A Systems Biology Workflow to Support the Diagnosis of Pyrimidine and Urea Cycle Disorders

Denise N. Slenter[a*], Irene M.G.M. Hemel[a,b], Chris T. Evelo[a,b], Jörgen Bierau[c,d], Egon L. Willighagen[a], Laura K.M. Steinbusch[c]

[a] Department of Bioinformatics (BiGCaT), NUTRIM, Maastricht University, Maastricht, The Netherlands
[b] Maastricht Centre for Systems Biology (MaCSBio), Maastricht University, Maastricht, The Netherlands
[c] Department of Clinical Genetics, Maastricht University Medical Center, Maastricht, The Netherlands
[d] Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, The Netherlands

[*] Corresponding author at: Department of Bioinformatics - BiGCaT, NUTRIM, Universiteitssingel 60, 6229 ER Maastricht, The Netherlands. E-mail address: denise.slenter@maastrichtuniversity.nl

Highlights

- Machine readable pathway models suitable for purine, pyrimidine, and urea cycle omics data analysis
- Visualizing clinically validated biomarkers with pathway knowledge and theoretical biomarker data
- Utilizing semantic web technology and network approaches to precisely represent the biochemistry underlying pyrimidine and urea cycle disorders
- Opens up the possibility to visualize the interactions between metabolic pathways, which can be used in the diagnostic workflow
- Approach is extendable to other inherited metabolic disorders

Abstract

Background: Pyrimidine (and purine) metabolism provides essential high-energy vehicles, universal throughout all life, which carry both activated metabolites that serve as fuel and building blocks for countless metabolic processes as well as the messenger molecules that steer these processes. Furthermore, these nitrogen-containing structures encode the alphabet needed to construct the genetic code. The pyrimidine pathway overlaps with the Urea Cycle through a common metabolite (carbamoyl phosphate), with the urea cycle being responsible for the production of several amino acids and removing ammonia. Various Inherited Metabolic Disorders (IMDs) disrupt these pathways, with considerably large phenotypic variation in affected patients, hindering diagnosis. Moreover, the biomarkers for these disorders overlap substantially between the IMDs and their underlying biological pathways, interfering with identifying the deficient protein. We developed a framework that allows us to combine clinical and theoretical biomarkers with pathway models through network approaches and semantic web technologies to support the diagnosis of pyrimidine and urea cycle IMDs.

Methods: We integrated literature and expert knowledge into machine-readable pathway models for pyrimidine and urea cycle disorders, including disease information and relevant downstream biomarkers. The theoretical change for each biomarker per disease was compiled based on a manual database search. The top three pathways of interest were retrieved through semantic web technologies, by selecting pathways that covered most unique markers as well as showing overlap with the theoretical marker data. These pathways and the corresponding clinical data were visualized through network analysis. Two expert laboratory scientists evaluated our approach by diagnosing a cohort of sixteen previously diagnosed patients with various pyrimidine and urea cycle disorders, based on a visualization of theoretical biomarker overlap with patient data, as well as top three pathways displaying the relationships between the individual biomarkers.

Results: The number of relevant biomarkers for each patient varies greatly (five to 48), and likewise the pathways covering most unique biomarkers differ for equivalent disorders. The two experts reached similar conclusions regarding the diagnosis of 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, which could be prioritized for further investigation. Three cases were found to be undiagnosable with the data available.

Conclusion: The presented workflow supports the diagnosis of several IMDs of pyrimidine and urea cycle pathways, by directly linking biological pathway knowledge and theoretical biomarker data to clinical cases. This workflow is adaptable to analyze different types of IMDs, difficult patient cases and functional assays in the future. Furthermore, the pathway models can be used as a basis to perform various other types of (omics) data analysis, e.g. transcriptomics, metabolomics, fluxomics.

Data Availability:

The data used in this study have been deposited at https://github.com/BiGCAT-UM/IMD-PUPY .

Keywords

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

1. Introduction

Many enzymes are critically involved in the synthesis, degradation, and transport of molecules implicated in metabolic processes [1]. Consequently, genetic disorders leading to deficiencies and malfunctioning of these enzymes can result in a lack of or (potentially) toxic levels of metabolites. In addition, inherited enzymatic disturbances can affect other sections of the pathway(s) in which the protein is active [2]. Collectively, these diseases are called Inherited Metabolic Disorders (IMDs) or Inborn Errors of Metabolism [3].

The limited therapies which are available [4] can only be initiated after a timely and accurate diagnosis. Currently, the diagnoses of IMDs are based on both symptoms and biomarkers, which are measured in various bodily fluids. Methods to detect genetic variants such as Whole Exome Sequencing are used to diagnose IMDs as well, although so far have been found less sensitive and less 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. This issue is especially relevant for pyrimidine disorders, which often present with nonspecific clinical symptoms and a lack of a clear genotype-phenotype correlation [9–11]. The symptoms related to urea cycle disorders are often more specific (e.g. hyperammonemia, lethargy, vomiting, coma) [12]. The pyrimidine pathway synthesizes thymine, cytosine and uracil, DNA and RNA building blocks, and other molecules pertinent to signal transduction and energy transport. The urea cycle removes ammonia from the body and is a direct upstream pathway for pyrimidine metabolism, with several metabolites converted to amino acids, mainly taking place in the liver. This overlap leads to elevated metabolic levels in both pathways for certain IMDs [13]. Even though recent advances in clinical biomarker measurements [14] in urine samples have aided in diagnosis for some IMDs, most markers are currently not used for newborn (dried blood spot) screening [15], due to limited detectability of the biomarkers in these spots and with most pyrimidine IMDs missing treatment options.

This paper provides a framework using Systems Biology approaches to integrate data related to pyrimidine and urea cycle IMDs, including metabolic reference and patient values, theoretical biomarkers, and biological pathway information. 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. The presented workflow also highlights chances for the IMD field as a whole, by showcasing that improving data and

identifier 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 IMDs and is adaptable to analyze different types of IMDs and functional assays in the future.

2. Material and Methods

2.1 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. Several data harmonization steps were needed, before arriving at the final data interpretation step. Also, knowledge from various databases had to be integrated within the workflow, which are summarized in Table 1. All data processing steps which could be automated were captured in an RMarkdown script [16] in the R programming language [17] (version 4.1.0), tested through RStudio [18] (version 1.4.1717), available at https://github.com/BiGCAT-UM/IMD-PUPY .

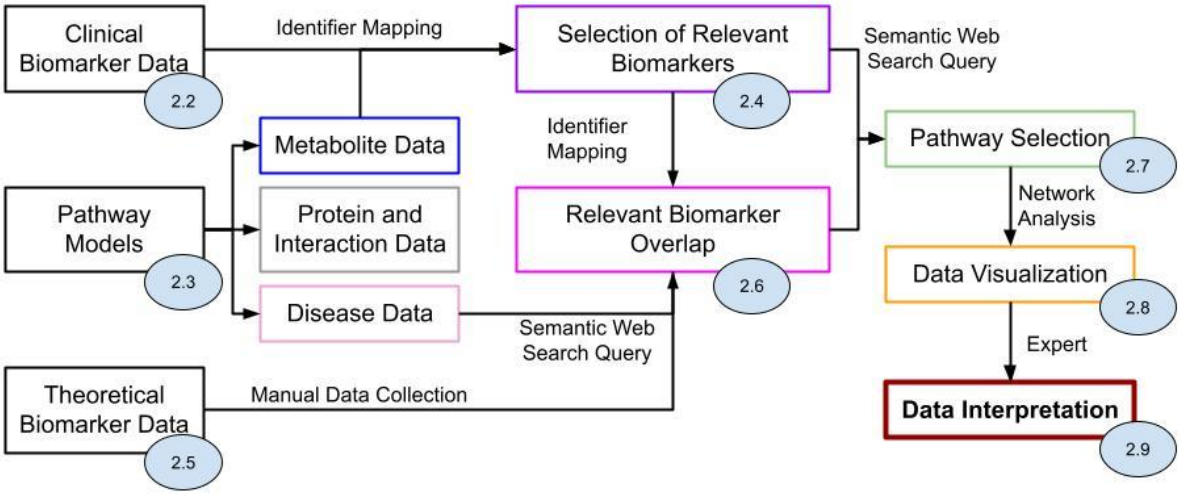


Figure 1: Depiction of workflow to interpret clinical data of IMDs through network analysis; each circle refers to a specific step within the Materials and Methods Section.

