A Systems Biology Workflow to Support the Diagnosis of Inherited Metabolic Disorders: a study on Pyrimidine and Urea Cycle disorders

Denise N. Slenter*, Irene M.G.M. Hemel, Chris T. Evelo, Jörgen Bierau, Egon L. Willighagen,

Laura K.M. Steinbusch

Abstract:

Background: Inherited Metabolic Disorders (IMDs) are rare diseases where one impaired protein leads to a cascade of changes in the adjacent chemical conversions. These disorders often present with non-specific symptoms, a lack of a clear genotype-phenotype correlation, and *de novo* mutations, complicating diagnosis. Furthermore, products of one metabolic conversion can be the substrate of another pathway (much further) down- or upstream, obscuring biomarker identification and causing overlapping biomarkers for different disorders. Last, current clinical diagnoses are lacking a visualization of the connections between individual metabolic biomarkers and the enzymes involved in their synthesis and degradation. The goal of this study was to provide a proof-of-concept systems biology framework for diagnostic support of IMDs by combining clinical and theoretical metabolic biomarkers with pathway models using network approaches and semantic web technologies. This workflow could support the diagnosis of other less understood IMDs, by building on two groups of well-studied metabolic disorders linked to the pyrimidine and urea cycle pathways.

Methods: Our framework integrates literature and expert knowledge into machine-readable pathways. The top three pathways of interest were retrieved covering most unique patient markers and overlapping with theoretical marker data, which was used to visualize clinical data of 16 previously diagnosed patients with various pyrimidine and urea cycle disorders. Two expert laboratory scientists evaluated the resulting visualizations to derive a diagnosis.

Results: The number of relevant biomarkers for each patient varied greatly (five to 48), and the pathways covering most unique biomarkers differ for equivalent disorders. The two experts reached the same conclusions for all samples with our proposed framework for nine patient samples, without knowledge about clinical symptoms or sex. For the remaining seven cases, four interpretations pointed in the direction of a subset of disorders, while three cases were found to be undiagnosable with the available data.

Conclusion: The presented workflow integrates chemical reactions with clinical data in a visual representation and is adaptable to analyze different types of IMDs (known and novel), difficult patient cases, and functional assays in the future. By using semantic web technologies and network approaches, the workflow could easily be extended with other OMICS datasets, as well as linked to knowledge captured as Linked Open Data.

Keywords: Clinical Metabolic Biomarkers; Purine and Pyrimidine metabolism; Urea Cycle; Semantic Web Technologies; Network Data Analysis; Systems Biology.

Background

Many enzymes are critically involved in the synthesis, degradation, and transport of molecules implicated in metabolic processes [1], with malfunctions of these enzymes or enzymatic disturbances resulting in a lack of or (potentially) toxic levels of metabolites, as well as affecting other (downstream) pathways [2]. Collectively, these diseases are called Inherited Metabolic Disorders (IMDs) or Inborn Errors of Metabolism [3].

A timely and accurate diagnosis of IMDs, currently based on both symptoms and biomarkers measured in various bodily fluids, is required to initiate therapies, which are sparsely available [4]. Methods to detect genetic variants (e.g. Whole Exome Sequencing) are also used to diagnose IMDs, although so far have been found less sensitive and specific as compared to metabolic measurements [5]. Furthermore, genetic profiles of patients can also contain variants of uncertain significance, which can only be classified as (likely) pathogenic when genomic, transcriptomic, proteomic, metabolomic, and fluxomic data are integrated through pathway or network analysis [6–8]. Unfortunately, diagnosing IMDs using these approaches is challenging due to the commonly observed overlap between their biomarkers, since the individual compounds are often involved in more than one pathway and can therefore be metabolized to various products. Furthermore, the diagnostic process can be quite time-consuming, requiring a manual inspection by an expert in the field, who needs to be familiar with all relevant metabolic conversion and their respective enzymes to point out the malfunctioning protein.

Therefore, this study provides a framework using Systems Biology approaches to integrate metabolic reference and patient values, theoretical biomarkers, and biological pathway information for IMDs. Semantic web technologies are used to accurately retrieve relevant pathways for network data visualization, providing an overview of the processes relevant to the patient-specific deficient protein. With this approach, the attention progresses from individual markers to changes at the process level, which enables linking biological pathway knowledge to clinical cases. This direct link shows which metabolic reactions are disturbed, which proteins are related to these reactions, and potentially which specific protein is impaired, aiding diagnosis. Furthermore, metabolic disturbances can be recognized which cannot be attributed directly to the disorder, revealing potential blind spots in existing clinical knowledge. Our workflow was tested on two groups of IMDs (pyrimidine metabolism and urea cycle) with a well-understood molecular mechanism, however with diagnostic issues due to a common metabolite (carbamoyl phosphate) creating substantial biomarker overlap for several IMDs within these pathways [9]. Furthermore, pyrimidine disorders often present with nonspecific clinical symptoms and a lack of a clear genotype-phenotype correlation [10–12], while urea cycle disorders are often more specific (e.g. hyperammonemia, lethargy, vomiting, coma) [13].

The presented workflow highlights chances for the IMD field as a whole, by showcasing that improving data and identifier (ID) harmonization increases machine readability and integration of clinical data with pathway knowledge and biomarker information. Furthermore, the automation of the workflow could aid in the diagnostic process of other (novel) IMDs and is adaptable to analyze different types of IMDs

and functional assays in the future, as well as integrating other types of (omics) data analysis, e.g. transcriptomics, metabolomics, and fluxomics. By using network approaches, the workflow could also be extended with information on drug targets or genetic variants, which could allow for personalized medicine.

1. Methods

Workflow

Figure 1 shows the proposed workflow to connect clinical data to pathway models and theoretical biomarker data, using semantic web and network analysis. The information captured through the pathway models was pivotal in connecting clinical data with biomarker data. ID and age category harmonization between the various databases used in this workflow were needed, as well as compatibility between patient biomarker data and theoretical markers before arriving at the final data interpretation step. Also, knowledge from various databases had to be integrated into the workflow, which is summarized in Table 1. All data processing steps which could be automated were captured in an RMarkdown script [14] in the R programming language (version 4.1.3) [15], tested through RStudio (version 2022.02.2) [16], available at https://github.com/BiGCAT-UM/IMD-PUPY.

Clinical Biomarker Data

All patient data were collected through two targeted chemical assays in urine - one for purines and pyrimidines (PUPY) and one for amino acids (AA) - as described previously [17,18], and were reported in µmol/mmol creatinine. All reference metabolic data were collected from patients suspected of having an IMD, with no

apparent IMD as assessed by selective metabolic screening. For analysis of PUPY, 4853 samples were selected over ten years, and for AA analysis 1872 samples were selected over five years; these values characterize the 95% interval limits describing the lower and upper limits of distribution in a control population. The metabolites in the reference data were annotated with corresponding ChEBI [19] (IDs) or Wikidata [20] IDs (when no ChEBI ID was available), see Data/Reference_values_PUPY.txt and Data/Reference_values_Urea.txt on GitHub. The reference data was categorized based on the following age ranges: 0 up to 1 year, 1 up to 5 years, 5 up to 16 years, and over 16 years old. The reference data from the overarching category 0 to 16 years was used for several PUPY biomarkers if no reference value was available for a specific age category (Data/Reference_values_PUPY_noNAs.txt). Patient data was anonymized and data regarding treatment (allopurinol, which is metabolized in the liver to the active form oxypurinol) was removed from the dataset, as well as argininosuccinic acid anhydride (ASA-anhydride) which became obsolete after switching the separation method from anion exchange chromatography to UHPLC-MS/MS for AA analysis [18]. Four patients were removed from this study, due to missing data for the AA panel (patients labeled B, C, P, and Q). Due to the rare nature of IMDs, it is impossible to correct for ethnicity, age range, or sex in this study. Table 2 details the sample size, diseases, and corresponding age ranges used in this study. No patients have objected to the anonymous use of their leftover material from routine diagnostics for laboratory development and validation purposes.

Pathway models

Machine-readable versions of the purine, pyrimidine, and urea cycle metabolic pathways were created using the pathway editor and curation tool PathVisio (version 3.3.0) [21], as well as pathway models (PWMs) on biomarkers, visualizing several markers missing from the main pathway models. All proteins were annotated with UniProt IDs [22], which linked to potentially relevant Rhea IDs [23] for the metabolic conversion interactions; the Rhea IDs were manually curated regarding the directionality of the reaction and incorporating detailed stereochemistry on substrate and product of the reactions if available. Corresponding ChEBI IDs [19] in Rhea were used to annotate the metabolites. IMDs were annotated with OMIM disease IDs [24]. Data on the created PWMs was deposited in WikiPathways [25] and retrieved from the WikiPathways SPARQL endpoint [26] (data from September 2021 [27]). This endpoint contains PWM data in a harmonized and semantic format, the Resource Description Framework (RDF) [28].

Selection of Relevant Biomarkers

For each patient, measured biomarkers from the clinical assay data (workflow step 1) were annotated automatically with the ChEBI IDs from the reference data by character vector matching, including the age-relevant reference values. Missing biomarker data (null-values) were disregarded, as well as patient or reference concentration data being equal to zero. Relevant biomarkers were detected by calculating how far the measured patient value was removed from either the lower or upper reference limits; below the lower limit indicates a negative change or decrease (values between 0 and 1) and above the upper limit indicates an increase or a positive change (values above 1). When a patient value was in between or exactly

equal to the reference values, the change was designated to be one. All resulting calculated values were log(2) transformed to show proportional changes.

