MetGENE: Gene-centric Metabolomics Information Retrieval Tool

Sumana Srinivasan, Mano Ram Maurya, Srinivasan Ramachandran, Eoin Fahy and Shankar Subramaniam

CFDE: Gene Working Group

Biomedical Data Commons Workbench (BDCW)

Metabolomics Workbench DCC

University of California, San Diego

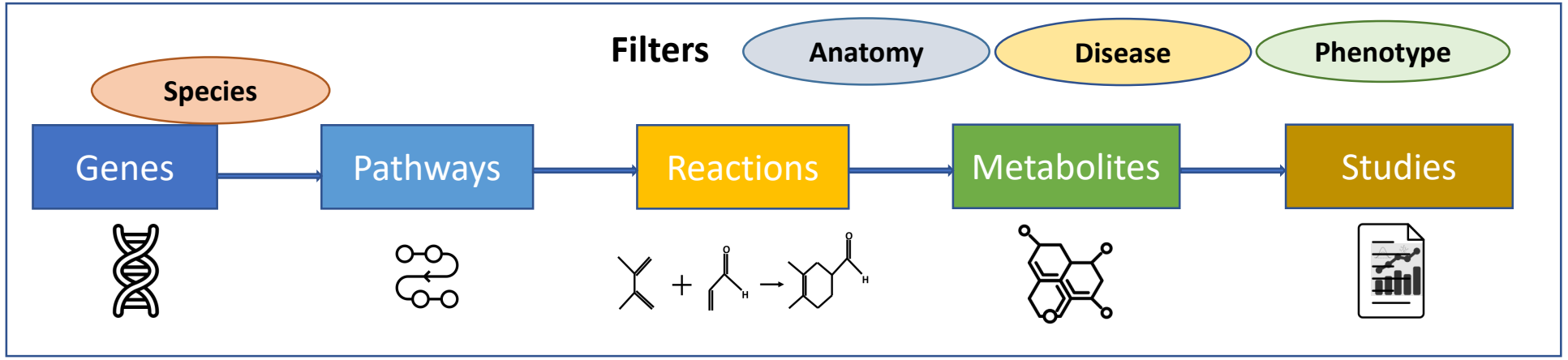
Funding: NIH Common Funds (Metabolomics Workbench and Common Fund Data Ecosystem (CFDE))

Motivation

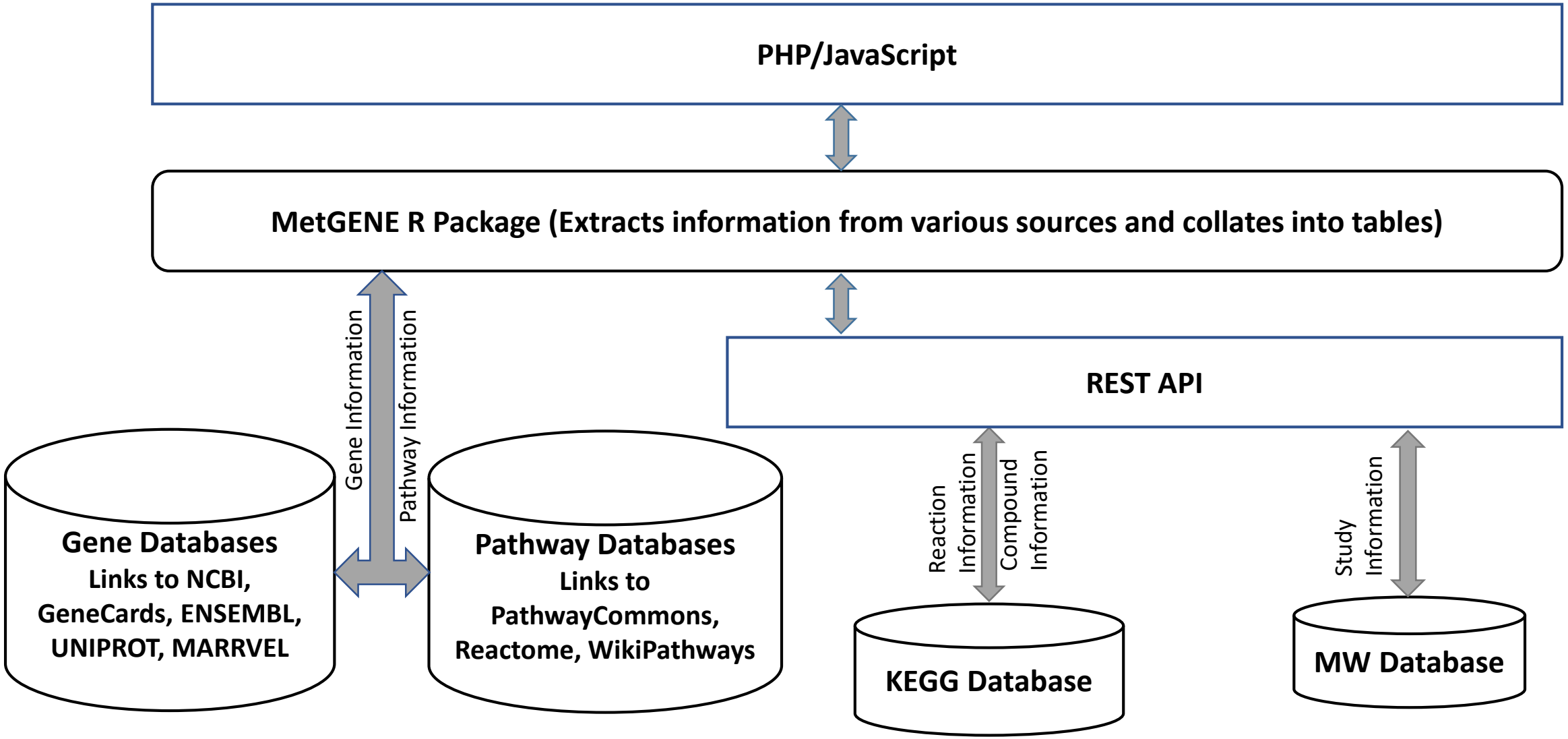
- Biomedical research involves integration of multiomics data contextually, to reconstruct networks, understand mechanisms, identify biomarkers and potential drug targets for drug design and therapy.
- Researchers need to access data and information from several diverse sources comprising of data in a variety of (and sometimes incomplete) formats.
- Next, the data and information are further analyzed and aggregated manually (or semi-automatically at best) with considerable effort in order to reconstruct networks, develop quantitative models and decipher mechanisms.
- MetGENE is a knowledge-based data integration tool that uses a hierarchical information retrieval approach to aggregate gene, pathway, reaction and metabolite and metabolomic study-centric data from various sources (such as KEGG and Metabolomics Workbench) under one dashboard, to enable ease of access by centralization of relevant information.