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.	2, 3, 4, 6
Wikidata [20]	Wikidata is a free and open knowledge base that can be read and edited by both humans and machines	2
UHPLC-MS/MS [21]	Ultra High Performance Liquid Chromatography - tandem Mass Spectrometer is a chemistry technique combining the physical separation capabilities of liquid	2

	chromatography with the mass analysis capabilities of mass spectrometry.	
PathVisio [22]	PathVisio is a free open-source pathway analysis and drawing software that allows drawing, editing, and analyzing biological pathways.	3
UniProt [23]	UniProt is a freely accessible database of protein sequence and functional information, many entries being derived from genome sequencing projects.	3
Rhea [24]	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.	3
OMIM [25]	Online Mendelian Inheritance in Man (OMIM) is a comprehensive, authoritative compendium of human genes and genetic phenotypes that is freely available.	3
WikiPathways [26]	WikiPathways is a database of biological pathways maintained by and for the scientific community.	
WikiPathways SPARQL Endpoint [27]	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.	3, 5, 7, 8
HGNC (gene ID) [28]	The HGNC is a resource for approved human gene nomenclature.	5
HMDB [29]	The Human Metabolome Database (HMDB) is a freely available electronic database containing detailed information about small molecule metabolites found in the human body.	5, 6
Cytoscape (REST API) [30,31]	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.	8
BridgeDb (Cytoscape App and webservice) [30,32]	BridgeDb is a framework to map identifiers between various biological databases, built into other tools such as the Cytoscape app, and available through a webservice.	8
KEGG [33]	KEGG is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism, and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome sequencing and other high-throughput experimental technologies.	8
GitHub (github.com)	GitHub is a web-based hosting service for version control using git.	1-9

2.2 Clinical Biomarker Data

All patient data were collected through two targeted chemical assays in urine - one for purine and pyrimidine (PUPY) and one for amino acids (AA) - as described previously [14,21], 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 limit of distribution in a control population. The metabolites in the reference data were annotated with corresponding ChEBI [19] identifiers (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 [21]. Four patients were disregarded from this study, due to missing data for the AA panel (patients labeled B, C, P and Q). Because most IMD disorders are quite rare, the collected data did not include a large sample size regarding ethnicity, age range or sex, and can therefore not be regarded as representative for the worldwide patient population. 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 left-over material from routine diagnostics for laboratory development and validation purposes.

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.

Age range (years) → Diseases ↓	**0-1**	**1-5**	**5-16**	**16+**	**0-16**
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)					

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

2.3 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) [22], as well as pathway models (PWMs) on biomarkers, visualizing several markers missing from the main pathway models. All proteins were annotated with UniProt IDs [23], which linked to potentially relevant Rhea IDs [24] 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 [25]. Data on the created PWMs was deposited in WikiPathways [26] and retrieved from the WikiPathways SPARQL endpoint [34] (data from September 2021 [35]). This endpoint contains PWM data in a harmonized and semantic format, the Resource Description Framework (RDF) [27].

2.4 Selection of Relevant Biomarkers

For each patient, measured biomarkers from the clinical assay data (step 2.2) 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 change-values were log(2) transformed to show proportional changes.

2.5 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 [28] first. Second, the corresponding potentially relevant biomarkers for these disorders were retrieved manually from IEMbase [36] V 2.0.0 (accessed on 2021-08-05) through their HGNC gene name, including sample matrix, HMDB

IDs [29], 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. 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.

2.6 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 (step 2.4) 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 added to the theoretical biomarker visualization (step 2.5) by matching on HMDB IDs, and visualized in a heatmap (with the same settings as step 2.5).

2.7 Pathway Selection

Relevant pathways for each patient were found through another semantic web query against the WikiPathways SPARQL endpoint (step 2.3), matching the remaining changed biomarkers (2.4), which have a log(2) change value not equal to zero. Relevant biomarkers which were not 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.

2.8 Data visualization

The data for each patient was visualized with the network analysis tool Cytoscape [37] (version 3.8.2), by using the Cytoscape REST API [31] (version v1) 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 “XrefID” column in Cytoscape. When this resulted in no matches, the BridgeDb App in Cytoscape [30] was used to convert the XrefID column (from either HMDB [29] or KEGG compound [33] IDs) to corresponding ChEBI IDs, relying on the BridgeDb web service for the metabolite ID matching [38].

2.9 Data Interpretation

For each patient, a network data visualization (step 2.8) and relevant biomarker overlap heatmap (step 2.8) were provided to two Laboratory Specialists Biochemical Genetics, after

which narrative feedback on a potential diagnosis was collected. For each patient, a Cytoscape session file (step 2.8) is available on GitHub (Cytoscape_networks folder), as well as an interpretation overview containing patient characteristics containing age category, relevant biomarkers, personalized heatmap (step 2.6), selected pathways, and clinical interpretation (Expert_interpretation folder).

3. Results

3.1 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 will be discussed in more detail below.

3.2 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 IDs 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 (step 2.4), the pathway's coverage of biomarkers relevant for each patient (pathway), and the highest total of markers covered by one pathway.

Table 3: Biomarker data per patient.

HGNC Name Disorder	**Patient Label**	**Biomarkers with reference data**	**Altered biomarkers**	**Amount of Biomarkers not in any pathway** (ChEBI ID)	**Total amount of relevant pathways** (First pathway biomarker coverage)
UPB1D	J	65	15	1 (27596)	37 (5)
DPYD	A	61	16	-	47 (6)
	E	54	12	2 (35621, 86498)	41 (5)
	I	60	7	2 (70744, 17261)	9 (4)
	K	45	5	-	10 (3)
	F	52	12	2 (27596, 86498)	23 (4)

ASL	L	50	15	2 (27596, 17261)	31 (4)
	M	56	30	2 (17261, 27596)	75 (14)
	T	58	32	3 (27596, 86498, 17261)	77 (12)
	N	61	36	2 (27596, 17261)	76 (15)
ASS1	D	59	16	1 (49015)	46 (7)
SLC25A15	H	61	48	6 (27596, 35621, 86498, 17261, 70744, 50599)	77 (16)
OTC	G	67	19	2 (35621, 27596)	26 (8)
	O	55	42	5 (50599, 86498, 27596, 35621, 17261)	77 (15)
	R	62	43	4 (27596, 17261, 86498, 35621)	78 (16)
	S	62	36	4 (27596, 86498, 17261, 35621)	74 (15)

3.3 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 [26] 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.

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 URL (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	-

3.4 Selection of Relevant Biomarkers

The annotated data from step 2.2 (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.

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

3.5 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, PRPPS, TYMP, TK2, RRM2B, ITPA, IMPDH1, DHODH, 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.

3.6 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.

3.7 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 on 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.

