Biological databases

Matthias König
https://mcp.readthedocs.io

livermetabolism.com
@konigmatt

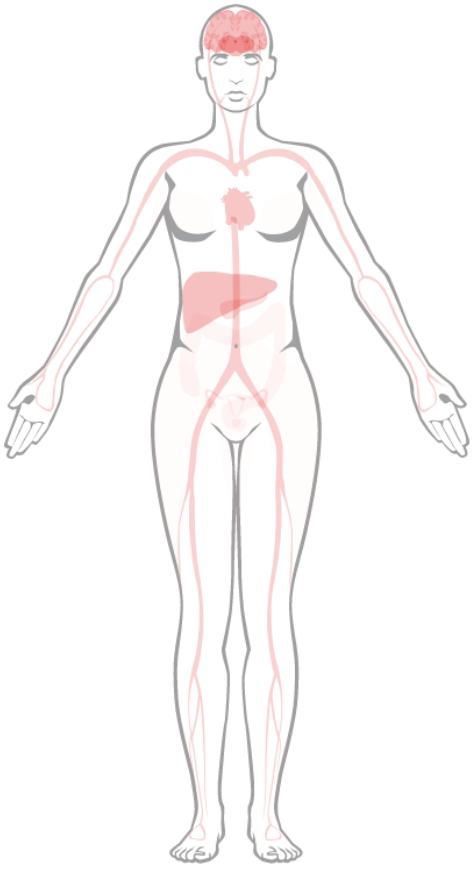
Entry points & portals

- **PathGuide**
  http://www.pathguide.org/
  - General entry point for pathway related databases
  - Good overall overview (many very specific databases exist)
- **BioNumbers**
  https://bionumbers.hms.harvard.edu/search.aspx
  - Database of useful biological numbers
- **EMBL-EBI services**
  https://www.ebi.ac.uk/services
  - Freely available and up-to-date molecular resources

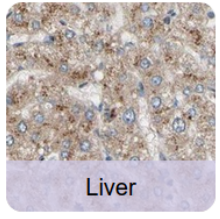
Protein databases

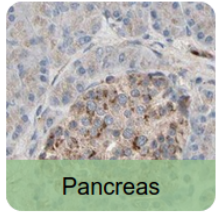
- **UniProt**/Swiss-Prot/TrEMBL
  https://www.uniprot.org/
- **Human Protein Atlas**
  https://proteinatlas.org
  https://www.proteinatlas.org/ENSG00000106633-GCK
- **InterPro**
  https://www.ebi.ac.uk/interpro/
  - Classification of protein families and predicting domains
- **PROSITE**
  https://prosite.expasy.org/
  - Protein families and domains
- **Pfam**
  https://pfam.xfam.org/
  - Protein families and domains

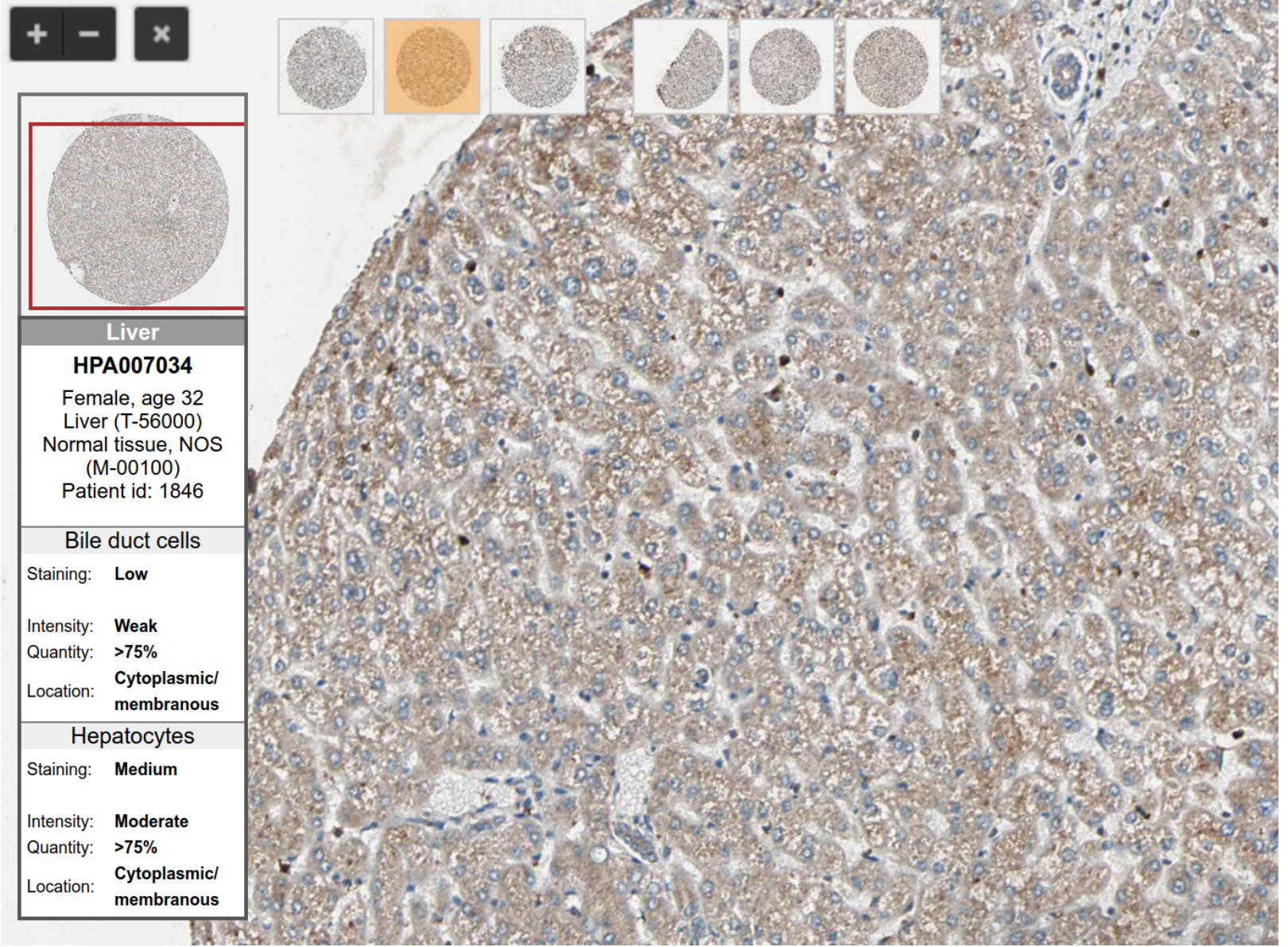
RNA AND PROTEIN EXPRESSION SUMMARY[i]



	RNA expression (NX)[i]	Protein expression (score)[i]
Brain		
Eye		
Endocrine tissues		
Lung		
Proximal digestive tract		
Gastrointestinal tract		
Liver & gallbladder		