Gene to Metabolomic Studies

- MetGENE is a hierarchical, knowledge-based gene-centric information retrieval tool.
- Given one or more genes, the MetGENE identifies associations between the gene(s) and the metabolites that are biosynthesized, metabolized, or transported by proteins coded by the genes.
- The gene(s) link to metabolites, the chemical transformations involving the metabolites through gene-specified proteins/enzymes, the functional association of these gene-associated metabolites and the pathways involving these metabolites.



MetGENE Architecture



Features

- Client-side error handling and validation - gene search input is validated to allow only alphanumeric symbols. Invalid gene IDs are recognized, and appropriate error messages are displayed.
- Supports filtering based on
  - Taxonomy - Uses terms (e.g. Human, Mouse), as per the NCBI taxonomy database
  - Anatomy – Uses terms (e.g. Blood, Liver), from Uber-anatomy Ontology (Uberon)
  - Disease - Uses terms (e.g. Diabetes, Fatty liver disease) from Disease Ontology database
  - Phenotype (in progress) - Uses terms (e.g. BMI) from the Human Phenotype Ontology database
  - JSON files for each of these categories are curated and updated regularly to keep them current and use it to generate a two-step selection menu based on anatomy, disease and phenotype slim terms in the first level and corresponding anatomy, disease and phenotype terms in the second level for ease of presenting the options to the user.

Web based UI

- User can specify the gene using several gene ID types, such as SYMBOL, ENTREZ gene ID and ENSEMBL ID.
- Gene ID conversion tool translates these into harmonized IDs that are basis at the computational end for metabolite associations (data integration).
- All studies involving the metabolites associated with the gene-coded proteins, as present in the Metabolomics Workbench (MW), the portal for the NIH Common Fund National Metabolomics Data Repository (NMDR), will be accessible to the user through the portal interface.

Example: PNPLA3

- The *PNPLA3* gene encodes for the protein **adiponutrin**, found in fat and liver cells.
- Regulates the development of adipocytes and the production and breakdown of fats (lipogenesis and lipolysis).
- Diseases associated with PNPLA3 include Fatty Liver Disease and Non-Alcoholic Steatohepatitis.
- Pathways: Metabolism and Glycerophospholipid biosynthesis.

