

Chiralify: A Framework for Fragmentation Modeling and Advanced Chiral Metabolite Annotation

Ruggiero Gorgoglione

Libera Università Mediterranea

Valeria Impedovo

The University of Texas at Austin

Yujin Lee

The University of Texas at Austin

Alessia Lodi

The University of Texas at Austin

Stefano Tiziani

`tiziani@austin.utexas.edu`

The University of Texas at Austin

Research Article

Keywords: Untargeted chiral metabolomics, enantiomers, Chiralify, metabolite annotation, mass spectrometry, computational metabolomics, Untargeted Metabolomics, Biomarkers Discovery

Posted Date: August 27th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10722928/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Metabolomics enables the comprehensive profiling of small molecules in biological systems and has become essential for investigating disease-associated metabolic alterations and biomarker discovery. Advances in high-resolution mass spectrometry and automated computational workflows have greatly expanded the scale and throughput of metabolite profiling. However, conventional untargeted metabolomics pipelines generally lack the capability to distinguish enantiomeric metabolites based on mass spectrometry data alone. This limitation is critical, as many metabolites exist as mirror-image forms, namely enantiomers, which are detected in human tissues and biofluids and may play distinct roles in disease-related metabolic alterations. Despite advances in chiral separation strategies using chiral selectors or derivatization reagents, the analysis of enantiomeric metabolites remains largely manual, limiting high-throughput analysis and biomarker discovery. Here we introduce Chiralify, an open-source cheminformatics framework designed to enable automated and scalable untargeted chiral metabolomics. Chiralify integrates fragmentation modeling and an expandable spectral library for metabolite annotation with a flexible framework for multiplexed analysis of chiral metabolomics datasets. The platform automates the detection, annotation, and comparative analysis of enantiomeric metabolites, providing a standardized and scalable solution for incorporating stereochemical information into metabolomics workflows.

Scientific Contribution

Chiralify is an open-source cheminformatics framework that automates and streamlines chiral metabolomics analysis, enabling systematic characterization of enantiomeric metabolic heterogeneity. Application to diverse cancer cell lines revealed numerous enantiomeric metabolite pairs and previously unexplored patterns of stereochemical diversity. Chiralify provides a standardized and scalable computational approach that extends cheminformatics to stereochemically resolved metabolomics and establishes stereochemistry as an additional dimension for metabolite annotation and biomarker discovery.

Introduction

Metabolism plays a central role in biological processes by providing energy, biosynthetic precursors, and regulatory signals required for cellular function and homeostasis. The metabolome, therefore, closely reflects the physiological state of an organism, and most diseases are characterized by metabolic alterations at both cellular and systemic levels[1]. Metabolomics, the comprehensive measurement of low-molecular weight molecules within biological systems, has therefore become an essential approach for characterizing metabolic changes associated with disease progression and for identifying disease-associated biomarkers [2–4].

Many metabolites exist as enantiomers, namely mirror-image forms with distinct biological properties, and their presence in human tissues[5] and body fluids[6, 7] is challenging the long-standing notion of a ‘stereoselective’ human metabolism, where only one canonical enantiomeric form exists. Therefore,

metabolic chirality is emerging as an additional, and thus far largely overlooked, layer of metabolic complexity[8–12]. Indeed, several ‘non-canonical’ enantiomers play crucial physiological roles and are associated with disease when their levels are altered or dysregulated[6, 13, 14]. For instance, D-serine and D-aspartate are important neuromodulators in the central nervous system[15, 16], and dysregulation of their homeostasis has been linked with schizophrenia[17–19] and Alzheimer's disease[20], underscoring the importance of investigating metabolic chirality in the pathophysiology of disease.

The field of metabolomics has seen a substantial development in computational platforms and analytical pipelines designed to automate, improve and standardize metabolite annotation[21–28]. These software packages typically rely on metabolite databases[29, 30] and spectral libraries[31–33] for the structural annotation of small molecules. While these platforms support a wide range of data and workflows for untargeted metabolomics, they cannot resolve enantiomeric molecules. Indeed, enantioselective analysis necessitates separation techniques that rely on chiral selectors and derivatizing agents[34–36], thus limiting the application of existing automated annotation workflows. As a result, chiral metabolomics methods usually require extensive manual curation of the results due to the absence of chiral-specific database entries. Although these strategies are effective for a targeted enantiomeric analysis restricted to only a few metabolites[37–39], they are not feasible for untargeted approaches that aim to characterize hundreds of chiral metabolites within a single analytical workflow[40–43].

To address this computational gap, we present Chiralify, an open-source cheminformatics framework for the automated and scalable analysis of untargeted chiral metabolomics data. Chiralify provides a highly automated and customizable framework for the annotation of enantiomeric metabolites. To do so, this platform is tailored for the untargeted detection of chiral amino- and hydroxy-acids simultaneously derivatized with (+)/(-)-diacetyl-tartaric anhydride (DATAN) and acquired using reverse-phase liquid chromatography high-resolution tandem mass spectrometry (RPLC-HR-MS/MS). To address the unique strategies of enantiomeric separation, we established an expandable spectral library of chiral metabolites and integrated it into an automated platform for the multiplexed analysis of chiral metabolomics data. The workflow consists of the following steps: (1) feature annotation based on spectral matching with the customized spectral library, (2) identification of candidate enantiomeric pairs in (+)- and (-)-DATAN derivatized samples based on spectral similarity and chromatographic retention behavior, (3) integration of the two derivatization modes to retain only the metabolite pairs that exhibit both high spectral similarity and a reversal of elution order between (+)- and (-)-DATAN derivatives. By establishing this pipeline, Chiralify considerably streamlines the computational analysis of enantiomeric datasets and provides a novel strategy to boost the untargeted exploration of the chiral metabolome (Fig. 1).

Results

Development of a chiral small-molecule database through in silico derivatization and MS2 fragmentation

To build a reference library of DATAN-derivatized metabolites, we generated a custom database based on the Human Metabolome Database (HMDB, version 5.0). HMDB v5.0 comprises an extensive metabolite collection, only a subset of which is suitable for DATAN derivatization. Compounds containing Mn, Ca, Zn, Mg, Fe, Cu, Br, Co, Mo, K, As, Na, I, Cd, B, Cl, F, Pt or Se were excluded from the dataset.

Subsequently, the RDKit package (version 2022.03.5; open-source cheminformatics toolkit, <https://www.rdkit.org>) was employed to identify compounds containing unconjugated hydroxyl and primary and secondary amine functional groups. Specifically, free hydroxyl groups, either aliphatic or aromatic, not conjugated or bound to carbonyl or other electrophilic centers were selected, thereby excluding carboxylic, esterified, enolic, etherified, and phosphorylated or sulfonated -OH functionalities. Primary amines, both aliphatic and aromatic, were retained only when free and not conjugated with carbonyl, imine, oxime or thiamine groups. Secondary amines (-NH-), both cyclic and acyclic, were included only when not involved in amide, urea or amidine/guanidine linkages. While DATAN reactivity toward hydroxyl and primary amine groups is well established in the literature[44], its reactivity toward secondary amines has not been thoroughly investigated. Nonetheless, in our previous studies we demonstrated the derivatization of several metabolites bearing secondary amine functionalities (e.g., proline and pipecolic acid), indicating that DATAN can also react with secondary amino groups[40]. On this basis, we also included metabolites bearing secondary amines to provide more comprehensive coverage.

Compounds bearing up to two DATAN-reactive sites were retained, resulting in a final selection of 6,187 metabolites. These metabolites were then processed in RDKit to generate DATAN-derivatized molecules. To account for all possible DATAN-derivatized forms in molecules with more than one reactive functional group, we generated both the structure for molecules fully derivatized at all sites, as well as molecules partially derivatized at one of the reactive groups. All DATAN-derivatized molecules were systematically renamed according to a standardized nomenclature scheme (Fig. 2) to indicate the number of bound DATAN and the degree of derivatization (*i.e.*, molecules derivatized at either one functional group or both). The resulting library comprised 9,624 distinct entries.

The *in-silico* MS² spectra of DATAN-derivatized metabolites were generated using CFM-ID 4.4.7[31, 45] (Competitive Fragmentation Modeling for Metabolite Identification) executed within Docker Desktop. The DATAN-derivatized SMILES structures from the RDKit processing were provided as input to the *cfm-predict* function to predict electrospray ionization (ESI) MS/MS spectra, as described in the CFM-ID documentation. Consistent with spectra previously acquired from standards using the analytical method described by Pandey *et al.*[40], the predicted spectra reproduced characteristic neutral losses and revealed distinct fragmentation patterns between -NH₂ and -OH derivatized compounds (**Supplementary Figure S1A**). Two MS² spectral libraries were generated: a reference library produced using the default untrained model available in CFM-ID 4.4.7 (Default-Lib), and a second library constructed with a custom model trained on experimental MS² spectra acquired in this study (Trained-Lib). Model training was performed using 27 experimental MS² spectra acquired from 9 DATAN-

derivatized reference metabolites, each analyzed at two nominal collision energies. Both libraries were constructed in positive electrospray ionization (ESI) mode using identical cfm-predict parameters. The MS² spectra of 94 (out of 9624) DATAN-derivatized molecules present in the RDKit-generated library could not be successfully predicted. Therefore, both the Default- and the Trained-Lib included 9,530 predicted MS² spectra.

This novel library broadens traditional databases for untargeted metabolomics studies by considering small-molecule chirality.

The Chiralify Graphic User Interface for Parameter Configuration

Chiralify features a user-friendly graphical user interface (GUI) designed to facilitate parameter configuration according to experimental conditions (Table 1). The interface guides the user through each step of the data analysis, from feature assignment to candidate enantiomeric pairs identification. Feature tables corresponding to both (+)- and (–)-DATAN-derivatized samples, exported from MS-DIAL, are used as input for the algorithm. The workflow begins by putatively annotating the features, for each derivatization mode, based on observed m/z values against the *in-silico* generated DATAN-derivatized spectral library using a user-defined mass tolerance and selectable adduct forms.

Table 1

Parameter definitions and default settings used for MS¹ and MS² data processing, feature annotation, and clustering into pairs and quartets. The table summarizes all processing parameters, their default or user-selectable values, and their functional roles within the workflow. These parameters are systematically recorded in the results file, which serves both as a user report and as a valid input template for iterative analyses.