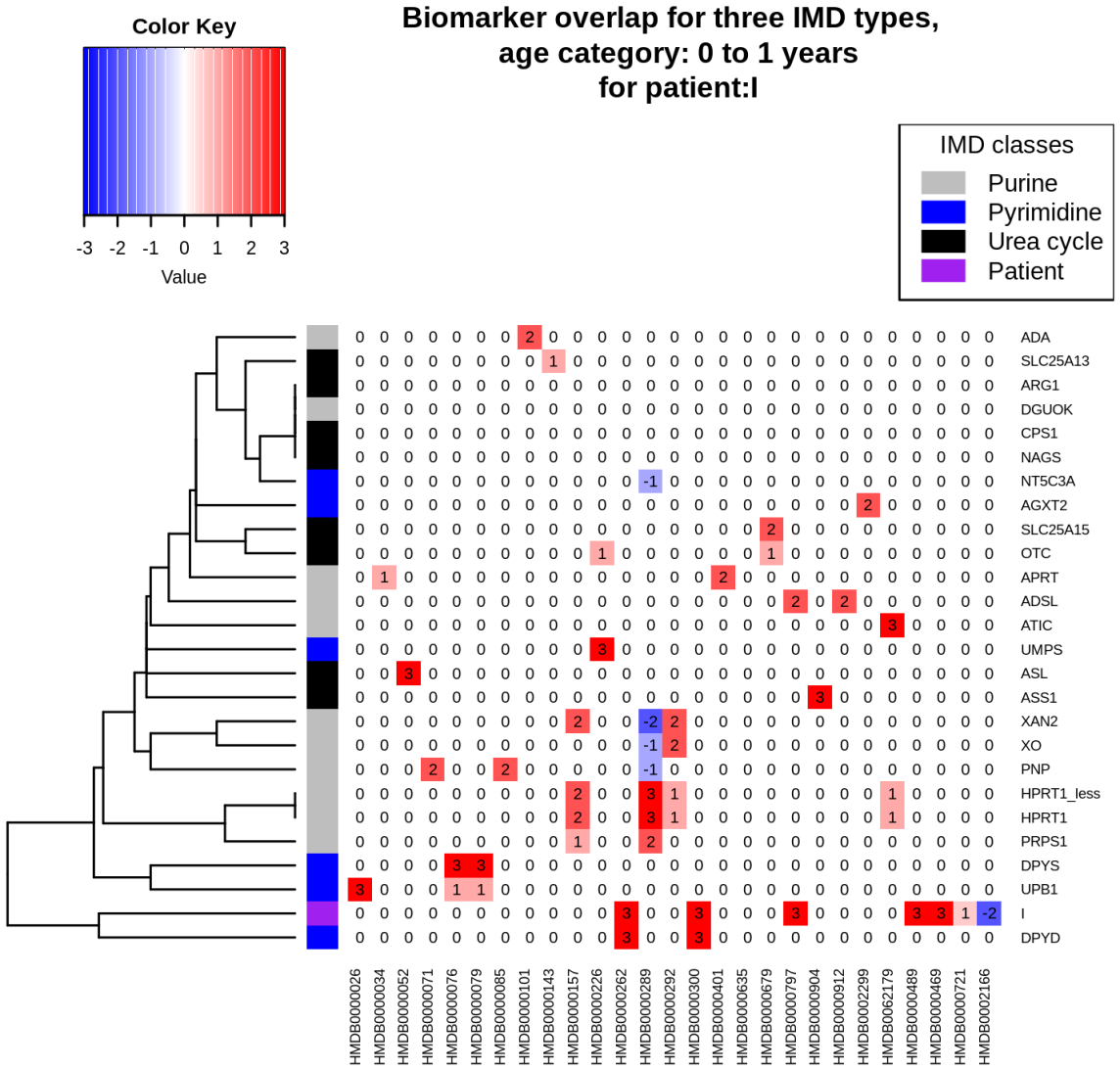


Figure 2: Visualization of how theoretical biochemical markers in urine overlap for individual IMDs in the purine, pyrimidine, and urea cycle pathways. Protein names correspond to the following disorders: PRPS1: Phosphoribosyl pyrophosphate synthetase 1 superactivity, XAN2: Xanthinuria, Type II, ADSL: Adenyl- succinate lyase deficiency, ATIC: AICAR transformylase/IMP cyclohydrolase deficiency, APRT: Adenine phosphoribosyltransferase deficiency, HPRT1: Lesch-Nyhan syndrome, ADA: Adenosine deaminase deficiency, HPRT1_less: Kelley-Seegmiller syndrome, PNP: Purine nucleoside phosphorylase deficiency, XO: Xanthinuria, Type I, DGUOK: Deoxyguanosine kinase deficiency, NT5C3A: Pyrimidine 5-nucleotidase superactivity, DPYD: Dihydropyrimidine dehydrogenase deficiency, UMPS: Orotic aciduria type I, TYMP: Thymidine phosphorylase deficiency, DPYS: Dihydropyrimidinase deficiency, NA: Beta-alanine alpha-ketoglutarate transaminase deficiency, UPB1: Beta-ureidopropionase deficiency, AGXT2: Beta-aminoisobutyrate-pyruvate transaminase deficiency, NAGS: N-Acetylglutamate synthase deficiency, OTC: Ornithine transcarbamylase deficiency, ASS1: Citrullinemia type I, ARG1: Arginase deficiency, SLC25A15: Ornithine transporter deficiency, CPS1: Carbamoyl phosphate synthetase I deficiency, SLC25A13: Citrin deficiency, ASL: Argininosuccinic aciduria. HMDB IDs resemble these metabolites: HMDB0000157: Hypoxanthine, HMDB0000289: Uric acid, HMDB0000292: Xanthine, HMDB0000797: SAICA riboside, HMDB0000912: Succinyladenosine, HMDB0062179: AICA riboside, HMDB0000401: 2,8-Dihydroxyadenine, HMDB0000034: Adenine, HMDB0000101: Deoxyadenosine, HMDB0000085: Purine nucleoside phosphorylase, HMDB0000071: Deoxyinosine, dIno, HMDB0000635: Succinylacetone, HMDB0000262: Thymine, HMDB0000300: Uracil, HMDB0000226: Orotic acid, HMDB0000012: Deoxyuridine, HMDB0000273: Thymidine, HMDB0000079: Dihydrothymine, HMDB0000076: Dihydrouracil, HMDB0000056: Beta-alanine, HMDB0000026: N-Carbamyl-beta-alanine, HMDB0002299: beta-aminoisobutyrate, HMDB0000679: Homocitrulline, HMDB0000052: Argininosuccinate, HMDB0000904: Citrulline, HMDB0000052: Argininosuccinate, HMDB0000143: Galactose. Additional biomarkers relevant for this specific patient (not part of IEMbase

data): HMDB0000489: Aspartylglycosamine; HMDB0000469: 5-Hydroxymethyluracil; HMDB0000721: Glycylproline; HMDB0002166: (S)-beta-Aminoisobutyric acid.

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**	**Pathways in Top 3 (WikiPathways ID)** [amount of covered biomarkers]	**Not covered biomarkers (ChEBI IDs)**	**Diagnosable?** Yes / No / Direction
UPB1D	J	Pyrimidine metabolism and related diseases (WP4225) [5] Amino acid metabolism (WP3925)# [3] Purine metabolism (WP4792) [3]	16283, 18019, 27596, 30913, 17368, 17775, 17712	
DPYD	A	Amino acid metabolism (WP3925) [6] Biomarkers for pyrimidine metabolism disorders (WP4584) [5] Urea cycle and related diseases (WP4571)# [1]	17553, 18237, 7274, 57743, 15727	
	E	Amino acid metabolism (WP3925) [5] Pyrimidine metabolism and related diseases (WP4225) [3] -	18237, 35621, 86498, 16000	
	I	Biomarkers for pyrimidine metabolism disorders (WP4584) [4] Purine metabolism (WP4792) [1] -	17261, 70744	
	K	Biomarkers for pyrimidine metabolism disorders (WP4584) [3] Urea cycle and related diseases (WP4571) [1] -	18237	
ASL	F	Biomarkers for urea cycle disorders (WP4583) [4] One-carbon metabolism and related pathways (WP3940) [3] Pyrimidine metabolism (WP4002) [1]	86498, 15727, 27596, 85981, 17568	
	L	Urea cycle and related diseases (WP4571) [4] Glycerophospholipid biosynthetic pathway (WP2533) [2] Amino acid metabolism (WP3925) [2]	18237, 27596, 85981, 16283, 17549, 17261, 17053, 17368, 16414	
	M	Amino acid metabolism (WP3925) [14] Urea cycle and associated pathways (WP4595) [7] -	85981, 15727, 17261, 16283, 27596, 18237, 16000, 18019, 16865	
	N	Amino acid metabolism (WP3925) [15] Urea cycle and associated pathways (WP4595) [7] -	17553, 18237, 85981, 27596, 15727, 16865, 17261, 16000, 37024, 58148, 18019, 15611, 27891, 16283	
	T	Amino acid metabolism (WP3925) [12] Urea cycle and associated pathways (WP4595) [6] Vitamin B6-dependent and responsive disorders (WP4228) [3]	30913, 16000, 27596, 85981, 17553, 17261, 18237, 16283, 17549, 16927, 16865, 16811, 37024, 86498	
ASS1	D	Amino acid metabolism (WP3925) [7] Pyrimidine metabolism and related diseases (WP4225) [5] Vitamin B6-dependent and responsive disorders (WP4228) [2]	49015, 7274, 32682, 30913, 18019	
SLC25A 15	H	Amino acid metabolism (WP3925) [16] Pyrimidine metabolism and related diseases (WP4225) [8] Biomarkers for urea cycle disorders (WP4583)# [3]	32682, 58148, 18019, 57472, 27596, 15611, 16000, 18323, 18237, 35621, 37024, 16283, 86498, 16865, 15727, 85981, 16927, 17261, 17549, 27891, 70744,	

			50599, 7274, 17553	
OTC	G	Pyrimidine metabolism and related diseases (WP4225) [8] Alanine and aspartate metabolism (WP106) [2] Purine metabolism (WP4792) [2]	35621, 58148, 27596, 17596, 37024, 7274, 16000, 17712, 16414	
	O	Amino acid metabolism (WP3925) [15] Biomarkers for urea cycle disorders (WP4583) [7] One-carbon metabolism and related pathways (WP3940) [4]	17553, 18323, 27891, 50599, 86498, 27596, 35621, 18019, 37024, 16000, 15727, 18237, 16283, 15891, 17261, 15611, 7274, 16865, 85981, 17568	
	R	Amino acid metabolism (WP3925) [16] Pyrimidine metabolism and related diseases (WP4225) [7] Biomarkers for urea cycle disorders (WP4583) [3]	17553, 16000, 18019, 15727, 16865, 27596, 18237, 16283, 57472, 16811, 17261, 86498, 85981, 27891, 17549, 35621, 32682, 15611, 37024, 58148	
	S	Amino acid metabolism (WP3925) [15] Pyrimidine metabolism and related diseases (WP4225) [6] Glycerolipids and glycerophospholipids (WP4722)[#] [2]	17553, 16000, 85981, 15727, 16811, 27596, 16865, 18019, 16283, 57472, 86498, 17261, 18237, 35621, 17549	

3.8 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.

