 2,206 metabolites were reported by a single study and are likely false positives representing the potential noise level of clinical metabolomics studies. A heatmap (Fig. S1) based on the 561 metabolites detected in two or more colorectal (CRC), esophageal, gastrointestinal (GI), hepatocellular carcinoma (HCC) or lung cancer studies and hierarchically clustered according to cancer type is clearly random and is not dictated by any biological similarity. Most of the clinical metabolomics studies indicated a relative change (*i.e.,* increasing or decreasing) in the metabolite's serum concentration for cancer patients. Troublingly, 4 studies corresponding to a total of 220 potential metabolite biomarkers did not indicate a relative concentration change and these metabolites were excluded from our study. Absolute metabolite concentrations were rarely reported.

The low reproducibility of cancer biomarkers may be a simple artifact of grouping together the data from the 19 distinct cancer groups (Table 1, Table S6). The metabolic diversity between and within cancer subtype may mask any cancer specific biomarkers, especially considering the large range in clinical studies (1–36) available per cancer group (Fig. 2b). Additionally, it is important to consider the technical variations inherent to each metabolomics study. As outlined in **Sections 1.3 and 1.4**, the lack of widely adopted best practices and the resulting large diversity in experimental protocols has likely contributed to these differences in identified metabolites and concentrations. Large, biologically relevant metabolite concentration ranges may also affect the detected metabolic profile. For example, glucose is the most abundant metabolite found in serum (5 mM) followed by urea (4 mM) and amino acids such as glutamine (500 μM), alanine (500 μM), glycine (350 μM), and lysine (350 μM) [69]. These highly abundant metabolites may interfere with the detection of other low abundant metabolites, especially considering the different choices of experimental protocols.

We assessed the consistency of these metabolite concentration changes across the data set for the 36 metabolites reported in 20 or more

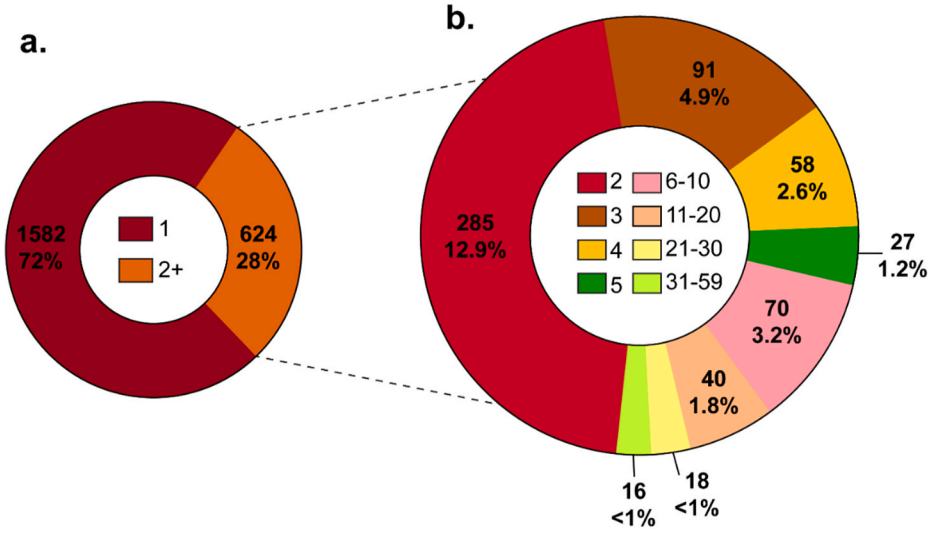


Fig. 4. Metabolite Report Frequency. a. Ring chart showing ratio of metabolite report frequency. Singly reported metabolites are shown in red (72%) and metabolites reported at least twice are shown in orange (28%). **b.** Ring chart showing a more detailed breakdown of the metabolites reported by two or more studies (2+). Metabolites counted exactly twice are shown in red (12.9%), 3x shown in orange (4.9%), 4x shown in yellow (2.6%), 5x shown in green (1.2%), 6–10x shown in pale red (3.2%), 11–20x shown in pale orange (1.8%), 21–30x shown in pale yellow (<1%), and 31–59x shown in pale green (<1%). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Summary of clinical metabolomics data set.

Cancer Group	No. Studies[b]	No. Significant Metabolites[a]			
		Mean	STDEV	Min	Max
Biliary Tract Cancer	3	22	18	9	34
Bone Cancer	2	49	9	42	55
Breast Cancer	12	18	16	3	57
Colorectal Cancer	35	27	44	2	240
Esophageal Cancer	18	20	26	1	103
Gastrointestinal Cancer	28	17	14	1	55
Glioblastoma (Brain Cancer)	1	3		3	3
Gynecological Cancer	17	23	26	4	100
Head and Neck Cancer	10	28	54	3	181
Hepatocellular Carcinoma	31	19	19	1	102
Kidney Cancer	12	17	17	2	64
Leukemia	5	14	10	5	30
Lung Cancer	36	26	35	2	149
Lymphoma	3	20	6	17	27
Myeloma	4	17	18	2	42
Prostate Cancer	13	16	16	2	59
Skin Cancer	2	22	5	18	25
Thyroid Cancer	4	26	26	5	64
Urothelial Carcinoma	8	12	10	5	33

[a] The mean, standard deviation, minimum, and maximum number of metabolites identified across the clinical metabolomics studies identified to be dysregulated in the associated cancer group.
[b] The number of manuscripts in the clinical metabolomics data set associated with the listed cancer type.

studies. While there is significant scatter in Fig. 5a, not surprisingly the average ratio of change across the values is centered around a 51% ± 16% increase and 49% ± 16% decrease, which is expected for a random outcome of two options. In general, as more studies reported the metabolite as changing in the serum of cancer patients, the percentage of studies reporting the same concentration trend decreased towards zero (Fig. 5b). The metabolites responsible for the spikes in the trend correspond to metabolites such as lactic acid and glutamic acid for CRC, glutamine for esophageal cancer, glutamic acid for HCC, and hippuric acid, palmitic acid, phenylalanine and LPC (16:1) for lung cancer. Interestingly, when analyzing consistency across all cancers, there were only two metabolites (pipecolic acid (PC (38:6)) and methyladenosine) that showed a consistent trend across 6 or 8 studies, respectively. Nevertheless, these metabolites were reported in less than 4% of the 244 studies and cannot be described as biologically significant but do warrant further considerations in future studies.

2.5. Impact of cancer type on metabolites reported as potential biomarkers

An analysis of the type of metabolites reported by two or more studies shows that amino acids were commonly dysregulated across all cancer types (Fig. 5a). These results are consistent with the well-known Warburg effect that has been shown to disrupt amino acid metabolism across multiple cancer types [70,71]. Nevertheless, and despite this expected outcome, even glutamine, which was the most commonly reported and abundant amino acid, was only reported 59 times across the 244 studies (24%). Other commonly reported metabolites included lactic acid (18%), glucose (18%), fatty acid amides (15%), and some phospholipid species (13%).