Parameter	Variable Name	Default
Minimum m/z MS1 (Da)	ms1_Min	240
Maximum m/z MS1 (Da)	ms1_Max	1000
MS1 annotation tolerance (ppm)	ms1_Toll_Ann	5
MS1 clustering tolerance (ppm)	ms1_Toll_Comp	5
Minimum MS2 similarity score	ms2Score_TRS	0.2
Minimum MS2 fragment count	Nmzlimit	3
MS2 bin width (Da)	t	0.01
Minimum MS2 relative intensity (%)	minInt	10
Minimum m/z MS2 (Da)	ms2_Min	140
Minimum Δ RT intra-QC (min)	RT_min_intra	0.25
Maximum Δ RT intra-QC (min)	RT_max_intra	8
Maximum Δ RT inter-QC (min)	RT_max_INTER	10
Minimum RT inversion score	RTinvScore_Min	0.2
Maximum RT inversion score	RTinvScore_Max	5
Library for annotation	Input_Library	Original/Trained
Selected adducts	adducts	[H] ⁺ [NH ₄] ⁺ [Na] ⁺ [K] ⁺
Adduct m/z shift	massShift	1.00728 18.0338 22.9892 38.9632
Collision energy	fragmentation	msms10HCD msms20HCD
Number of parpool workers	Workers	Max Allowance – 1

Candidate annotations are ranked based on MS² spectral similarity (**Supplementary Figure S1B**). Spectral similarity is evaluated using both a cosine score (*ms2scoreDot*) and reference-restricted cosine score (*ms2scoreRev1*). With this second approach, only the *m/z* values present in the reference spectrum are compared to the experimental spectrum; therefore, the score reflects how closely the experimental spectrum reproduces the fragmentation pattern of the reference. To ensure comprehensive coverage of multiple possible annotations, the algorithm is designed to retain more than one candidate assignment per feature.

Following the putative feature annotation, candidate enantiomeric pairs are independently identified within (+)- and (–)-DATAN-derivatized samples based on m/z match, retention time (RT) proximity and MS² spectral similarity.

RT proximity serves to define the range of allowable difference in elution times between two chromatographic peaks with the same m/z in order for the two peaks to be considered an enantiomeric pair. To determine the minimum and maximum values of the RT range, we implemented two user-defined parameters, namely the *minimum* $\Delta_{RT-intra-QC}$ and the *maximum* $\Delta_{RT-intra-QC}$. The first establishes the minimum distance between two chromatographic peaks to be considered a pair, while the latter establishes the maximum distance between two peaks to be considered a pair (Fig. 3A). A classical cosine similarity metric is used at this stage to evaluate spectral similarity between the two experimental spectra within each candidate enantiomeric pair.

Following the identification of candidate enantiomeric pairs within each derivatization mode, quartets are generated by pairing the corresponding enantiomeric pairs from (+)- and (–)-DATAN-derivatized samples.

The reversal of elution order observed for chiral pairs derivatized with (+)- or (–)-DATAN provides additional support for the untargeted identification of candidate enantiomeric pairs[40, 43]. Therefore, in addition to concordance in mean m/z and mean RT across paired features, the quartet assembly relies on RT inversion and MS² spectral similarity. The *maximum* $\Delta_{RT-INTER-QC}$ defines the tolerated difference in average RT between the two derivatization modes (Fig. 3B) and can be adjusted to account for RT drift across chromatographic runs.

The RT inversion score is calculated by comparing the integrated peak areas between derivatization modes. Only candidate pairs with scores within a user-defined threshold are retained (Fig. 3C). Finally, all features within each quartet are evaluated for MS² spectral similarity.

Overall, the manual configuration of feature assignment and enantiomeric pairing parameters offers a flexible and robust framework for the computational analysis of chiral metabolomics data.

Manual Curation of the Results Using the ChiralApp

We developed the ChiralApp to address the need for extensive manual curation of Chiralify results. The Chiralify output reports in table format all the annotated features and candidate enantiomeric pairs identified by the algorithm based on the user-defined parameters. However, these results require careful manual inspection to ensure the selection of the most plausible annotations, retain the pairs of interest and assess whether the chosen parameters are appropriate or require further optimization through iterative reanalysis. This curation process is challenging because each enantiomeric pair can be associated with multiple ranked assignments, and because the presence of multiple adducts and putative *in-source* fragments generates large and complex datasets that are difficult to compare using

static tables alone. To overcome these limitations, the ChiralApp provides a dedicated graphical interface that facilitates metabolite searching, annotation inspection, and visualization of chromatographic and MS-related information, thereby reducing processing time and supporting informed decision-making during result curation, ultimately improving reproducibility.

The app includes a pairs table displaying annotated enantiomeric pairs, and two feature-specific tables that report parameters for each feature in a pair separately (Fig. 4A), with a metabolite search bar to search for compounds of interest. These interactive tables are linked to a visualization tab, whereby the selection of a candidate pair enables direct inspection of the corresponding chromatograms (Fig. 4B), adducts, and *in-source* fragments (Fig. 4C), allowing the user to efficiently compare multiple annotations for each feature and select the most reliable enantiomeric pairs. A scatter plot and a summary table enable the user to visualize the Chiralify output and manage the manual refinement of the results. The ChiralApp includes a user manual, accessible through the 'Help' button, that provides a detailed description of each feature offered by the app.

Overall, the ChiralApp complements the Chiralify pipeline by transforming large, complex outputs into a curated and easily interpretable set of high-confidence enantiomeric annotations.

Model Training Enhances MS Spectra Prediction and Annotation Accuracy

To validate the chiral small molecule database and the Chiralify pipeline, we evaluated their performance with a standard mixture of 43 chiral and non-chiral metabolites (**Supplementary Table S1**) at two normalized collision energies, namely NCE10 and NCE20. Chiralify parameters were optimized to identify all chiral standards while minimizing the number of false positives and retaining all the correct enantiomeric pairs (**Supplementary Figure S2A-C**).

Our first goal was to assess the ability of Chiralify to annotate metabolites. All the standards in the mixture were annotated according to the established parameters at both collision energies (**Supplementary Table S2**). Moreover, Chiralify removed approximately 50% of the features that did not meet the specified criteria (**Supplementary Figure S3A**). Since 'non-canonical' D-enantiomers are often present in low abundances, we also confirmed the efficacy of Chiralify in identifying and annotating enantiomeric pairs with extreme ratios (**Supplementary Figure S3B**).

We next evaluated the quality of metabolite annotation by estimating the true-positive annotation rate. All the chiral metabolites in the standard mix were correctly identified as chiral pairs. As shown in Fig. 5A, these true positives represented approximately 5% of all the annotated features at NCE10 with the optimized parameters. The remaining annotated features were false positives, which were partially attributed to *in-source* fragmentation and alternative adduct forms not included in the analysis. Notably, Chiralify identifies *in-source* fragments, and all adducts associated with each parent ion, and groups them into clusters (Fig. 5B), thereby facilitating their removal from the results.

Lastly, we examined the annotation performed with the *in-silico* MS² spectral library before and after the model training with experimental MS² spectra. To do so, we compared the MS² annotation score that Chiralify attributed to each metabolite in the standard mix. Overall, both the cosine score and reference-restricted cosine score improved after model training at both collision energies (Fig. 5C-G, **Supplementary Figure S3C-F**), with only a few metabolites showing unchanged or reduced scores. Notably, the MS² annotation score for carnitine decreased after training, possibly due to its fully protonated state at the working pH. Given that multiple annotations per feature are ranked according to their MS² scores from highest to lowest, we further evaluated the ranking position of each standard annotation. The cumulative relative frequency (CRF) calculated for each rank position shows that model training consistently improves annotation ranking at both collision energies, with NCE10 (Fig. 5E) performing better than NCE20 (**Supplementary Figure S3D**), thereby enhancing result interpretability by increasing the proportion of true annotations ranked among the top candidates. Notably, in most cases, when the correct molecule did not score the top rank, the compound assigned a higher score was a structural isomer of the correct molecule. To assess whether the post-training MS² scoring improvement reflected instrument-specific overfitting, we further evaluated the standard mix acquired on an independent platform, an Orbitrap IQ-X Tribrid mass spectrometer. Model training similarly improves the MS² scores and annotation ranking in this dataset (Fig. 5E-F, **Supplementary Figure S3E-F**), supporting the robustness of the training strategy and its effectiveness in enhancing fragmentation prediction.

These results indicate that model training improves the MS² spectra prediction and enhances the annotation accuracy, thus addressing a key limitation of tandem mass spectrometry in resolving subtle structural differences among isomers[46].

Taken together, these results demonstrate the accuracy of the *in-silico* prediction of MS² spectra for DATAN-derivatized metabolites and the efficacy of Chiralify in identifying the enantiomeric pairs. Moreover, we also established a custom training model to improve prediction with empirical MS² spectra and overcome potential instrument-related spectral variation.

Untargeted Analysis of the Enantiomeric Metabolome in Cancer Cells

Chiralify enables the untargeted identification and assignment of novel chiral metabolite pairs in biological samples, providing a powerful tool for biomarker discovery. To assess its performance, the algorithm was applied to a panel of cancer cell lines representing both hematological and solid malignancies (**Supplementary Table S3**). Indeed, chiral small molecules have recently emerged as novel diagnostic biomarkers[37, 43, 47, 48] and therapeutic targets[12, 49] in cancer.

To validate the ability of Chiralify to annotate chiral datasets acquired on cellular samples, we evaluated the quality of the annotation of chiral metabolite pairs included in our in-house standard library[40], for which we know the identity. Overall, 80 to 90% of these metabolites were annotated within the top five ranked candidates (Fig. 6A), with only a small number of metabolites assigned to lower-ranking positions. Compared to the standard validation, these results exhibit lower ranking performance,

primarily due to the increased complexity of MS² deconvolution in biological samples containing a large number of metabolites[50]. Notably, the known chiral metabolites accounted for only a small fraction of the candidate enantiomeric pairs identified by Chiralify, whereas approximately 90% of the pairs corresponded to newly annotated metabolites, underscoring the potential for the discovery of novel chiral metabolites. Moreover, approximately 30% of the identified candidate pairs were classified by the algorithm as in-source fragments or alternative ion adducts and were therefore excluded from downstream analysis.