Theoretical Biomarker Data

Since the coverage of biomarkers in chemical assays was highly dependent on the chemical analysis used, all PUPY IMDs were queried through the WikiPathways SPARQL endpoint by their OMIM IDs and HGNC gene name [29] first. Second, the corresponding potentially relevant biomarkers for these disorders were retrieved manually from IEMbase [30] V 2.0.0 (accessed on 2021-08-05) through their HGNC gene name, including sample matrix, HMDB IDs [31], and positive or negative concentration change. The latter was converted to a numeric scale for each provided age category: neonatal (birth-1mth), infancy (1-18mths), childhood (1.5-11yrs), adolescence (11-16yrs), and adulthood (>16yrs). If a disorder from the PWMs could not be linked to an HGNC gene name, a manual inspection was performed to include the full name and abbreviation of said disorder in the search criteria. The biomarkers were retrieved as HMDB IDs, with their reported positive or negative changed concentration information, and unified to the HMDB 4.0 structure. The visual depiction of relevant biomarkers using arrows as well as other characters indicating a value range was converted to a numeric scale (from -3 to +3) according to these rules: arrows depicting a singular value were represented as one numeric representation (e.g. three downward pointing arrows were transformed to -3), representations of arrow value ranges were converted to a numeric range (e.g. a range between two and one downward pointing arrow signifies a range '-2 to -1'); and descriptions of other characters (n and ± both indicating normal range, assumed

to be equal to zero) in combination with arrows (e.g. ‘n to 1 arrow pointing up’ as +0.5). Correlations between individual metabolic biomarkers and diseases were visualized in a heatmap (Euclidean distance) with the gplots package (version 3.1.1, https://cran.r-project.org/package=gplots); positively changed biomarkers were colored red (using three shades to show mildly, high, and very high), negatively changed markers blue (again three shades), and markers which were not altered for a disease were colored white.

Relevant Biomarker overlap

All biomarkers were mapped manually from ChEBI IDs to their corresponding HMDB IDs to match with the theoretical biomarker data; L enantiomers were chosen over D enantiomers for the annotations when this was unclear from the metabolite name. The changed biomarkers (workflow step 3) relevant for each patient were first converted to the same scale as the theoretical biomarkers (values for the log(2) change above 3 or below -3 were set at 3 and -3, respectively). Second, the patient data was visualized together with the theoretical biomarker visualization (workflow step 4) by matching HMDB IDs.

Pathway Selection

Relevant pathways for each patient were found through another semantic web query against the WikiPathways SPARQL endpoint (workflow step 2), matching the remaining changed biomarkers (workflow step 3), which have a log(2) change value not equal to zero. Relevant biomarkers which were not present in any pathway model were reported. The relevant pathways were sorted based on the highest

number of matching biomarkers. In total three pathways were automatically selected, based on including most unique biomarkers.

Data visualization

The data for each patient was visualized with the network analysis tool Cytoscape [32] (version 3.9.1), by using the Cytoscape REST API [33] (version v1) and WikiPathways App for Cytoscape [34] (version 3.3.10) through R. The absolute highest value for the log(2)change was used to determine the color scale, using a five-point scale to accommodate for small changes (values between -1.5 and 1.5) and high (abnormal) biomarker values. If no value was available for a node within the network, the fill color was set to gray. The patient data were linked to the IDs in the network by matching on the unified “ChEBI” column in Cytoscape.

Data Interpretation

For each patient, a network data visualization (workflow step 7) and relevant biomarker overlap heatmap (step 5) were provided to two Laboratory Specialists in Biochemical Genetics, after which narrative feedback on a potential diagnosis was collected. For each patient, a Cytoscape session file (step 7) is available on GitHub (Cytoscape_networks folder), as well as an interpretation overview containing patient characteristics containing age category, relevant biomarkers, personalized heatmap (workflow step 5), selected pathways, and clinical interpretation (Expert_interpretation folder).

2. Results

Workflow

A systems biology workflow was designed to connect clinical biomarker data for IMDs to pathway models. The biomarker data from several patients were visualized through network data analysis, which is discussed in more detail below.

Clinical Biomarker Data

Data from 16 patients with a variety of pyrimidine and urea cycle IMDs were considered to test our workflow. In total 88 markers were measured, 34 through the PUPY panel and 54 by the AA panel. Out of these biomarkers, two could not be annotated with one unified database ID from ChEBI. The first biomarker, CysHCys, could not be annotated with one unique ID; the second marker (2,8-dihydroxyadenine) could not be found in the ChEBI database and was therefore annotated with a Wikidata ID. Treatment compounds (oxypurinol and allopurinol) or compounds not measured by the panel anymore (argininosuccinic acid anhydride) were removed, which leaves 83 potential biomarkers. Due to missing reference data, the markers n-carbamyl-aspartate (CHEBI:32814), allantoin (CHEBI:15676), cytosine (CHEBI:16040), and cytidine (CHEBI:17562) were disregarded for further data analysis. Table 3 shows the number of remaining biomarkers for each patient after connecting the data to reference data, as well as the number of altered biomarkers per patient. This table further details how many biomarkers were missing from any PWM (workflow step 3), the pathway's coverage of biomarkers relevant for each patient (pathway), and the highest total of markers covered by one pathway.

Pathways Models

Since the chemical assay for pyrimidine metabolites also measures purine compounds, we created PWMs for both, as well as for the urea cycle. The final PUPY and urea cycle IMD models have been distributed through WikiPathways[25] with IDs wikipathways:WP4224 (purine), wikipathways:WP4225 (pyrimidine), and wikipathways:WP4571 (urea cycle), the latter was later extended with associated pathway information and diseases (wikipathways:WP4595). Last, two biomarker data pathways were developed, pyrimidine metabolism disorders (wikipathways:WP4584) and urea cycle disorders (wikipathways:WP4583); the biomarkers were connected to at least one reaction, either as substrate or product. Table 4 describes the content of each pathway, regarding proteins, metabolites, interactions, and described disorders.

Selection of Relevant Biomarkers

The annotated data from workflow step 1 (Figure 1) was used to find relevant pathways for visualization, which led to 171 pathways in total, including one or more distinct biomarkers. Fifteen of these pathways contain 10 or more markers, which are displayed in Table 5. There were seven markers not part of any pathway model (with the corresponding ChEBI IDs: 28315, 40279, 17755, 43433, 89698, 49015, 61511). For individual patients, the number of relevant biomarkers and biomarkers not in any PWM are depicted in Table 3, as well as the number of biomarkers captured by the highest-scoring PWM.

Theoretical Biomarker Data

Since the chemical assay for pyrimidine metabolites also measures purine compounds, we collected theoretical biomarker data for both pathways, in addition to

the urea cycle. In total, 17 unique purine, 10 pyrimidine, and 8 urea cycle IMDs were included in this study (a total of 35 disorders). For these three classes of IMDs, 27 unique biomarkers were found to be relevant for the sample matrix urine. Taking available reference data into account and the availability of an HMDB ID, a maximum of 22 unique biomarkers could be linked to 23 IMDs; Figure 2 shows an example visualization for the age category 0 up to 1 year old, which correlates the disorders based on their theoretical biomarkers change, sorted on their respective IMD class. For the following IMDs, no data was available: TPMT, PRPS1, TYMP, TK2, RRM2B, ITPA, IMPDH1, DHODH, and AMPD1. Three disorders in the urea cycle category - CPS1, ARG1, NAGS - and one IMD in the purine pathway - DGUOK - were indistinguishable based on the retrieved biomarkers from IEMBase. Two additional disorders - SLC25A13 (urea cycle) and NT5C3A (pyrimidine) - were the closest match to these previous four indistinguishable IMDs. The group of pyrimidine IMDs (AGXT2, UMPS, DPYD, DPYS, UBP1), urea cycle (ASL, ASS1), and purine (ADA, APRT, ADSL, ATIC) contained some discriminating values for the biomarkers, enabling distinction between these individual disorders based on only biochemical markers in theory. OTC and SLC25A15 (both urea cycle disorders) can be difficult to differentiate from one another based on theoretical biomarkers. A group of purine IMDs - XAN2, XO, PNP, HPRT1 (with two known phenotypes), PRSP1 - can be considered control cases, since these too should allow for distinction between IMDs based on theoretically relevant biomarkers.

Relevant Biomarker overlap

Figure 2 also shows biomarker data for patient I (bottom row in purple), who was diagnosed with DPYD, and how this data correlates with theoretical biochemical markers in urine for individual IMDs in the purine, pyrimidine, and urea cycle pathways. The sample obtained from this patient showed three additionally changed biomarkers as compared to the theoretical values, as well as a distinct correlation to the theoretical biomarker profile of DPYD.

Pathway Selection

By only querying the pathway data for relevant biomarkers (instead of all markers in a panel), a customized visualization was possible for each patient. The selection of the top three pathways containing the highest number of unique markers was performed automated. However, this selection can also be performed manually (for example by an expert with knowledge of relevant pathways), aiming to include pathways containing biomarkers with the largest change. Table 3 shows how many pathways were relevant for each patient, as well as how many unique biomarkers the highest-scoring pathway covers. Table 6 shows which pathways ended up in the top three for each patient, as well as which biomarkers were not part of the visualization after selecting the top three.

Data Visualization

Figure 3 shows the network visualization for patient I on the ‘Biomarkers for pyrimidine metabolism disorders’ (left) and ‘Purine metabolism’ (right). Two metabolites (thymine, uracil) which are directly converted by DPD show elevated levels; one direct downstream metabolite of thymine (5-OH-methyluracil) also shows

a high concentration. Two downstream metabolites of DPD (dihydrouracil, beta-alanine) are found to be within the healthy reference values, whereas (S)-beta-aminoisobutyrate shows a decreased concentration. Last, the elevated level of SAICA-riboside was not expected for this disorder and might suggest immaturity of the patient.

Data Interpretation

The two laboratory specialists in biochemical genetics combined the heatmap from the 'Relevant Biomarker overlap' section and the network biomarker data visualization from the 'Data Visualization' section to arrive at an IMD diagnosis. Nine disorders out of the 16 patient samples were diagnosed with the correct IMD, whereas for four patients the visualization helped to suggest further informative assays (the last column of Table 6 summarizes these results). These latter patients included one case of Dihydropyrimidine dehydrogenase deficiency (DPD - patient E), and three cases of Ornithine Transcarbamylase deficiency (OTC - patients G, R, and S).