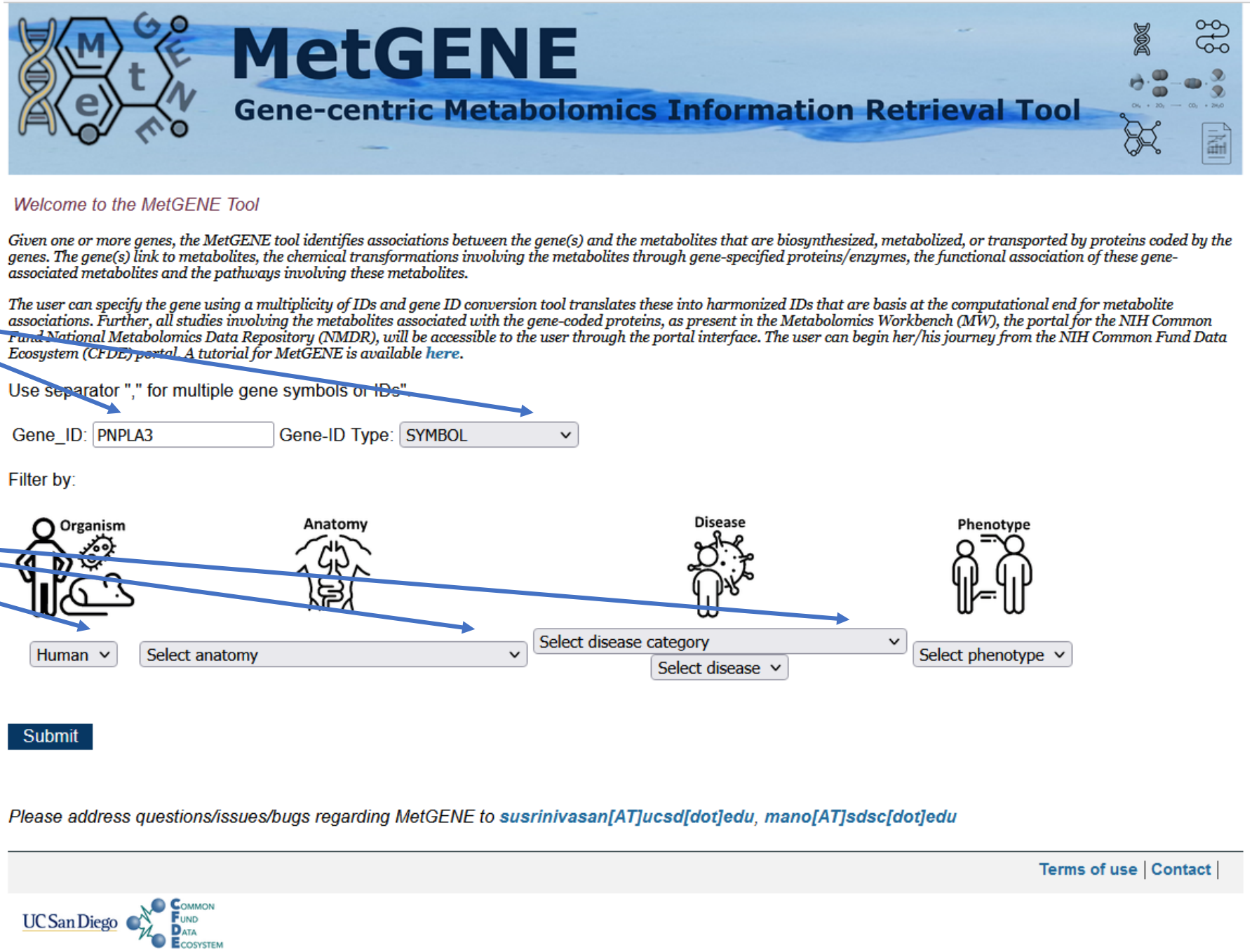
Kohen, … Hobbs, Human fatty liver disease: old questions and new insights , Science, 2011
https://pubmed.ncbi.nlm.nih.gov/21700865/
Biomarkers of NAFLD progression: a lipidomics approach to an epidemic
https://pubmed.ncbi.nlm.nih.gov/25598080/
The role of PNPLA3 in health and disease: https://pubmed.ncbi.nlm.nih.gov/29935383/
PNPLA3—A Potential Therapeutic Target for Personalized Treatment of Chronic Liver Disease: https://www.frontiersin.org/articles/10.3389/fmed.2019.00304/full

Landing page

https://bdcw.org/MetGENE

Gene ID and Type

Filters

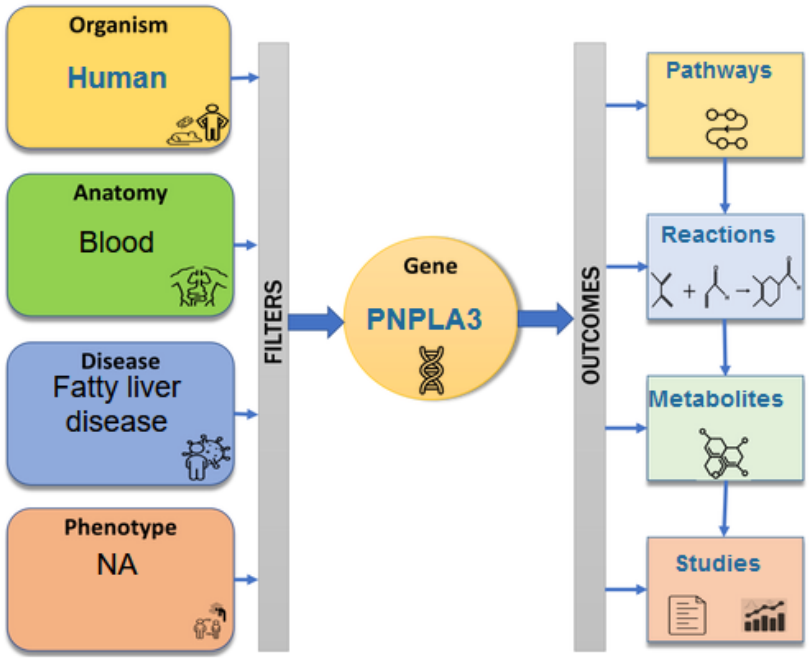


MetGENE

Gene-centric Metabolomics Information Retrieval Tool

Home | Genes | Pathways | Reactions | Metabolites | Studies

Provides context and can be navigated



In the MetGENE tool, information about the gene(s) PNPLA3 is presented in **Genes**, the corresponding pathways in **Pathways** and the reactions in **Reactions** tabs. The metabolites participating in the reactions are presented in **Metabolites** tab. For each metabolite, the studies containing the metabolite are identified from the **Metabolomics Workbench** (MW) and presented in **Studies** tab.

The data from MW studies are presented as table(s), with the metabolite names hyperlinked to MW **RefMet** page (or to the corresponding **KEGG** entry in the absence of a RefMet name) for the metabolite, reaction hyperlinked to its KEGG entry and MW studies hyperlinked to their respective pages. The user also has access to the metabolite statistics via **MetStat**. Further, the user has the option to select more than one metabolite to list only those studies in which all the selected metabolites appear and can download the table as a text, HTML or JSON file.