The distribution of identified enantiomeric pairs across cancer types (Fig. 6B) revealed that glioblastoma (GBM) and pancreatic cancer (PanC) exhibited a higher number of candidate chiral metabolites compared with acute lymphoblastic leukemia (ALL) and prostate cancer (PC). Principal component analysis (PCA) was then applied to reduce data dimensionality and to highlight the major sources of variance across samples. The first two components accounted for 50.8% of the total variance [principal component 1 (PC1): 30.2%, principal component 2 (PC2): 20.6%]. The PCA scores plot (Fig. 6C) demonstrates a clear separation of cell lineages across PC1 and PC2. The corresponding biplot (Fig. 6D) highlights the top metabolites that contributed to the separation of the two PCs. For annotated metabolites with known RT, we could establish the chiral form (L-carnitine, D- and L-glutamic acid and D-leucine), whereas others were reported as D-/L- metabolites due to unknown elution orders. Notably, for some major contributors only one chiral form is responsible for driving the variability across the two PCs, underscoring the importance of stereochemical resolution in characterizing cancer-associated metabolic phenotypes.

Altogether, these results highlight the potential of Chiralify to expand the detection of chiral metabolites in biological samples and reveal them as potential drivers of the metabolic heterogeneity of cancer.

Validation of novel small molecule enantiomers annotations generated by Chiralify

Chiralify holds strong potential for the discovery of novel chiral biomarkers. When coupled with Untargeted Chiral Metabolomics (UCM) of different cancer cell lines, it identified and assigned hundreds of novel candidate enantiomeric pairs. However, new chiral pairs identification requires validation through spectral and retention time matching with a reference standard. For this purpose, we selected one of the newly annotated chiral metabolites and confirmed the assignment with a reference standard. We select 2-aminoadipic acid, a lysine-derivative that has already been proposed as potential biomarker of cancer[51], oxidative stress[52], and pre-diabetic state[53]. Hence, the enantiomeric discrimination of this small molecule may improve the understanding of its perturbation in disease settings.

Our algorithm systematically annotated multiple features as aminoadipic acid across cell lines. Each feature had multiple assignments, including not only aminoadipic acid, but also its structural isomers N-acetyl threonine, acetyl homoserine and hydroxy-butryl glycine. These results demonstrate Chiralify's ability to detect and putatively annotate metabolites from their MS/MS spectra, while also highlighting

an inherent limitation of spectral matching: structurally related isomers can produce highly similar fragmentation patterns, preventing unambiguous identification when additional structural evidence is unavailable[54].

To confirm the correct assignment and attribute the elution order to the two enantiomeric forms, we separately analyzed L- and D-aminoadipic acid standards derivatized with both (+)- and (–)-DATAN. Figure 7A shows that the two chiral forms elute at 14.55 and 15.22 minutes, with L-aminoadipic acid eluting first when derivatized with (+)-DATAN. These results confirm only the feature in Panc10.05 cells eluting at 14.6 minutes in (+)-DATAN-derivatized mode and at 15.4 in (–)-DATAN-derivatized mode is L-aminoadipic acid (Fig. 7B).

These findings highlight that the annotation of structural isomers remains challenging and ultimately requires validation against authentic reference standards. However, future training of our algorithm using larger collections of experimental MS/MS spectra, may enhance scoring performance and further improve discrimination between structural isomers. Thus, this work establishes a framework in which computational identification and prioritization are complemented by targeted experimental validation, providing a robust strategy for advancing chiral metabolomics toward biomarker discovery.

Conclusions

Untargeted metabolomics enables unbiased metabolic profiling, making it a powerful tool for biomarker discovery[13, 55]. For this reason, a rapid improvement of metabolite databases and standardized computational pipelines for data processing and annotation has substantially improved in recent years[22, 24, 56]. However, these platforms are generally not suitable for chiral metabolomics, limiting the systematic investigation of metabolic chirality and often requiring laborious targeted analyses of a small number of enantiomeric pairs[37, 38, 43, 57]. We previously showed that derivatization with DATAN enables the simultaneous separation of hundreds of chiral metabolite pairs containing amino- and hydroxy- groups[40]. Nevertheless, this analytical development has not been accompanied by an improvement in the computational analysis of the results, thus limiting the potential of the method for the untargeted exploration of the chiral metabolome[43, 57, 58]. In this study, we developed Chiralify, an open-source cheminformatics framework for the automated processing of chiral metabolomics data that addresses the requirement for sample derivatization in untargeted enantiomeric analysis.

The chiral small-molecule library provides an expandable resource for enantiomer-specific metabolite annotation, containing more than 9,000 entries comprising both chiral and non-chiral molecules. We established a framework for the *in-silico* derivatization of small molecules and their MS² spectra prediction. To date, the *in-silico* derivatization is limited to DATAN, given it is a well-established derivatizing agent for untargeted chiral metabolomics[40]. However, the modular design of the library and workflow enables the future incorporation of additional derivatization chemistries and metabolite classes, thereby extending the chemical space accessible to Chiralify [41, 42][59]. The current implementation is also limited to compounds containing a maximum of two DATAN-reactive sites and

has been evaluated using data generated with a single analytical method, restricting the current application primarily to amino- and hydroxy-acids. These limitations define the current scope of the platform rather than fundamental constraints of the framework and can be addressed through expansion of the library, derivatization models, and analytical workflows.

We also generated the theoretical ion-dissociation spectra of derivatized molecules through an *in-silico* fragmentation algorithm to support structural annotation and reduce ambiguity in spectral matching. Given the limitations of *in-silico* spectrum prediction and scoring methods[60], we further implemented machine-learning-based model training to improve MS/MS spectral prediction. Although MS/MS prediction overall improved, model training did not fully overcome the challenge of distinguishing structural isomers, which often generate highly similar fragment ions. Consequently, the *in-silico* MS/MS spectral prediction methods often show low accuracy in interpreting structural differences among isomers, especially in complex biological matrices. Accordingly, Chiralify provides a robust first-pass annotation of structural isomers by reporting candidate matches, thereby enabling systematic downstream prioritization and validation. Future work will focus on improving isomers discrimination employing existing models that better capture the differences in fragment ions relative abundances[61] and fragmentation patterns of different classes of molecules[62]. In addition, the generation of MS/MS spectra with higher resolution will increase the training data and further improve model performance.

Chiralify provides a novel and open-source platform for untargeted chiral metabolomics. The user-friendly GUI contributes to the automation of the data processing by facilitating parameters setting and results visualization. Chiralify offers a variety of parameters to ensure results optimization and reproducibility, representing the first framework explicitly dedicated to standardizing and automating the processing of chiral metabolomics data. This framework establishes a foundation for extending untargeted metabolomics and biomarker discovery to enantiomer-resolved small-molecule analysis.

Overall, Chiralify addresses an important computational gap in chiral metabolomics by providing, to our knowledge, the first standardized framework specifically designed for the automated processing and annotation of untargeted chiral metabolomics data. This tool will facilitate the exploration of the chiral metabolome across a wide variety of applications and expand our understanding of metabolic chirality.

Materials and Methods

Chemicals and Reagents

LC-MS grade acetic acid, acetonitrile (ACN), ammonium formate, formic acid (FA), methanol (MeOH) and water were purchased from Fisher Scientific (Hampton, NH). (+)-Diacetyl-L-tartaric Anhydride (purity > 97%) and (-)-Diacetyl-D-tartaric Anhydride (purity > 98%) were purchased from TCI America, Inc. (Portland, OR). Calcium nitrate, magnesium sulfate, potassium chloride, sodium bicarbonate, sodium chloride and sodium phosphate dibasic anhydrous were purchased from Thermo Fisher Scientific.

Glycine, L-Hydroxyvaleric Acid, D-Serine, L-Serine, DL-Threonine were purchased from Fisher Scientific (Hampton, NH). D-Carnitine, L-Carnitine, D-Hydroxyglutaric Acid, L-Hydroxyglutaric Acid, D-Ornithine, L-Ornithine were purchased from Cayman Chemical (Washtenaw County, MI). L-Lactic Acid was purchased from MP Biomedicals. D-Hydroxyvaleric Acid was purchased from Sigma-Aldrich (St. Louis, MO). All the other DL-Amino acids, DL-Malic acid and D-Lactic acid were purchased from TCI America, Inc. (Portland, OR).

Sample Preparation and Chiral Derivatization

All standards were prepared at a stock concentration of 1.5 mg/mL, and then serially diluted to a final concentration of 1.5 mg/L (1.5 ppm) for the DATAN derivatization. For the evaluation of Chiralify performance with metabolites present in trace levels L-Lactate and L-Serine were prepared at a concentration of 1.5 mg/L, while D-Lactate and D-Serine were prepared at a concentration of 1.5 µg/L.

The cell lines CCRF-CEM, LnCaP, Panc04.03, Panc10.05, PC-3 (all from ATCC) and PF-382 (DSMZ) were grown in RPMI 1640 medium (no glutamine) supplemented with 10% v/v of characterized FBS and 2 mM L-glutamine. The cell lines A172 and U251 (all from ATCC) were grown in DMEM (high glucose, pyruvate, no glutamine) supplemented with 10% v/v of characterized FBS and 2 mM L-glutamine. All cell lines were routinely checked to ensure that they were mycoplasma free and were grown at 37°C with 5% CO₂.

Cells were harvested and washed twice in cold PBS. Washed cell pellets were flash-frozen in liquid nitrogen and stored at -80°C until UCM analysis. Intracellular polar metabolites were extracted using a modified Bligh-Dyer method optimized for polar and non-polar metabolites partitioning, as previously described[63]. For this purpose, frozen cellular pellet was resuspended in 1:1:2 methanol:water:chloroform with the addition of 10mM of ammonium bicarbonate and 2,6-Di-tert-butyl-4-methylphenol (Fisher Scientific, Hampton, NH) to preserve redox-sensitive metabolites. Polar fraction was collected for subsequent DATAN-derivatization and analysis.

Quality control (QC) samples were prepared for each cell line by mixing 40 µL aliquots from each of 4 replicates. The total 160 µL were then aliquoted in two separate glass vials: one for derivatization with (+)-DATAN, the other for derivatization with (-)-DATAN.