Samples from patients under treatment, e.g. patient H (diagnosed with hyperornithinemia- hyperammonemia- homocitrullinuria (HHH) syndrome, also known as ornithine translocase (SLC25A15) deficiency) receiving citrulline, were difficult to diagnose since the workflow did not distinguish between abnormal biomarker values due to treatment or caused by the IMD. This result highlights the importance of combining metabolite levels with clinical information. The computational workflow should only be used on data from untreated persons. Patient J (diagnosed with Beta-ureidopropionase deficiency, UPB1D) was not correctly

diagnosed by both experts, which we attribute to the very mild disturbances in the biomarker patterns. Last, patient O (diagnosed with Ornithine Transcarbamylase deficiency, OTC), showed an unexpectedly higher value for arginine rather than ornithine.

3. Discussion

Workflow

Our proposed workflow was based on combining semantic web and network analysis techniques. Together, these techniques created the possibility to visualize clinical biochemical assay data on the pathway level, allowing for a more detailed interpretation of the connections between the different markers. The developed workflow was designed for individual patient analysis and optimized for pyrimidine and urea cycle disorders with biomarkers measured through targeted assays. We believe this workflow can be used to construct a similar system for other IMDs, however, the following limitations should be taken into account for each step of the workflow.

Clinical Biomarker Data

First, all chemicals in the assays were annotated manually based on their (Dutch) name; using persistent IDs to annotate data is a key aspect to enable open science [35], and ultimately leads to FAIR data [36]. We advise the (rare) disease community to include an ID for each entry, based on a publically available database, in order to support the development of computational models. This process led to two biomarkers with different annotations. The first, CysHCys, is a disulfide from cysteine

and homocysteine; annotating this compound was not possible based on the name provided. Drawing the chemical structure and converting the structure to a SMILES [37] could be used to annotate the compound. The second biomarker, 2,8-dihydroxyadenine, could not be located in the database we used for harmonization (ChEBI), however, could be annotated with a Wikidata ID.
Second, several biomarkers which were recently added to the targeted clinical assays (n-carbamoyl-aspartate, allantoin, cytosine, and cytidine) had not been measured in enough control samples to calculate a reliable healthy control concentration range; these values are expected to be available in the future. The HMDB database [31] includes information on these biomarkers which might be useful to integrate, however the first biomarker (known in HMDB as ureidosuccinic acid) only has reference data available for saliva samples. The latter three compounds all have reference data available in HMDB for urine samples, however, these values were partly derived from NMR measurements, which are not directly comparable to the mass spectrometer data used in this study.

Pathways Models

Creating PWMs annotated with resolvable IDs for the entities within the pathway [38] were crucial for data analysis and visualization [39]. Several initiatives merge pathway information [40,41] based on gene and protein content, rather than metabolite and chemical conversion levels, which makes them unsuitable for IMD metabolic data analysis. Several biomarkers of PUPY and urea cycle IMDs are numerous reaction steps removed from the original malfunctioning protein; covering all steps up to the originally affected protein will differ for each disorder and each

marker. Also, this reaction information is scattered over publications in images and text [42] as well as various databases [43], which requires dedicated curation time to arrive at a PWM covering all relevant interactions. Lacking automatable access to pathway content or licenses prohibiting data extraction was another hurdle in acquiring all relevant interactions for biomarkers. Last, completing the models presented in this study was a ponderous task, due to the lack of naming standardization in databases and papers alike, as also has been noted in systems pharmacology approaches [44]. This is despite the IUPAC-nomenclature rules [45] and available software to translate IUPAC names to chemical structures [46] and vice versa [47].

Selection of Relevant Biomarkers

Only after dedicated manual annotations of clinical biomarkers, missing data in pathway models could be evaluated. For this study, seven out of 83 markers could not be found in any of the consulted pathway data. In order to overcome a mismatch of biomarkers and pathway data, clinical biomarker IDs could be converted to their corresponding neutral molecular structure InChIKey [48] ID or substructure matching [49].

Theoretical Biomarker Data

Connecting all disease IDs to their counterpart protein ID and data from the IEMbase [30] could only be performed manually. The IDs of the biomarkers were a mixture of HMDB 3.0 and 4.0, which led to mapping issues from HMDB to the ChEBI database. We also found that some syndromes (Lesch-Nyhan versus Kelley-Seegmiller, also

known as HPRT-related gout hyperuricemia) were treated as individual disorders by one database (OMIM), while the other combines the information on both disorders in one entry (IEMBase). For another two disorders, pyrimidine 5'-nucleotidase superactivity versus pyrimidine 5'-nucleotidase I deficiency, IEMbase has two separate entries (with different biomarkers), while OMIM only includes pyrimidine 5'-nucleotidase I deficiency. To facilitate data integration and comparison on an automated basis, we advise providing programmatic access to biomedical databases, for example through an API [50] or SPARQL endpoint [51]. We also want to encourage harmonizing the information in databases, for example by using existing ontologies such as the Human Disease Ontology [52], Human Phenotype Ontology [53], or Nosology on Inherited Metabolic Disorders [54]. Last, several disorders could not be distinguished based on the available biomarkers in IEMBase; the workflow can be extended with in-house biomarker data, which could lead to a clearer separation between certain disorders, or with additional information from other databases and literature; however absolute changes might be difficult to compare with the existing data.

Relevant Biomarker overlap

Several IMDs could not be connected to one or more metabolic biomarkers measured in urine. Furthermore, the log2(change) data of patients was converted to the -3 to +3 scale from IEMBase, where the contribution of highly altered biomarkers to the correlation might get obscured.

Pathway Selection

Selecting the top three pathways covering the most unique biomarkers could leave out markers very distinctive for a disorder, or not take into account the amount of change; biomarkers with a high change might be more descriptive for a disorder as compared to markers showing marginal changes. Several pathways overlap in terms of content, which could conceal potentially relevant pathways. The query to select the top three pathways covering most unique biomarkers did not take into account which proteins were catalyzing the reactions. The workflow leaves room for manual selection of potentially relevant pathways by experts, which could be aided by reviewing the patient-specific heatmap (step 6). Finally, some pathways were relatively large in terms of node size and captured reactions, leading to higher coverage of biomarkers but also to a larger distance between the individual markers, which diminishes visualizing a clear biological cause and effect path between the affected protein and the relevant biomarkers.

Data Visualization

Connecting data to pathway and network models often requires a form of ID mapping. Unifying the IDs to one database per biological type was a crucial step to avoid mismatches or one-to-many mapping issues. When a biomarker was present at the outer border of a pathway and depicted solely as a reactant, data interpretation might be hindered since this leads to dispersed data visualization. Furthermore, since not all clinical data could be visualized directly in one PWM (due to the biomarkers being spread over multiple models), other approaches will be needed to overcome the boundaries imposed by individual models. Even though the WikiPathways RDF contains pathway data from Reactome pathways, we decided to

not use these pathways in our workflow. The model conversion [55] from the native Reactome pathway modeling language SBGN to the language behind PathVisio, GPML, leads to quite some biomarkers captured as unconnected nodes. Other possibilities to visualize pathway data in a network tool such as Cytoscape include the Reactome Cytoscape Plugin [56], which supports automation however has been designed for cancer and other types of diseases network analysis focussing on gene and protein content. Other pathway data is made available in Cytoscape for KEGG [57], which can be automated. Finding relevant pathways, however, was not possible due to a lack of an API or SPARQL endpoint. Two other pathway apps, CyPath2 [41] and cy3sabiork [58] could not be automated.

Data Interpretation

Even though recent advances in clinical biomarker measurements [17] in urine samples have aided in the diagnosis of some IMDs, most markers are currently not used for newborn (dried blood spot) screening [59], due to limited detectability of the biomarkers in these spots and with most pyrimidine IMDs missing treatment options. Within our workflow, we only took the urinary biomarkers into account, which could be complemented with additional data from blood samples or other relevant matrices.

Currently, diagnostic laboratories for inborn errors of metabolism often report exact metabolite concentrations in diagnostic patient reports; Z-scores are (rarely) used as a measure to compare analyte values to the analyte values from a control population. Pathway and network models, on the contrary, normally report (log)-fold changes to compare two groups (diseased versus control) or z-scores and pathway

names with probabilities as a measure for over-representation of genes, proteins, and/or metabolites. We believe that laboratory specialists would benefit from learning to interpret new diagnostic tools. In this study, the network model helped to easily diagnose 9 out of 16 patient samples, and pointed in the correct direction or suggested follow-up analysis for 4 patients. Patient H (diagnosed with hyperornithinemia-hyperammonemia-homocitrullinuria (HHH) syndrome) was found to have more elevated biomarkers than noted in IEMbase (which mentions increased excretion of homocitrulline and orotic acid), while the network visualization showed increased excretion of several amino acids in the urea cycle including argininosuccinate. The latter usually is not present in urine and as such is a biomarker for ASL deficiency. However, this patient received citrulline treatment which can lead to an increased presence of argininosuccinate. This example shows that laboratory experts are needed to interpret network and pathway data visualizations and combine their knowledge of patient treatment and symptoms.
For the validation of a recently implemented diagnostic tool called targeted urine metabolomics (TUM) [60], similar samples were used as discussed in our study. When comparing the data interpretation from TUM and the systems biology workflow, we can deduce that similar interpretations were reached. This agreement was also found for the Ornithine Transcarbamylase deficiency (OTC) cases, which remains difficult to diagnose in women since the disorder is X-linked causing an atypical biomarker pattern. Patient O is an example of such, where we hypothesize that the cyclic metabolic conversion of arginine, argininosuccinate, and ornithine into one another could be the cause of this unexpected pattern. To understand these atypical cases of OTC and corresponding biomarker patterns, data from more

patients is required and we recommend other laboratories to share their data on IMDs if possible. For future studies, the interest should shift to measuring metabolite fluxes [2] over a longer timespan to better understand how protein intake triggers disorders [61] and if fasting could be used to treat patients with hypoglycemic symptoms [62].