Liver

HPA007034

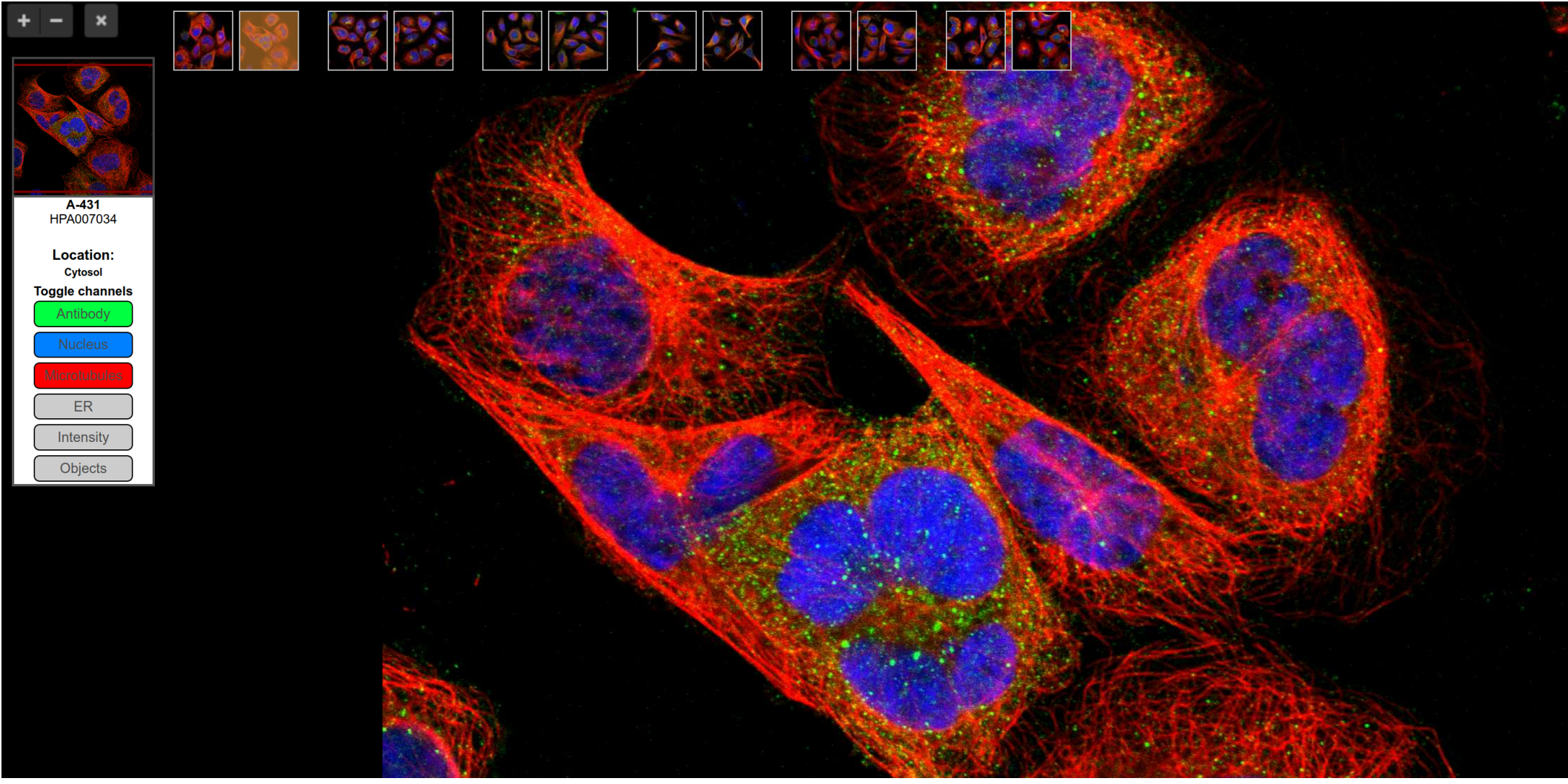
Female, age 32
Liver (T-56000)
Normal tissue, NOS (M-00100)
Patient id: 1846

Bile duct cells

Staining:	**Low**
Intensity:	**Weak**
Quantity:	**>75%**
Location:	**Cytoplasmic/ membranous**

Hepatocytes

Staining:	**Medium**
Intensity:	**Moderate**
Quantity:	**>75%**
Location:	**Cytoplasmic/ membranous**


A-431
HPA007034
Location:
Cytosol
Toggle channels
Antibody
Nucleus
Microtubules
ER
Intensity
Objects

Pathology & Biotech

Involvement in disease

Maturity-onset diabetes of the young 2 (MODY2) 20 Publications

The disease is caused by mutations affecting the gene represented in this entry.

Disease description: A form of diabetes that is characterized by an autosomal dominant mode of inheritance, onset in childhood or early adulthood (usually before 25 years of age), a primary defect in insulin secretion and frequent insulin-independence at the beginning of the disease.

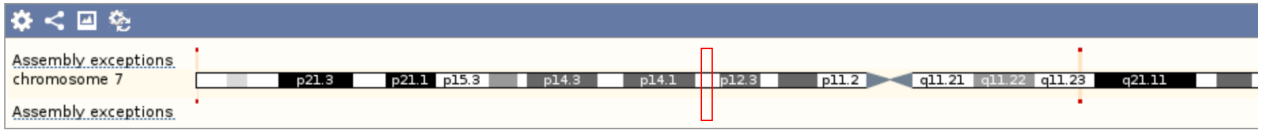
Related information in OMIM

Feature key	Position(s)	Description	Actions	Graphical view	Length
Natural variant (VAR_079430)	16	V → E in MODY2. 1 Publication			1
Natural variant (VAR_079431)	19	I → N in MODY2. 1 Publication			1
Natural variant (VAR_079432)	20	L → P in MODY2. 1 Publication			1
Natural variant (VAR_010584)	36	R → W in MODY2. 3 Publications Corresponds to variant dbSNP:rs762263694	Ensembl, ClinVar.		1
Natural variant (VAR_075220)	43	R → H in MODY2; unknown pathological significance; no change in glucokinase activity. 1 Publication Corresponds to variant dbSNP:rs764232985	Ensembl, ClinVar.		1
Natural variant (VAR_079435)	43	R → S in MODY2. 1 Publication			1
Natural variant (VAR_079436)	44	G → S in MODY2. 1 Publication Corresponds to variant dbSNP:rs267601516	Ensembl.		1
Natural variant (VAR_010585)	53	A → S in MODY2. 1 Publication			1
Natural variant (VAR_079438)	61 – 465	Missing in MODY2. 1 Publication	Add BLAST		405
Natural variant (VAR_079439)	61	Y → S in MODY2; decreased glucokinase activity; decreased affinity for glucose; increased affinity for ATP. 2 Publications			1
Natural variant (VAR_075221)	68	G → D in MODY2; unknown pathological significance; mildly increases glucokinase activity. 1 Publication Corresponds to variant dbSNP:rs373418736	Ensembl.		1
Natural variant (VAR_003693)	70	E → K in MODY2; decreased affinity for glucose. 2 Publications			1
Natural variant (VAR_079440)	72	G → R in MODY2 and PNDM; decreased stability; no effect on glucokinase activity; no effect on affinity for glucose. 2 Publications Corresponds to variant dbSNP:rs193922289	Ensembl, ClinVar.		1