A further analysis of these commonly reported metabolites by cancer group once again showed no clear pattern (Fig. 5c). For clarification, Fig. 5c is an expanded view of Fig. 5a, color coded by cancer type studies that indicate either an increase or decrease in the concentration of the metabolite. Both Mosaic plots in Fig. 5c were normalized to 100% (compared to the original percentages plotted in Fig. 5a) to enhance the visualization of low percentage cancer groups. These and other metabolites exhibited an equal likelihood to be increased or decreased in the serum of different cancer patients. For example, Glu is commonly reported as increasing in the serum of 14 different types of cancer, but it has also been identified as decreasing in 9 other cancers. LPC (14:0) appears to only be increasing in lung and biliary tract cancer, but it has been shown to decrease in eight different cancers. Carnitine is only reported as decreasing in esophageal cancer, but it has been reported to increase in 9 other cancers. Overall, the lack of consistency in reporting potential cancer biomarkers across the entire data set or relative to any specific cancer type, or the high variability in concentration trends, raises serious concerns of the robustness and utility of these metabolites as diagnostic markers of cancer (Fig. 5).

2.6. Experimental factors correlated with cancer biomarkers

A thorough analysis of the effects of study parameters on metabolite reproducibility revealed several notable trends (Fig. 6). Unsurprisingly, studies that used both a targeted and untargeted approach to metabolomics reported an increase in the total number of statistically significant metabolite changes compared to studies that only relied on a targeted or untargeted approach (Fig. 6a). However, the total number of reported metabolites was completely independent of several other study design factors including instrumentation method (Fig. 6c), cohort size (Fig. 6e), journal impact factor (Fig. 6g), and multiple hypothesis testing (MHT) usage (Fig. 6i). Instead, only a large variability was observed in the total number of reported metabolites. While these trends are important to note, the total number of significant metabolites may not be the best indicator of consistency.

The number of metabolites reported by multiple studies may be a better surrogate for the reproducibility and reliability of potential cancer biomarkers. In this regard, the total number of dysregulated metabolites was normalized to the number of studies in each category, which produced several additional trends. The normalized number of metabolites increased as the methodology changed from untargeted, to targeted, and then to a combination of both untargeted and targeted (Fig. 6b). A similar statistically significant increase occurred as the analytical technique changed from MS, to NMR, and then to a combination of both NMR and MS (Fig. 6d). NMR and targeted metabolomics provide for an absolute quantification of metabolite changes in biofluids obtained from cancer patients, presumably leading to a higher accuracy and precision in the identified cancer metabolite biomarkers relative to MS and untargeted metabolomics.

Other factors such as cohort size, journal impact factor and MHT usage were also assessed to ascertain their contributions to cancer biomarker reproducibility. A larger cohort provides for a greater statistical power, which is expected to lead to a robust outcome and more reliable cancer biomarkers. All cohorts with <200 participants exhibited similarly low levels of metabolite reproducibility, but a dramatic increase occurred as the number of cohort participants surpassed 200 patients (Fig. 6f). A surprisingly small but statistically significant increase was seen with MHT usage (Fig. 6j). The application of MHT or a false discovery rate correction would be expected to increase the reproducibility of cancer biomarkers by decreasing type I errors. Unfortunately, MHT was only employed by 45% of the studies (Fig. 3c). It is interesting to note that the use of MHT was correlated with journal impact factor (Fig. 6k), where MHT usage improved as the impact factor reached 3–4 and higher. Unexpectedly and aside from affecting MHT implementation rate, the journal impact factor had no meaningful influence on biomarker reproducibility as articles published in a journal with an impact factor of <1 or 10–20 had the same number of multiply reported metabolites (Fig. 6k).

2.7. Biomarkers need a minimum of three or more independent reports of significance

If a metabolite is a true diagnostic biomarker of cancer, it would be expected to be reported as significantly altered in a high percentage of comparative studies. General diagnostic guidelines indicate that a