The standard mix and the cellular polar fractions were evaporated to dryness in a CentriVap refrigerated vacuum concentrator (Labconco, Kansas City, MO) at 4 °C. Samples were then derivatized by adding either 50 µL of (+)-DATAN solution or 50 µL of (-)-DATAN solution, at 75 mg/mL in acetonitrile: glacial acetic acid (ACN: AcOH, v/v 4:1), as previously described[40, 43, 57].

Untargeted Chiral Metabolomics Analytical Method

Chiral metabolic analysis was performed on a Hybrid quadrupole-Orbitrap mass spectrometer (Q-Exactive, Thermo Scientific, Waltham, MA) and with a Tribrid mass spectrometer (IQ-X, Thermo Scientific, Waltham, MA) each coupled to a Vanquish Flex ultrahigh pressure liquid chromatography (UHPLC) system equipped with an ACQUITY BEH C18 150 × 2.1 mm (1.7 µm, 130 Å) column (Waters, Milford,

MA). Chromatographic separation was achieved as previously described[40, 43, 57] with a ternary gradient elution method with the following mobile phases: (A) 0.06% formic acid (FA) and 10 mM ammonium formate, pH 3.6; (B) 0.1% FA in ACN; (C) 0.1% FA in MeOH. With the Q-Exactive, detection was achieved in full MS/all ion fragmentation scan (AIF) in positive ionization mode under the following conditions: micro scans 1; AGC target $1e^6$; maximum injection time 200 ms; m/z range 250–1000; m/z range MS/MS 70-1000, NCE 10 and 20. With the IQ-X, detection was achieved in full MS/data independent acquisition (DIA) in positive ionization mode under the following conditions: AGC target $4e^5$; auto maximum injection time, m/z range 250/1000; m/z range MS/MS 70-1000, NCE 10 and 20.

Peak Detection, Alignment and Statistical Analysis

Raw Mass-Spectrometry/All-Ion Fragmentation (MS/AIF) data were processed with the open-source software MS-DIAL[21] for peak detection and alignment. Detected features were annotated using an in-house MATLAB script based on m/z and retention-time matching against a library of previously acquired standards, and the resulting annotations were compared with those generated by Chiralify. For statistical analysis, data were normalized using Probabilistic Quotient Normalization (PQN)[64], log₁₀-transformed, and filtered to retain features with a coefficient of variation (CV) below 0.3 across biological replicates to ensure data reproducibility. Principal Component Analysis (PCA) was performed in MATLAB using mean-centered data and the default Singular Value Decomposition (SVD) algorithm.

Abbreviations

ACN

acetonitrile

ALL

acute lymphoblastic leukemia

CRF

cumulative relative frequency

DATAN

diacetyl tartaric anhydride

ESI

electrospray ionization

FA

formic acid

GBM

glioblastoma

GUI

graphic user interface

HMDB

Human Metabolome Database

NCE

normalized collision energy

PanC

pancreatic cancer

PCA

principal component analysis

PC

prostate cancer

RPLC-HR-MS/MS

reverse-phase liquid chromatography high-resolution tandem mass spectrometry

RT

retention time

UCM

untargeted chiral metabolomics.

Declarations

Author Contribution

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript. ‡These authors contributed equally.

Notes

The authors declare no competing interests.

Acknowledgments

This work was supported by The Welch Foundation (Grant 713478, ST), and the CNS Spark Award (ST) and CNS Catalyst 2.0 Award (AL) from the College of Natural Sciences, The University of Texas at Austin.

Graphical Abstract, Figure 1 and Figure 3 created in BioRender. Impedovo, V. (2026)

<https://BioRender.com/s58u952>, <https://BioRender.com/isw40yw>, <https://BioRender.com/kz4shz0>, <https://BioRender.com/temchlw>

Availability of data and materials

The datasets supporting the conclusions of this article are available in the Zenodo repository [DOI 10.5281/zenodo.21384300; <https://zenodo.org/records/21384300>]. The Chiralify source code is available at Zenodo (10.5281/zenodo.21972550; <https://zenodo.org/records/21972550>).

References

1. Meacham C, DeVilbiss A, Morrison S (2022) Metabolic regulation of somatic stem cells in vivo. *Nat Rev Mol Cell Biol* 23(6). 10.1038/s41580-022-00462-1
2. DeBerardinis R, Keshari K (2022) Metabolic analysis as a driver for discovery, diagnosis, and therapy. *Cell* 185(15). 10.1016/j.cell.2022.06.029
3. Bujak R, Struck-Lewicka W, Markuszewski M, Kaliszan R (2015) Metabolomics for laboratory diagnostics. *J Pharm Biomed Anal* 113. 10.1016/j.jpba.2014.12.017
4. AC S-R SGC (2016) Untargeted Metabolomics Strategies-Challenges and Emerging Directions. *J Am Soc Mass Spectrom* 27(12). 10.1007/s13361-016-1469-y
5. Kiriya Y, Nochi H (2016) D-Amino Acids in the Nervous and Endocrine Systems. *Scientifica*. ;2016. 10.1155/2016/6494621
6. Lin C, Yang H, Chiu C, Lane H (2017) Blood levels of D-amino acid oxidase vs. D-amino acids in reflecting cognitive aging. *Sci Rep* 7(1). 10.1038/s41598-017-13951-7
7. Armstrong D, Gasper M, Lee S, Zukowski J, Ercal N (1993) D-amino acid levels in human physiological fluids. *Chirality* 5(5). 10.1002/chir.530050519
8. DW A, M G (1993) D-amino acid levels in human physiological fluids. *Chirality* 5(5). 10.1002/chir.530050519
9. Armstrong D, Berthod A (2023) Occurrence of D-amino acids in natural products. *Nat Prod bioprospecting* 13(1). 10.1007/s13659-023-00412-0
10. Du S, Wey M, Armstrong D (2023) d-Amino acids in biological systems. *Chirality* 35(9). 10.1002/chir.23562
11. Nagata Y, Higashi M, Ishii Y, Sano H, Tanigawa M, Nagata K et al (2006) The presence of high concentrations of free D-amino acids in human saliva. *Life Sci* 78(15). 10.1016/j.lfs.2005.08.009
12. Zangari J, Stehling O, Freibert S, Bhattacharya K, Rouaud F, Serre-Beinier V et al (2025) D-cysteine impairs tumour growth by inhibiting cysteine desulfurase NFS1. *Nat metabolism* 7(8). 10.1038/s42255-025-01339-1
13. Liu Y, Wu Z, Armstrong DW, Wolosker H, Zheng Y (2023) Detection and analysis of chiral molecules as disease biomarkers. *Nat Reviews Chem* 1–19. 10.1038/s41570-023-00476-z
14. Bastings J, van Eijk H, Olde Damink S, Rensen S (2019) d-amino Acids in Health and Disease: A Focus on Cancer. *Nutrients* 11(9). 10.3390/nu11092205
15. Kasper BH, Feng Y, Riley EP, Hiro F, Lonnie PW, Alasdair JG et al (2025) Structure, function, and allosteric modulation of NMDA receptors. *J Gen Physiol* 150(8):1081. 10.1085/jgp.201812032
16. Mothet J, Parent A, Wolosker H, Brady R, Linden D, Ferris C et al (2000) D-serine is an endogenous ligand for the glycine site of the N-methyl-D-aspartate receptor. *Proc Natl Acad Sci USA* 97(9). 10.1073/pnas.97.9.4926
17. Errico F, Mothet J, Usiello A, D-Aspartate (2015) An endogenous NMDA receptor agonist enriched in the developing brain with potential involvement in schizophrenia. *J Pharm Biomed Anal* 116. 10.1016/j.jpba.2015.03.024

18. Burnet P, Eastwood S, Bristow G, Godlewska B, Sikka P, Walker M et al (2008) D-amino acid oxidase activity and expression are increased in schizophrenia. *Mol Psychiatry* 13(7). 10.1038/mp.2008.47
19. Hashimoto K, Fukushima T, Shimizu E, Komatsu N, Watanabe H, Shinoda N et al (2003) Decreased serum levels of D-serine in patients with schizophrenia: evidence in support of the N-methyl-D-aspartate receptor hypofunction hypothesis of schizophrenia. *Arch Gen Psychiatry* 60(6). 10.1001/archpsyc.60.6.572
20. Piubelli L, Pollegioni L, Rabattoni V, Mauri M, Princiotta CL, Versino M et al (2021) Serum D-serine levels are altered in early phases of Alzheimer's disease: towards a precocious biomarker. *Translational psychiatry* 11(1). 10.1038/s41398-021-01202-3
21. Tsugawa H, Cajka T, Kind T, Ma Y, Higgins B, Ikeda K et al (2015) MS-DIAL: data-independent MS/MS deconvolution for comprehensive metabolome analysis. *Nat Methods* 12(6):523–526. 10.1038/nmeth.3393
22. Tautenhahn R, Patti G, Rinehart D, Siuzdak G (2012) XCMS Online: a web-based platform to process untargeted metabolomic data. *Anal Chem* 84(11). 10.1021/ac300698c
23. Heuckeroth S, Damiani T, Smirnov A, Mokshyna O, Brungs C (2024) Reproducible mass spectrometry data processing and compound annotation in MZmine 3. *Nat Protoc* 19(9). 10.1038/s41596-024-00996-y
24. Pang Z, Lu Y, Zhou G, Hui F, Xu L, Viau C et al (2024) *Nucleic Acids Res* 52(W1). 10.1093/nar/gkae253. MetaboAnalyst 6.0: towards a unified platform for metabolomics data processing, analysis and interpretation
25. Shen X, Wang R, Xiong X, Yin Y, Cai Y, Ma Z et al (2019) Metabolic reaction network-based recursive metabolite annotation for untargeted metabolomics. *Nat Commun* 10(1). 10.1038/s41467-019-09550-x
26. Adams K, Pratt B, Bose N, Dubois L, St John-Williams L, Perrott K et al (2020) Skyline for Small Molecules: A Unifying Software Package for Quantitative Metabolomics. *J Proteome Res* 19(4). 10.1021/acs.jproteome.9b00640
27. Li S, Siddiqi A, Thapa M, Chi Y, Zheng S (2023) Trackable and scalable LC-MS metabolomics data processing using asari. *Nat Commun* 14(1). 10.1038/s41467-023-39889-1
28. Faizan-Khan M, Giné R, Badia J, Pérez-Ribera M, Ruiz-Botella M, Junza A et al (2026) ChemEmbed: a deep learning framework for metabolite identification using enhanced MS/MS data and multidimensional molecular embeddings. *Brief Bioinform* 27(1). 10.1093/bib/bbag054
29. Wishart D, Knox C, Guo A, Eisner R, Young N, Gautam B et al (2009) HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* 37(Database issue). 10.1093/nar/gkn810
30. Kanehisa M, Furumichi M, Tanabe M, Sato Y, Morishima K (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res* 45(D1). 10.1093/nar/gkw1092
31. Allen F, Pon A, Wilson M, Greiner R, Wishart D CFM-ID: a web server for annotation, spectrum prediction and metabolite identification from tandem mass spectra. *Nucleic acids research*. 2014;42(Web Server issue). 10.1093/nar/gku436