Genes

MetGENE

Gene-centric Metabolomics Information Retrieval Tool

Home | Genes | Pathways | Reactions | Metabolites | Studies

Gene Information for ***Human*** gene(s) ***PNPLA3***

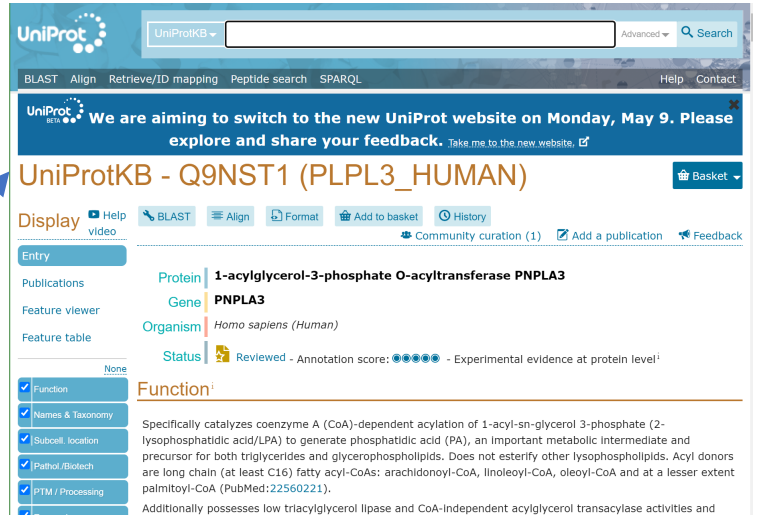
Symbol	KEGG	GeneCards	NCBI	Ensembl	UniProt	MARRVEL
PNPLA3	80339	PNPLA3	80339	ENSG00000100344	Q9NST1	80339

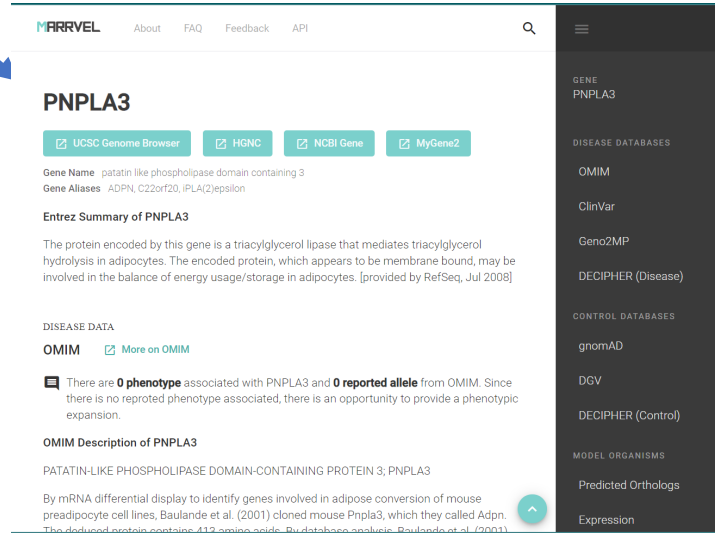
TO JSON | TO CSV

Back to Gene Query

Terms of use | Contact |

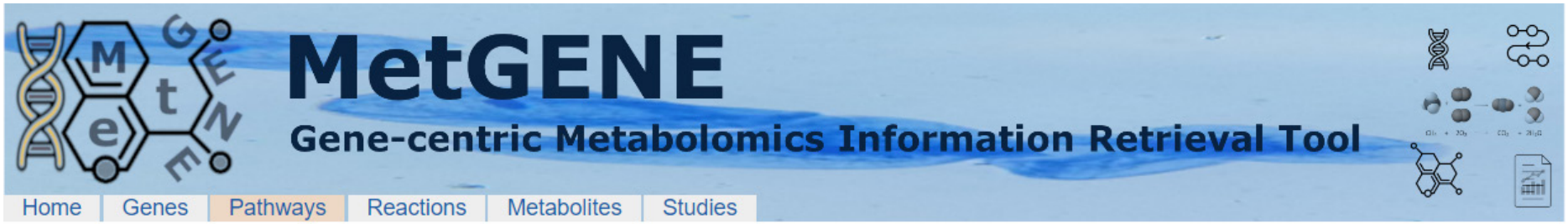
UC San Diego | Common Fund Data Ecosystem