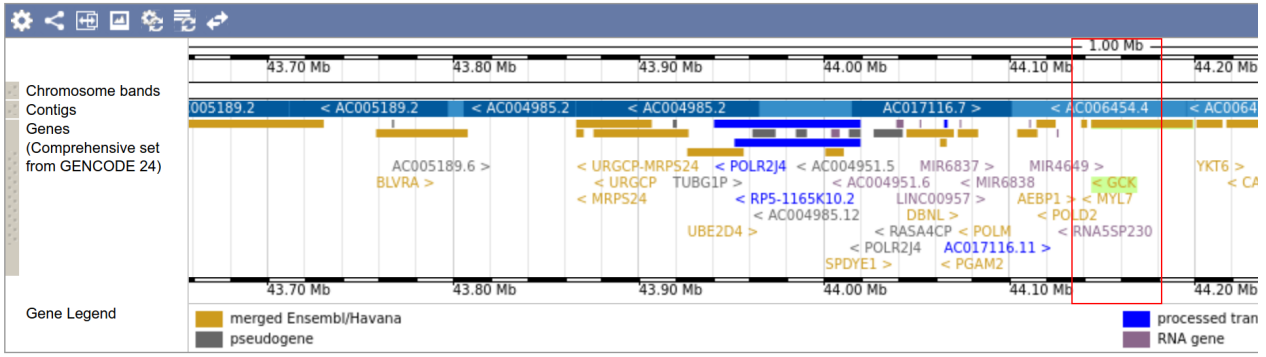
Nucleotide sequence databases

- **ensembl**
  https://www.ensembl.org/index.html
  - Ensembl is a genome browser for vertebrate genomes
- **Entrez**
  - provides integrated access to nucleotide and protein sequence data
  - Federated search engine of NCBI
- **GeneCards** (Human Gene Database)
  https://www.genecards.org
  - Integrative database that provides comprehensive information on annotated and predicted human genes
- **GeneBank**/RefSeq/UniGene
  - Genetic sequence database

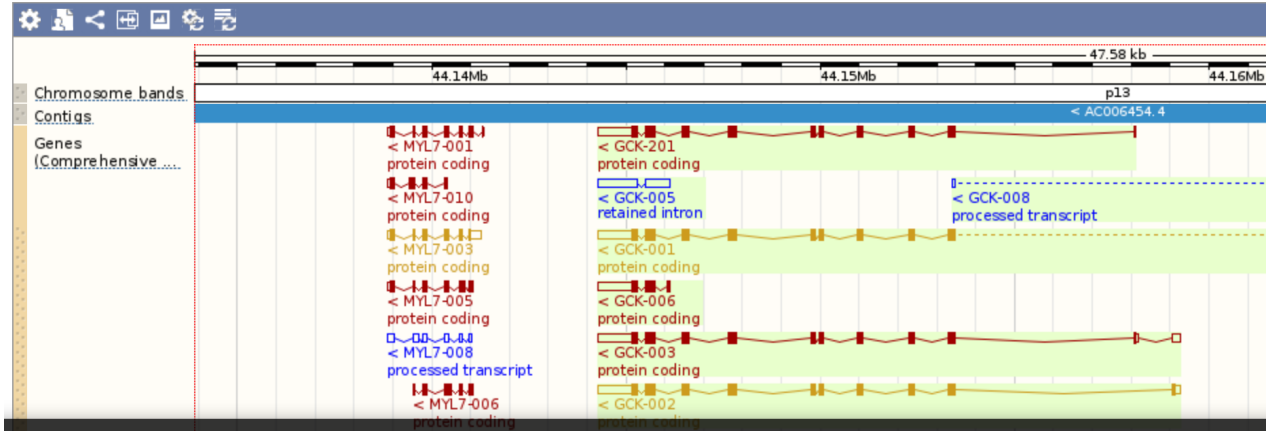
Chromosome 7: 44,133,886-44,181,469



Region in detail



Location: 7:44133886-44181469 Go Gene: Go



Genome Data Viewer

Homo sapiens: GRCh38.p13 (GCF_000001405.39) **Chr 7 (NC_000007.14): 44,138,590 - 44,194,062**

Reset All | Share this page | FAQ | Help | Browser Agreement | Version 4.8.4

We now support table downloads of gene annotation! Check out the Download menu on the graphical viewer toolbar!

Pick Assembly

Switch organism

GCF_000001405.39 (GRCh38.p13)