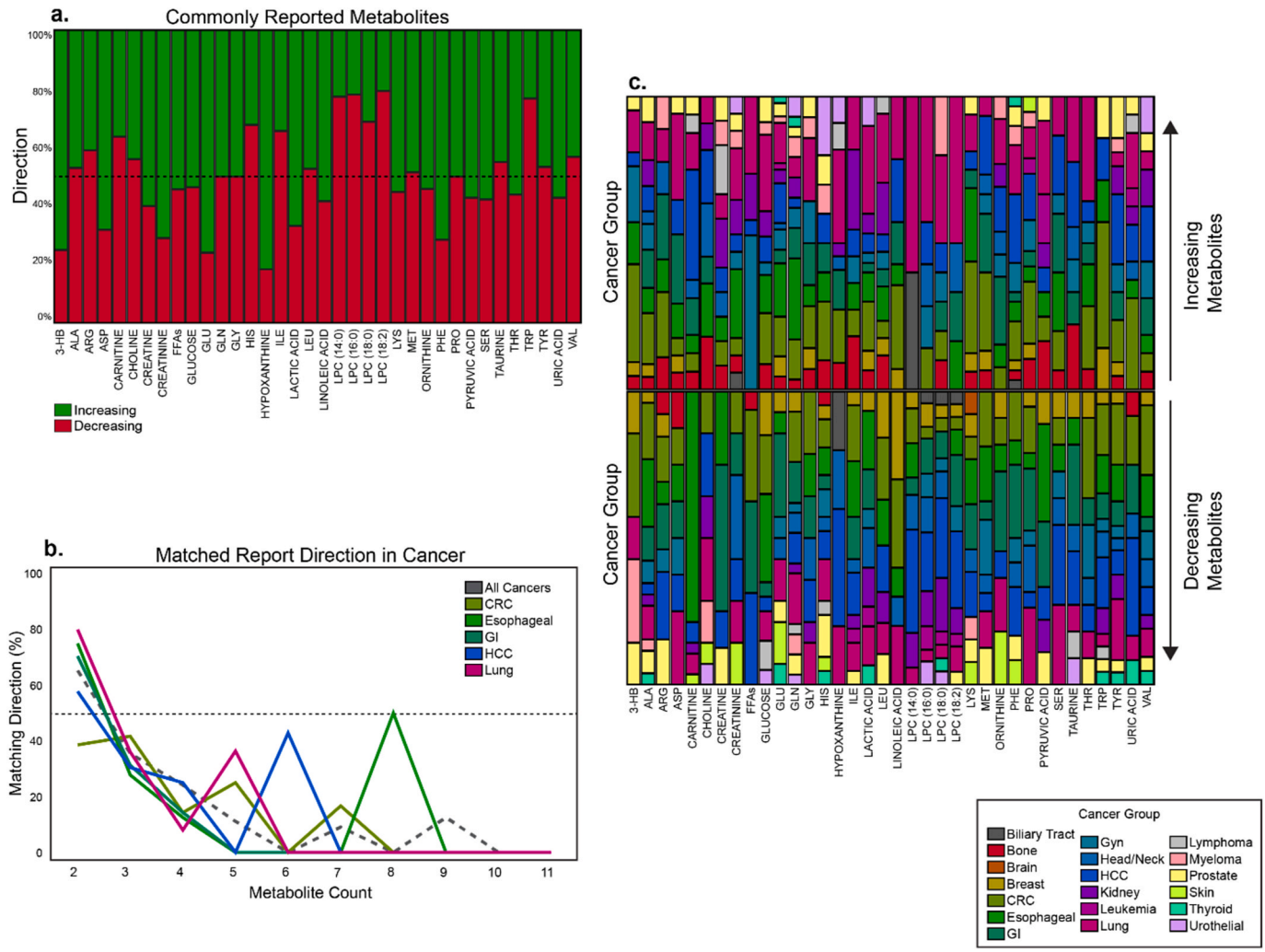
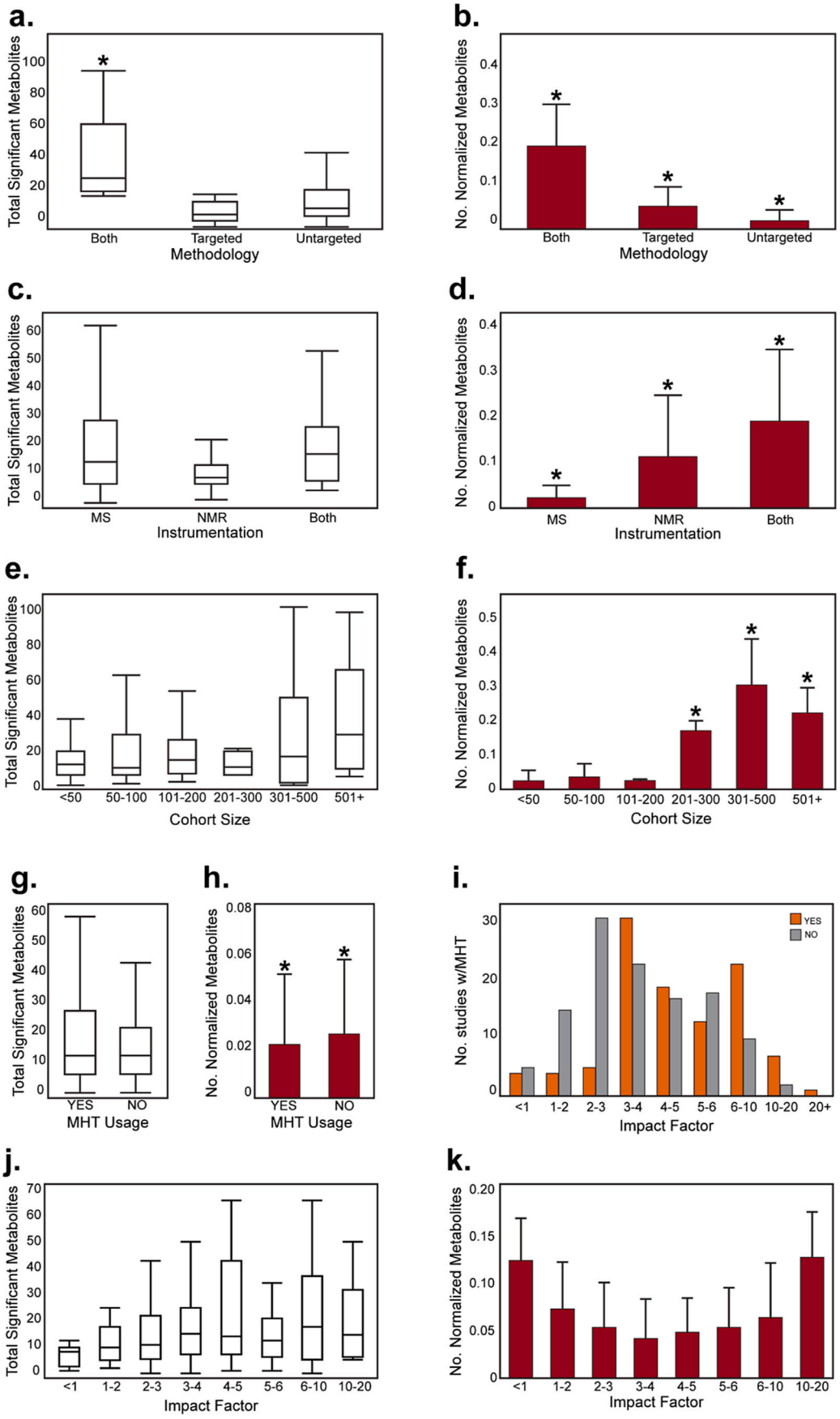


Fig. 5. Common Metabolites by Direction and Cancer Group. a. Mosaic plot for the 36 metabolites that were detected by at least 20 clinical metabolomics studies. The percentage of studies that reported an increase in the metabolite's concentration in cancer are colored green. A reported decrease in the metabolite's concentration is colored red. Metabolite order is alphabetical left-to-right. There is a dashed line across the chart at 50% for reference. **b.** Line plot depicting the percentage of relative concentration changes (*i.e.,* increasing or decreasing) that were matched for metabolites detected by 2–11 different clinical metabolomics studies. All 19 cancers are depicted by a grey, dashed line with individual cancer types plotted as: CRC (olive), esophageal (green), GI (teal), HCC (blue), and lung (pink). The dotted line indicates 50%. **c.** An expanded view of the Mosaic plot in **c** color coded by the percentage of cancer type studies that identified the metabolites as (*top*) increasing or (*bottom*) decreasing. Please note, the y-axis was normalized to range from 0 to 100% for both the increasing and decreasing metabolite Mosaic plots to clearly visualize the low percentage cancer types. The metabolites are listed in alphabetical order. Abbreviations: 3-HB – 3-hydroxybutyrate, ALA – alanine, ARG – arginine, ASP – aspartic acid, CRC – colorectal cancer, FFAs – free fatty acids, GI – Gastrointestinal, GLU – glutamic acid, GLN – glutamine, GLY – glycine, Gyn – gynecological, HCC – hepatocellular carcinoma, HIS – histidine, ILE – isoleucine, Leu – leucine, LPC – lysophosphatidylcholine, MET – methionine, PHE – phenylalanine, PRO – proline, SER – serine, THR – threonine, TRP – tryptophan, TYR – tyrosine, VAL – valine. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