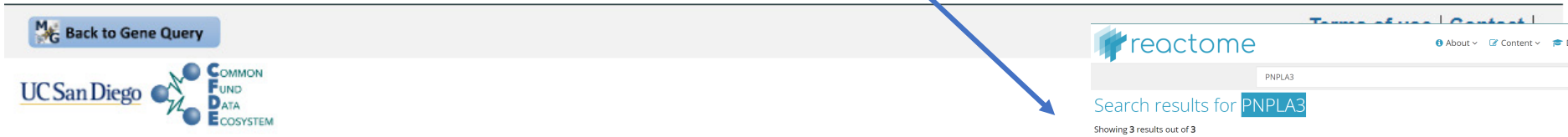
Takes the user to the genomic databases

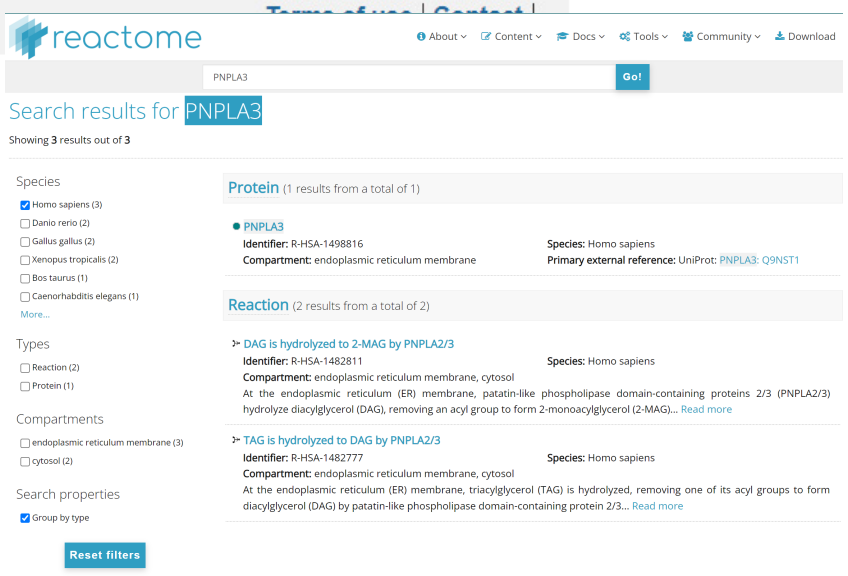
Pathways



Pathway Information for ***Human*** gene(s) ***PNPLA3***

PNPLA3	PNPLA3	PNPLA3	PNPLA3





Reactome Pathway page for PNPLA3 showing DAG and TAG reactions

reactome

About | Content | Docs | Tools | Community | Download

e.g. O95631, NTN1, signaling by EGFR, glucose, GO:0043293 Go!

TAG is hydrolyzed to DAG by PNPLA2/3

Stable Identifier	R-HSA-1482777
Type	Reaction
Species	Homo sapiens
Compartment	endoplasmic reticulum membrane, cytosol

Locations in the PathwayBrowser

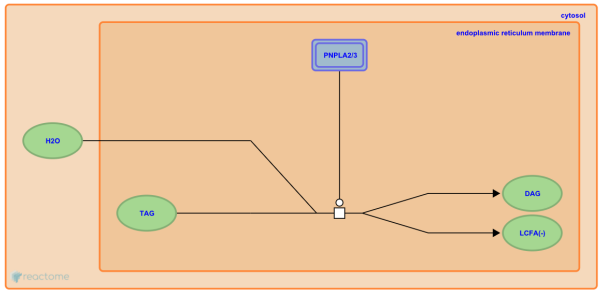
Expand All

Metabolism (Homo sapiens)

General

SBML | BioPAX | PDF

SVG | PNG | PPTX | SBGN



Click the image above or here to open this reaction in the Pathway Browser

The layout of this reaction may differ from that in the pathway view due to the constraints in pathway layout

At the endoplasmic reticulum (ER) membrane, triacylglycerol (TAG) is hydrolyzed, removing one of its acyl groups to form diacylglycerol (DAG) by patatin-like phospholipase domain-containing protein 2/3 (PNPLA2/3) (He et al. 2010, Jenkins et al. 2004, Basantani et al. 2011).

reactome

About | Content | Docs | Tools | Community | Download

e.g. O95631, NTN1, signaling by EGFR, glucose, GO:0043293 Go!

DAG is hydrolyzed to 2-MAG by PNPLA2/3

Stable Identifier	R-HSA-1482811
Type	Reaction
Species	Homo sapiens
Compartment	endoplasmic reticulum membrane, cytosol

Locations in the PathwayBrowser

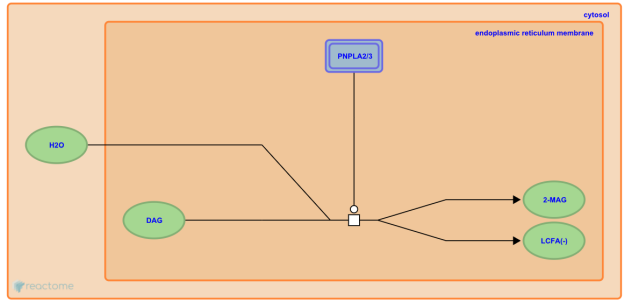
Expand All

Metabolism (Homo sapiens)

General

SBML | BioPAX | PDF

SVG | PNG | PPTX | SBGN



Click the image above or here to open this reaction in the Pathway Browser

The layout of this reaction may differ from that in the pathway view due to the constraints in pathway layout

At the endoplasmic reticulum (ER) membrane, patatin-like phospholipase domain-containing proteins 2/3 (PNPLA2/3) hydrolyze diacylglycerol (DAG), removing an acyl group to form 2-monoacylglycerol (2-MAG) (He et al. 2010, Jenkins et al. 2004, Basantani et al. 2011).