4. Conclusions

With this study, we show the potential of a Systems Biology approach combining semantic web technologies for data linking and network analysis for data visualization, to directly connect biological pathway knowledge to clinical cases and biomarker data. The presented workflow is adaptable to analyze different types of IMDs, difficult patient cases, and functional assays in the future, which opens up the possibility for usage in the diagnostic workflow. Information on treatment and clinical conditions remains important for accurate diagnosis, as well as expert interpretation of all information combined into this workflow. Furthermore, several steps in the workflow are now highly dependent on the manual curation of data and databases requiring expert knowledge of the information therein. The issues highlighted in the discussion section should be overcome in the future to allow our developed workflow to be easily used for other IMDs, by adding persistent identifiers to (clinical) biomarker data, allowing automatable data downloads from relevant databases, and creating computer-readable pathway models from pathway figures.

Declarations:

Ethics approval and consent to participate

The Medical Research Involving Human Subjects Act (WMO) did not apply to the study mentioned above, official approval of this study by our committee was not required according to the Medical-Ethical Review committee (METC) azM/UM.

Consent for publication

Not applicable.

Availability of data and materials

The datasets supporting the conclusions of this article are available in the IMD-PUPY repository, https://bigcat-um.github.io/IMD-PUPY .

Funding

Research reported in this manuscript was supported by the European Union's Horizon 2020 research and innovation program under the EJP RD COFUND-EJP N° 825575.

Competing interests

The authors declare that they have no competing interests.

Authors and Affiliations

Department of Bioinformatics (BiGCaT), NUTRIM, Maastricht University, Maastricht, The Netherlands
Denise N. Slenter, Irene M.G.M. Hemel, Chris T. Evelo, Egon L. Willighagen
Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands
Irene M.G.M. Hemel, Chris T. Evelo
Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands
Jörgen Bierau, Laura K.M. Steinbusch
Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands
Jörgen Bierau
Consortia
The European Joint Programme on Rare Diseases (EJP RD).
Corresponding author
Correspondence to Denise Slenter: denise.slenter@maastrichtuniversity.nl

Contributions

DS: Conception and design, analysis and interpretation of data, drafted the manuscript with tables and figures (correspondence).
IH: analysis of data, performed initial review, and critically reviewed the manuscript.
CE: critically reviewed the manuscript.
JB: Conception and design, provided data, interpretation of data, performed initial review, and critically reviewed the manuscript.
EW: conducted the initial review, and critically reviewed the manuscript.
LS: conception and design, provided data, interpretation of data, performed initial review, and critically reviewed the manuscript.

Acknowledgments

We are grateful to all participating patients and their families. Four patient samples were kindly provided by Dr. P. Fitzsimons, Dr. A. Monavari, and Dr. E. Crushell from

the Children's University Hospital and the National Centre for Inherited Metabolic Disorders, Temple Street, Ireland. We also want to acknowledge the work done by the laboratory technicians Huub Waterval, Marjon Ortmans-Ploemen, Sandra Busch, Jean-Pierre Bollen, and Karin Habets-van der Poel from the laboratory of Clinical Genetics at Maastricht University Medical Center (MUMC+).

Abbreviations

Workflow:

AA: Amino acids (panel)

IDs: Identifiers

IMDs: Inherited Metabolic Disorders

PWMs: Pathway Models

PUPY: purine and pyrimidine (panel)

RDF: Resource Description Framework

Disorders:

ADA: Adenosine deaminase

ADSL: Adenylosuccinate lyase

AGXT2: Alanine--glyoxylate aminotransferase 2

AMPD1: Adenosine monophosphate deaminase 1

APRT: Adenine phosphoribosyltransferase

ARG1: Arginase 1

ASL: Argininosuccinate lyase

ASS1: Argininosuccinate synthase 1

ATIC: 5-aminoimidazole-4-carboxamide ribonucleotide formyltransferase/IMP cyclohydrolase

CPS1: Carbamoyl-phosphate synthase 1

DGUOK: Deoxyguanosine kinase

DHODH: Dihydroorotate dehydrogenase (quinone)

DPYD: Dihydropyrimidine dehydrogenase dihydropyrimidine

DPYS: Dihydropyrimidinase

HPRT1: Hypoxanthine phosphoribosyltransferase 1

IMPDH1: Inosine monophosphate dehydrogenase 1

ITPA: Inosine triphosphatase

NAGS: N-acetylglutamate synthase

NT5C3A: 5'-nucleotidase, cytosolic IIIA

OTC: Ornithine carbamoyltransferase

PNP: Purine nucleoside phosphorylase

PRPPs (class of enzymes): Ribose-phosphate diphosphokinase

PRPS1: Phosphoribosyl pyrophosphate synthetase 1

RRM2B: Ribonucleotide reductase regulatory TP53 inducible subunit M2B

SLC25A13: Solute carrier family 25 member 13

SLC25A15: Solute carrier family 25 member 15 (ornithine translocase)

TK2: Thymidine kinase 2

TPMT: Thiopurine S-methyltransferase

TYMP: Thymidine phosphorylase

UBP1: Upstream binding protein 1

UMPS: Uridine monophosphate synthetase

XAN2 (gene: XDH): Xanthine dehydrogenase

XO (gene: MOCOS): Molybdenum cofactor sulfurase

References

1. Fukao T, Nakamura K. Advances in inborn errors of metabolism. J Hum Genet [Internet]. 2019;64:65. Available from: http://dx.doi.org/10.1038/s10038-018-0535-7

2. Lanpher B, Brunetti-Pierri N, Lee B. Inborn errors of metabolism: the flux from Mendelian to complex diseases. Nat Rev Genet [Internet]. 2006;7:449–60. Available from: http://dx.doi.org/10.1038/nrg1880

3. Ferreira CR, Rahman S, Keller M, Zschocke J, ICIMD Advisory Group. An international classification of inherited metabolic disorders (ICIMD). J Inherit Metab Dis [Internet]. 2021;44:164–77. Available from: http://dx.doi.org/10.1002/jimd.12348

4. Burton BK. Inborn errors of metabolism in infancy: a guide to diagnosis. Pediatrics [Internet]. 1998;102:E69. Available from: http://dx.doi.org/10.1542/peds.102.6.e69

5. Adhikari AN, Gallagher RC, Wang Y, Currier RJ, Amatuni G, Bassaganyas L, et al. The role of exome sequencing in newborn screening for inborn errors of metabolism. Nat Med [Internet]. 2020;26:1392–7. Available from: http://dx.doi.org/10.1038/s41591-020-0966-5

6. Coene KLM, Kluijtmans LAJ, van der Heeft E, Engelke UFH, de Boer S, Hoegen B, et al. Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients [Internet]. Journal of Inherited Metabolic Disease. 2018. p. 337–53. Available from: http://dx.doi.org/10.1007/s10545-017-0131-6

7. Stenton SL, Kremer LS, Kopajtich R, Ludwig C, Prokisch H. The diagnosis of inborn errors of metabolism by an integrative “multi-omics” approach: A perspective encompassing genomics, transcriptomics, and proteomics. J Inherit Metab Dis [Internet]. 2020;43:25–35. Available from: http://dx.doi.org/10.1002/jimd.12130

8. Arnold GL. Inborn errors of metabolism in the 21 century: past to present. Ann Transl Med [Internet]. 2018;6:467. Available from: http://dx.doi.org/10.21037/atm.2018.11.36

9. Burrage LC, Thistlethwaite L, Stroup BM, Sun Q, Miller MJ, Nagamani SCS, et al. Untargeted metabolomic profiling reveals multiple pathway perturbations and new clinical biomarkers in urea cycle disorders. Genet Med [Internet]. 2019;21:1977–86. Available from: http://dx.doi.org/10.1038/s41436-019-0442-0

10. Jurecka A. Inborn errors of purine and pyrimidine metabolism. J Inherit Metab Dis [Internet]. 2009;32:247–63. Available from: http://dx.doi.org/10.1007/s10545-009-1094-z

11. Balasubramaniam S, Duley JA, Christodoulou J. Inborn errors of purine metabolism: clinical update and therapies. J Inherit Metab Dis [Internet]. 2014;37:669–86. Available from: http://dx.doi.org/10.1007/s10545-014-9731-6

12. Balasubramaniam S, Duley JA, Christodoulou J. Inborn errors of pyrimidine metabolism: clinical update and therapy. J Inherit Metab Dis [Internet]. 2014;37:687–98. Available from: http://dx.doi.org/10.1007/s10545-014-9742-3

13. Smith W, Kishnani PS, Lee B, Singh RH, Rhead WJ, Sniderman King L, et al. Urea cycle disorders: clinical presentation outside the newborn period. Crit Care Clin [Internet]. 2005;21:S9–17. Available from: http://dx.doi.org/10.1016/j.ccc.2005.05.007

14. Allaire JJ, Xie Y, McPherson J, Luraschi J, Ushey K, Atkins A, et al. rmarkdown: Dynamic Documents for R [Internet]. 2021. Available from: https://github.com/rstudio/rmarkdown

15. R Core team. R: A Language and Environment for Statistical Computing [Internet]. 2020 [cited 2021 Sep 11]. Available from: https://www.r-project.org/

16. RStudio Team. RStudio: Integrated Development Environment for R [Internet]. 2019 [cited 2021 Sep 11]. Available from: https://www.rstudio.com/

17. Monostori P, Klinke G, Hauke J, Richter S, Bierau J, Garbade SF, et al. Extended diagnosis of purine and pyrimidine disorders from urine: LC MS/MS assay development and clinical validation. PLoS One [Internet]. 2019;14:e0212458. Available from: http://dx.doi.org/10.1371/journal.pone.0212458