"good" molecular biomarker achieves an 80% sensitivity where a 60% threshold is considered "acceptable" [72]. The best-known cancer biomarker is the prostate-specific antigen,

which has an 86% detection sensitivity for prostate cancer [73]. However, another common cancer biomarker, cancer antigen 15–3 (CA 15–3), which is used to detect the presence of breast cancer, only has a reported sensitivity of 55.6% [74]. Unfortunately, our meta-analysis indicates that the 2,206 potential metabolic biomarkers for cancer fall far short of this range of values. In fact, the best performing metabolites in the clinical metabolomics data set, glutamate, glutamine, alanine, and lactic acid, were only reported as being statistically dysregulated in less than 26% of the clinical metabolomics studies. This is despite the well-established fact that these metabolites are disrupted in cancer [70, 75–77]. The clear lack of reproducibility across the 244 clinical metabolomics studies warrants a further analysis to determine if the observed outcomes are the result of a truly random process or if any of the potential biomarkers are statistically relevant. To address this possibility, a statistical analysis of the entire metabolomics data set was conducted to identify the number of times a metabolite needs to be detected across multiple studies to be classified as a statistically relevant or as statistical noise.

A lower threshold of significance was determined by implementing two independent statistical approaches: empirically modeling the metabolite count by bootstrapping and fitting the one-inflated positive Poisson generalized linear model to the metabolite count. The 95th percentile was calculated to determine the threshold value. The models were fit to the entire data set as well as to the top five most abundant cancer subsets present in the population (Table 2). The models showed that any metabolite detected only 1 to 2 times across all 244 studies as irrelevant at the $\alpha = 0.05$ level and should not be classified as a statistically significant metabolic biomarker for cancer. This represents the statistical noise in the data set and reduces the pool of total metabolites



(caption on next page)

Fig. 6. Study Parameter Effects on Reproducibility. Box and Whisker plot and bar chart comparing the **a.** Total number of significant metabolites and **b.** Total number of significant metabolites normalized to the total number of studies reporting the metabolites to the metabolomics methodology. Box and Whisker plot and bar chart comparing the **c.** Total number of significant metabolites and **d.** Total number of significant metabolites normalized to the total number of studies reporting the metabolites to instrumentation method. Box and Whisker plot and bar chart comparing the **e.** Total number of significant metabolites and **f.** Total number of significant metabolites normalized to the total number of studies reporting the metabolites to cohort size. Bar chart comparing the **g.** total number of significant metabolites and **h.** Total number of significant metabolites normalized to the total number of studies reporting the metabolites to multiple hypothesis testing (MHT) usage. **i.** Bar chart plotting the number of studies with (orange) or without (grey) the usage of MHT plotted against the journal impact factor. Box and Whisker plot and bar chart comparing the **j.** Total number of significant metabolites and **k.** Total number of significant metabolites normalized to the total number of studies reporting the metabolites to journal impact factor. * - denotes significance at $p < 0.05$ across all groups. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Lower thresholds for statistical significance.

Cancer Type	No. Studies	No. Metabolites	Lower Threshold	Threshold Metabolites	% Total
All	244	2206	2	339	15.4
Lung	36	622	2	73	11.7
Colorectal	35	654	4	21	3.2
Hepatocellular	31	341	4	14	4.1
Gastrointestinal	28	325	3	15	4.6
Esophageal	18	281	4	9	3.2

by 85%, from 2,206 to only 339 (Fig. 4). When considering only the subset of metabolites associated with a specific cancer type, the reduction in the potential metabolite pool is far greater, a decrease by 88.3–96.8% occurs with the study threshold increasing to 4 to 5 replicate detections (Table 2).

It is important to note that while the model suggests that any metabolite reported three or more times out of 244 studies is statistically significant, the model does not account for any form of biological relevancy. The model's purpose is to determine a lower threshold of statistical insignificance, rather than providing an upper threshold of biological relevance. The impact of these findings on the clinical metabolomics field shows that at least three independent studies need to detect the same metabolite as significantly altered before it should begin to be considered as a statistically relevant cancer biomarker. Quantifying the noise level in metabolomics data is an important piece of information that can guide future studies and better inform the community on when it is appropriate to designate a metabolite as a potential biomarker. However, it is important to realize that while these findings provide interesting insight into the noise of this data set, the model was built from this specific data set and as the field continues to expand and grow, these numbers may change. Thus, it is important to continue to complete systematic analyses of metabolomics biomarker studies to keep expanding our understanding of the collected data.

3. Conclusions

A meta-analysis of 244 clinical metabolomics studies identified a total of 2,206 potential serum biomarkers from 19 different cancer groups. Only 28% of these metabolites were reported by more than one study, where the vast majority, consisting of 1,582 metabolites, were detected by a single study. Only 36 metabolites (1.6%) were detected by 20 or more studies (8%), but even when detected by multiple studies the typical serum concentration change in cancer patients (*i.e.,* increasing or decreasing) was approximately random (~50%) (Figs. 4 and 5a). Our meta-analysis clearly demonstrates that a general metabolic response to cancer does not currently exist in the available data sets. In essence, the metabolic changes observed differ across and within cancer types (Figs. 4b and 5c-d). There was also no definitive evidence of any cancer specific metabolic biomarker. However, we were able to establish a lower detection rate threshold of statistical significance ranging from 3 to 5 replicate detections across all 244 studies that identifies an effective noise level.

Again, the extremely low reproducibility of the 2,206 metabolites reported across the 244 metabolic studies negates any reliable serum cancer biomarker. Instead, our meta-analysis essentially identified an exhaustive and large list of metabolites that are not a robust or best choice for a cancer biomarker. The reasons behind the lack of biomarker reproducibility as partly summarized herein are many-fold and are likely a combination of the wide variety of metabolomics protocols employed by the community leading to inconsistencies in sample collection, handling and storage, lack of widely adopted standard protocols, inappropriate application, and interpretation of statistical models, unreliable or inaccurate nomenclature, and fundamental limitations and discrepancies in software performance. However, these issues are not unique to metabolomics and can be found in other "-omics"s research as well.

We are not the first to conclude that there is a desperate need for standardization across clinical metabolomics studies to improve the reliability and robustness of their outcomes [39,40,42,43]. Hopefully, our meta-analysis provides further evidence to encourage the community to establish and adopt best practices to ensure future successes. COSMOS, MSI, mQACC, and MANA are valuable resources for metabolomics investigators, and are actively providing a variety of recommendations for these best practices [38–43]. For example, MANA and mQACC have published a recent series of manuscripts that provides guidance on the future directions of NMR-based metabolomics, a perspective on minimal reporting standards, and a summary of current best practices employed by the NMR metabolomics community, among other recommendations [78–80]. Thus, one path to addressing the lack of biomarker reproducibility is for the metabolomics community to actively engage with these organizations, help evaluate and develop standard protocols, and readily adopt validated recommendations. The overall poor reproducibility of the metabolite biomarkers identified by these clinical metabolomics studies also strongly identifies the important need to replicate studies with a second, independent laboratory that analyzes the same metabolomics samples to verify the identical dysregulated metabolites are detected with a similar cancer-induced concentration change.