Reactions

Reaction IDs are linked to KEGG reactions

Reaction Name is KEGG reaction name

Reaction Information for ***Human*** gene ***PNPLA3***

KEGG_REACTION_ID	KEGG_REACTION_NAME
R02250	triacylglycerol acylhydrolase
R02687	1,2-diacyl-sn-glycerol acylhydrolase

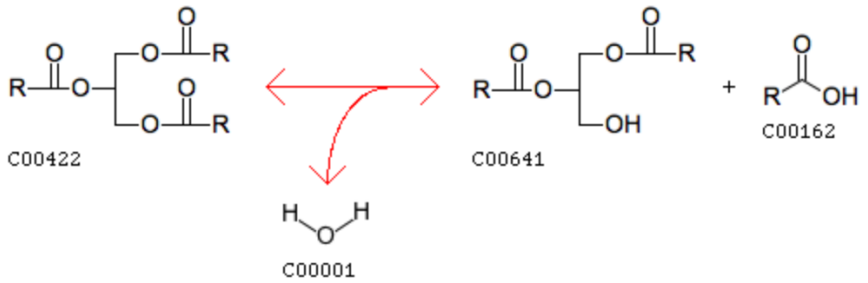
TO JSON | TO CSV

KEGG page for Reaction R02250

KEGG

REACTION: R02250

Help

Entry	R02250 Reaction
Name	triacylglycerol acylhydrolase
Definition	Triacylglycerol + H2O <=> 1,2-Diacyl-sn-glycerol + Fatty acid
Equation	C00422 + C00001 <=> C00641 + C00162
	
Reaction class	RC00020 C00162_C00422 RC00041 C00422_C00641
Enzyme	3.1.1.3
Pathway	rn00561 Glycerolipid metabolism rn01100 Metabolic pathways
Module	M00098 Acylglycerol degradation
Orthology	K01046 triacylglycerol lipase [EC:3.1.1.3] K12298 bile salt-stimulated lipase [EC:3.1.1.3 3.1.1.13] K13534 patatin-like phospholipase domain-containing protein 3 [EC:3.1.1.3 2.3.1.-] K14073 pancreatic triacylglycerol lipase [EC:3.1.1.3] K14074 pancreatic lipase-related protein 1 [EC:3.1.1.3] K14075 pancreatic lipase-related protein 2 [EC:3.1.1.3] K14076 pancreatic lipase-related protein 3 [EC:3.1.1.3] K14452 gastric triacylglycerol lipase [EC:3.1.1.3] K14674 TAG lipase / steryl ester hydrolase / phospholipase A2 / LPA acyltransferase [EC:3.1.1.3 3.1.1.13 3.1.1.4 2.3.1.51] K14675 TAG lipase / lysophosphatidylethanolamine acyltransferase [EC:3.1.1.3 2.3.1.-] K16816 patatin-like phospholipase domain-containing protein 2 [EC:3.1.1.3] K17900 lipase ATG15 [EC:3.1.1.3] K22283 hepatic triacylglycerol lipase [EC:3.1.1.3] K22284 endothelial lipase [EC:3.1.1.3]

Metabolites

MetGENE
Gene-centric Metabolomics Information Retrieval Tool

Home | Genes | Pathways | Reactions | Metabolites | Studies

Metabolite Information for ***Human*** gene(s) ***PNPLA3*** anatomy ***Blood*** disease ***Fatty liver disease***

KEGGMETABID	REFMETNAME	REACTIONS	METSTAT
C00001	H2O;	R02250 R02687	
C00162	Fatty acid	R02250 R02687	
C00422	TG 14:0_16:0_18:1	R02250	
C00422	TG 14:0_18:1_18:1	R02250	
C00422	TG 14:0_18:2_18:2	R02250	
C00422	TG 16:0_16:0_16:0	R02250	
C00422	TG 16:0_18:0_18:1	R02250	
C00422	TG 16:0_18:2_18:2	R02250	
C00422	TG 16:0_18:2_18:3	R02250	
C00422	TG 16:0_18:2_22:6	R02250	
C00422	TG 18:0/18:0/18:0	R02250	
C00422	TG 18:0_18:1_20:4	R02250	
C00422	TG 18:1_18:1_18:1	R02250	
C00422	TG 18:1_18:1_22:6	R02250	

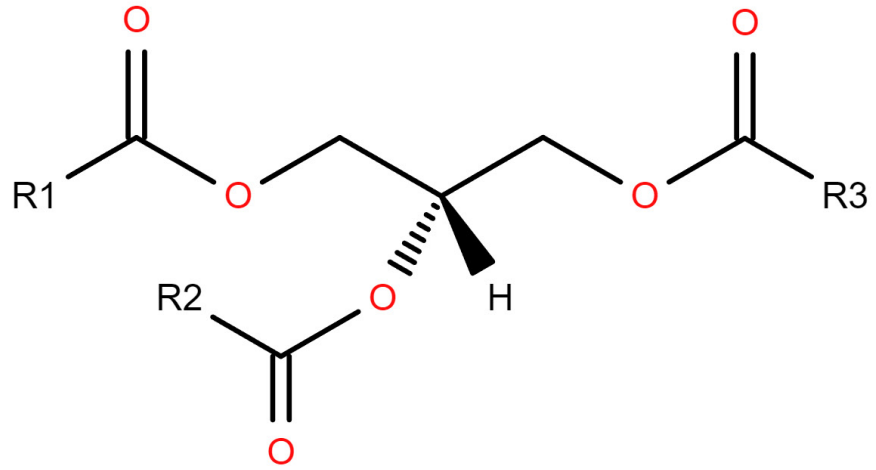
Linked to KEGG (Reaction IDs))