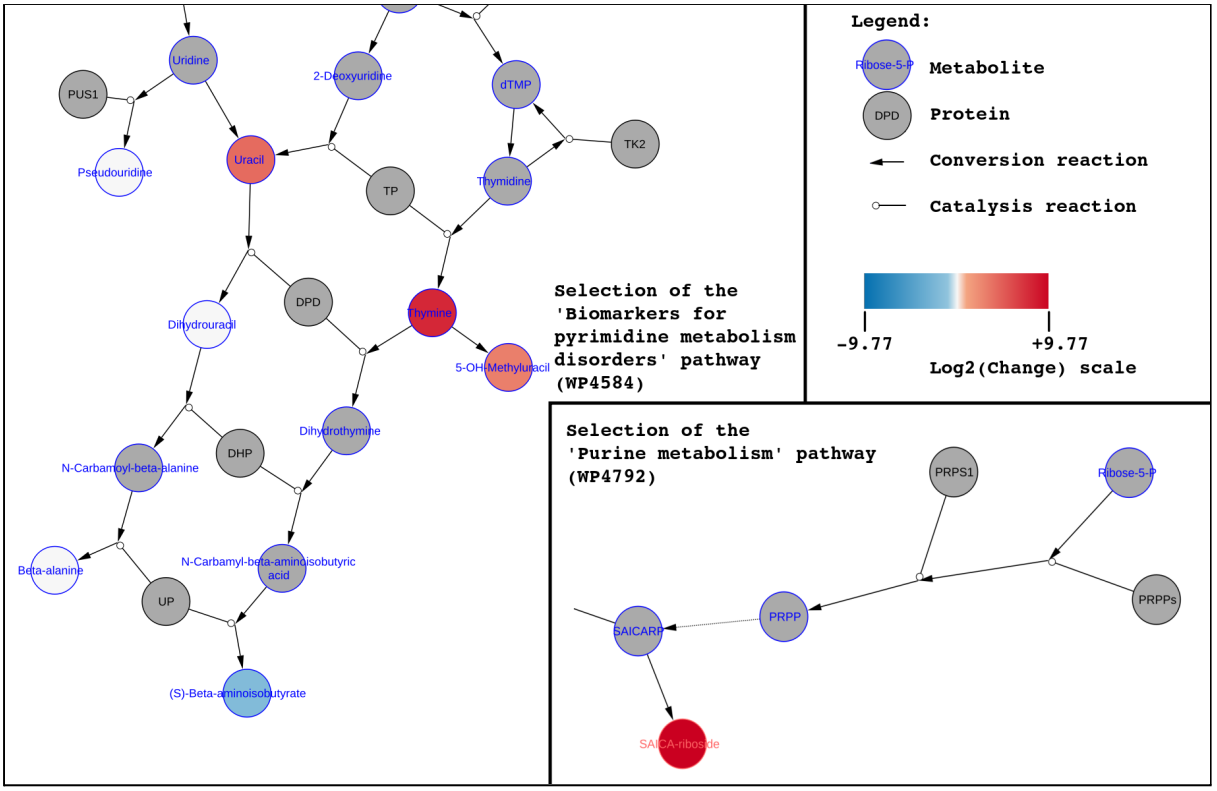


Figure 3: Network visualization of biochemical markers for patient I (diagnosed with DPYD). *Left:* selected section of pyrimidine biomarkers pathway; *Right:* selection of purine metabolism pathway.

3.9 Data Interpretation

The two laboratory specialists in biochemical genetics combined the heatmap from step 3.6 and the network biomarker data visualization from step 3.8 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. 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 unexpected higher value for arginine rather than ornithine.

4. Discussion

4.1 Workflow

Our proposed workflow was based on combining several data analysis techniques, e.g. semantic web and network analysis. 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 PUPY and urea cycle disorders with biomarkers measured through targeted assays. We believe this workflow can be used as an inspiration to construct a similar system for other IMDs, however, the following limitations should be taken into account for each step of the workflow.

4.2 Clinical Biomarker Data

First, all chemicals in the assays were annotated manually based on their (Dutch) name; using persistent identifiers to annotate data is a key aspect to enable open science [39], and ultimately lead to FAIR data [40]. This process led to two biomarkers differentiating in annotation. 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 [41] could be used to annotate the compound. This conversion to SMILES could be performed in several chemical drawing programs, out of which CDK Depict (https://github.com/cdk/depict) is a free web-based version implementing the Chemistry Development Kit [42], an open-source cheminformatics toolkit. 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. We would recommend annotating chemical biomarkers with a database reference along with the name of the compound by laboratories and vendors of instruments alike.

Second, since we combined data from two clinical assays (which is common practice in Clinical Genetics laboratories when a patient is suspected of having an IMD with nonspecific symptoms), we were also faced with different age categories for each assay. A solution would be to (re)structure data to allow for comparing all assays available in a lab [43].

Third, not each age category could be linked to reference data, therefore we had to substitute missing values with data from a broader age range. Several biomarkers have not been measured in enough non-patients to calculate a reliable healthy concentration range; reference data were missing for n-carbamoyl-aspartate, allantoin, cytosine, and cytidine. The first biomarker (known in HMDB [29] 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 tandem mass spectrometer data used in this study.

4.3 Pathways Models

Creating PWMs annotated with resolvable identifiers for the entities within the pathway [44] were crucial for data analysis and visualization [45]. The biomarker data models however only incorporate the markers as substrates or reaction products, without showing the metabolites more than one step removed from the original marker. Several biomarkers of PUPY and urea cycle IMDs are numerous steps removed from the original malfunctioning protein; covering all steps up to the original affected reaction will differ for each disorder and each marker. Also, this reaction information is scattered over publications in images and text [46] as well as various databases [47], which requires dedicated curation time to arrive at a PWM covering all relevant interactions. Several initiatives merge pathway information [48,49] based on gene and protein content, rather than metabolite and chemical conversion level. Lacking automatable access to pathway content or licenses prohibiting data extraction were 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 [50]. This despite the IUPAC-nomenclature rules [51] and available software to translate IUPAC names to chemical structures [52] and vice versa [53].

4.4 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. Due to the scattered nature of metabolic reaction mechanism knowledge in pathway databases, network approaches were difficult to implement in an automated fashion. By annotating the metabolic interactions in the PWMs with Rhea [24], information from various databases can be connected to the reaction level rather than the compound level in the future. This annotation also provides stoichiometry and charge-balanced reactions. However, the participating metabolite protonation states were based on a calculation of the charge close to physiological pH, leading to different IDs for the molecules within the PMWs as compared to the clinical assay data. The Rhea database provides an extension to the ChEBI ontology to overcome this issue, by adding the "has_major_microspecies_at_pH_7_3" connection. Another approach to overcome the mismatching of biomarkers could be by converting the clinical biomarker IDs to their corresponding neutral molecular structure InChIKey [54] ID or substructure matching [55].

4.5 Theoretical Biomarker Data

Connecting all disease IDs to their counterpart protein ID and data from the IEMbase [36] could only be performed manually. Data could be downloaded as CSV or PDF data, where only the latter includes information on the biomarkers and their positive or negative change status. The IDs of the biomarkers were a mixture of HMDB 3.0 and 4.0, which led to mapping issues. 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 to provide programmatic access to the aforementioned databases, for example through an API [ref] or SPARQL endpoint [ref]. We also want to encourage harmonizing the information in databases, for example by using existing ontologies such as the Human Disease Ontology [56], human phenotype ontology [57], or nosology on Inherited Metabolic Disorders [58].

4.6 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.

4.7 Pathway Selection

Selecting the top three pathways covering 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 of a disorder than 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.

4.8 Data Visualization

Connecting data to pathway and network models often require a form of identifier 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 (as described in step 4.3). Even though the WikiPathways RDF contains pathway data from Reactome pathways, we decided to not use these pathways in our workflow. The model conversion 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 [59] 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 was made available in Cytoscape for KEGG [60], 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 [49] and cy3sabiork [61] could not be automated.

4.9 Data Interpretation

Currently, diagnostic laboratories for inborn errors of metabolism report exact metabolite concentrations in diagnostic patient reports. Pathway and network models on the contrary, report log- or z-scores and pathway names with probabilities. We believe that laboratory specialists first have to learn the interpretation of these new diagnostic tools. In this study the network model helped to easily diagnose 9 out of 16 patient samples, and pointed into the correct direction or suggested follow up analysis for 4 samples. As mentioned in the results section, patient H (diagnosed with hyperornithinemia-hyperammonemia-homocitrullinuria (HHH) syndrome) was found to have more raised biomarkers than noted in IEMbase. IEMbase 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

increased presence of argininosuccinate. This example shows that laboratory experts are needed to interpret network and pathway models and combine their knowledge on patient treatment and symptoms.
For the validation of a recently implemented diagnostic tool called targeted urine metabolomics (TUM) used many of the samples discussed in this current study [43]. When comparing the data interpretation from TUM and the network and pathway workflow here, 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 especially 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. This disorder can be further investigated by measuring OTC activity in blood circulation [62], however this approach also did not seem feasible for asymptomatic heterozygotes. Other approaches entail genetic screening [63] (which can miss *de novo* cases) or a protein loading test, however, when this provocative test fails a detailed family history and additional molecular testing is advised [64].
For future studies, the interest should shift to measuring metabolite fluxes [2] over a longer timespan to better understand how protein intake triggers disorders [65] and if fasting could be used to treat patients with hypoglycemic symptoms [66].

5. 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. Information on treatment and clinical condition remains important for accurate diagnosis, as well as expert interpretation of all information combined into this workflow. 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.

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+).

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

[1] T. Fukao, K. Nakamura, Advances in inborn errors of metabolism, J. Hum. Genet. 64 (2019) 65. https://doi.org/10.1038/s10038-018-0535-7.

[2] B. Lanpher, N. Brunetti-Pierri, B. Lee, Inborn errors of metabolism: the flux from Mendelian to complex diseases, Nat. Rev. Genet. 7 (2006) 449–460. https://doi.org/10.1038/nrg1880.