Locations for Gene GCK

Sequence	Location
NC_000007.14	44,143,213 - 44,189,439

Select an assembly to change view

Ideogram View

Unplaced/unlocalized scaffolds: 168

Alt loci/patches: 446

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

16 17 18 19 20 21 22 X Y MT

Search

Location, gene or phenotype

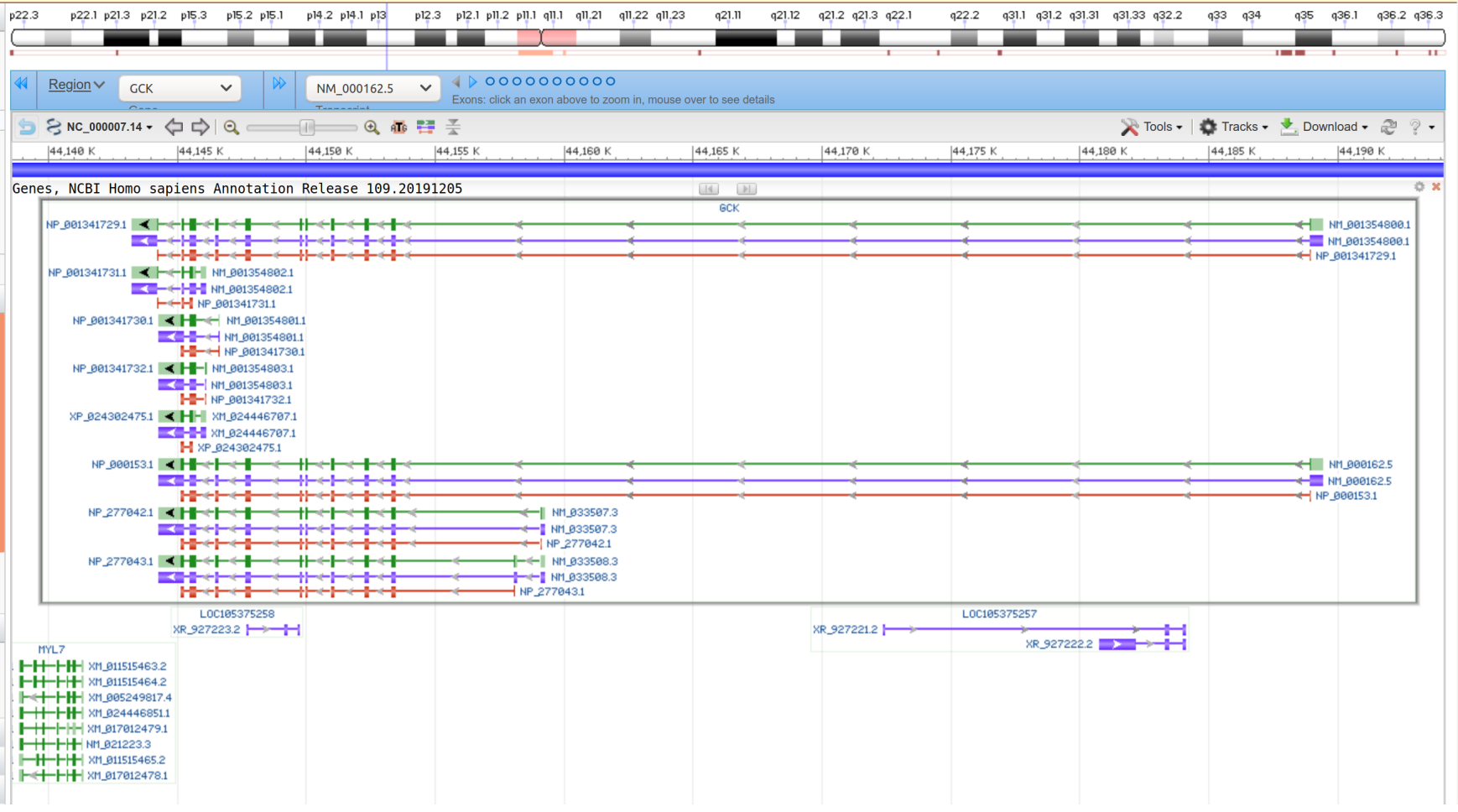
Enter a location, gene name or phenotype

Search examples:

User Data and Track Hubs

BLAST

Add Tracks



Protein-protein interaction databases

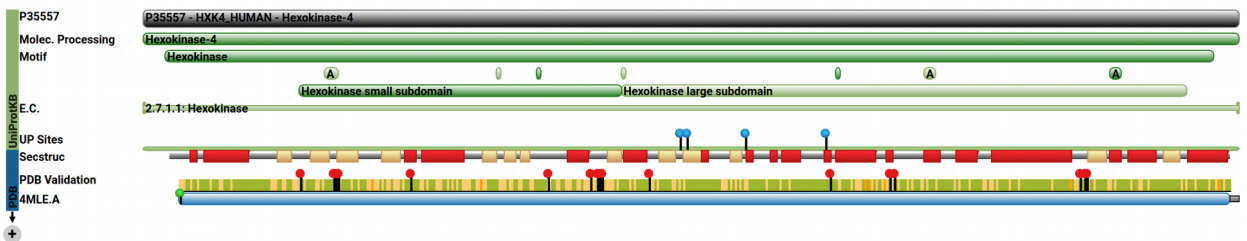
- **Intact**
  https://www.ebi.ac.uk/intact
  https://www.ebi.ac.uk/intact/interactors/id:P35557
  - Manually curated interaction database

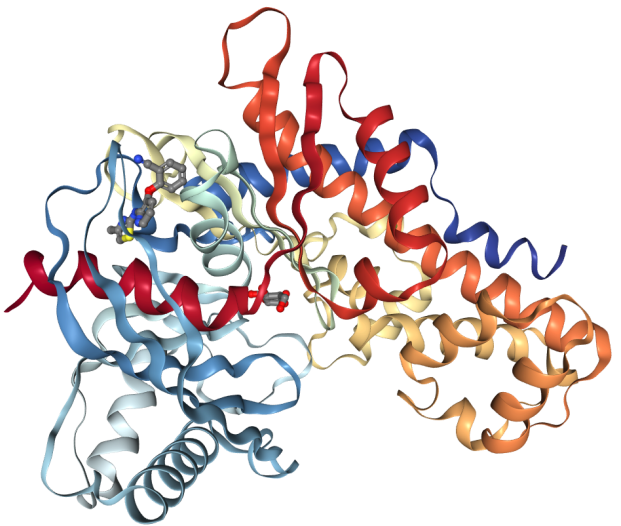
Customize view | Select format to Download | Download

	Dts	Molecule 'A'	Links 'A'	Molecule 'B'	Links 'B'	Interaction Detection Method	Interaction AC	Source Database
		GCK	P35557 EBI-709928	GCKR	Q14397 EBI-709948	phage display	EBI-710326	IntAct
						enzyme linked immunosorbent assay	EBI-710154	IntAct
						two hybrid prey pooling approach	EBI-23685784 **imex :** IM-25472-75812	IntAct
						two hybrid array	EBI-23978961 **imex :** IM-25472-94079	IntAct
						validated two hybrid	EBI-24669263 **imex :** IM-25472-147696	IntAct
		GCK	P35557 EBI-709928	SPDYE4	A6NLX3 EBI-12047907	two hybrid prey pooling approach	EBI-22322661 **imex :** IM-25472-666	IntAct
						two hybrid array	EBI-22781133 **imex :** IM-25472-25700	IntAct
						validated two hybrid	EBI-24280666 **imex :** IM-25472-111483	IntAct

Protein structures

- **PDB** (Protein database)
  https://www.rcsb.org/
  - 3D shapes of proteins, nucleic acids and complex assemblies





Compound and drug databases

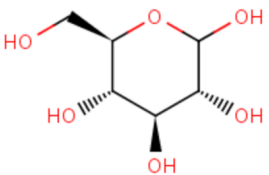
- **ChEBI**
  https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:4167
  - chemical entities of biological interest

- **ChEMBL**
  https://www.ebi.ac.uk/chembl/
  - Manually curated bioactive molecules

- **PubChem**
  https://pubchem.ncbi.nlm.nih.gov/
  - Quickly find chemical information

CHEBI:4167 - D-glucopyranose

Main | ChEBI Ontology | Automatic Xrefs | Reactions | Pathways | Models



ChEBI Name	**D-glucopyranose**
ChEBI ID	**CHEBI:4167**
ChEBI ASCII Name	D-glucopyranose
Definition	A glucopyranose having D-configuration.
Stars	⭐⭐⭐ This entity has been manually annotated by the ChEBI Team.
Supplier Information	eMolecules:711823, eMolecules:29536451, MolPort-021-782-999
Download	Molfile XML SDF