Linked to MW RefMet

Linked to MetStat (metabolite statistics)

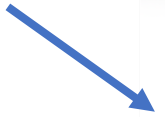
RefMet page for TG 14:0_18:1_18:1

RefMet Compound Details



RefMet name	**TG 14:0_18:1_18:1**	
Alternative name	**TG(14:0_18:1_18:1)**	
Sum Composition	**TG 50:2**	View other entries in RefMet with this sum composition
Exact mass	**830.736340 (neutral)**	**Calculate m/z:** (Choose adduct) **View other RefMet entries with this exact (neutral) mass:** +/- 0.05 amu +/- 0.1 amu +/- 0.2 amu +/- 0.5 amu
Formula	**C53H98O6**	View other entries in RefMet with this formula
Super Class	**Glycerolipids**	
Main Class	Triradylglycerols	
Sub Class	TAG (Triacylglycerols)	
Annotation level	**3** (1:Known structure; 2:Known regiochemistry; 3:Partial structure; 4:Sum-composition)	
Human quantitation	View measurements in targeted assays on human samples	

MetStat page for TG 14:0_16:0_18:1



Insufficent data points (1) to generate a histogram of RSD values

TG 14:0_16:0_18:1(Name:TG 14:0_16:0_18:1 Source:Fatty liver disease Source:Blood Species:Human)
ANOVA results for this metabolite where p-value <=0.05

Name	Study_id	Analysis_id	ANOVA p-Value	FDR	Experimental Conditions (factors)
TG(14:0/16:0/18:1)	ST001710	AN002785	4.105E-4	9.639E-4	T2DM;NA;N;Y
TG(14:0/16:0/18:1)	ST001710	AN002785	1.817E-2	8.446E-2	NAFLD.Category;1;2;4;NA;-
TG(14:0/16:0/18:1)	ST001710	AN002785	4.985E-2	2.257E-1	Kleiner.Steatosis;1;2;3;NA;-

TG 14:0_16:0_18:1(Name:TG 14:0_16:0_18:1 Source:Fatty liver disease Source:Blood Species:Human)
List of studies, analyses, submitted metabolite names, experimental conditions and RSD values

Replicate numbers reflect the number of replicates for each experimental condition where the measured value was not null

Name	Study_id	Analysis_id	Factors	Range(RSD)	Replicates
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:NA | Kleiner.Steatosis:3 | Inflammation:1	84.64	2
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:NA | Kleiner.Steatosis:3 | Inflammation:2	22.64	4
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:N | Kleiner.Steatosis:1 | Inflammation:-	96.44	2
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:N | Kleiner.Steatosis:1 | Inflammation:1	74.63	6
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:N | Kleiner.Steatosis:2 | Inflammation:1	76.32	2
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:N | Kleiner.Steatosis:3 | Inflammation:1	84.28	2
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:N | Kleiner.Steatosis:3 | Inflammation:2	91.08	3
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:1 | T2DM:Y | Kleiner.Steatosis:1 | Inflammation:1	41.36	3
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:2 | T2DM:N | Kleiner.Steatosis:1 | Inflammation:-	26.74	2
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:2 | T2DM:N | Kleiner.Steatosis:2 | Inflammation:1	34.11	4
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:2 | T2DM:N | Kleiner.Steatosis:2 | Inflammation:2	85.85	8
TG(14:0/16:0/18:1)	ST001710	AN002785	Data type:Serum lipidomics | Sex:F | NAFLD.Category:2 | T2DM:N | Kleiner.Steatosis:3 | Inflammation:1	50.93	5

Studies

Metabolomic Studies Information for ***Human*** gene(s) ***PNPLA3*** anatomy ***Blood*** disease ***Fatty liver disease***

Use check boxes to select metabolites to combine their studies.

SELECT	KEGGMETABID	REFMETNAME	STUDIES
	C00001	H2O;	*No studies found*
	C00162	Fatty acid	*No studies found*
☐	C00422	TG 14:0_16:0_18:1	ST001710
☐	C00422	TG 14:0_18:1_18:1	ST0017
☐	C00422	TG 14:0_18:2_18:2	ST001710
☐	C00422	TG 16:0_16:0_16:0	ST001710
☐	C00422	TG 16:0_18:0_18:1	ST001710
☐	C00422	TG 16:0_18:2_18:2	ST001710
☐	C00422	TG 16:0_18:2_18:3	ST001710
☐	C00422	TG 16:0_18:2_22:6	ST001710
☐	C00422	TG 18:0/18:0/18:0	ST001710
☐	C00422	TG 18:0_18:1_20:4	ST001710
☐	C00641	DG 18:2_20:4	ST002091
☐	C00641	DG 18:2_22:6	ST002091
☐	C00641	DG 20:0_20:0	ST002091
	C01885	1-Acylglycerol;	*No studies found*

Metabolic signatures of NAFLD - Lipidomics data (part 1 of 3)