Despite the discouraging outcome of our meta-analysis, it is still plausible that common cancer metabolites and diagnostic biomarkers for cancer may be identifiable from these and other clinical metabolomics data. Simply put, the "metabolic noise" that is currently prevalent in clinical studies and is masking real outcomes and needs to be removed to reveal true metabolic biomarkers of cancer.

CRediT authorship contribution statement

Darcy Cochran: Writing – original draft, Visualization, Formal analysis, Data curation. **Mai Noureldein:** Formal analysis, Data curation. **Dominika Bezdeková:** Formal analysis, Data curation. **Aaron Schram:** Writing – original draft, Validation, Methodology, Formal analysis. **Réka Howard:** Validation, Methodology, Formal analysis. **Robert Powers:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

All data are available as supplementary data.

Acknowledgements

This work was supported in part by funding from the Nebraska Center for Integrated Biomolecular Communication (P20 GM113126, NIGMS). The research was performed in facilities renovated with support from the National Institutes of Health (RR015468-01). The statistical modeling was provided by the Statistical Cross-disciplinary Collaboration and Consulting Lab (SC3L) in the Department of Statistics at the University of Nebraska-Lincoln.

Appendix A Methods

A.1 Selection of Clinical Metabolomics Studies

Clinical metabolomics studies focused on the identification of cancer biomarkers were sourced from four databases: Scopus (https://www.scopus.com), PubMed (https://pubmed.ncbi.nlm.nih.gov/), Web of Science (https://www.webofscience.com/wos), and SciFinder (https://scifinder.cas.org/scifinder/). Separate keyword and abstract searches were submitted to each of the four databases. Queries included all the following terms "biofluid", "biomarkers", "blood", "cancer", "carcinoma", "diagnostic biomarkers", "metabolomics", "metabonomics", "plasma", "sarcoma", "serum", and "tumor", which were separated by the Boolean "or" operator. The exact search parameters used for each individual database are listed in Table S1. The database search results were exported as a standardized tag format file (RIS file format) for the exchange of literature citations and then imported into Endnote 20 (Fig. 1). The initial database search yielded approximately 1,000 manuscripts. After a cursory manual examination of just titles and abstracts, the total number of manuscripts was reduced to approximately 300 studies. Initial exclusion criteria consisted of eliminating non-human and non-serum studies, review articles, methods papers, and studies that did not rely on either NMR or MS for the analysis of the metabolome, and exclusively lipidomics studies. Metabolomics studies that included lipids were included in the meta-analysis, but studies that analyzed lipids and no other metabolites were excluded. The number of manuscripts was reduced to a final total of 234 papers. An exhaustive reading of each of these manuscripts revealed that several studies analyzed multiple cancers simultaneously and, accordingly, each analysis was treated as a separate and unique study, bringing the total number of studies to 244. A complete list of literature citations for the 234 manuscripts is provided in Table S2. A general overall inclusion criterion consisted of diagnostic biomarker studies using human serum to distinguish between cancerous and non-cancerous individuals using an NMR and/or MS analytical platform. Exclusion criteria consisted of removing clinical studies that were exclusively reliant on lipidomics, used animal models, or were studies focused on the identification of prognostic biomarkers or differentiating between different stages of cancer. Manuscripts that utilized both NMR and MS to separately identify potential cancer biomarkers were treated as two distinct clinical studies. Similarly, manuscripts that used NMR or MS and an additional analytical method such as Fourier transform infrared (FT-IR) spectroscopy were also separated into distinct metabolomics projects. Accordingly, and if possible, metabolites reported as significantly dysregulated in cancer patients were cataloged by the analytical method used to identify the metabolite (*i.e.,* NMR or MS detected).

A.2 Data Extraction

Each manuscript was examined at least twice by 2-3 individuals to ensure accurate data extraction. The information recorded from each manuscript included author, journal, impact factor, publication year, cancer type, metabolomics method, statistical methods and the list of metabolites reported to be significantly dysregulated in cancer patients (Fig. 1) . Table S3 lists by manuscript number all the information extracted from each paper. Metabolites were identified as either increasing or decreasing in cancer patients. It was also noted if the manuscript reported an absolute concentration or fold change. An alphabetical list of all metabolites identified as potential cancer biomarkers is provided in Table S4, which includes the number of times each metabolite was identified as increasing or decreasing in a cancer group.

Correlating data across the 234 manuscripts was challenged by the lack of a uniform or consistent nomenclature for cancer type, metabolite name, or experimental protocol. Cancer types were homogenized to form 19 cancer groups to simplify the analysis at the cohort level and to maximize the number of replicate studies. For example, acute lymphoblastic leukemia (ALL) and acute myeloid leukemia (AML) were grouped into the broader leukemia cohort. Similarly, non-small cell lung cancer (NSCLC) and lung adenocarcinoma were placed into the lung cancer group. In regards to metabolite nomenclature, many publications did not provide HMDB [81], KEGG [82], ChemSpider [83] or any other database identification number. Instead, only common names were provided, requiring metabolites and groups to be homogenized manually by name only. Furthermore, specific metabolite structural information such as stereochemistry and regiochemistry was removed due to inconsistency in reporting across the studies. There were 2,876 unique metabolite names reported in the 244 studies, but after grouping similar metabolites together and removing structural information, the list was reduced to 2,206 unique metabolite names. Examples of homogenization included removing abbreviations and stereochemistry, Phe, PHE and L-Phenylalanine were converted to Phenylalanine, and β-D-glucose was converted to glucose; removing bond location information from lipids, 8z, 14z-eicosadienoic acid was converted to eicosadienoic acid; and merging lipid nomenclature, PC(16:0/0:0) was converted to LPC (16:0). Lipid nomenclature followed the protocol recently published by Lipid Maps (https://www.lipidmaps.org/) [84]. Notably, all carnitines and fatty acid amides of various lengths were grouped as the more generic name "carnitine" and "fatty acid amide". Similar nomenclature homogenization was completed for statistical methods. For example, Lilliefors, Shapiro-Wilk, and Kolmogrov-Smirnov tests were grouped into a more general category of normality testing. Similarly, GC-TOF-MS, GCxGC-TOF-MS, and GC-TQ-MS were all added to the GC-MS group. Tables S5–S8 contain the original reported names for cancer types, metabolites, and experimental protocols and the corresponding manually assigned homogenized groups.