32. Wolf S, Schmidt S, Müller-Hannemann M, Neumann S (2010) In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* 11. 10.1186/1471-2105-11-148
33. Dührkop K, Shen H, Meusel M, Rousu J, Böcker S (2015) Searching molecular structure databases with tandem mass spectra using CSI:FingerID. *Proc Natl Acad Sci USA* 112(41). 10.1073/pnas.1509788112
34. Berthod A, Liu Y, Bagwill C, Armstrong D (1996) Facile liquid chromatographic enantioresolution of native amino acids and peptides using a teicoplanin chiral stationary phase. *J Chromatogr A* 731(1–2). 10.1016/0021-9673(95)01198-6
35. Calderón C, Lämmerhofer M (2022) Enantioselective metabolomics by liquid chromatography-mass spectrometry. *J Pharm Biomed Anal* 207. 10.1016/j.jpba.2021.114430
36. Yashima E (2001) Polysaccharide-based chiral stationary phases for high-performance liquid chromatographic enantioseparation. *J Chromatogr A* 906(1–2). 10.1016/s0021-9673(00)00501-x
37. Du S, Wang Y, Alatrash N, Weatherly C, Roy D, MacDonnell F et al (2019) Altered profiles and metabolism of L- and D-amino acids in cultured human breast cancer cells vs. non-tumorigenic human breast epithelial cells. *J Pharm Biomed Anal* 164. 10.1016/j.jpba.2018.10.047
38. Suh E, Geraldine C, Chirayil S, Faubert B, Ayala R, DeBerardinis R et al (2021) Detection of glucose-derived D- and L-lactate in cancer cells by the use of a chiral NMR shift reagent. *Cancer metabolism* 9(1). 10.1186/s40170-021-00267-4
39. Mochizuki T, Takayama T, Todoroki K, Inoue K, Min J, Toyo'oka T (2015) Towards the chiral metabolomics: Liquid chromatography-mass spectrometry based DL-amino acid analysis after labeling with a new chiral reagent, (S)-2,5-dioxopyrrolidin-1-yl-1-(4,6-dimethoxy-1,3,5-triazin-2-yl)pyrrolidine-2-carboxylate, and the application to saliva of healthy volunteers. *Anal Chim Acta* 875. 10.1016/j.aca.2015.02.054
40. Pandey R, Collins M, Lu X, Sweeney S, Chiou J, Lodi A et al (2021) Novel Strategy for Untargeted Chiral Metabolomics using Liquid Chromatography-High Resolution Tandem Mass Spectrometry. *Anal Chem* 93(14). 10.1021/acs.analchem.0c05325
41. Takayama T, Mochizuki T, Todoroki K, Min J, Mizuno H, Inoue K et al (2015) A novel approach for LC-MS/MS-based chiral metabolomics fingerprinting and chiral metabolomics extraction using a pair of enantiomers of chiral derivatization reagents. *Anal Chim Acta* 898. 10.1016/j.aca.2015.10.010
42. Lkhagva A, Tai H (2021) Dimethylcysteine (DiCys)/ o-Phthalaldehyde Derivatization for Chiral Metabolite Analyses: Cross-Comparison of Six Chiral Thiols. *Molecules* 26(24). 10.3390/molecules26247416
43. Collins M, Gorgoglione R, Impedovo V, Pan X, Chakkarai S, Yi S et al (2024) Exploration of the intracellular chiral metabolome in pediatric BCP-ALL: a pilot study investigating the metabolic phenotype of IgH locus aberrations. *Front Oncol* 14. 10.3389/fonc.2024.1413264
44. Bogos L, Pralea I, Moldovan R, Iuga C (2022) Indirect Enantioseparations: Recent Advances in Chiral Metabolomics for Biomedical Research. *Int J Mol Sci* 23(13). 10.3390/ijms23137428

45. Wang F, Liigand J, Tian S, Arndt D, Greiner R, Wishart D (2021) CFM-ID 4.0: More Accurate ESI-MS/MS Spectral Prediction and Compound Identification. *Anal Chem* 93(34).
10.1021/acs.analchem.1c01465
46. Claes B, Takeo E, Fukusaki E, Shimma S, Heeren R (2019) Imaging Isomers on a Biological Surface: A Review. *Mass spectrometry (Tokyo. Japan)* 8(1). 10.5702/massspectrometry.A0078
47. Yun T, Yanran C, Yiwei L, Yide X, Faqiong Z, Baizhao Z (2022) Detection of Gastric Cancer-Associated d-Amino Acids and Carcinoembryonic Antigen by Colorimetric and Immuno ECL Sensing Platform Based on the Catalysis of N/S-Doped Carbon Dots @ N-Rich Porous Carbon Nanoenzyme.
10.1021/acs.analchem.2c03433
48. Liu C, Zhang J, Zheng Q, Zhao L, Kong P, Yang H et al (2023) Surface-Enhanced Raman Scattering Enantioselective Detection of Gastric Cancer-Related d-Amino Acids in Saliva Based on Enzyme-Mediated Cascade Reaction. *Anal Chem* 95(35). 10.1021/acs.analchem.3c01030
49. de Bari L, Moro L, Passarella S (2013) Prostate cancer cells metabolize d-lactate inside mitochondria via a D-lactate dehydrogenase which is more active and highly expressed than in normal cells. *FEBS Lett* 587(5). 10.1016/j.febslet.2013.01.011
50. Andrade J, Costa G, Siqueira C, Carbonezi C, Binotto R, Kartnaller V (2025) Trends and Innovations in Tools for Processing Chromatographic Data Using Mass Spectrometry Detection: A Systematic Review. *Critical reviews in analytical chemistry*. 10.1080/10408347.2025.2528134
51. Jung K, Reszka R, Kamlage B, Bethan B, Stephan C, Lein M et al (2013) Tissue metabolite profiling identifies differentiating and prognostic biomarkers for prostate carcinoma. *Int J Cancer* 133(12).
10.1002/ijc.28303
52. Yuan W, Zhang J, Li S, Edwards J (2011) Amine metabolomics of hyperglycemic endothelial cells using capillary LC-MS with isobaric tagging. *J Proteome Res* 10(11). 10.1021/pr200815c
53. Wijekoon E, Skinner C, Brosnan M, Brosnan J (2004) Amino acid metabolism in the Zucker diabetic fatty rat: effects of insulin resistance and of type 2 diabetes. *Can J Physiol Pharmacol* 82(7).
10.1139/y04-067
54. Emma LS, Junho J, Rebekka G, Kathrin F, Matthias R, Heinz PS et al (2014) Identifying Small Molecules via High Resolution Mass Spectrometry: Communicating Confidence.
10.1021/es5002105
55. Kaushik A, DeBerardinis R (2018) Applications of metabolomics to study cancer metabolism. *Biochim et Biophys acta Reviews cancer* 1870(1). 10.1016/j.bbcan.2018.04.009
56. Martano G, Leone M, D'Oro P, Matafora V, Cattaneo A, Masseroli M et al (2020) SMfinder: Small Molecules Finder for Metabolomics and Lipidomics Analysis. *Anal Chem* 92(13).
10.1021/acs.analchem.0c00585
57. Lodi A, Pandey R, Chiou J, Bhattacharya A, Huang S, Pan X et al (2022) Circulating metabolites associated with tumor hypoxia and early response to treatment in bevacizumab-refractory glioblastoma after combined bevacizumab and evofosfamide. *Front Oncol* 12.
10.3389/fonc.2022.900082

58. Reifenberg P, Rosenberger L, Mignon M, Zimmer A (2025) Replacement of Isoleucine and Leucine by their Keto Acids leads to increased formation of α -Hydroxy Acids in Chinese Hamster Ovary cells. *Biotechnol J* 20(6). 10.1002/biot.70041
59. Sun Y, Hu Q, Zuo J, Wang H, Guo Z, Wang Y et al (2024) Simultaneous Quantification of Carboxylate Enantiomers in Multiple Human Matrices with the Hydrazide-Assisted Ultrahigh-Performance Liquid Chromatography Coupled with Tandem Mass Spectrometry. *Anal Chem* 96(45). 10.1021/acs.analchem.4c04187
60. de Jonge N, Mildau K, Meijer D, Louwen J, Bueschl C, Huber F et al (2022) Good practices and recommendations for using and benchmarking computational metabolomics metabolite annotation tools. *Metabolomics: Official J Metabolomic Soc* 18(12). 10.1007/s11306-022-01963-y
61. Nothias L, Petras D, Schmid R, Dührkop K, Rainer J, Sarvepalli A et al (2020) Feature-based molecular networking in the GNPS analysis environment. *Nat Methods* 17(9). 10.1038/s41592-020-0933-6
62. Wang X, Guo A, Wang R, Gao W, Yang H (2024) AnnoSM: An Automated Annotation Tool for Determining the Substituent Modes on the Parent Skeleton Based on a Characteristic MS/MS Fragment Ion Library. *Anal Chem* 96(9). 10.1021/acs.analchem.3c04946
63. Wu H, Southam A, Hines A, Viant M (2008) High-throughput tissue extraction protocol for NMR- and MS-based metabolomics. *Anal Biochem* 372(2). 10.1016/j.ab.2007.10.002
64. Dieterle F, Ross A, Schlotterbeck G, Senn H (2006) Probabilistic quotient normalization as robust method to account for dilution of complex biological mixtures. Application in ^1H NMR metabolomics. *Anal Chem* 78(13). 10.1021/ac051632c