18. Waterval WAH, Scheijen JLJM, Ortmans-Ploemen MMJC, Habets-van der Poel CD, Bierau J. Quantitative UPLC-MS/MS analysis of underivatised amino acids in body fluids is a reliable tool for the diagnosis and follow-up of patients with inborn errors of metabolism. Clin Chim Acta [Internet]. 2009;407:36–42. Available from: http://dx.doi.org/10.1016/j.cca.2009.06.023

19. Hastings J, Owen G, Dekker A, Ennis M, Kale N, Muthukrishnan V, et al. ChEBI in 2016: Improved services and an expanding collection of metabolites [Internet]. Nucleic Acids Research. 2016. p. D1214–9. Available from: http://dx.doi.org/10.1093/nar/gkv1031

20. Waagmeester A, Stupp G, Burgstaller-Muehlbacher S, Good BM, Griffith M, Griffith OL, et al. Wikidata as a knowledge graph for the life sciences. Elife [Internet]. 2020;9. Available from: http://dx.doi.org/10.7554/eLife.52614

21. Kutmon M, van Iersel MP, Bohler A, Kelder T, Nunes N, Pico AR, et al. PathVisio 3: an extendable pathway analysis toolbox. PLoS Comput Biol [Internet]. 2015;11:e1004085. Available from: http://dx.doi.org/10.1371/journal.pcbi.1004085

22. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. Nucleic Acids Res [Internet]. 2021;49:D480–9. Available from: http://dx.doi.org/10.1093/nar/gkaa1100

23. Lombardot T, Morgat A, Axelsen KB, Aimo L, Hyka-Nouspikel N, Niknejad A, et al. Updates in Rhea: SPARQLing biochemical reaction data. Nucleic Acids Res [Internet]. 2019;47:D596–600. Available from: http://dx.doi.org/10.1093/nar/gky876

24. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. Nucleic Acids Res [Internet]. 2019;47:D1038–43. Available from: http://dx.doi.org/10.1093/nar/gky1151

25. Martens M, Ammar A, Riutta A, Waagmeester A, Slenter DN, Hanspers K, et al. WikiPathways: connecting communities. Nucleic Acids Res [Internet]. 2021;49:D613–21. Available from: http://dx.doi.org/10.1093/nar/gkaa1024

26. Waagmeester A, Kutmon M, Riutta A, Miller R, Willighagen EL, Evelo CT, et al. Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources. PLoS Comput Biol [Internet]. 2016;12:e1004989. Available from: http://dx.doi.org/10.1371/journal.pcbi.1004989

27. Slenter D, WikiPathways Sept 2021 Release-RDF Data. WikiPathways Sept 2021 Release - RDF data [Internet]. Zenodo; 2021. Available from: https://zenodo.org/record/5632921

28. Galgonek J, Hurt T, Michlíková V, Onderka P, Schwarz J, Vondrášek J. Advanced SPARQL querying in small molecule databases. J Cheminform [Internet]. 2016;8:31. Available from: http://dx.doi.org/10.1186/s13321-016-0144-4

29. Tweedie S, Braschi B, Gray K, Jones TEM, Seal RL, Yates B, et al. Genenames.org: the HGNC and VGNC resources in 2021. Nucleic Acids Res [Internet]. 2021;49:D939–46. Available from: http://dx.doi.org/10.1093/nar/gkaa980

30. Lee JJY, Wasserman WW, Hoffmann GF, van Karnebeek CDM, Blau N. Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism. Genet Med [Internet]. 2018;20:151–8. Available from: http://dx.doi.org/10.1038/gim.2017.108

31. Wishart DS, Feunang YD, Marcu A, Guo AC, Liang K, Vázquez-Fresno R, et al. HMDB 4.0: the human metabolome database for 2018. Nucleic Acids Res [Internet]. 2018;46:D608–17. Available from: http://dx.doi.org/10.1093/nar/gkx1089

32. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res [Internet]. 2003;13:2498–504. Available from: http://dx.doi.org/10.1101/gr.1239303

33. Otasek D, Morris JH, Bouças J, Pico AR, Demchak B. Cytoscape Automation: empowering workflow-based network analysis. Genome Biol [Internet]. 2019;20:185. Available from: http://dx.doi.org/10.1186/s13059-019-1758-4

34. Kutmon M, Lotia S, Evelo CT, Pico AR. WikiPathways App for Cytoscape: Making biological pathways amenable to network analysis and visualization. F1000Res [Internet]. 2014;3:152. Available from: http://dx.doi.org/10.12688/f1000research.4254.2

35. Dappert A, Farquhar A, Kotarski R, Hewlett K. Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research [Internet]. Data Science Journal. 2017. Available from: http://dx.doi.org/10.5334/dsj-2017-028

36. Wilkinson MD, Dumontier M, Aalbersberg IJJ, Appleton G, Axton M, Baak A, et al. The FAIR Guiding Principles for scientific data management and stewardship. Sci Data [Internet]. 2016;3:160018. Available from: http://dx.doi.org/10.1038/sdata.2016.18

37. Weininger D. SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules [Internet]. Journal of Chemical Information and Modeling. 1988. p. 31–6. Available from: http://dx.doi.org/10.1021/ci00057a005

38. Kyle JE, Aimo L, Bridge AJ, Clair G, Fedorova M, Helms JB, et al. Interpreting the lipidome: bioinformatic approaches to embrace the complexity. Metabolomics [Internet]. 2021;17:55. Available from: http://dx.doi.org/10.1007/s11306-021-01802-6

39. Hanspers K, Kutmon M, Coort SL, Digles D, Dupuis LJ, Ehrhart F, et al. Ten simple rules for creating reusable pathway models for computational analysis and visualization. PLoS Comput Biol [Internet]. Public Library of Science; 2021 [cited 2021 Sep 23];17:e1009226. Available from: https://journals.plos.org/ploscompbiol/article/file?id=10.1371/journal.pcbi.1009226&type=printable

40. Domingo-Fernández D, Hoyt CT, Bobis-Álvarez C, Marín-Llaó J, Hofmann-Apitius M. ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases. NPJ Syst Biol Appl [Internet]. 2019;5:3. Available from: http://dx.doi.org/10.1038/s41540-018-0078-8

41. Rodchenkov I, Babur O, Luna A, Aksoy BA, Wong JV, Fong D, et al. Pathway Commons 2019 Update: integration, analysis and exploration of pathway data. Nucleic Acids Res [Internet]. 2020;48:D489–97. Available from: http://dx.doi.org/10.1093/nar/gkz946

42. Hanspers K, Riutta A, Summer-Kutmon M, Pico AR. Pathway information extracted from 25 years of pathway figures. Genome Biol [Internet]. 2020;21:273. Available from: http://dx.doi.org/10.1186/s13059-020-02181-2

43. Stobbe MD, Houten SM, Jansen GA, van Kampen AHC, Moerland PD. Critical assessment of human metabolic pathway databases: a stepping stone for future integration. BMC Syst Biol [Internet]. 2011;5:165. Available from: http://dx.doi.org/10.1186/1752-0509-5-165

44. Benson HE, Watterson S, Sharman JL, Mpamhanga CP, Parton A, Southan C, et al. Is systems pharmacology ready to impact upon therapy development? A study on the cholesterol biosynthesis pathway. Br J Pharmacol [Internet]. 2017;174:4362–82. Available from: http://dx.doi.org/10.1111/bph.14037

45. Hellwich K-H, Hartshorn RM, Yerin A, Damhus T, Hutton AT. Brief Guide to the Nomenclature of Organic Chemistry (IUPAC Technical Report) [Internet]. IUPAC Standards Online. Available from:

http://dx.doi.org/10.1515/iupac.92.0027

46. Lowe DM, Corbett PT, Murray-Rust P, Glen RC. Chemical name to structure: OPSIN, an open source solution. J Chem Inf Model [Internet]. 2011;51:739–53. Available from: http://dx.doi.org/10.1021/ci100384d

47. Rajan K, Zielesny A, Steinbeck C. STOUT: SMILES to IUPAC names using neural machine translation. J Cheminform [Internet]. 2021;13:34. Available from: http://dx.doi.org/10.1186/s13321-021-00512-4

48. Heller SR, McNaught A, Pletnev I, Stein S, Tchekhovskoi D. InChI, the IUPAC International Chemical Identifier. J Cheminform [Internet]. 2015;7:23. Available from: http://dx.doi.org/10.1186/s13321-015-0068-4

49. Kratochvíl M, Vondrášek J, Galgonek J. Sachem: a chemical cartridge for high-performance substructure search. J Cheminform [Internet]. 2018;10:27. Available from: http://dx.doi.org/10.1186/s13321-018-0282-y

50. Tarkowska A, Carvalho-Silva D, Cook CE, Turner E, Finn RD, Yates AD. Eleven quick tips to build a usable REST API for life sciences. PLoS Comput Biol [Internet]. 2018;14:e1006542. Available from: http://dx.doi.org/10.1371/journal.pcbi.1006542

51. Marshall MS, Boyce R, Deus HF, Zhao J, Willighagen EL, Samwald M, et al. Emerging practices for mapping and linking life sciences data using RDF — A case series. Web Semant [Internet]. Elsevier BV; 2012;14:2–13. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1570826812000376

52. Schriml LM, Mitraka E, Munro J, Tauber B, Schor M, Nickle L, et al. Human Disease Ontology 2018 update: classification, content and workflow expansion. Nucleic Acids Res [Internet]. 2019;47:D955–62. Available from: http://dx.doi.org/10.1093/nar/gky1032

53. Köhler S, Gargano M, Matentzoglu N, Carmody LC, Lewis-Smith D, Vasilevsky NA, et al. The Human Phenotype Ontology in 2021. Nucleic Acids Res [Internet]. 2021;49:D1207–17. Available from: http://dx.doi.org/10.1093/nar/gkaa1043