- Find compounds which contain this structure
- Find compounds which resemble this structure
- Take structure to the Advanced Search

more structures >>

Formula	C6H12O6
Net Charge	0
Average Mass	180.15588
Monoisotopic Mass	180.06339
InChI	InChI=1S/C6H12O6/c7-1-2-3(8)4(9)5(10)6(11)12-2/h2-11H,1H2/t2-,3-,4+,5-,6?/m1/s1
InChIKey	WQZGKKKJIJFFOK-GASJEMHNSA-N
SMILES	OC[C@H]1OC(O)[C@H](O)[C@@H](O)[C@@H]1O

Reaction databases

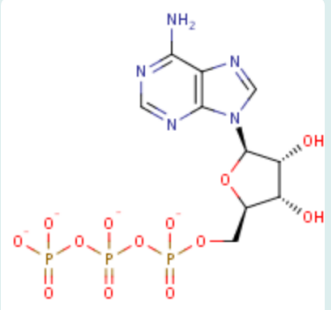
- **Rhea**
  https://www.rhea-db.org/
  - Expert curated source of biochemical reactions

RHEA:17825 (APPROVED)

ATP + D-glucose = ADP + D-glucose 6-phosphate + H(+)

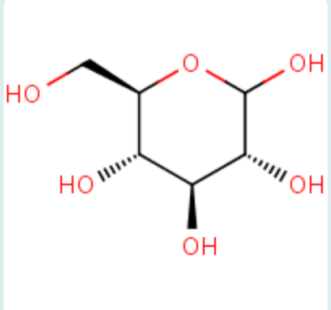
Last modified: 2019-11-04. **Chemically balanced:** yes.

ATP
CHEBI:30616
Formula: C10H12N5O13P3
Charge: -4



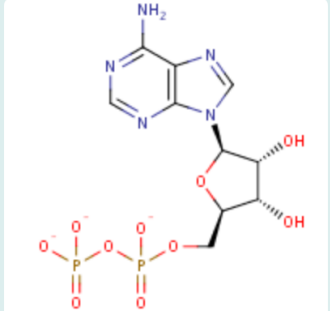
+

D-glucose
CHEBI:4167
Formula: C6H12O6
Charge: 0



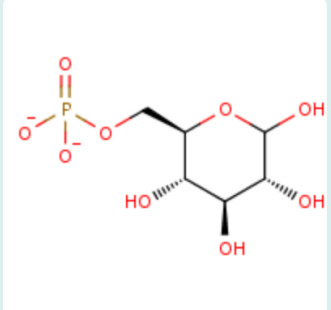
=

ADP
CHEBI:456216
Formula: C10H12N5O10P2
Charge: -3



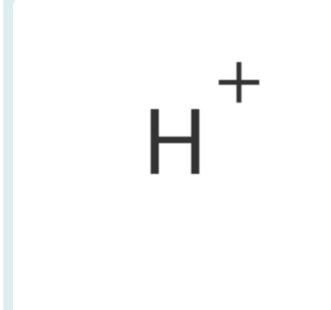
+

D-glucose 6-phosphate
CHEBI:61548
Formula: C6H11O9P
Charge: -2



+

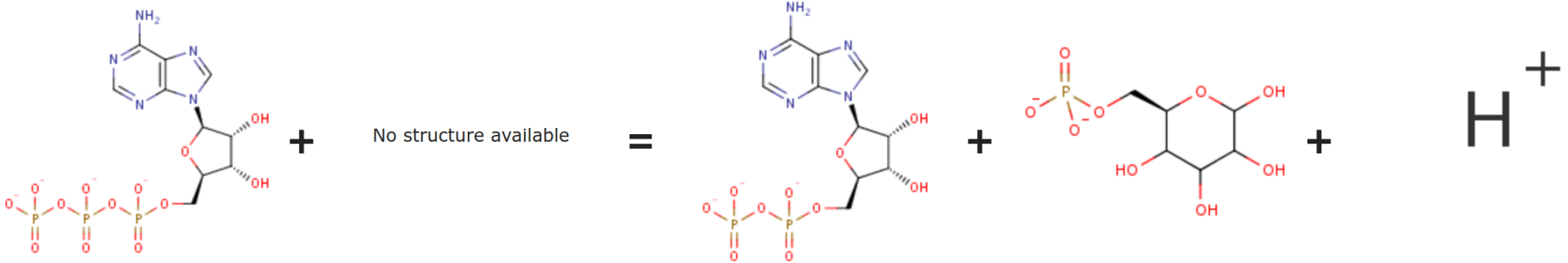
H+
CHEBI:15378
Formula: H
Charge: 1



Catalytic activity

- ATP + D-hexose = ADP + D-hexose 6-phosphate + H^+ 7 Publications
  EC:2.7.1.1 7 Publications
  This reaction proceeds in the forward 7 Publications direction.
  Source: Rhea. « Hide

ATP + D-hexose = ADP + D-hexose 6-phosphate + H^+



zoom

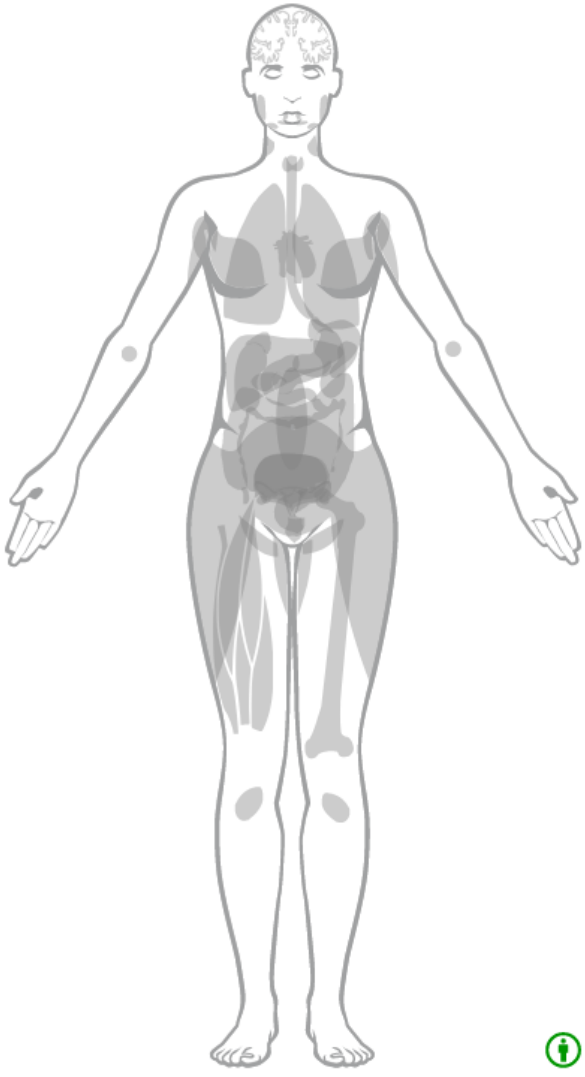
- ATP + D-fructose = ADP + D-fructose 6-phosphate + H^+ 1 Publication
  EC:2.7.1.1 1 Publication
  This reaction proceeds in the forward 1 Publication direction.
  Source: Rhea. Show »