[3] C.R. Ferreira, S. Rahman, M. Keller, J. Zschocke, ICIMD Advisory Group, An international classification of inherited metabolic disorders (ICIMD), J. Inherit. Metab. Dis. 44 (2021) 164–177. https://doi.org/10.1002/jimd.12348.

[4] B.K. Burton, Inborn errors of metabolism in infancy: a guide to diagnosis, Pediatrics. 102 (1998) E69. https://doi.org/10.1542/peds.102.6.e69.

[5] A.N. Adhikari, R.C. Gallagher, Y. Wang, R.J. Currier, G. Amatuni, L. Bassaganyas, F. Chen, K. Kundu, M. Kvale, S.D. Mooney, R.L. Nussbaum, S.S. Randi, J. Sanford, J.T. Shieh, R. Srinivasan, U. Sunderam, H. Tang, D. Vaka, Y. Zou, B.A. Koenig, P.-Y. Kwok, N. Risch, J.M. Puck, S.E. Brenner, The role of exome sequencing in newborn screening for inborn errors of metabolism, Nat. Med. 26 (2020) 1392–1397. https://doi.org/10.1038/s41591-020-0966-5.

[6] K.L.M. Coene, L.A.J. Kluijtmans, E. van der Heeft, U.F.H. Engelke, S. de Boer, B. Hoegen, H.J.T. Kwast, M. van de Vorst, M.C.D. Huigen, I.M.L. Keularts, M.F. Schreuder, C.D.M. van Karnebeek, S.B. Wortmann, M.C. de Vries, M.C.H. Janssen, C. Gilissen, J. Engel, R.A. Wevers, Next-generation metabolic screening: targeted and untargeted metabolomics for the diagnosis of inborn errors of metabolism in individual patients, J. Inherit. Metab. Dis.. 41 (2018) 337–353. https://doi.org/10.1007/s10545-017-0131-6.

[7] S.L. Stenton, L.S. Kremer, R. Kopajtich, C. Ludwig, H. Prokisch, The diagnosis of inborn errors of metabolism by an integrative “multi-omics” approach: A perspective encompassing genomics, transcriptomics, and proteomics, J. Inherit. Metab. Dis. 43 (2020) 25–35. https://doi.org/10.1002/jimd.12130.

[8] G.L. Arnold, Inborn errors of metabolism in the 21 century: past to present, Ann. Transl. Med. 6 (2018) 467. https://doi.org/10.21037/atm.2018.11.36.

[9] A. Jurecka, Inborn errors of purine and pyrimidine metabolism, J. Inherit. Metab. Dis. 32 (2009) 247–263. https://doi.org/10.1007/s10545-009-1094-z.

[10] S. Balasubramaniam, J.A. Duley, J. Christodoulou, Inborn errors of purine metabolism: clinical update and therapies, J. Inherit. Metab. Dis. 37 (2014) 669–686. https://doi.org/10.1007/s10545-014-9731-6.

[11] S. Balasubramaniam, J.A. Duley, J. Christodoulou, Inborn errors of pyrimidine metabolism: clinical update and therapy, J. Inherit. Metab. Dis. 37 (2014) 687–698. https://doi.org/10.1007/s10545-014-9742-3.

[12] W. Smith, P.S. Kishnani, B. Lee, R.H. Singh, W.J. Rhead, L. Sniderman King, M. Smith, M. Summar, Urea cycle disorders: clinical presentation outside the newborn period, Crit. Care Clin. 21 (2005) S9–17. https://doi.org/10.1016/j.ccc.2005.05.007.

[13] L.C. Burrage, L. Thistlethwaite, B.M. Stroup, Q. Sun, M.J. Miller, S.C.S. Nagamani, W. Craigen, F. Scaglia, V.R. Sutton, B. Graham, A.D. Kennedy, Members of the UCDC, A. Milosavljevic, B.H. Lee, S.H. Elsea, Untargeted metabolomic profiling reveals multiple pathway perturbations and new clinical biomarkers in urea cycle disorders, Genet. Med. 21 (2019) 1977–1986. https://doi.org/10.1038/s41436-019-0442-0.

[14] P. Monostori, G. Klinke, J. Hauke, S. Richter, J. Bierau, S.F. Garbade, G.F. Hoffmann, C.-D. Langhans, D. Haas, J.G. Okun, Extended diagnosis of purine and pyrimidine disorders from urine: LC MS/MS assay development and clinical validation, PLoS One. 14 (2019) e0212458. https://doi.org/10.1371/journal.pone.0212458.

[15] P. Kruszka, D. Regier, Inborn Errors of Metabolism: From Preconception to Adulthood, Am. Fam. Physician. 99 (2019) 25–32.

[16] J.J. Allaire, Y. Xie, J. McPherson, J. Luraschi, K. Ushey, A. Atkins, H. Wickham, J. Cheng, W. Chang, R. Iannone, rmarkdown: Dynamic Documents for R, (2021). https://github.com/rstudio/rmarkdown.

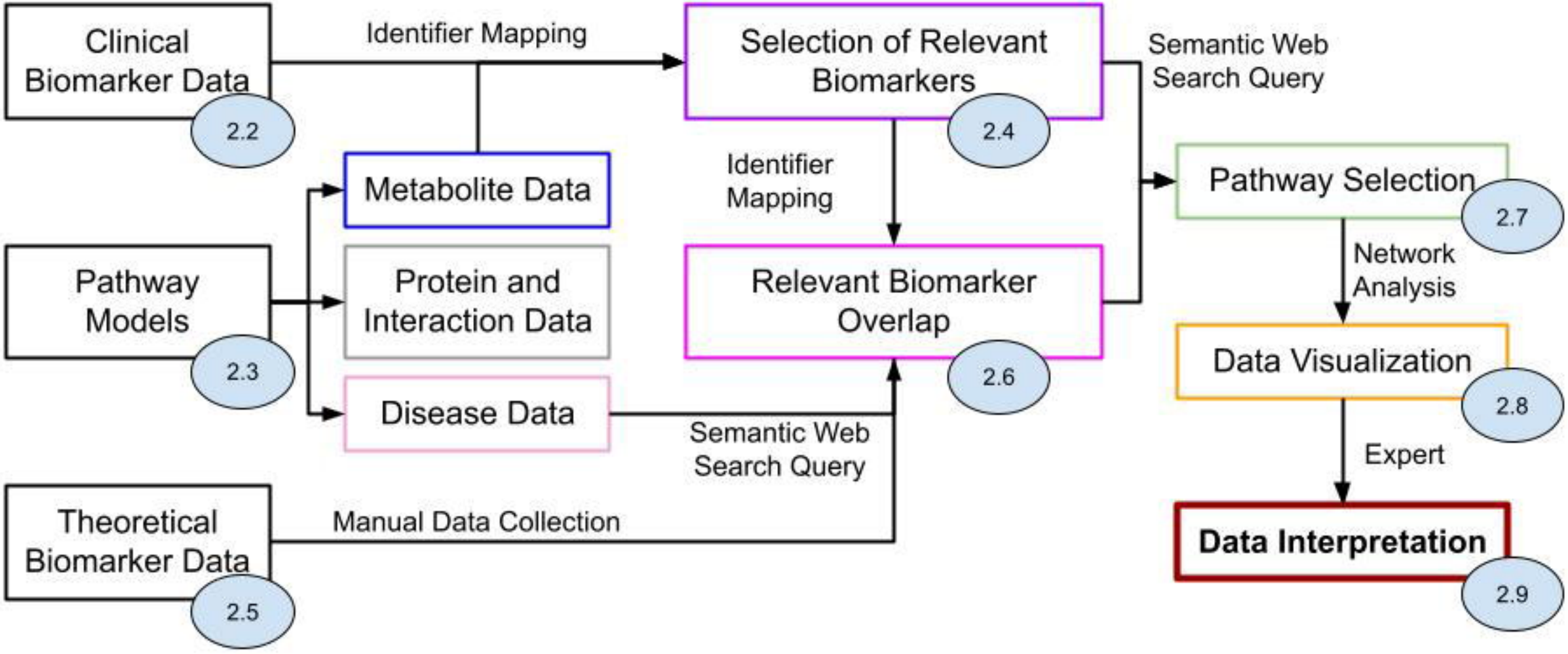
[17] R Core team, R: A Language and Environment for Statistical Computing, (2020). https://www.r-project.org/ (accessed September 11, 2021).