Figures

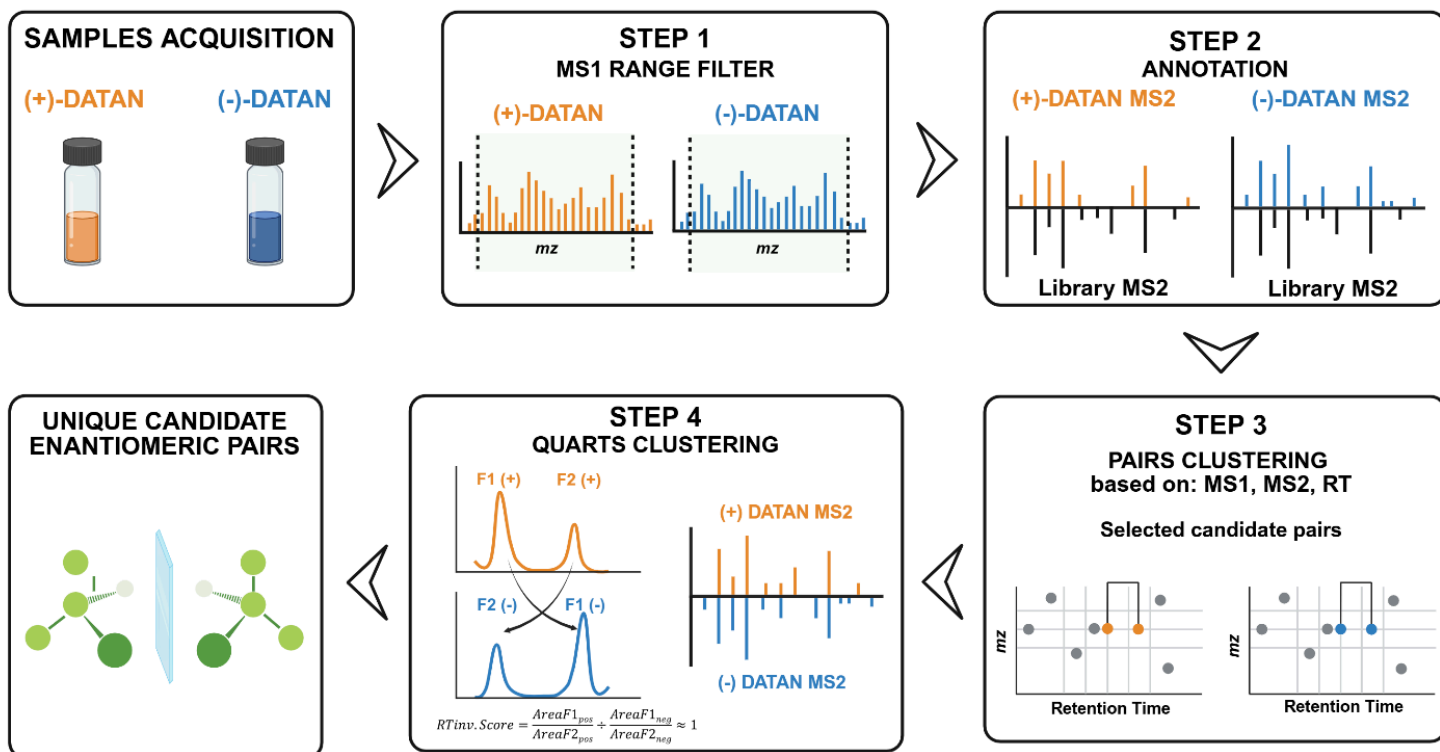


Figure 1

The Chiralify pipeline. Features from (+)- and (-)-DATAN-derivatized samples are filtered according to the *m/z* range and then annotated by matching their MS1 to the in-silico DATAN-derivatized metabolites library. A score is attributed to each assignment for the same features by MS2 spectra similarity with the predicted MS2 library. The features are then paired based on MS1 and MS2 similarity and retention time proximity, and (+)-DATAN-derivatized features are matched with the (-)-DATAN ones to identify quartets. Finally, the algorithm leverages the reversal elution order between the two derivatives to identify the unique candidate enantiomeric pairs.

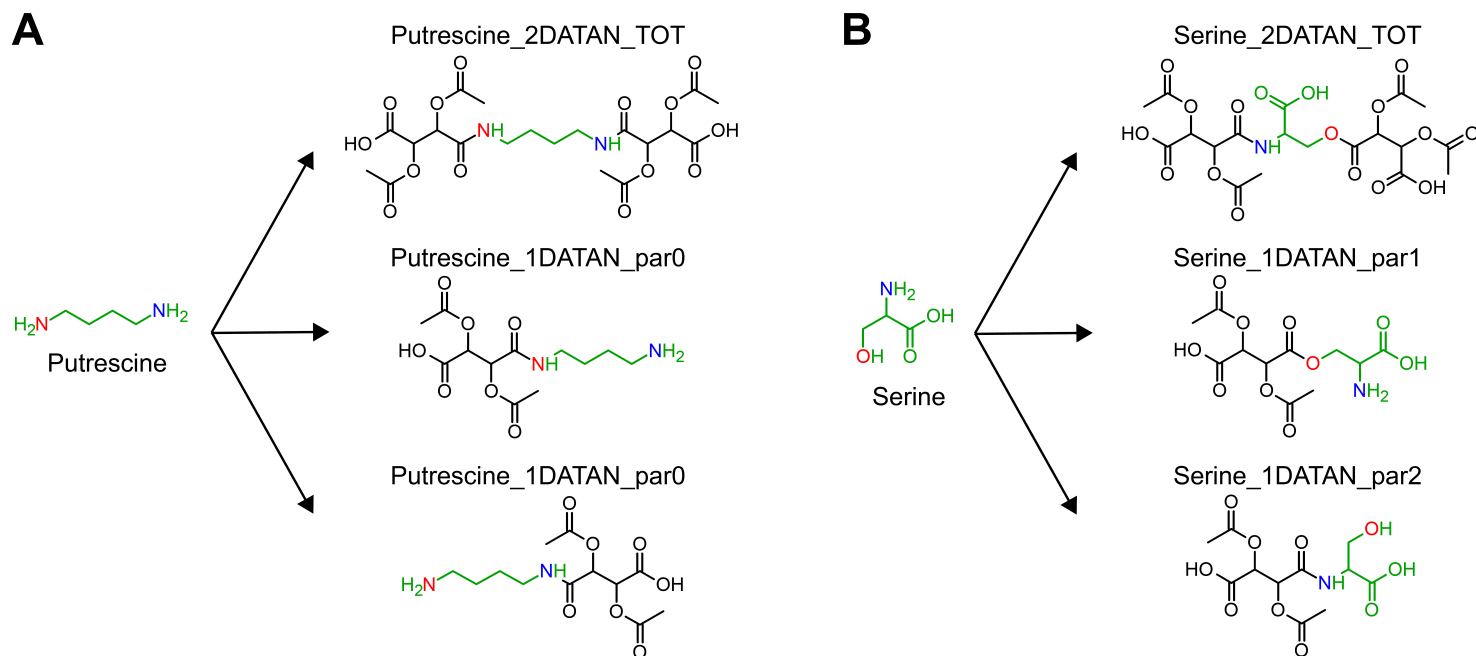


Figure 2

The in-silico library nomenclature. The name of each library entry indicates the number of DATAN molecules attached to the metabolite (i.e., 1DATAN or 2DATAN) and discriminates fully derivatized molecules (TOT) from partially derivatized ones (par). The partial derivatives are indicated as **(A)** _par0 when the two derivatized forms are structurally identical, and **(B)** _par1/_par2 when their configurations differ.

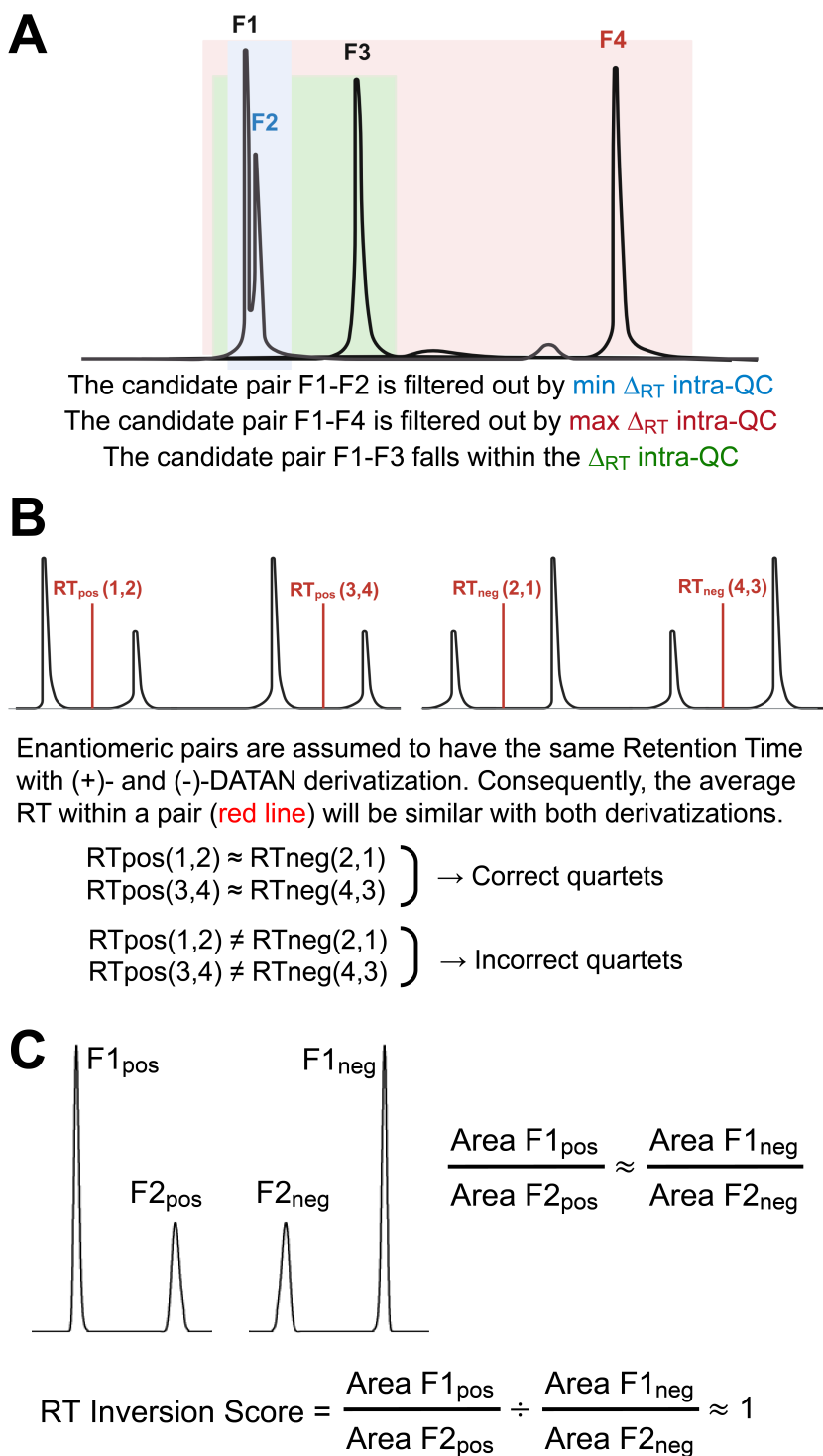


Figure 3

Chiralityfy identifies candidate enantiomeric pairs according to a sequence of user-defined parameters. **(A)** Candidate enantiomeric pairs are identified according to their RT proximity. **(B)** The pairs identified in (+)-DATAN derivatized samples are matched with the ones in (-)-DATAN derivatized samples to create quartets of peaks. **(C)** The quartets are filtered to retain only those with an RT inversion score within the selected threshold range.