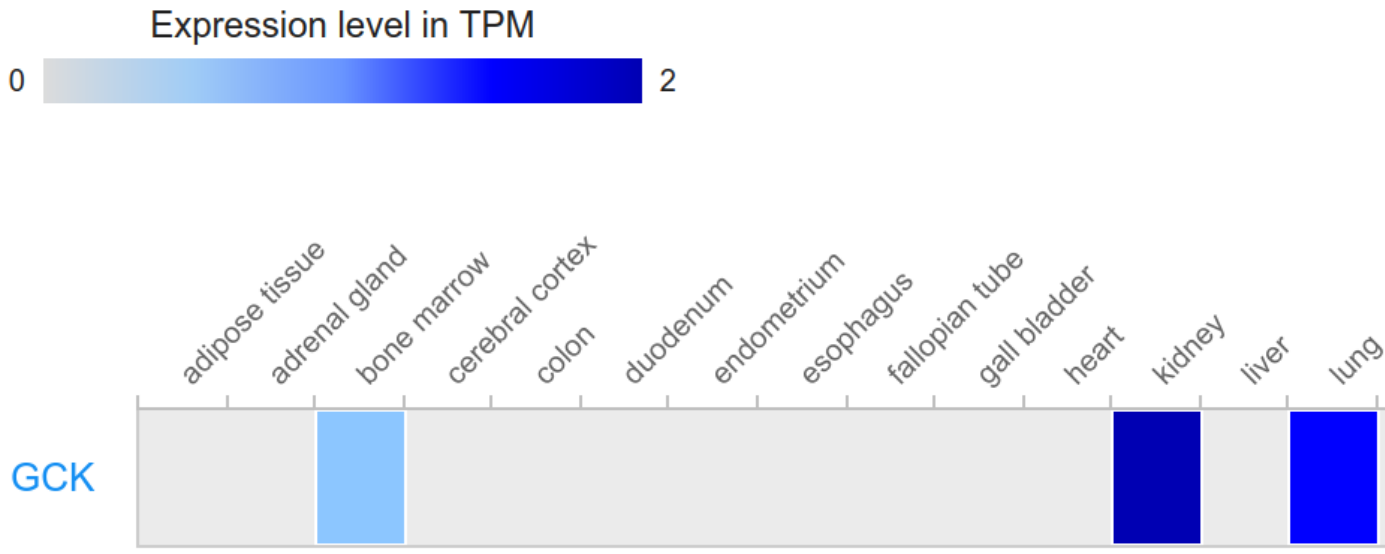
- ATP + D-glucose = ADP + D-glucose 6-phosphate + H^+ 1 Publication
  EC:2.7.1.1 1 Publication
  This reaction proceeds in the forward 1 Publication direction.
  Source: Rhea. Show »

- ATP + D-mannose = ADP + D-mannose 6-phosphate + H^+ 1 Publication
  EC:2.7.1.1 1 Publication
  This reaction proceeds in the forward 1 Publication direction.
  Source: Rhea. Show »

Microarray & Expression Databases

- **Expression atlas**
  https://www.ebi.ac.uk/gxa/home
- **ArrayExpress**
  https://www.ebi.ac.uk/arrayexpress/
  - functional genomics data
- **GTExPortal**
  https://www.gtexportal.org
- **GEO** (Gene Expression Omnibus)
  https://www.ncbi.nlm.nih.gov/geo/
  - Array and sequence based data




Expression level in TPM
0
2
adipose tissue
adrenal gland
bone marrow
cerebral cortex
colon
duodenum
endometrium
esophagus
fallopian tube
gall bladder
heart
kidney
liver
lung
GCK

Pathway databases

- **KEGG Pathways**
  https://www.genome.jp/kegg/pathway.html
  - collection of manually drawn pathway maps
  - part of larger KEGG database (compounds, genes, drugs, ...)
- **Reactome**
  https://reactome.org/
  - Free, open-source, curated and peer-reviewed pathway database
- **SMPDB** (small molecule pathway database)
  http://smpdb.ca/

Pathway meta-databases

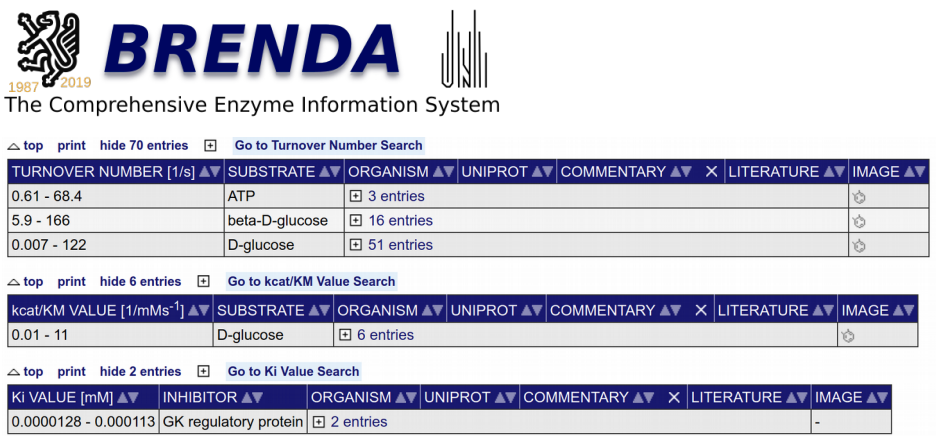
- **PathCards** (Pathway unification database)
  https://pathcards.genecards.org/
  - Integrated database of human biological pathways