54. Ferreira CR, van Karnebeek CDM, Vockley J, Blau N. A proposed nosology of inborn errors of metabolism. Genet Med [Internet]. 2019;21:102–6. Available from: http://dx.doi.org/10.1038/s41436-018-0022-8

55. Bohler A, Wu G, Kutmon M, Pradhana LA, Coort SL, Hanspers K, et al. Reactome from a WikiPathways Perspective. PLoS Comput Biol [Internet]. 2016;12:e1004941. Available from: http://dx.doi.org/10.1371/journal.pcbi.1004941

56. Wu G, Feng X, Stein L. A human functional protein interaction network and its application to cancer data analysis [Internet]. Genome Biology. 2010. p. R53. Available from: http://dx.doi.org/10.1186/gb-2010-11-5-r53

57. Nishida K, Ono K, Kanaya S, Takahashi K. KEGGscape: a Cytoscape app for pathway data integration. F1000Res [Internet]. 2014;3:144. Available from: http://dx.doi.org/10.12688/f1000research.4524.1

58. König M. cy3sabiork: A Cytoscape app for visualizing kinetic data from SABIO-RK [Internet]. F1000Res. (approved with reservations). 2016. Available from: http://dx.doi.org/10.1101/062091

59. Kruszka P, Regier D. Inborn Errors of Metabolism: From Preconception to Adulthood. Am Fam Physician [Internet]. 2019;99:25–32. Available from: https://www.ncbi.nlm.nih.gov/pubmed/30600976

60. Steinbusch LKM, Wang P, Waterval HWAH, Stassen FAPM, Coene KLM, Engelke UFH, et al. Targeted urine metabolomics with a graphical reporting tool for rapid diagnosis of inborn errors of metabolism. J Inherit Metab Dis [Internet]. 2021;44:1113–23. Available from:

http://dx.doi.org/10.1002/jimd.12385

61. Deodato F, Boenzi S, Santorelli FM, Dionisi-Vici C. Methylmalonic and propionic aciduria. Am J Med Genet C Semin Med Genet [Internet]. 2006;142C:104–12. Available from: http://dx.doi.org/10.1002/ajmg.c.30090

62. Scholl-Bürgi S, Höller A, Pichler K, Michel M, Haberlandt E, Karall D. Ketogenic diets in patients with inherited metabolic disorders. J Inherit Metab Dis [Internet]. 2015;38:765–73. Available from: http://dx.doi.org/10.1007/s10545-015-9872-2

Tables and Figures

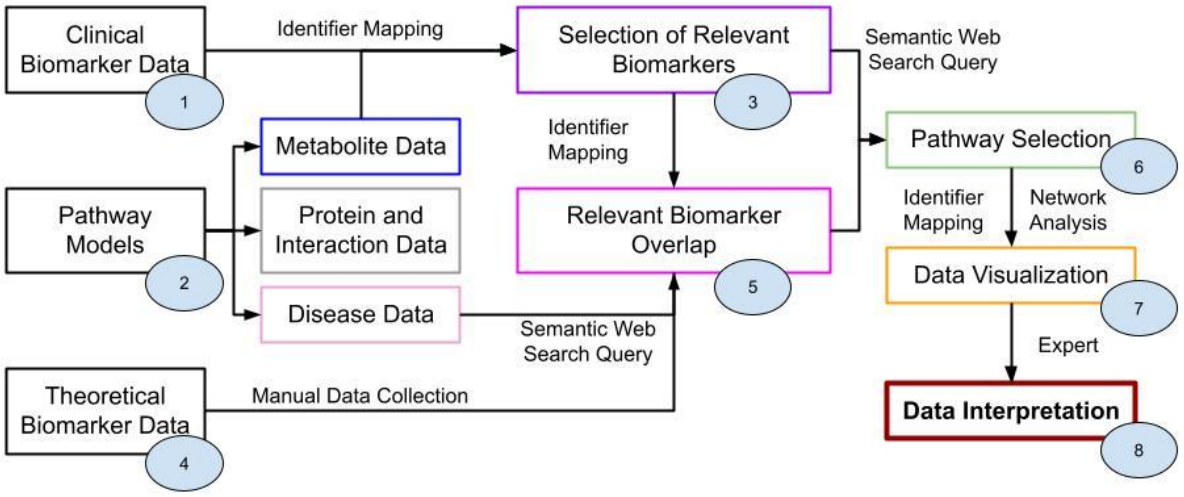

Figure 1: Depiction of workflow to interpret clinical data of IMDs through network analysis; each circle refers to a specific step within the Materials and Methods Section.