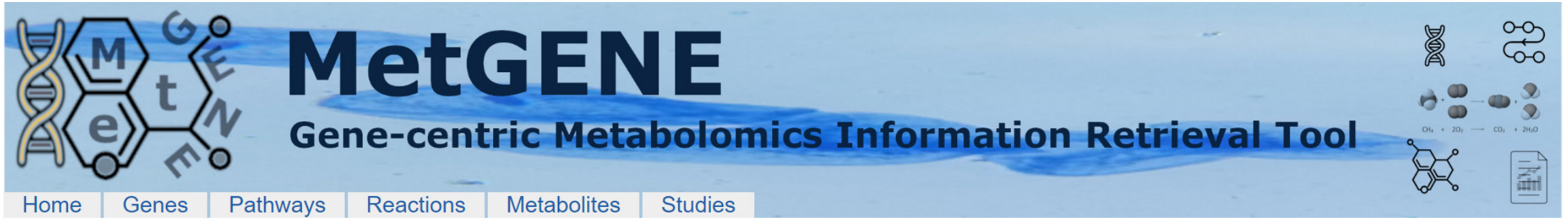
Combine Studies

TO JSON TO CSV

Studies can be combined across different metabolites as selected by the user

Hyperlinked study IDs with hover on study title

Combined list of studies



Combined studies for the selected metabolites

TG 14:0_16:0_18:1,DG 20:0_20:0	ST001710, ST002091

TO JSON | TO CSV

Gene Query

Terms of use | Contact |

UC San Diego

Common Fund Data Ecosystem

https://www.metabolomicsworkbench.org/data/DRCCMetadata.php?Mode=Study&StudyID=ST001710

Summary of Study ST001710

This data is available at the NIH Common Fund's National Metabolomics Data Repository (NMDR) website, the Metabolomics Workbench, https://www.metabolomicsworkbench.org, where it has been assigned Project ID PR001095. The data can be accessed directly via it's Project DOI: **10.21228/M85976**

This work is supported by NIH grant, U2C- DK119886.

See: **https://www.metabolomicsworkbench.org/about/howtocite.php**

This study contains a large results data set and is not available in the mwTab file. It is only available for download via FTP as data file(s) **here**.

Perform statistical analysis | **Show all samples** | **Show named metabolites** | **Download named metabolite data**
Download mwTab file (text) | **Download mwTab file(JSON)** | **Download data files (Contains raw data)**

Study ID	ST001710
Study Title	Metabolic signatures of NAFLD - Lipidomics data (part 1 of 3)
Study Summary	Serum samples were randomized and extracted using a modified version of the previously-published Folch procedure, as applied recently [20]. The maternal samples were analysed as one batch and the cord blood samples as a second batch. In short, 10 µL of 0.9% NaCl and, 120 µL of CHCl3: MeOH (2:1, v/v) containing the internal standards (c = 2.5 µg/mL) was added to 10 µL of each serum sample. The standard solution contained the following compounds: 1,2-diheptadecanoyl-sn-glycero-3-phosphoethanolamine (PE(17:0/17:0)), N-heptadecanoyl-D-erythro-sphingosylphosphorylcholine (SM(d18:1/17:0)), N-heptadecanoyl-D-erythro-sphingosine (Cer(d18:1/17:0)), 1,2-diheptadecanoyl-sn-glycero-3-phosphocholine (PC(17:0/17:0)), 1-heptadecanoyl-2-hydroxy-sn-glycero-3-phosphocholine (LPC(17:0)) and 1-palmitoyl-d31-2-oleoyl-sn-glycero-3-phosphocholine (PC(16:0/d31/18:1)), were purchased from Avanti Polar Lipids, Inc. (Alabaster, AL, USA), and, triheptadecanoylglycerol (TG(17:0/17:0/17:0)) was purchased from Larodan AB (Solna, Sweden). The samples were vortex mixed and incubated on ice for 30 min after

REST API

- MetGENE output tables can be accessed using Smart APIs (https://smart-api.info/registry?q=342e4cec92030d74efd84b61650fb0ea)
- Information can also be accessed via the REST API from the browser, for example,

  https://bdcw.org/MetGENE/rest/reactions/species/hsa/GeneIDType/SYMBOL/GeneInfoStr/HK1/anatomy/Blood/disease/Diabetes/phenotype/BMI/viewType/json

  https://bdcw.org/MetGENE/rest/metabolites/species/hsa/GeneIDType/SYMBOL/GeneInfoStr/HK1/anatomy/Blood/disease/Diabetes/phenotype/BMI/viewType/json

  https://bdcw.org/MetGENE/rest/studies/species/hsa/GeneIDType/SYMBOL/GeneInfoStr/HK1/anatomy/Blood/disease/Diabetes/phenotype/BMI/viewType/json

REST API for Summary

- Single gene case:

https://bdcw.org/MetGENE/mgSummary.php?species=hsa&ENSEMBL=ENSG00000000419&viewType=all

https://bdcw.org/MetGENE/mgSummary.php?species=hsa&GeneSym=ALDOB&GeneID=229

https://bdcw.org/MetGENE/mgSummary.php?species=hsa&GeneSym=RPE&GeneID=6120&viewType=PIE

https://bdcw.org/MetGENE/mgSummary.php?species=hsa&GeneSym=RPE&GeneID=6120&viewType=BAR

- Multiple genes case:

https://bdcw.org/MetGENE/mgSummary.php?species=hsa&GeneSym=RPE__ALDOB__GPI&GeneID=6120__229__2821

https://bdcw.org/MetGENE/mgSummary.php?species=hsa&GeneSym=RPE__ALDOB__GPI&GeneID=6120__229__2821&viewType=PIE

https://bdcw.org/MetGENE/mgSummary.php?species=hsa&GeneSym=RPE__ALDOB__GPI&GeneID=6120__229__2821&viewType=BAR