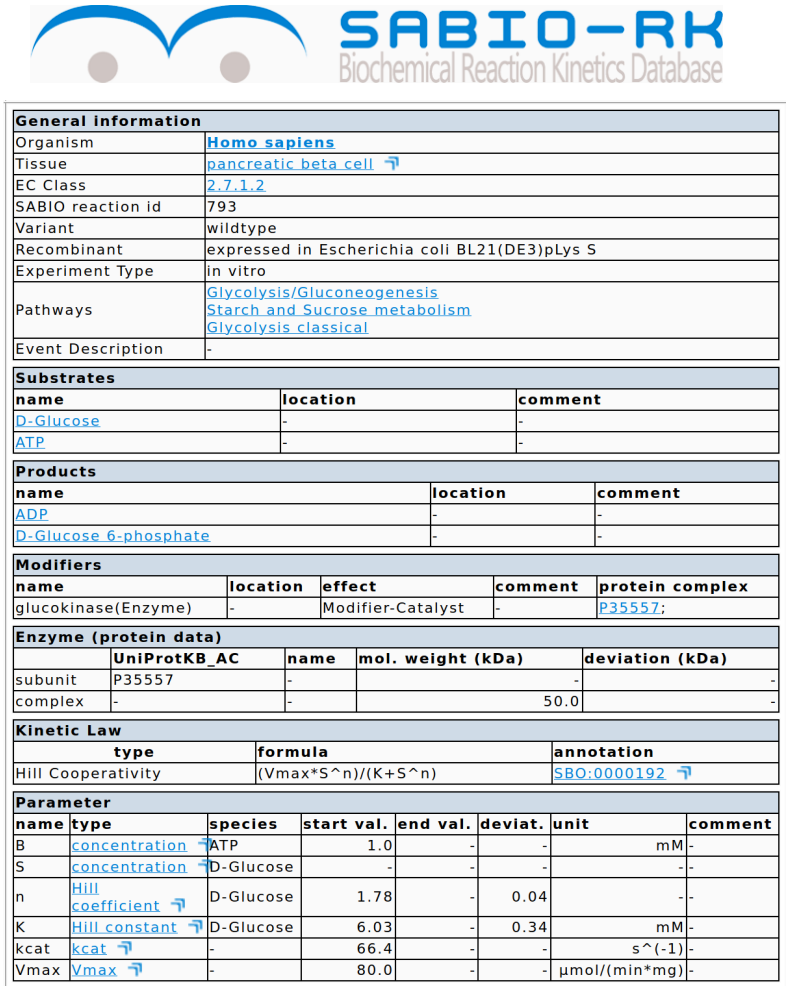
- **Pathway Commons**
  http://www.pathwaycommons.org
  - Collects information from other pathway databases and provides in a standardized format

- **WikiPathways**
  http://wikipathways.org
  - A community-driven collection of pathways that also includes pathways from other databases

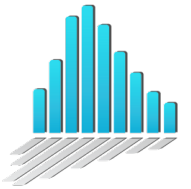
Enzyme Reaction Kinetics Databases

- **BRENDA**
  https://www.brenda-enzymes.org
  - enzyme information database
- **SABIO-RK**
  http://sabiork.h-its.org/
- reaction kinetics database





Metabolomics



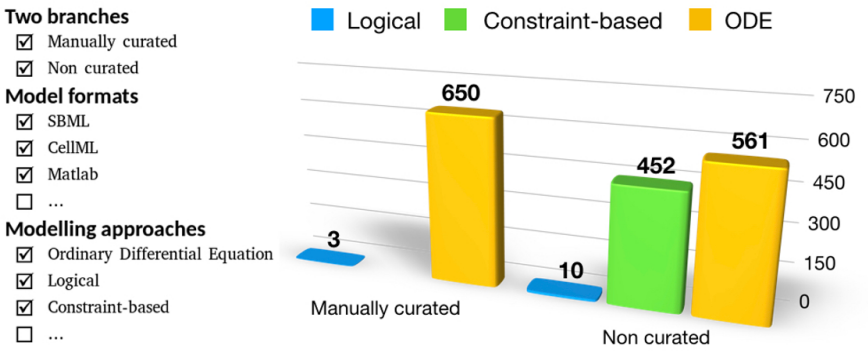
- **MetaboLights**
  https://www.ebi.ac.uk/metabolights/
  - database for Metabolomics experiments and derived information
- **HMDB** (Human metabolome database)
  http://www.hmdb.ca/
  - small molecule metabolites found in the human body

- **Metabolic Atlas**
  https://www.metabolicatlas.org/
  - Integrates genome-scale metabolic models for easy browsing

- **VMH** (Virtual metabolic human)
  https://www.vmh.life/
  - VMH database captures information on human metabolism

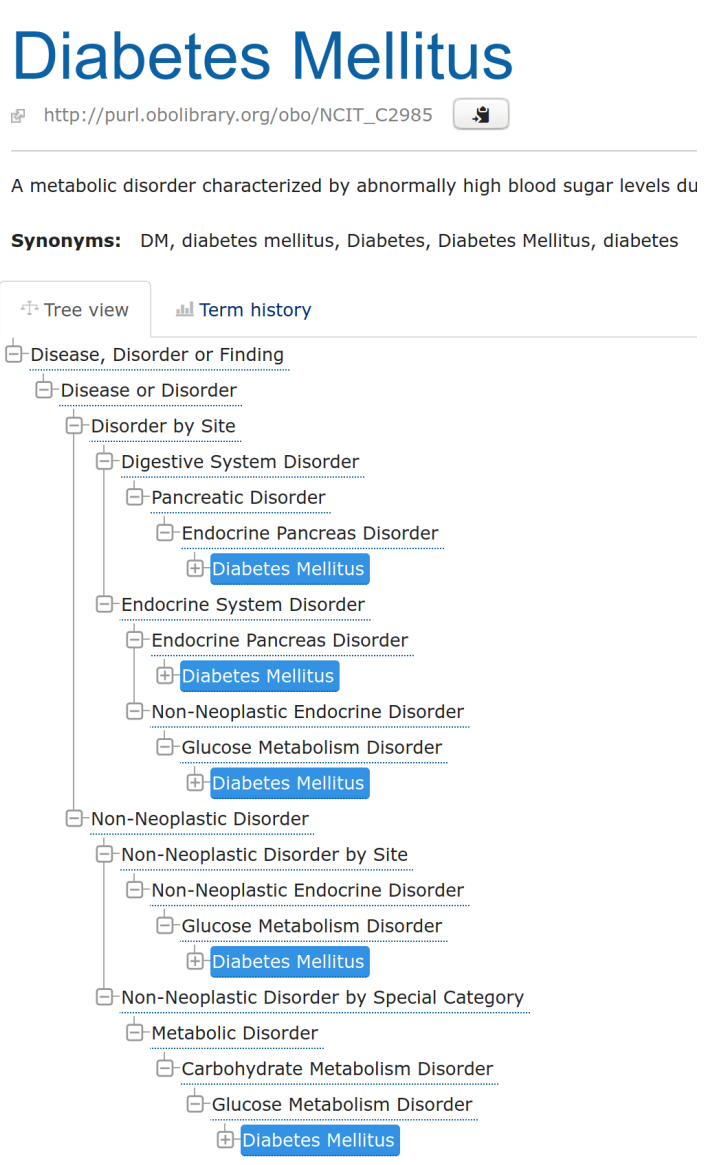
Model Collections

- **BioModels**
  https://www.ebi.ac.uk/biomodels/
  - SBML models (computation models)
- **Physiome model repository**
  https://models.physiomeproject.org/welcome
  - CellML models (curated)
- **BiGG database**
  http://bigg.ucsd.edu/
  - Genome-scale metabolic networks



Ontology database

- **The Gene Ontology** (GO)
  http://amigo.geneontology.org/amigo/gene_product/UniProtKB:P35557
  - Knowledge base on functions of genes
- **Ontology lookup service** (OLS)
  https://www.ebi.ac.uk/ols/index
  - Search in ontologies and terms



GO - Molecular function[i]