A.3 Statistical Methods and Modeling

Statistical analysis of data and figure generation were completed in Microsoft Excel and JMP 17.2.0 (https://www.jmp.com/). Comparisons of groups were completed via Student's *t*-test or one-way ANOVA followed by Tukey's post hoc test. A p-value <0.05 were considered significant. Hierarchical two-way clustering was completed with the Ward's minimum variance method and using standardized data. To evaluate the effects of study parameters on metabolite reproducibility, frequency normalization was completed on the metabolites that were detected multiple times by dividing the metabolite count by the number of studies that reported each metabolite. In this manner, frequency normalization accounted for unequal distributions of study characteristics across the 244 studies.

Two different approaches were employed to determine the threshold of statistical significance for the number of times a metabolite was

documented in a study. Empirical modeling of metabolite data was done using classical bootstrapping of the sample means generated from repeated resampling of the data with replacement. The bootstrapping was implemented using R 4.3.2 (R Core Team, 2023) functions. Threshold values were generated from the resulting empirical distributions. Furthermore, Generalized Linear Models were fit in contrast with the bootstrap techniques [85]. Specifically, One-Inflated Positive Poisson distributions were fit using the vlgm function in addition to the *roipospois* function from the VGAM package (v1.1-9) in R to generate random samples [86]. Finally, the *qoipospois* function in the VGAM package in R was implemented to calculate the percentiles of the theoretical distribution.

Appendix B. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.trac.2024.117918.