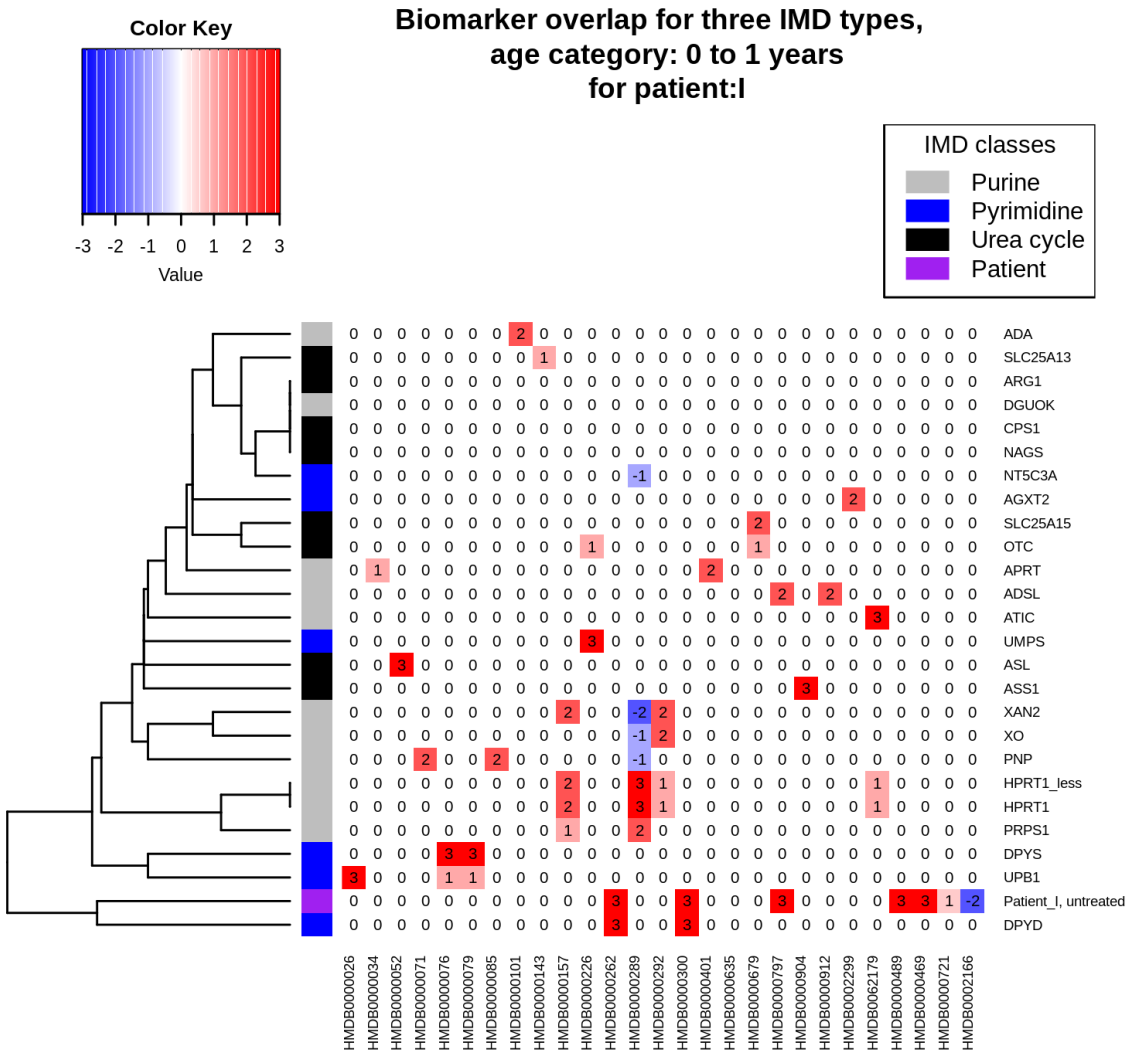

Figure 2: Visualization of how theoretical biochemical markers in urine overlap for individual IMDs in the purine, pyrimidine, and urea cycle pathways. Protein names correspond to the following disorders: ADA: Adenosine deaminase deficiency, SLC25A13: Citrin deficiency, ARG1: Arginase deficiency, DGUOK: Deoxyguanosine kinase deficiency, CPS1: Carbamoyl phosphate synthetase I deficiency, NAGS: N-Acetylglutamate synthase deficiency, NT5C3A: Pyrimidine 5-nucleotidase superactivity, AGXT2: Beta-aminoisobutyrate-pyruvate transaminase deficiency, SLC25A15: Ornithine transporter deficiency, OTC: Ornithine transcarbamylase deficiency, APRT: Adenine phosphoribosyltransferase deficiency, ADSL: Adenyl- succinate lyase deficiency, ATIC: AICAr transformylase/IMP cyclohydrolase deficiency, UMPS: Orotic aciduria type I, ASL: Argininosuccinic aciduria, ASS1: Citrullinemia type I, XAN2: Xanthinuria, Type II, XO: Xanthinuria, Type I, PNP: Purine nucleoside phosphorylase deficiency, HPRT1_less: Kelley-Seegmiller syndrome, HPRT1: Lesch-Nyhan syndrome, PRPS1: Phosphoribosyl pyrophosphate synthetase 1 superactivity, DPYD: Dihydropyrimidine dehydrogenase deficiency, DPYS: Dihydropyrimidinase deficiency, UPB1: Beta-ureidopropionase deficiency HMDB IDs resemble these metabolites: HMDB0000026: N-Carbamyl-beta-alanine, HMDB0000034: Adenine, HMDB0000052: Argininosuccinate, HMDB0000071: Deoxyinosine, dIno, HMDB0000076: Dihydrouracil, HMDB0000079: Dihydrothymine, HMDB0000085: Purine nucleoside phosphorylase, HMDB0000101: Deoxyadenosine, HMDB0000143: Galactose, HMDB0000157: Hypoxanthine, HMDB0000226: Orotic acid, HMDB0000262: Thymine, HMDB0000289: Uric acid, HMDB0000292: Xanthine, HMDB0000300: Uracil, HMDB0000401: 2,8-Dihydroxyadenine, HMDB0000635: Succinylacetone, HMDB0000679: Homocitrulline, HMDB0000797: SAICA riboside, HMDB0000904: Citrulline, HMDB0000912: Succinyladenosine, HMDB0002299: beta-aminoisobutyrate, HMDB0062179: AICA riboside. Additional biomarkers relevant for this specific patient (not part of IEMbase data): HMDB0000489: N-Aspartylglucosamine, HMDB0000469: 5-(Hydroxymethyl)uracil, HMDB0000721: Gly-pro, HMDB0002166: (S)-Beta-aminoisobutyrate.

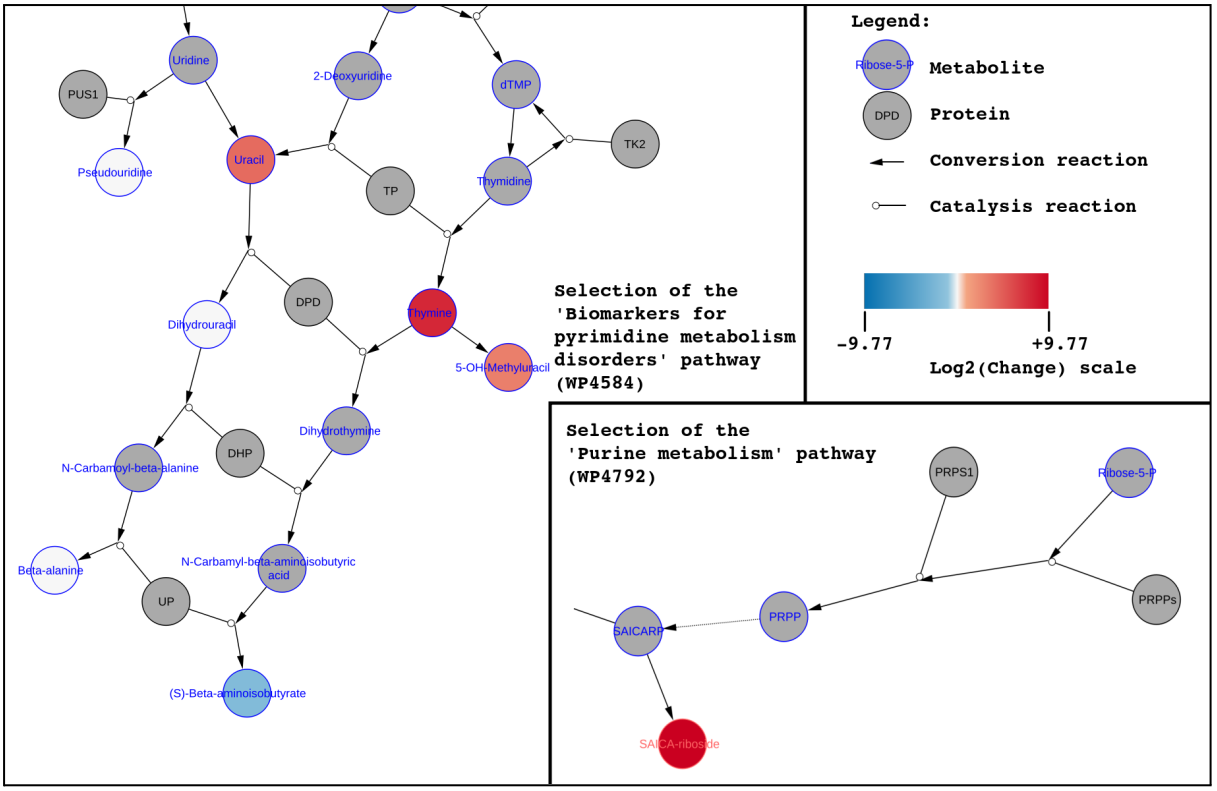

Figure 3: Network visualization of biochemical markers for patient I (diagnosed with DPYD). *Left:* selected section of pyrimidine biomarkers pathway; *Right:* selection of purine metabolism pathway.

Table 1: Overview of Databases and tools used within the workflow

Database/tool [reference]	Description	Workflow Step
ChEBI [19]	Chemical Entities of Biological Interest (ChEBI) is a freely available dictionary of molecular entities focused on 'small' chemical compounds.	1-3, 5-7
Wikidata [20]	Wikidata is a free and open knowledge base that can be read and edited by both humans and machines	1
UHPLC-MS/MS [18]	Ultra High-Performance Liquid Chromatography-Tandem Mass Spectrometer is a chemical technique combining the physical separation capabilities of liquid chromatography with the mass analysis capabilities of mass spectrometry.	1
PathVisio [21]	PathVisio is a free open-source pathway analysis and drawing software that allows drawing, editing, and analyzing biological pathways.	2
UniProt [22]	UniProt is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects.	2
Rhea [23]	Rhea is an expert-curated knowledge base of chemical and transport reactions of biological interest - and the standard for enzyme and transporter annotation in UniProtKB.	2

OMIM [24]	Online Mendelian Inheritance in Man (OMIM) is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available.	2
WikiPathways [25]	WikiPathways is a database of biological pathways maintained by and for the scientific community.	2, 4, 6, 7
WikiPathways SPARQL Endpoint [26]	The semantic web format of the WikiPathways database, using the Resource Description Framework (RDF) as data format which can be queried through the SPARQL-query language.	2, 4, 6, 7
HGNC (gene ID) [29]	The HGNC is a resource for approved human gene nomenclature.	4
HMDB [31]	The Human Metabolome Database (HMDB) is a freely available electronic database containing detailed information about small molecule metabolites found in the human body.	4, 5
Cytoscape (REST API) [33]	Cytoscape is an open-source software platform for visualizing complex networks and integrating these with any type of attribute data. cyREST is a language-agnostic, programmer-friendly RESTful API module for Cytoscape, allowing for programmatic access to Cytoscape features.	7
GitHub (github.com)	GitHub is a web-based hosting service for version control using git.	1-8

Table 2: Sample size sorted on disease classification and age range. The patient labels (A to T) are included per relevant age category. The HGNC name for the gene linked to the disorder is provided if deviating from the common abbreviation. The # indicates if a patient has received treatment.

Age range (years) → Diseases ↓	**0-1**	**1-5**	**5-16**	**16+**
Pyrimidine - Full name (Abbreviation, OMIM-ID, HGNC name)				
Beta-ureidopropionase deficiency (UPB1D, 613161, UPB1)	1 (J)			
Dihydropyrimidine dehydrogenase deficiency (DPD, 274270, DPYD)	1 (I)	1 (A)	1 (E)	1 (K)
Urea cycle - Full name (Abbreviation, OMIM-ID, HGNC name)				
Argininosuccinic aciduria (ASLD, 207900, ASL)		2 (M, N)	2 (L, T)	1 (F)
Citrullinemia Type I (ASS1, 215700)	1 (D)			
Hyperornithinemia-hyperammonemia-homocitrullinuria syndrome (HHHS,			1 (H#)	

238970, SLC25A15)				
Ornithine Transcarbamylase deficiency (OTCD, 311250, OTC)		1 (G)	2 (R, S)	1 (O)

Table 3: Overview of biomarker data per patient, including number of altered biomarkers and coverage thereof in pathways models.

HGNC Name Disorder	Patient Label	Biomarkers with reference data	Altered biomarkers	Amount of Biomarkers not in any pathway (ChEBI IDs; corresponding names)	Total amount of relevant pathways (First pathway biomarker coverage)
UPB1D	J	65	15	1 (27596; 3-Methyl-histidine)	37 (5)
DPYD	A	61	16	-	47 (6)
	E	54	12	2 (35621, 86498; AABA, Hydroxylysine)	41 (5)
	I	60	7	2 (17261, 70744; N-Aspartylglucosamine, Gly-pro)	9 (4)
	K	45	5	-	10 (3)
ASL	F	52	12	2 (86498, 27596; Hydroxylysine, 3-Methyl-histidine)	23 (4)
	L	50	15	2 (27596, 17261; 3-Methyl-histidine, N-Aspartylglucosamine)	31 (4)
	M	56	30	2 (17261, 27596; N-Aspartylglucosamine, 3-Methyl-histidine)	75 (14)
	N	61	36	2 (27596, 17261; 3-Methyl-histidine, N-Aspartylglucosamine)	76 (15)
	T	58	32	3 (27596, 17261, 86498; 3-Methyl-histidine, N-Aspartylglucosamine, Hydroxylysine)	77 (12)
ASS1	D	59	16	1 (49015; Piperideine carboxylic acid)	46 (7)
SLC25A15	H	61	48	6 (27596, 35621, 86498, 17261, 70744, 50599; 3-Methyl-histidine, AABA, Hydroxylysine, N-Aspartylglucosamine, Gly-pro, 1-Methyl-histidine)	77 (16)
OTC	G	67	19	2 (35621, 27596; AABA, 3-Methyl-histidine)	26 (8)
	O	55	42	5 (50599, 86498, 27596, 35621, 17261; 1-Methyl-histidine, Hydroxylysine, 3-Methyl-histidine, AABA, N-Aspartylglucosamine)	77 (15)
	R	62	43	4 (27596, 17261, 86498, 35621; 3-Methyl-histidine,	78 (16)

				N-Aspartylglucosamine, Hydroxylysine, AABA)	
	S	62	36	4 (27596, 86498, 17261, 35621; 3-Methyl-histidine, Hydroxylysine, N-Aspartylglucosamine, AABA)	74 (15)

Table 4: Amount of unique DataNodes (proteins, metabolites) and disorders captured by each pathway model. Uniqueness counts based on unification to HGNC names (Proteins), ChEBI IDs (Metabolites), and OMIM URLs (Disorders).