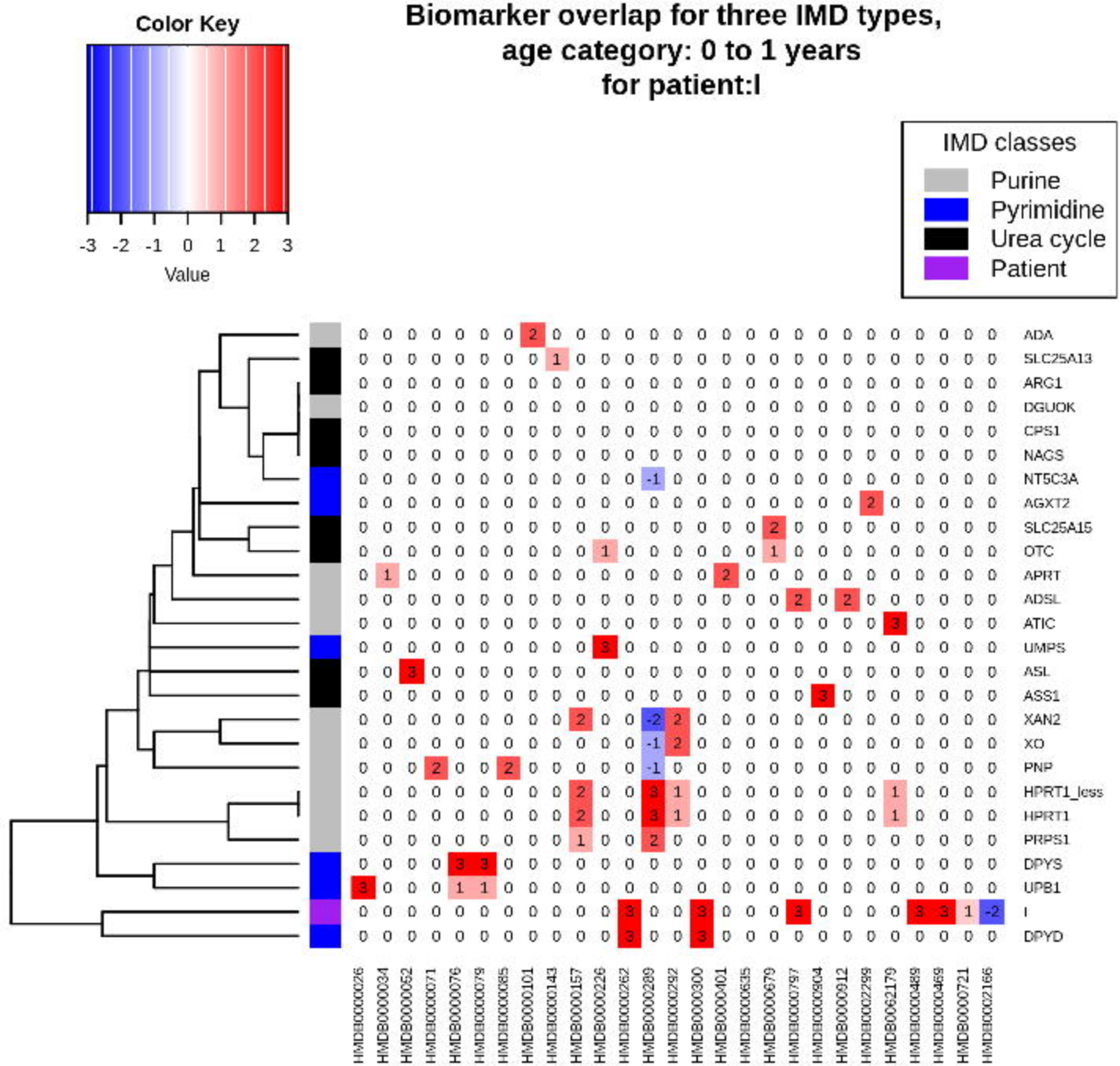
[18] RStudio Team, RStudio: Integrated Development Environment for R, (2019). https://www.rstudio.com/ (accessed September 11, 2021).

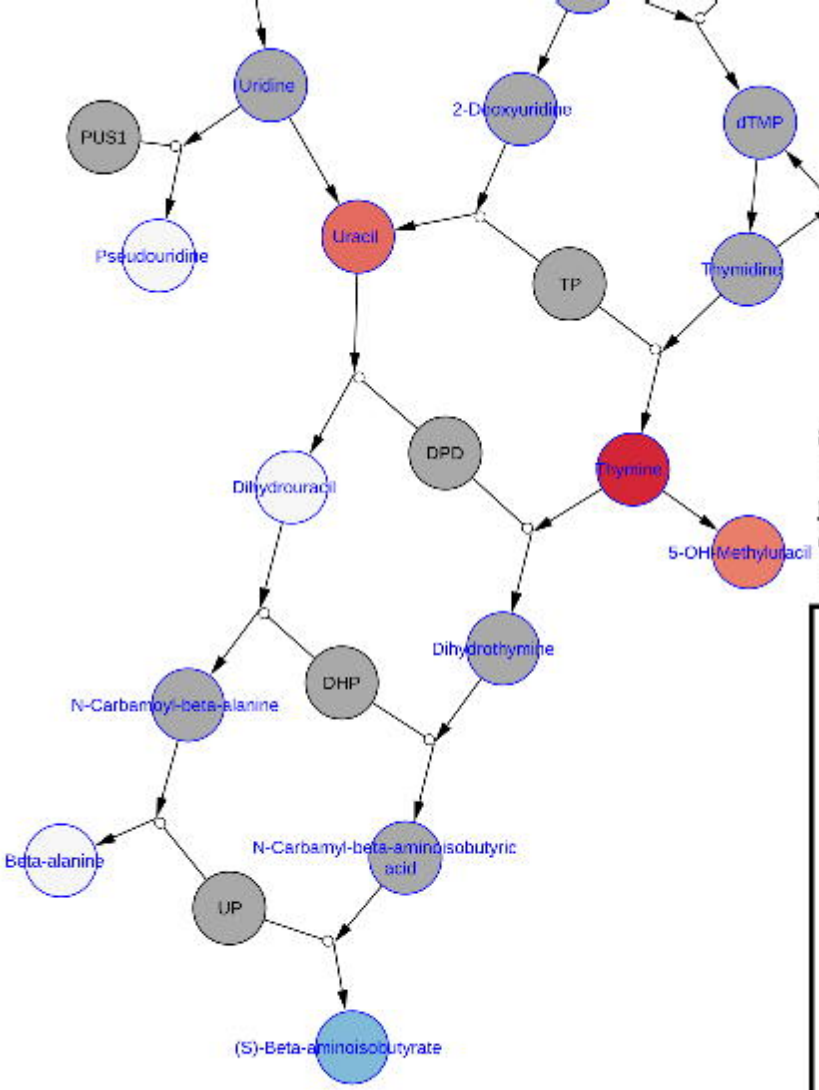
[19] J. Hastings, G. Owen, A. Dekker, M. Ennis, N. Kale, V. Muthukrishnan, S. Turner, N. Swainston, P. Mendes, C. Steinbeck, ChEBI in 2016: Improved services and an expanding collection of metabolites, Nucleic Acids Res. 44 (2016) D1214–D1219. https://doi.org/10.1093/nar/gkv1031.
[20] A. Waagmeester, G. Stupp, S. Burgstaller-Muehlbacher, B.M. Good, M. Griffith, O.L. Griffith, K. Hanspers, H. Hermjakob, T.S. Hudson, K. Hybiske, S.M. Keating, M. Manske, M. Mayers, D. Mietchen, E. Mitraka, A.R. Pico, T. Putman, A. Riutta, N. Queralt-Rosinach, L.M. Schriml, T. Shafee, D. Slenter, R. Stephan, K. Thornton, G. Tsueng, R. Tu, S. Ul-Hasan, E. Willighagen, C. Wu, A.I. Su, Wikidata as a knowledge graph for the life sciences, Elife. 9 (2020). https://doi.org/10.7554/eLife.52614.
[21] W.A.H. Waterval, J.L.J.M. Scheijen, M.M.J.C. Ortmans-Ploemen, C.D. Habets-van der Poel, J. Bierau, 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. 407 (2009) 36–42. https://doi.org/10.1016/j.cca.2009.06.023.
[22] M. Kutmon, M.P. van Iersel, A. Bohler, T. Kelder, N. Nunes, A.R. Pico, C.T. Evelo, PathVisio 3: an extendable pathway analysis toolbox, PLoS Comput. Biol. 11 (2015) e1004085. https://doi.org/10.1371/journal.pcbi.1004085.
[23] UniProt Consortium, UniProt: the universal protein knowledgebase in 2021, Nucleic Acids Res. 49 (2021) D480–D489. https://doi.org/10.1093/nar/gkaa1100.
[24] T. Lombardot, A. Morgat, K.B. Axelsen, L. Aimo, N. Hyka-Nouspikel, A. Niknejad, A. Ignatchenko, I. Xenarios, E. Coudert, N. Redaschi, A. Bridge, Updates in Rhea: SPARQLing biochemical reaction data, Nucleic Acids Res. 47 (2019) D596–D600. https://doi.org/10.1093/nar/gky876.
[25] V.A. McKusick, Mendelian Inheritance in Man: A Catalog of Human Genes and Genetic Disorders, JHU Press, 1998.
[26] M. Martens, A. Ammar, A. Riutta, A. Waagmeester, D.N. Slenter, K. Hanspers, R. A Miller, D. Digles, E.N. Lopes, F. Ehrhart, L.J. Dupuis, L.A. Winckers, S.L. Coort, E.L. Willighagen, C.T. Evelo, A.R. Pico, M. Kutmon, WikiPathways: connecting communities, Nucleic Acids Res. 49 (2021) D613–D621. https://doi.org/10.1093/nar/gkaa1024.
[27] A. Waagmeester, M. Kutmon, A. Riutta, R. Miller, E.L. Willighagen, C.T. Evelo, A.R. Pico, Using the Semantic Web for Rapid Integration of WikiPathways with Other Biological Online Data Resources, PLoS Comput. Biol. 12 (2016) e1004989. https://doi.org/10.1371/journal.pcbi.1004989.
[28] S. Tweedie, B. Braschi, K. Gray, T.E.M. Jones, R.L. Seal, B. Yates, E.A. Bruford, Genenames.org: the HGNC and VGNC resources in 2021, Nucleic Acids Res. 49 (2021) D939–D946. https://doi.org/10.1093/nar/gkaa980.
[29] D.S. Wishart, Y.D. Feunang, A. Marcu, A.C. Guo, K. Liang, R. Vázquez-Fresno, T. Sajed, D. Johnson, C. Li, N. Karu, Z. Sayeeda, E. Lo, N. Assempour, M. Berjanskii, S. Singhal, D. Arndt, Y. Liang, H. Badran, J. Grant, A. Serra-Cayuela, Y. Liu, R. Mandal, V. Neveu, A. Pon, C. Knox, M. Wilson, C. Manach, A. Scalbert, HMDB 4.0: the human metabolome database for 2018, Nucleic Acids Res. 46 (2018) D608–D617. https://doi.org/10.1093/nar/gkx1089.
[30] J. Gao, C. Zhang, M. van Iersel, L. Zhang, D. Xu, N. Schultz, A. R Pico, BridgeDb app: unifying identifier mapping services for Cytoscape, F1000Res. 3 (2014) 148. https://doi.org/10.12688/f1000research.4521.1.
[31] D. Otasek, J.H. Morris, J. Bouças, A.R. Pico, B. Demchak, Cytoscape Automation: empowering workflow-based network analysis, Genome Biol. 20 (2019) 185. https://doi.org/10.1186/s13059-019-1758-4.
[32] M.P. van Iersel, A.R. Pico, T. Kelder, J. Gao, I. Ho, K. Hanspers, B.R. Conklin, C.T. Evelo, The BridgeDb framework: standardized access to gene, protein and metabolite identifier mapping services, BMC Bioinf. 11 (2010) 5. https://doi.org/10.1186/1471-2105-11-5.
[33] M. Hattori, Y. Okuno, S. Goto, M. Kanehisa, Development of a chemical structure comparison method for integrated analysis of chemical and genomic information in the metabolic pathways, J. Am. Chem. Soc. 125 (2003) 11853–11865. https://doi.org/10.1021/ja036030u.
[34] J. Galgonek, T. Hurt, V. Michlíková, P. Onderka, J. Schwarz, J. Vondrášek, Advanced SPARQL querying in small molecule databases, J. Cheminform. 8 (2016) 31. https://doi.org/10.1186/s13321-016-0144-4.
[35] D. Slenter, WikiPathways Sept 2021 Release-RDF Data, WikiPathways Sept 2021 Release - RDF data, (2021). https://doi.org/10.5281/ZENODO.5632921.
[36] J.J.Y. Lee, W.W. Wasserman, G.F. Hoffmann, C.D.M. van Karnebeek, N. Blau, Knowledge base and mini-expert platform for the diagnosis of inborn errors of metabolism, Genet. Med. 20 (2018)