References

[1] L.F. Ervik M, M. Laversanne, J. Ferlay, F. Bray, Global Cancer Observatory: Cancer over Time, International Agency for Research on Cancer, Lyon, France, 2021.
[2] F. Bray, M. Laversanne, H. Sung, J. Ferlay, R.L. Siegel, I. Soerjomataram, A. Jemal, CA A Cancer J. Clin. 74 (2024) 229.
[3] R.L. Siegel, K.D. Miller, N.S. Wagle, A. Jemal, CA A Cancer J. Clin. 73 (2023) 17.
[4] D. Mapes, Cancer Care: from 'sledgehammer' to Precision Cellular Therapy, Fred Hutchison Cancer Center, 2023.
[5] D. Crosby, S. Bhatia, K.M. Brindle, L.M. Coussens, C. Dive, M. Emberton, S. Esener, R.C. Fitzgerald, S.S. Gambhir, P. Kuhn, T.R. Rebbeck, S. Balasubramanian, Science 375 (2022) eaay9040.
[6] G. Lyratzopoulos, P. Vedsted, H. Singh, Br. J. Cancer 112 (Suppl 1) (2015) S84.
[7] G. Rubin, A. Berendsen, S.M. Crawford, R. Dommett, C. Earle, J. Emery, T. Fahey, L. Grassi, E. Grunfeld, S. Gupta, W. Hamilton, S. Hiom, D. Hunter, G. Lyratzopoulos, U. Macleod, R. Mason, G. Mitchell, R.D. Neal, M. Peake, M. Roland, B. Seifert, J. Sisler, J. Sussman, S. Taplin, P. Vedsted, T. Voruganti, F. Walter, J. Wardle, E. Watson, D. Weller, R. Wender, J. Whelan, J. Whitlock, C. Wilkinson, N. de Wit, C. Zimmermann, Lancet Oncol. 16 (2015) 1231.
[8] D.G. Fryback, J.R. Thornbury, Medical Decision Making (1991) 88.
[9] S. Jd, F. Pg, G. P, Am Soc Clin Oncol Educ Book 35 (2015) 57.
[10] K.A. Phillips, S.V. Bebber, A.M. Issa, Nat. Rev. Drug Discov. 5 (2006) 463.
[11] N.L. Henry, D.F. Hayes, Mol. Oncol. 6 (2012) 140.
[12] A.A. Crook, R. Powers, Molecules 25 (2020).
[13] M.J. Jeppesen, R. Powers, Magn. Reson. Chem. 61 (2023) 628.
[14] M. Saoi, P. Britz-McKibbin, Metabolites (2021) 11.
[15] M. Cuperlovic-Culf, D.A. Barnett, A.S. Culf, I. Chute, Drug Discov. Today 15 (2010) 610.
[16] J.L. Griffin, R.A. Kauppinen, J. Proteome Res. 6 (2007) 498.
[17] L.M. Poisson, H. Suhail, J. Singh, I. Datta, A. Denic, K. Labuzek, M.N. Hoda, A. Shankar, A. Kumar, M. Cerghet, S. Elias, R.P. Mohney, M. Rodriguez, R. Rattan, A.K. Mangalam, S. Giri, J. Biol. Chem. 290 (2015) 30697.
[18] M.G. Adam, G. Beyer, N. Christiansen, B. Kamlage, C. Pilarsky, M. Distler, T. Fahlbusch, A. Chromik, F. Klein, M. Bahra, W. Uhl, R. Grützmann, U. M. Mahajan, F.U. Weiss, J. Mayerle, M.M. Lerch, Gut 70 (2021) 2150.
[19] G.D. Tredwell, J.A. Miller, H.H. Chow, P.A. Thompson, H.C. Keun, J. Proteome Res. 13 (2014) 883.
[20] D.J. Panyard, K.M. Kim, B.F. Darst, Y.K. Deming, X. Zhong, Y. Wu, H. Kang, C. M. Carlsson, S.C. Johnson, S. Asthana, C.D. Engelman, Q. Lu, Commun. Biol. 4 (2021) 63.
[21] Y. Ma, P. Zhang, F. Wang, W. Liu, J. Yang, H. Qin, Ann. Surg. 255 (2012) 720.
[22] T. Buergel, J. Steinfeldt, G. Ruyoga, M. Pietzner, D. Bizzarri, D. Vojinovic, J. Upmeier Zu Belzen, L. Loock, P. Kittner, L. Christmann, N. Hollmann, H. Strangalies, J.M. Braunger, B. Wild, S.T. Chiesa, J. Spranger, F. Klostermann, E. B. van den Akker, S. Trompet, S.P. Mooijaart, N. Sattar, J.W. Jukema, B. Lavrijssen, M. Kavousi, M. Ghanbari, M.A. Ikram, E. Slagboom, M. Kivimaki, C. Langenberg, J. Deanfield, R. Eils, U. Landmesser, Nat. Med. 28 (2022) 2309.
[23] S. Xu, R.A. Panettieri Jr., J. Jude, Mol. Aspect. Med. 85 (2022) 100990.
[24] E.D.C. Nunes, A.M.B. Filippis, T. Pereira, N. Faria, A. Salgado, C.S. Santos, T. Carvalho, J.I. Calcagno, F.L.L. Chalhoub, D. Brown, M. Giovanetti, L.C. J. Alcantara, F.K. Barreto, I.C. de Siqueira, G.A.B. Canuto, Pathogens 10 (2021).
[25] D.R. Schmidt, R. Patel, D.G. Kirsch, C.A. Lewis, M.G. Vander Heiden, J.W. Locasale, CA A Cancer J. Clin. 71 (2021) 333.
[26] A. Le Gouellec, C. Plazy, B. Toussaint, Frontiers in Analytical Science 3 (2023).
[27] J.D. Odom, V.R. Sutton, Clin. Chem. 67 (2021) 1606.
[28] J.A. Kirwan, Nature Reviews Bioengineering 1 (2023) 228.
[29] A. Bodaghi, N. Fattahi, A. Ramazani, Heliyon 9 (2023) e13323.
[30] E. Drucker, K. Krapfenbauer, The EPMA, Journal 4 (2013).
[31] R. Simon, J. Natl. Cancer Inst. 107 (2015).
[32] F.S. Ou, S. Michiels, Y. Shyr, A.A. Adjei, A.L. Oberg, J. Thorac. Oncol. 16 (2021) 537.
[33] H.J. Issaq, T.J. Waybright, T.D. Veenstra, Electrophoresis 32 (2011) 967.
[34] A. Kaushal, N. Kaur, S. Sharma, A.K. Sharma, D. Kala, H. Prakash, S. Gupta, Vaccines 10 (2022).
[35] M. Palma, E. Ristori, E. Ricevuto, G. Giannini, A. Gulino, Crit. Rev. Oncol. Hematol. 57 (2006) 1.
[36] N.I.o. Health, Types of Blood Tests, U.S. Department of Health and Human Services, National Institutes of Health, 2022.
[37] S. Qiu, Y. Cai, H. Yao, C. Lin, Y. Xie, S. Tang, A. Zhang, Signal Transduct. Targeted Ther. 8 (2023) 132.
[38] R.D. Beger, W. Dunn, M.A. Schmidt, S.S. Gross, J.A. Kirwan, M. Cascante, L. Brennan, D.S. Wishart, M. Oresic, T. Hankemeier, D.I. Broadhurst, A.N. Lane, K. Suhre, G. Kastenmuller, S.J. Sumner, I. Thiele, O. Fiehn, R. Kaddurah-Daouk, M. for "precision, I. Pharmacometabolomics task group"-metabolomics society, Metabolomics 12 (2016) 149.
[39] O. Fiehn, D. Robertson, J. Griffin, M. van der Werf, B. Nikolau, N. Morrison, L. W. Sumner, R. Goodacre, N.W. Hardy, C. Taylor, J. Fostel, B. Kristal, R. Kaddurah-Daouk, P. Mendes, B. van Ommen, J.C. Lindon, S.-A. Sansone, Metabolomics 3 (2007) 175.
[40] D.W. Bearden, R.D. Beger, D. Broadhurst, W. Dunn, A. Edison, C. Guillou, R. Trengove, M. Viant, I. Wilson, Metabolomics 10 (2014) 539.
[41] R.D. Beger, Metabolomics 15 (2018) 1.
[42] R.D. Beger, W.B. Dunn, A. Bandukwala, B. Bethan, D. Broadhurst, C.B. Clish, S. Dasari, L. Derr, A. Evans, S. Fischer, T. Flynn, T. Hartung, D. Herrington, R. Higashi, P.C. Hsu, C. Jones, M. Kachman, H. Karuso, G. Kruppa, K. Lippa, P. Maruvada, J. Mosley, I. Ntai, C. O'Donovan, M. Playdon, D. Raftery, D. Shaughnessy, A. Souza, T. Spaeder, B. Spalholz, F. Tayyari, B. Ubhi, M. Verma, T. Walk, I. Wilson, K. Witkin, D.W. Bearden, K.A. Zanetti, Metabolomics 15 (2019) 4.
[43] J.A. Kirwan, H. Gika, R.D. Beger, D. Bearden, W.B. Dunn, R. Goodacre, G. Theodoridis, M. Witting, L.R. Yu, I.D. Wilson, A. metabolomics Quality, C. Quality Control, Metabolomics 18 (2022) 70.
[44] F. Bhinderwala, N. Wase, C. DiRusso, R. Powers, J. Proteome Res. 17 (2018) 4017.
[45] I.T. Sakallioglu, A.S. Maroli, A.D.L. Leite, R. Powers, J. Chromatogr. A 1662 (2022) 462739.
[46] C. Martias, N. Baroukh, S. Mavel, H. Blasco, A. Lefevre, L. Roch, F. Montigny, J. Gatien, L. Schibler, D. Dufour-Rainfray, L. Nadal-Desbarats, P. Emond, Molecules 26 (2021).
[47] G.A. Nagana Gowda, D. Raftery, Anal. Chem. 86 (2014) 5433.
[48] H.E. Roth, R. Powers, Cancers 14 (2022).
[49] J. Goveia, A. Pircher, L.C. Conradi, J. Kalucka, V. Lagani, M. Dewerchin, G. Eelen, R.J. DeBerardinis, I.D. Wilson, P. Carmeliet, EMBO Mol. Med. 8 (2016) 1134.
[50] A.J. Smits, L. Botros, M.A.E. Mol, K.A. Ziesemer, M.R. Wilkins, A. Vonk Noordegraaf, H.J. Bogaard, J. Aman, ERJ Open Res 8 (2022).
[51] A. Roointan, M. Ghaeidamini, S. Shafieizadegan, K.L. Hudkins, A. Gholaminejad, Sci. Rep. 13 (2023) 20325.
[52] J. Wang, Y. Sun, S. Teng, K. Li, BMC Med. 18 (2020) 83.
[53] K. Smith, Nature 617 (2023).
[54] D.I. Broadhurst, D.B. Kell, Metabolomics 2 (2006) 171.
[55] V. Tolstikov, A.J. Moser, R. Sarangarajan, N.R. Narain, M.A. Kiebish, Metabolites 10 (2020).
[56] D. Wolrab, R. Jirásko, E. Cífková, M. Höring, D. Mei, M. Chocholoušková, O. Peterka, J. Idkowiak, T. Hrnčiarová, L. Kuchař, R. Ahrends, R. Brumarová, D. Friedecký, G. Vivo-Truyols, P. Škrha, J. Škrha, R. Kučera, B. Melichar, G. Liebisch, R. Burkhardt, M.R. Wenk, A. Cazenave-Gassiot, P. Karásek, I. Novotný, K. Greplová, R. Hrstka, M. Holčapek, Nat. Commun. 13 (2022).
[57] W.B. Dunn, W. Lin, D. Broadhurst, P. Begley, M. Brown, E. Zelena, A.A. Vaughan, A. Halsall, N. Harding, J.D. Knowles, S. Francis-McIntyre, A. Tseng, D.I. Ellis, S. O'Hagan, G. Aarons, B. Benjamin, S. Chew-Graham, C. Moseley, P. Potter, C. L. Winder, C. Potts, P. Thornton, C. McWhirter, M. Zubair, M. Pan, A. Burns, J. K. Cruickshank, G.C. Jayson, N. Purandare, F.C. Wu, J.D. Finn, J.N. Haselden, A. W. Nicholls, I.D. Wilson, R. Goodacre, D.B. Kell, Metabolomics 11 (2015) 9.
[58] J.M. Taylor, Stat. Med. 5 (1986) 29.
[59] M. Iwagami, T. Shinozaki, Annals of Clinical Epidemiology 4 (2022) 33.
[60] F. Mauvais-Jarvis, N. Bairey Merz, P.J. Barnes, R.D. Brinton, J.J. Carrero, D. L. DeMeo, G.J. De Vries, C.N. Epperson, R. Govindan, S.L. Klein, A. Lonardo, P. M. Maki, L.D. McCullough, V. Regitz-Zagrosek, J.G. Regensteiner, J.B. Rubin, K. Sandberg, A. Suzuki, Lancet 396 (2020) 565.
[61] W.S. Noble, Nat. Biotechnol. 27 (2009) 1135.
[62] R. Powers, E.R. Andersson, A.L. Bayless, R.B. Brua, M.C. Chang, L.L. Cheng, C. S. Clendinen, D. Cochran, V. Copié, J.R. Cort, A.A. Crook, H.R. Eghbalnia, A. Giacalone, G.J. Gouveia, J.C. Hoch, M.J. Jeppesen, A.S. Maroli, M.E. Merritt, W. Pathmasiri, H.E. Roth, A. Rushin, I.T. Sakallioglu, S. Sarma, T.B. Schock, L. W. Sumner, P. Takis, M. Uchimiya, D.S. Wishart, TrAC, Trends Anal. Chem. 171 (2024).
[63] N.R. Cook, Circulation 115 (2007) 928.
[64] F.S. Nahm, Korean J Anesthesiol 75 (2022) 25.
[65] Y. Chen, E.M. Li, L.Y. Xu, Metabolites 12 (2022).
[66] S. Ren, A.A. Hinzman, E.L. Kang, R.D. Szczesniak, L.J. Lu, Metabolomics 11 (2015) 1492.
[67] R.P.B. Worley, Current Metabolomics 1 (2013) 92.
[68] A. Cambiaghi, M. Ferrario, M. Masseroli, Briefings Bioinf. 18 (2017) 498.
[69] N. Psychogios, D.D. Hau, J. Peng, A.C. Guo, R. Mandal, S. Bouatra, I. Sinelnikov, R. Krishnamurthy, R. Eisner, B. Gautam, N. Young, J. Xia, C. Knox, E. Dong, P. Huang, Z. Hollander, T.L. Pedersen, S.R. Smith, F. Bamforth, R. Greiner, B. McManus, J.W. Newman, T. Goodfriend, D.S. Wishart, PLoS One 6 (2011) e16957.
[70] E.L. Lieu, T. Nguyen, S. Rhyne, J. Kim, Exp. Mol. Med. 52 (2020) 15.