Model Name	**PWM ID**	**#Proteins** (HGNC)	**#Metabolites** (ChEBI)	**#Interactions** (Rhea)	**#Disorders** (OMIM)
Purine metabolism and related disorders	WP4224	22	61	51	17
Pyrimidine metabolism and related diseases	WP4225	17	41	33	15
Urea cycle and related diseases	WP4571	9	16	7	9
Urea cycle and associated pathways	WP4595	24	35	23	12
Biomarkers for urea cycle disorders	WP4583	12	29	10	-
Biomarkers for pyrimidine metabolism disorders	WP4584	15	39	29	-

Table 5: A total of fifteen pathways contain a high amount of biomarkers for both assays, with at best 46 markers included in one individual pathway model. %-sign indicates which pathways were not found in the top three relevant pathways for any patient.

Pathway ID	**Title of Pathway**	**ChEBI IDs In PW**
WP3604	Biochemical pathways: part I%	46
WP4042	Nucleobase catabolism%	21
WP4225	Pyrimidine metabolism and related diseases	18
WP1936	Transport of inorganic cations/anions and amino acids/oligopeptides%	17
WP4584	Biomarkers for pyrimidine metabolism disorders	17
WP1935	Transport of bile salts and organic acids, metal ions and amine compounds%	17
WP3925	Amino acid metabolism	15
WP4224	Purine metabolism and related disorders	14
WP4082	Nucleotide salvage%	14

WP1938	tRNA aminoacylation[%]	13
WP1937	Transport of vitamins, nucleosides, and related molecules[%]	12
WP4792	Purine metabolism	12
WP4595	Urea cycle and associated pathways	11
WP4022	Pyrimidine metabolism	10
WP4583	Biomarkers for urea cycle disorders	10

Table 6: Top three pathways containing most unique biomarkers, as well as which biomarkers were not covered by the top three pathways (however were part of a PWM). [#] indicates different pathways cover the same amount of unique metabolites leading to a manual selection of the most relevant pathway. - indicates no third pathway was selected.

HGNC Name Disorder	**Patient Label; diagnosis possible?** Yes / No / Direction	**Pathways in Top 3 (WikiPathways ID)** [amount of covered biomarkers]	**Not covered biomarkers (ChEBI IDs; corresponding names)**
UPB1D	J	Pyrimidine metabolism and related diseases (WP4225) [5] Amino acid metabolism (WP3925)[#] [3] Purine metabolism (WP4792) [3]	4 (16283, 18019, 27596, 30913; Cystine, Lysine, 3-Methyl-histidine, Pipecolic acid)
DPYD	A	Amino acid metabolism (WP3925) [6] Biomarkers for pyrimidine metabolism disorders (WP4584) [5] Urea cycle and related diseases (WP4571)[#] [1]	4 (17553, 18237, 7274, 15727; Phosphoethanolamine, Glutamic acid, Formiminoglutamic acid, Carnosine)
	E	Amino acid metabolism (WP3925) [5] Pyrimidine metabolism and related diseases (WP4225) [3] -	4 (18237, 35621, 86498, 16000; Glutamic acid, AABA, Hydroxylysine, Monoethanolamine)
	I	Biomarkers for pyrimidine metabolism disorders (WP4584) [4] Purine metabolism (WP4792) [1] -	2 (17261, 70744; N-Aspartylglucosamine, Gly-pro)
	K	Biomarkers for pyrimidine metabolism disorders (WP4584) [3] Urea cycle and related diseases (WP4571) [1] -	1 (18237; Glutamic acid)
ASL	F	Biomarkers for urea cycle disorders (WP4583) [4] One-carbon metabolism and related pathways (WP3940) [3] Pyrimidine metabolism (WP4002) [1]	4 (86498, 15727, 27596, 85981; Hydroxylysine, Carnosine, 3-Methyl-histidine, Homocarnosine)
	L	Urea cycle and related diseases (WP4571) [4] Glycerophospholipid biosynthetic pathway (WP2533) [2] Amino acid metabolism (WP3925) [2]	7 (18237, 27596, 85981, 16283, 17549, 17261, 17368; Glutamic acid, 3-Methyl-histidine, Homocarnosine, Cystine, Aminolevulinic acid, N-Aspartylglucosamine, Hypoxanthine)
	M	Amino acid metabolism (WP3925) [14] Urea cycle and associated pathways (WP4595) [7] -	9 (85981, 15727, 17261, 16283, 27596, 18237, 16000, 18019, 16865; Homocarnosine, Carnosine, N-Aspartylglucosamine, Cystine, 3-Methyl-histidine, Glutamic acid, Monoethanolamine, Lysine, GABA)
	N	Amino acid metabolism (WP3925) [15] Urea cycle and associated pathways (WP4595) [7] -	14 (17553, 18237, 85981, 27596, 15727, 16865, 17261, 16000, 37024, 58148, 18019, 15611, 27891, 16283; Phosphoethanolamine, Glutamic acid, Homocarnosine, 3-Methyl-histidine, Carnosine, GABA, N-Aspartylglucosamine, Monoethanolamine, 2-Aminoadipic acid, Homocitrulline, Lysine, Sarcosine, S-sulfocysteine, Cystine)

	T	Amino acid metabolism (WP3925) [12] Urea cycle and associated pathways (WP4595) [6] Vitamin B6-dependent and responsive disorders (WP4228) [3]	11 (16000, 27596, 85981, 17553, 17261, 18237, 16283, 17549, 16865, 16811, 86498; Monoethanolamine, 3-Methyl-histidine, Homocarnosine, Phosphoethanolamine, N-Aspartylglucosamine, Glutamic acid, Cystine, Aminolevulinic acid, GABA, Methionine, Hydroxylysine)
ASS1	D	Amino acid metabolism (WP3925) [7] Pyrimidine metabolism and related diseases (WP4225) [5] Vitamin B6-dependent and responsive disorders (WP4228) [2]	3 (49015, 7274, 32682; Piperideine carboxylic acid, Formiminoglutamic acid, Arginine)
SLC25A15	H	Amino acid metabolism (WP3925) [16] Pyrimidine metabolism and related diseases (WP4225) [8] Biomarkers for urea cycle disorders (WP4583)# [3]	21 (18019, 27596, 15611, 16000, 18323, 18237, 35621, 37024, 16283, 86498, 16865, 15727, 85981, 16927, 17261, 17549, 27891, 70744, 50599, 7274, 17553; Lysine, 3-Methyl-histidine, Sarcosine, Monoethanolamine, Anserine, Glutamic acid, AABA, 2-Aminoadipic acid, Cystine, Hydroxylysine, GABA, Carnosine, Homocarnosine, Saccharopine, N-Aspartylglucosamine, Aminolevulinic acid, S-sulfocysteine, Gly-pro, 1-Methyl-histidine, Formiminoglutamic acid, Phosphoethanolamine)
OTC	G	Pyrimidine metabolism and related diseases (WP4225) [8] Alanine and aspartate metabolism (WP106) [2] Purine metabolism (WP4792) [2]	7 (35621, 58148, 27596, 37024, 7274, 16000, 16414; AABA, Homocitrulline, 3-Methyl-histidine, 2-Aminoadipic acid, Formiminoglutamic acid, Monoethanolamine, Valine)
	O	Amino acid metabolism (WP3925) [15] Biomarkers for urea cycle disorders (WP4583) [7] One-carbon metabolism and related pathways (WP3940) [4]	16 (18323, 27891, 50599, 86498, 27596, 35621, 18019, 37024, 15727, 18237, 16283, 17261, 7274, 16865, 85981, 17568; Anserine, S-sulfocysteine, 1-Methyl-histidine, Hydroxylysine, 3-Methyl-histidine, AABA, Lysine, 2-Aminoadipic acid, Carnosine, Glutamic acid, Cystine, N-Aspartylglucosamine, Formiminoglutamic acid, GABA, Homocarnosine, Uracil)
	R	Amino acid metabolism (WP3925) [16] Pyrimidine metabolism and related diseases (WP4225) [7] Biomarkers for urea cycle disorders (WP4583) [3]	17 (17553, 16000, 18019, 15727, 16865, 27596, 18237, 16283, 16811, 17261, 86498, 85981, 27891, 17549, 35621, 15611, 37024; Phosphoethanolamine, Monoethanolamine, Lysine, Carnosine, GABA, 3-Methyl-histidine, Glutamic acid, Cystine, Methionine, N-Aspartylglucosamine, Hydroxylysine, Homocarnosine, S-sulfocysteine, Aminolevulinic acid, AABA, Sarcosine, 2-Aminoadipic acid)
	S	Amino acid metabolism (WP3925) [15] Pyrimidine metabolism and related diseases (WP4225) [6] Glycerolipids and glycerophospholipids (WP4722)# [2]	13 (85981, 15727, 16811, 27596, 16865, 18019, 16283, 57472, 86498, 17261, 18237, 35621, 17549; Homocarnosine, Carnosine, Methionine, 3-Methyl-histidine, GABA, Lysine, Cystine, Argininosuccinate, Hydroxylysine, N-Aspartylglucosamine, Glutamic acid, AABA, Aminolevulinic acid)