- carbohydrate binding Source: GO_Central
- enzyme binding Source: GO_Central
- enzyme inhibitor activity Source: GO_Central
- fructose-6-phosphate binding Source: UniProtKB

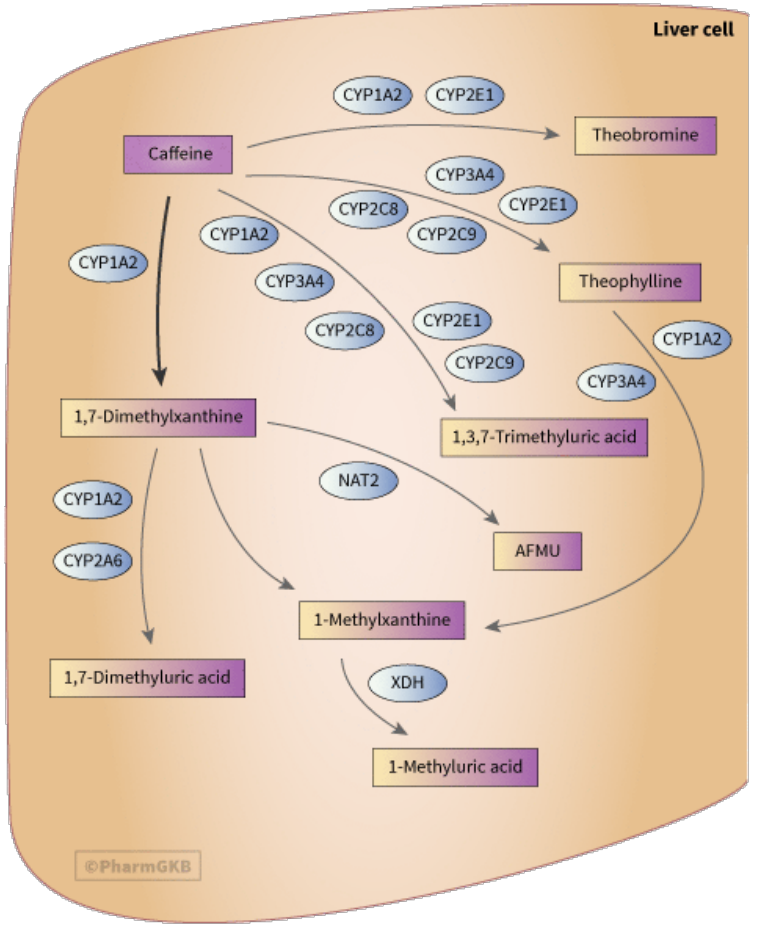
Complete GO annotation on QuickGO ...

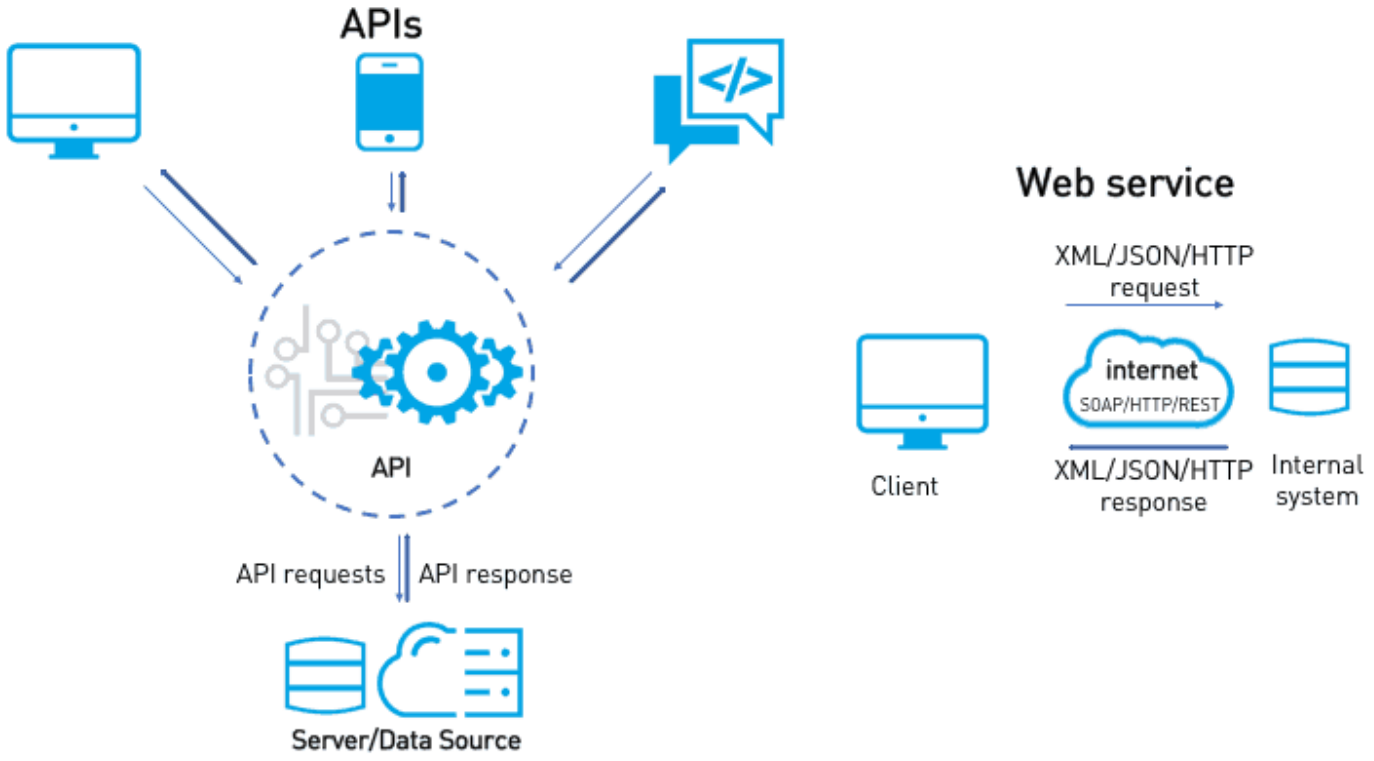
GO - Biological process[i]

- carbohydrate derivative metabolic process Source: InterPro
- carbohydrate metabolic process Source: UniProtKB-KW
- glucose homeostasis Source: GO_Central
- negative regulation of glucokinase activity Source: UniProtKB
- protein import into nucleus Source: BHF-UCL
- regulation of glycolytic process Source: Reactome
- response to fructose Source: BHF-UCL
- triglyceride homeostasis Source: BHF-UCL
- urate metabolic process Source: BHF-UCL

Pharmacokinetics

- **PHARMGKB**
  https://www.pharmgkb.org/
  - Annotated drugs, curated pathway, ...
- **PK-DB**
  https://pk-db.com
  - Pharmacokinetic data



Web services

- https://www.ebi.ac.uk/proteins/api/doc/