[71] R.M. Pascale, D.F. Calvisi, M.M. Simile, C.F. Feo, F. Feo, Cancers 12 (2020).
[72] M.S. Pepe, H. Janes, G. Longton, W. Leisenring, P. Newcomb, Am. J. Epidemiol. 159 (2004) 882.
[73] A.M. Wolf, R.C. Wender, R.B. Etzioni, I.M. Thompson, A.V. D'Amico, R.J. Volk, D. D. Brooks, C. Dash, I. Guessous, K. Andrews, C. DeSantis, R.A. Smith, C. American Cancer Society Prostate Cancer Advisory, CA A Cancer J. Clin. 60 (2010) 70.
[74] P. Stieber, D. Nagel, I. Blankenburg, V. Heinemann, M. Untch, I. Bauerfeind, D. Di Gioia, Clin. Chim. Acta 448 (2015) 228.
[75] S.Y. Choi, C.C. Collins, P.W. Gout, Y. Wang, J. Pathol. 230 (2013) 350.
[76] C.T. Hensley, A.T. Wasti, R.J. DeBerardinis, J. Clin. Invest. 123 (2013) 3678.
[77] A. Stepulak, R. Rola, K. Polberg, C. Ikonomidou, J. Neural. Transm. 121 (2014) 933.
[78] G.J. Gouveia, T. Head, L.L. Cheng, C.S. Clendinen, J.R. Cort, X. Du, A.S. Edison, C. C. Fleischer, J. Hoch, N. Mercaldo, W. Pathmasiri, D. Raftery, T.B. Schock, L. W. Sumner, P.G. Takis, V. Copié, H.R. Eghbalnia, R. Powers, Metabolomics 20 (2024) 41.
[79] J.A. Kirwan, L. Brennan, D. Broadhurst, O. Fiehn, M. Cascante, W.B. Dunn, M. A. Schmidt, V. Velagapudi, Clin. Chem. 64 (2018) 1158.
[80] D.S. Wishart, L.L. Cheng, V. Copié, A.S. Edison, H.R. Eghbalnia, J.C. Hoch, G. J. Gouveia, W. Pathmasiri, R. Powers, T.B. Schock, L.W. Sumner, M. Uchimiya, Metabolites 12 (2022).
[81] D.S. Wishart, A. Guo, E. Oler, F. Wang, A. Anjum, H. Peters, R. Dizon, Z. Sayeeda, S. Tian, B.L. Lee, M. Berjanskii, R. Mah, M. Yamamoto, J. Jovel, C. Torres-Calzada, M. Hiebert-Giesbrecht, V.W. Lui, D. Varshavi, D. Varshavi, D. Allen, D. Arndt, N. Khetarpal, A. Sivakumaran, K. Harford, S. Sanford, K. Yee, X. Cao, Z. Budinski, J. Liigand, L. Zhang, J. Zheng, R. Mandal, N. Karu, M. Dambrova, H.B. Schioth, R. Greiner, V. Gautam, Nucleic Acids Res. 50 (2022) D622.
[82] M. Kanehisa, S. Goto, Nucleic Acids Res. 28 (2000) 27.
[83] K.M. Hettne, A.J. Williams, E.M. van Mulligen, J. Kleinjans, V. Tkachenko, J. A. Kors, J. Cheminf. 2 (2010) 3.
[84] G. Liebisch, E. Fahy, J. Aoki, E.A. Dennis, T. Durand, C.S. Ejsing, M. Fedorova, I. Feussner, W.J. Griffiths, H. Kofeler, A.H. Merrill Jr., R.C. Murphy, V. B. O'Donnell, O. Oskolkova, S. Subramaniam, M.J.O. Wakelam, F. Spener, J. Lipid Res. 61 (2020) 1539.
[85] W.W. Stroup, Generalized Linear Mixed Models: Modern Concepts, Methods and Applications, CRC Press, 2016.
[86] T.W. Yee, R Package version 1.1–9 (2023).