151–158. https://doi.org/10.1038/gim.2017.108.
[37] P. Shannon, A. Markiel, O. Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. 13 (2003) 2498–2504. https://doi.org/10.1101/gr.1239303.
[38] D. Slenter, Metabolite_BridgeDb_ID_Mapping_Database, Metabolite BridgeDb ID Mapping Database (20210109), (2021). https://doi.org/10.6084/m9.figshare.13550384.v1.
[39] A. Dappert, A. Farquhar, R. Kotarski, K. Hewlett, Connecting the Persistent Identifier Ecosystem: Building the Technical and Human Infrastructure for Open Research, Data Sci. J. 16 (2017). https://doi.org/10.5334/dsj-2017-028.
[40] M.D. Wilkinson, M. Dumontier, I.J.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, J. Bouwman, A.J. Brookes, T. Clark, M. Crosas, I. Dillo, O. Dumon, S. Edmunds, C.T. Evelo, R. Finkers, A. Gonzalez-Beltran, A.J.G. Gray, P. Groth, C. Goble, J.S. Grethe, J. Heringa, P.A.C. 't Hoen, R. Hooft, T. Kuhn, R. Kok, J. Kok, S.J. Lusher, M.E. Martone, A. Mons, A.L. Packer, B. Persson, P. Rocca-Serra, M. Roos, R. van Schaik, S.-A. Sansone, E. Schultes, T. Sengstag, T. Slater, G. Strawn, M.A. Swertz, M. Thompson, J. van der Lei, E. van Mulligen, J. Velterop, A. Waagmeester, P. Wittenburg, K. Wolstencroft, J. Zhao, B. Mons, The FAIR Guiding Principles for scientific data management and stewardship, Sci Data. 3 (2016) 160018. https://doi.org/10.1038/sdata.2016.18.
[41] D. Weininger, SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules, J. Chem. Inf. Model. 28 (1988) 31–36. https://doi.org/10.1021/ci00057a005.
[42] E.L. Willighagen, J.W. Mayfield, J. Alvarsson, A. Berg, L. Carlsson, N. Jeliazkova, S. Kuhn, T. Pluskal, M. Rojas-Chertó, O. Spjuth, G. Torrance, C.T. Evelo, R. Guha, C. Steinbeck, The Chemistry Development Kit (CDK) v2.0: atom typing, depiction, molecular formulas, and substructure searching, J. Cheminform. 9 (2017). https://doi.org/10.1186/s13321-017-0220-4.
[43] L.K.M. Steinbusch, P. Wang, H.W.A.H. Waterval, F.A.P.M. Stassen, K.L.M. Coene, U.F.H. Engelke, D.D.J. Habets, J. Bierau, I.M.L.W. Körver-Keularts, Targeted urine metabolomics with a graphical reporting tool for rapid diagnosis of inborn errors of metabolism, J. Inherit. Metab. Dis. 44 (2021) 1113–1123. https://doi.org/10.1002/jimd.12385.
[44] J.E. Kyle, L. Aimo, A.J. Bridge, G. Clair, M. Fedorova, J.B. Helms, M.R. Molenaar, Z. Ni, M. Orešič, D. Slenter, E. Willighagen, B.-J.M. Webb-Robertson, Interpreting the lipidome: bioinformatic approaches to embrace the complexity, Metabolomics. 17 (2021) 55. https://doi.org/10.1007/s11306-021-01802-6.
[45] K. Hanspers, M. Kutmon, S.L. Coort, D. Digles, L.J. Dupuis, F. Ehrhart, F. Hu, E.N. Lopes, M. Martens, N. Pham, W. Shin, D.N. Slenter, A. Waagmeester, E.L. Willighagen, L.A. Winckers, C.T. Evelo, A.R. Pico, Ten simple rules for creating reusable pathway models for computational analysis and visualization, PLoS Comput. Biol. 17 (2021) e1009226. https://doi.org/10.1371/journal.pcbi.1009226.
[46] K. Hanspers, A. Riutta, M. Summer-Kutmon, A.R. Pico, Pathway information extracted from 25 years of pathway figures, Genome Biol. 21 (2020) 273. https://doi.org/10.1186/s13059-020-02181-2.
[47] M.D. Stobbe, S.M. Houten, G.A. Jansen, A.H.C. van Kampen, P.D. Moerland, Critical assessment of human metabolic pathway databases: a stepping stone for future integration, BMC Syst. Biol. 5 (2011) 165. https://doi.org/10.1186/1752-0509-5-165.
[48] D. Domingo-Fernández, C.T. Hoyt, C. Bobis-Álvarez, J. Marín-Llaó, M. Hofmann-Apitius, ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases, NPJ Syst. Biol. Appl. 5 (2019) 3. https://doi.org/10.1038/s41540-018-0078-8.
[49] I. Rodchenkov, O. Babur, A. Luna, B.A. Aksoy, J.V. Wong, D. Fong, M. Franz, M.C. Siper, M. Cheung, M. Wrana, H. Mistry, L. Mosier, J. Dlin, Q. Wen, C. O'Callaghan, W. Li, G. Elder, P.T. Smith, C. Dallago, E. Cerami, B. Gross, U. Dogrusoz, E. Demir, G.D. Bader, C. Sander, Pathway Commons 2019 Update: integration, analysis and exploration of pathway data, Nucleic Acids Res. 48 (2020) D489–D497. https://doi.org/10.1093/nar/gkz946.
[50] H.E. Benson, S. Watterson, J.L. Sharman, C.P. Mpamhanga, A. Parton, C. Southan, A.J. Harmar, P. Ghazal, Is systems pharmacology ready to impact upon therapy development? A study on the cholesterol biosynthesis pathway, Br. J. Pharmacol. 174 (2017) 4362–4382. https://doi.org/10.1111/bph.14037.
[51] K.-H. Hellwich, R.M. Hartshorn, A. Yerin, T. Damhus, A.T. Hutton, Brief Guide to the Nomenclature of Organic Chemistry (IUPAC Technical Report), IUPAC Standards Online. (n.d.).