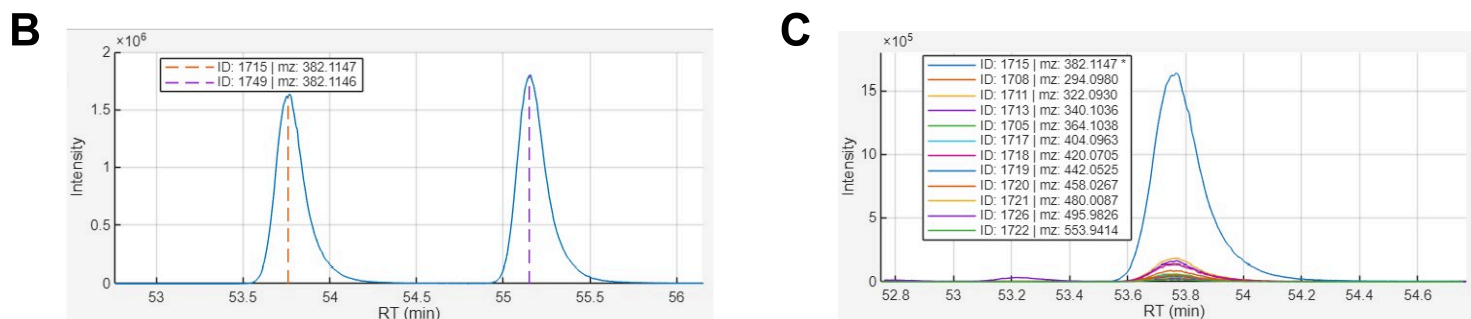
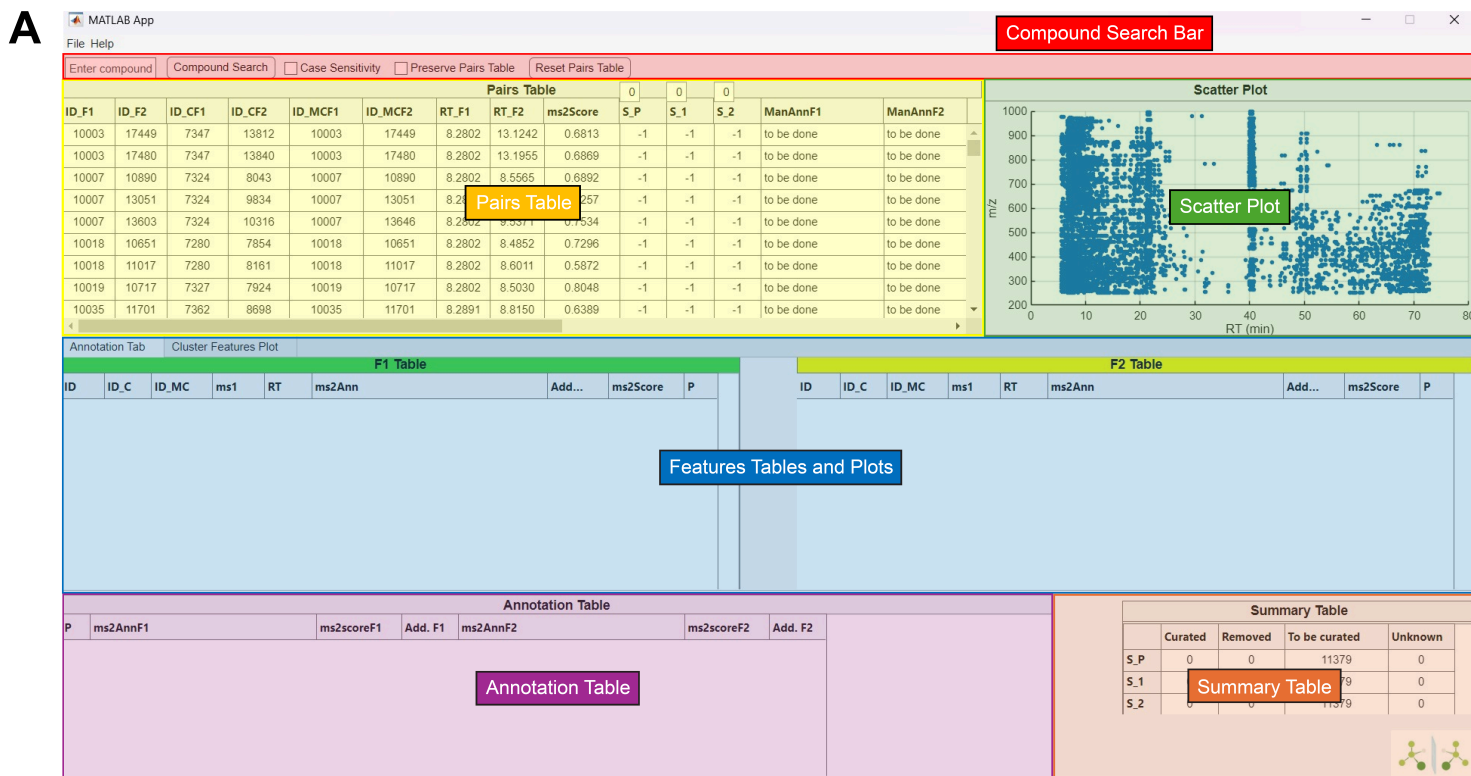


Figure 4

The ChiralAPP provides a user-friendly graphic interface to facilitate the manual curation of the results. **(A)** The ChiralApp sections include interactive tables and plots. The Pairs Table, listing candidate enantiomeric feature pairs, allows rapid navigation and selection for annotation review. By selecting a pair in this table, the Annotation Table will display all the annotations for each feature in the selected pair, while the Scatter Plot will highlight the pair in red. The F1 and F2 Tables are automatically populated with the annotation results retrieved through the compound search tool. The pair selection is performed on the Pairs Table, and the Summary Table provides a report of the curation results. **(B)** The Features Work Area displays the features selected in the Pairs Table. **(C)** The cluster work area displays chromatograms for the features belonging to the same selected cluster.

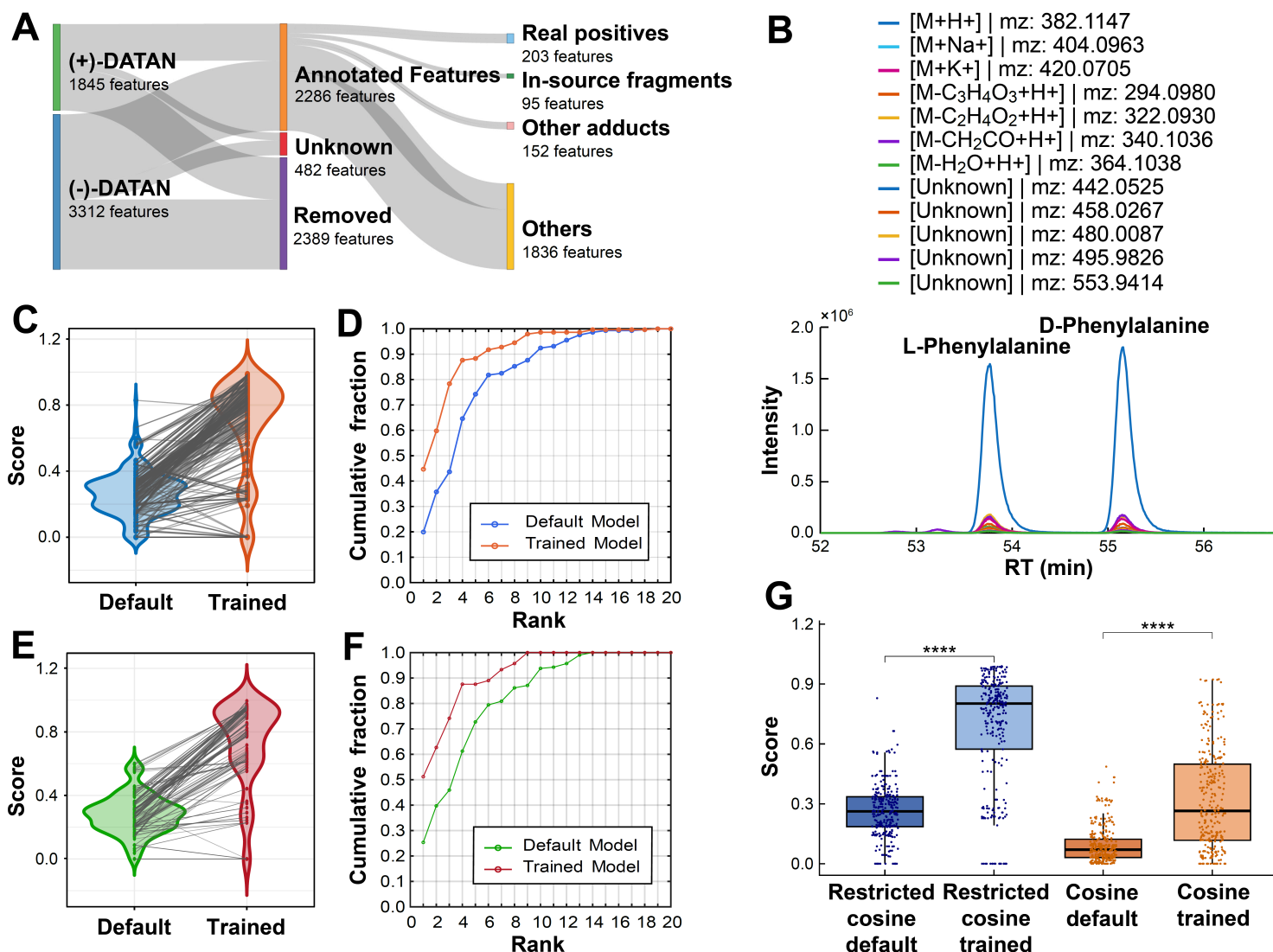


Figure 5

Validation of the Chiralify output using a standard mix of 43 chiral and non-chiral metabolites. **(A)** The Sankey plot shows the results of feature assignment in (+)- and (-)-DATAN derivatized samples. With the selected parameters, Chiralify annotated approximately half of the features, including all the standards in the mix. Moreover, for each standard, it identified its putative in-source fragments and other adducts. **(B)** The chromatogram shows the features co-eluting with the protonated phenylalanine. The parental ion (m/z 382.1147) displays higher intensities compared to other adducts and in-source fragments. **(C)** MS² scores and **(D)** cumulative relative frequency (CRF) plot of assignment ranking position for the standard mix acquired on a Q-Exactive Hybrid quadrupole-Orbitrap MS annotated using either the Default-Lib or the Trained-Lib at NCE10. Most standard assignment scores as well as the distribution of assignment ranking positions improve following the use of the Trained-Lib during annotation, thus proving the efficacy of the model training in enhancing the assignment accuracy. **(E)** MS² scores and **(F)** assignment ranking positions for the standard mix acquired at NCE10 on an Orbitrap IQ-X Tribrid MS and annotated using the Default-Lib or the Trained-Lib. After model training, most of the standard annotations exhibit increased MS² scores and CRF shows consistent improvement of the annotation ranking. **(G)**

Distribution of the cosine score and reference-restricted cosine score are shown before and after model training. For both scoring methods, post-training scores are overall higher than pre-training scores, indicating improved annotation quality.

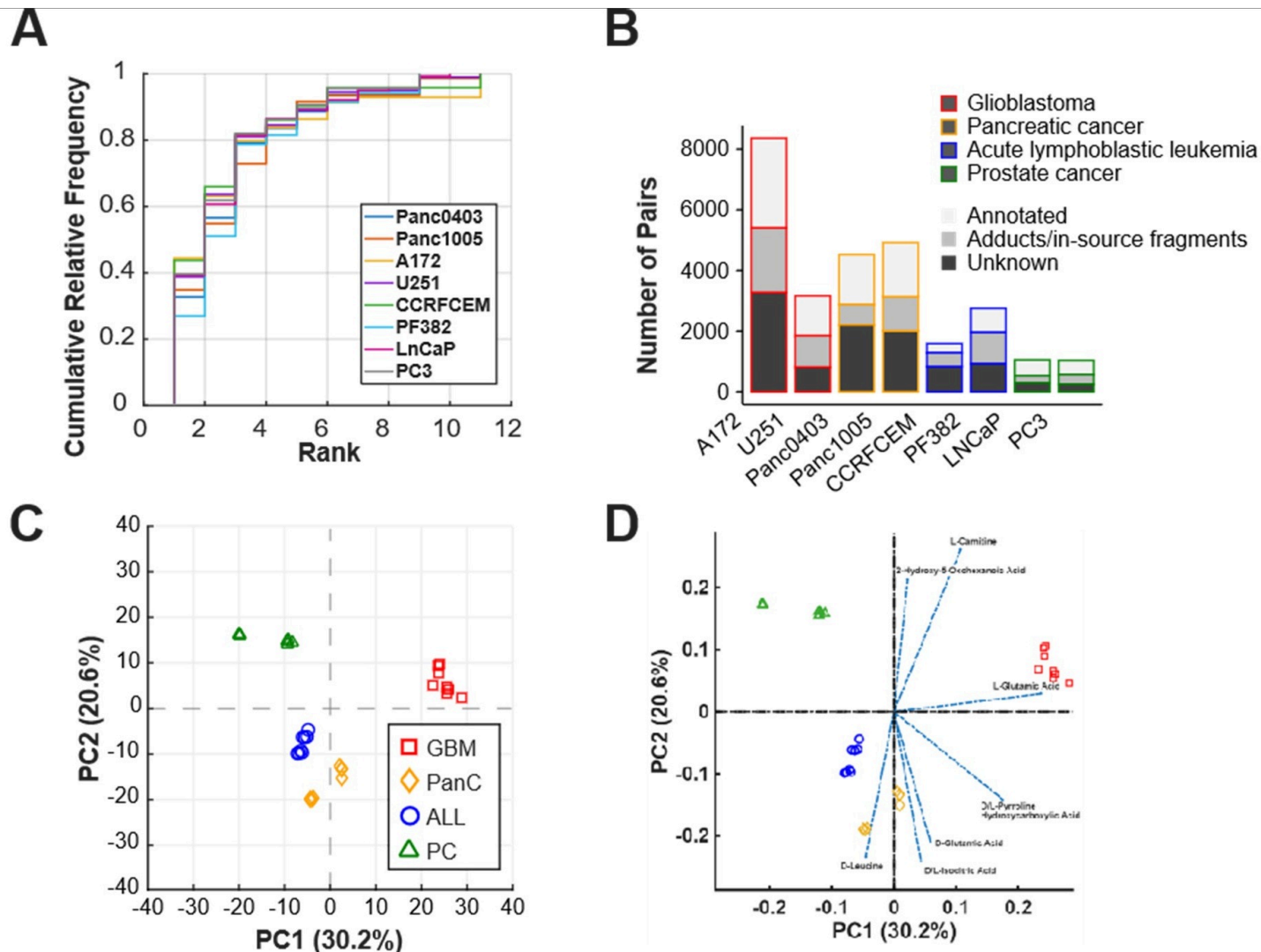


Figure 6

Untargeted identification of chiral metabolites in cancer cell lines. **(A)** The CRF plot compares the distribution of assignment ranking positions across cancer cell lines. **(B)** Number of annotated pairs, predicted adducts, in-source fragments, and unknown candidate enantiomeric pairs across the cell lines. Glioblastoma and pancreatic cancer display the highest number of candidate pairs out of all the cancer types included in the analysis. **(C)** PCA of cancer cell lines generated from the Chiralify output. The first two principal components of the scores plot capture a clear separation of cell lines according to their lineage of origin. **(D)** The biplot shows the loadings of major metabolic contributors to the separation observed along the first two principal components. Chiral metabolites are labeled according to their D- or L-configuration when retention times are known or as D/L when retention-time information is unknown

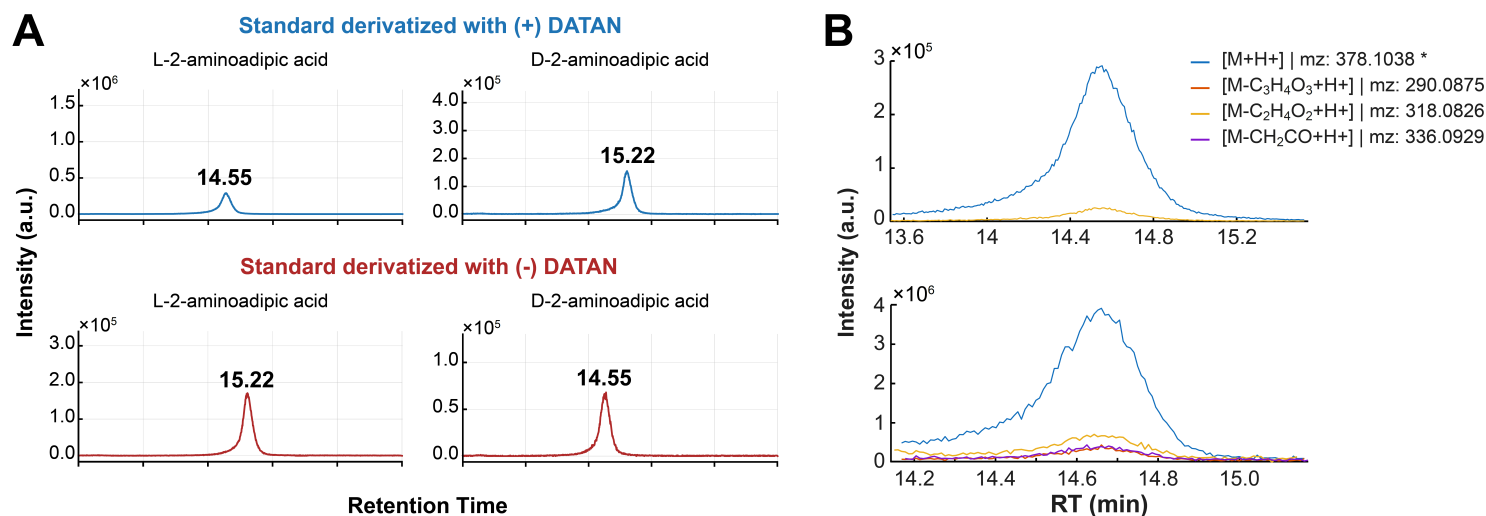


Figure 7

Standard validation of aminoadipic acid assignment. **(A)** Chromatograms of L-2-aminoadipic acid (left) and D-2-aminoadipic acid (right) derivatized with (+)-DATAN (in blue) and (-)-DATAN (in red). Elution order reversal for L-/D-2-aminoadipic acid occurs in the two derivatization modes with the L-form eluting first in (+)-DATAN derivatized samples and the D-form eluting second. **(B)** Comparison of (+)-DATAN derivatized L-aminoadipic acid standard (top chromatogram) and in Panc10.05 (bottom chromatogram). In both samples, the metabolite shows the same ionization and some common in-source fragments.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [AdditionalFileslegends.docx](#)
- [AdditionalFile1.docx](#)