https://doi.org/10.1515/iupac.92.0027.
[52] D.M. Lowe, P.T. Corbett, P. Murray-Rust, R.C. Glen, Chemical name to structure: OPSIN, an open source solution, J. Chem. Inf. Model. 51 (2011) 739–753. https://doi.org/10.1021/ci100384d.
[53] K. Rajan, A. Zielesny, C. Steinbeck, STOUT: SMILES to IUPAC names using neural machine translation, J. Cheminform. 13 (2021) 34. https://doi.org/10.1186/s13321-021-00512-4.
[54] S.R. Heller, A. McNaught, I. Pletnev, S. Stein, D. Tchekhovskoi, InChI, the IUPAC International Chemical Identifier, J. Cheminform. 7 (2015) 23. https://doi.org/10.1186/s13321-015-0068-4.
[55] M. Kratochvíl, J. Vondrášek, J. Galgonek, Sachem: a chemical cartridge for high-performance substructure search, J. Cheminform. 10 (2018) 27. https://doi.org/10.1186/s13321-018-0282-y.
[56] L.M. Schriml, E. Mitraka, J. Munro, B. Tauber, M. Schor, L. Nickle, V. Felix, L. Jeng, C. Bearer, R. Lichenstein, K. Bisordi, N. Campion, B. Hyman, D. Kurland, C.P. Oates, S. Kibbey, P. Sreekumar, C. Le, M. Giglio, C. Greene, Human Disease Ontology 2018 update: classification, content and workflow expansion, Nucleic Acids Res. 47 (2019) D955–D962. https://doi.org/10.1093/nar/gky1032.
[57] S. Köhler, M. Gargano, N. Matentzoglu, L.C. Carmody, D. Lewis-Smith, N.A. Vasilevsky, D. Danis, G. Balagura, G. Baynam, A.M. Brower, T.J. Callahan, C.G. Chute, J.L. Est, P.D. Galer, S. Ganesan, M. Griese, M. Haimel, J. Pazmandi, M. Hanauer, N.L. Harris, M.J. Hartnett, M. Hastreiter, F. Hauck, Y. He, T. Jeske, H. Kearney, G. Kindle, C. Klein, K. Knoflach, R. Krause, D. Lagorce, J.A. McMurry, J.A. Miller, M.C. Munoz-Torres, R.L. Peters, C.K. Rapp, A.M. Rath, S.A. Rind, A.Z. Rosenberg, M.M. Segal, M.G. Seidel, D. Smedley, T. Talmy, Y. Thomas, S.A. Wiafe, J. Xian, Z. Yüksel, I. Helbig, C.J. Mungall, M.A. Haendel, P.N. Robinson, The Human Phenotype Ontology in 2021, Nucleic Acids Res. 49 (2021) D1207–D1217. https://doi.org/10.1093/nar/gkaa1043.
[58] C.R. Ferreira, C.D.M. van Karnebeek, J. Vockley, N. Blau, A proposed nosology of inborn errors of metabolism, Genet. Med. 21 (2019) 102–106. https://doi.org/10.1038/s41436-018-0022-8.
[59] G. Wu, X. Feng, L. Stein, A human functional protein interaction network and its application to cancer data analysis, Genome Biol. 11 (2010) R53. https://doi.org/10.1186/gb-2010-11-5-r53.
[60] K. Nishida, K. Ono, S. Kanaya, K. Takahashi, KEGGscape: a Cytoscape app for pathway data integration, F1000Res. 3 (2014) 144. https://doi.org/10.12688/f1000research.4524.1.
[61] M. König, cy3sabiork: A Cytoscape app for visualizing kinetic data from SABIO-RK, F1000Res. (approved with Reservations). (2016). https://doi.org/10.1101/062091.
[62] J. Krijt, J. Sokolová, P. Ješina, L. Dvořáková, M. Řeboun, K. Brennerová, M. Mistrík, J. Zeman, T. Honzík, V. Kožich, Activity of the liver enzyme ornithine carbamoyltransferase (OTC) in blood: LC-MS/MS assay for non-invasive diagnosis of ornithine carbamoyltransferase deficiency, Clin. Chem. Lab. Med. 55 (2017) 1168–1177. https://doi.org/10.1515/cclm-2016-0715.
[63] S. Yamaguchi, L.L. Brailey, H. Morizono, A.E. Bale, M. Tuchman, Mutations and polymorphisms in the human ornithine transcarbamylase (OTC) gene, Hum. Mutat. 27 (2006) 626–632. https://doi.org/10.1002/humu.20339.
[64] K.M. Stepien, T. Geberhiwot, C.J. Hendriksz, E.P. Treacy, Challenges in diagnosing and managing adult patients with urea cycle disorders, J. Inherit. Metab. Dis. 42 (2019) 1136–1146. https://doi.org/10.1002/jimd.12096.
[65] F. Deodato, S. Boenzi, F.M. Santorelli, C. Dionisi-Vici, Methylmalonic and propionic aciduria, Am. J. Med. Genet. C Semin. Med. Genet. 142C (2006) 104–112. https://doi.org/10.1002/ajmg.c.30090.
[66] S. Scholl-Bürgi, A. Höller, K. Pichler, M. Michel, E. Haberlandt, D. Karall, Ketogenic diets in patients with inherited metabolic disorders, J. Inherit. Metab. Dis. 38 (2015) 765–773. https://doi.org/10.1007/s10545-015-9872-2.


Clinical Biomarker Data
2.2
Identifier Mapping
Selection of Relevant Biomarkers
2.4
Semantic Web Search Query
Metabolite Data
Identifier Mapping
Pathway Selection
2.7
Pathway Models
2.3
Protein and Interaction Data
Relevant Biomarker Overlap
2.6
Network Analysis
Data Visualization
2.8
Disease Data
Semantic Web Search Query
Expert
Theoretical Biomarker Data
2.5
Manual Data Collection
Data Interpretation
2.9


Biomarker overlap for three IMD types,
age category: 0 to 1 years
for patient:I
Color Key
-3 -2 -1 0 1 2 3
Value
IMD classes
Purine
Pyrimidine
Urea cycle
Patient
ADA
SLC25A13
ARG1
DGUOK
CPS1
NAGS
NT5C3A
AGXT2
SLC25A15
OTC
APRT
ADSL
ATIC
UMPS
ASL
ASS1
XAN2
XO
PNP
HPRT1_loss
HPRT1
PRPS1
DPYS
UPB1
I
DPYD
HMDB00000026
HMDB00000034
HMDB00000052
HMDB00000071
HMDB00000076
HMDB00000079
HMDB00000085
HMDB00000101
HMDB00000143
HMDB00000157
HMDB00000226
HMDB00000262
HMDB00000289
HMDB00000292
HMDB00000300
HMDB00000401
HMDB00000635
HMDB00000679
HMDB00000797
HMDB00000904
HMDB00000912
HMDB00002299
HMDB00062179
HMDB00000489
HMDB00000469
HMDB00000721
HMDB00002166



Selection of the 'Biomarkers for pyrimidine metabolism disorders' pathway (WP4584)

Legend:

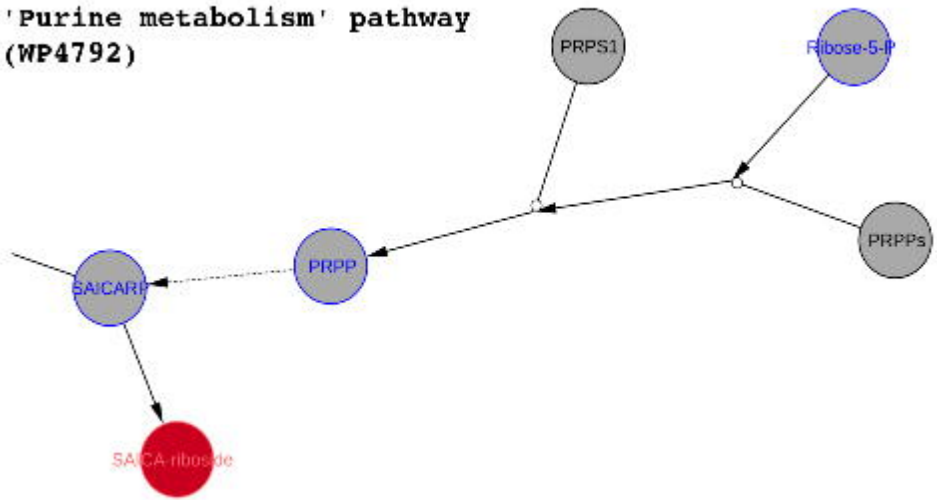
Ribose-5-P Metabolite

DPD Protein

Conversion reaction

Catalysis reaction

-9.77 +9.77

Log2(Change) scale

Selection of the 'Purine metabolism' pathway (WP4792)