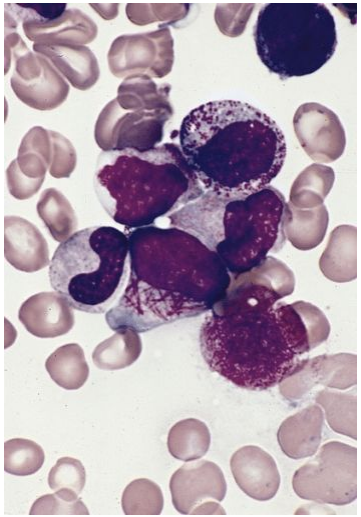
(Personal) use-cases for Figshare in a Wikidata narrative

Andra Waagmeester

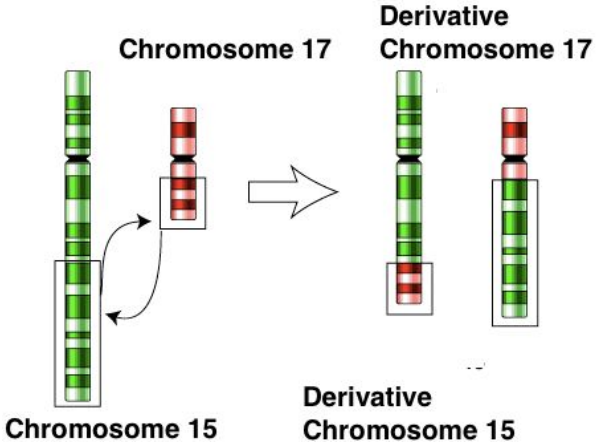
Micelio, Antwerp, Belgium | Email: andra@micelio.be, Twitter: @andrawaag

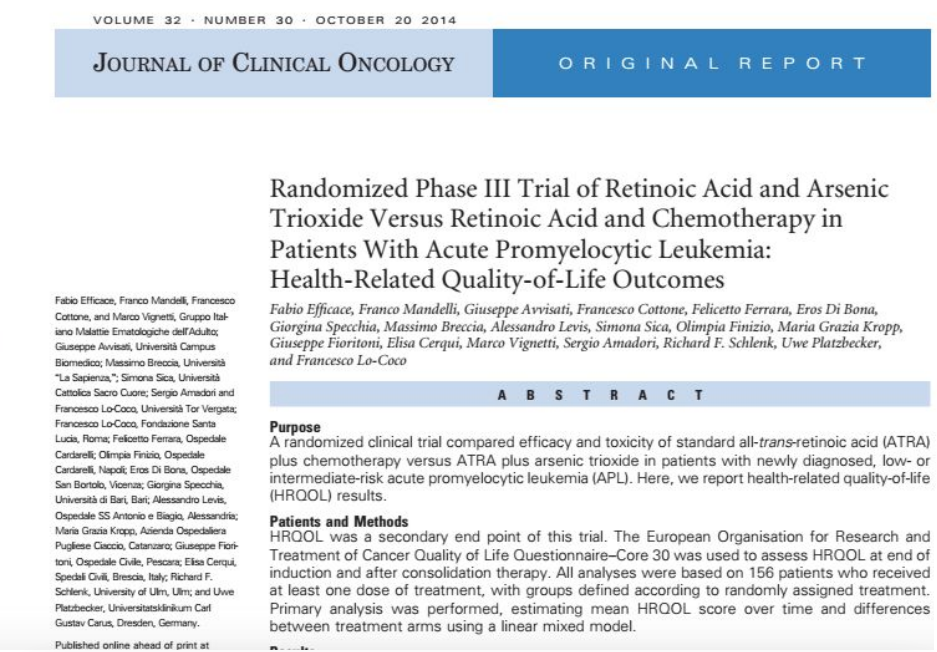
Acute promyelocytische leukemie

Cancer of the white blood cells



Chromosomal translocation on the RARα gene





VOLUME 32 · NUMBER 30 · OCTOBER 20 2014

JOURNAL OF CLINICAL ONCOLOGY

ORIGINAL REPORT

Randomized Phase III Trial of Retinoic Acid and Arsenic Trioxide Versus Retinoic Acid and Chemotherapy in Patients With Acute Promyelocytic Leukemia: Health-Related Quality-of-Life Outcomes

Fabio Efficace, Franco Mandelli, Giuseppe Avvisati, Francesco Cottone, Felicetto Ferrara, Eros Di Bona, Giorgina Specchia, Massimo Breccia, Alessandro Levis, Simona Sica, Olimpia Finizio, Maria Grazia Kropp, Giuseppe Fioritoni, Elisa Cerqui, Marco Vignetti, Sergio Amadori, Richard F. Schlenk, Uwe Platzbecker, and Francesco Lo-Coco

Fabio Efficace, Franco Mandelli, Francesco Cottone, and Marco Vignetti, Gruppo Italiano Malattie Ematologiche dell'Adulto; Giuseppe Avvisati, Università Campus Biomedico; Massimo Breccia, Università "La Sapienza,"; Simona Sica, Università Cattolica Sacro Cuore; Sergio Amadori and Francesco Lo-Coco, Università Tor Vergata; Francesco Lo-Coco, Fondazione Santa Lucia, Roma; Felicetto Ferrara, Ospedale Cardarelli; Olimpia Finizio, Ospedale Cardarelli, Napoli; Eros Di Bona, Ospedale San Bortolo, Vicenza; Giorgina Specchia, Università di Bari, Bari; Alessandro Levis, Ospedale SS Antonio e Biagio, Alessandria; Maria Grazia Kropp, Azienda Ospedaliera Pugliese Ciaccio, Catanzaro; Giuseppe Fioritoni, Ospedale Civile, Pescara; Elisa Cerqui, Spedali Civili, Brescia, Italy; Richard F. Schlenk, University of Ulm, Ulm; and Uwe Platzbecker, Universitatsklinikum Carl Gustav Carus, Dresden, Germany.

Published online ahead of print at

A B S T R A C T

Purpose

A randomized clinical trial compared efficacy and toxicity of standard all-*trans*-retinoic acid (ATRA) plus chemotherapy versus ATRA plus arsenic trioxide in patients with newly diagnosed, low- or intermediate-risk acute promyelocytic leukemia (APL). Here, we report health-related quality-of-life (HRQOL) results.

Patients and Methods

HRQOL was a secondary end point of this trial. The European Organisation for Research and Treatment of Cancer Quality of Life Questionnaire–Core 30 was used to assess HRQOL at end of induction and after consolidation therapy. All analyses were based on 156 patients who received at least one dose of treatment, with groups defined according to randomly assigned treatment. Primary analysis was performed, estimating mean HRQOL score over time and differences between treatment arms using a linear mixed model.

Effects of arsenic trioxide known for decades in China

Original papers were published in the Chinese language and in journals that are obscure even to most Chinese readers

SCIENCE CHINA
Life Sciences

• REVIEW •

June 2013 Vol.56 No.6: 495–502
doi: 10.1007/s11427-013-4487-z

Findings support the use of retinoic acid plus arsenic trioxide as prefered first-line treatment

A drug from poison: how the therapeutic effect of arsenic trioxide on acute promyelocytic leukemia was discovered

RAO Yi[1*], LI RunHong[2] & ZHANG DaQing[2]

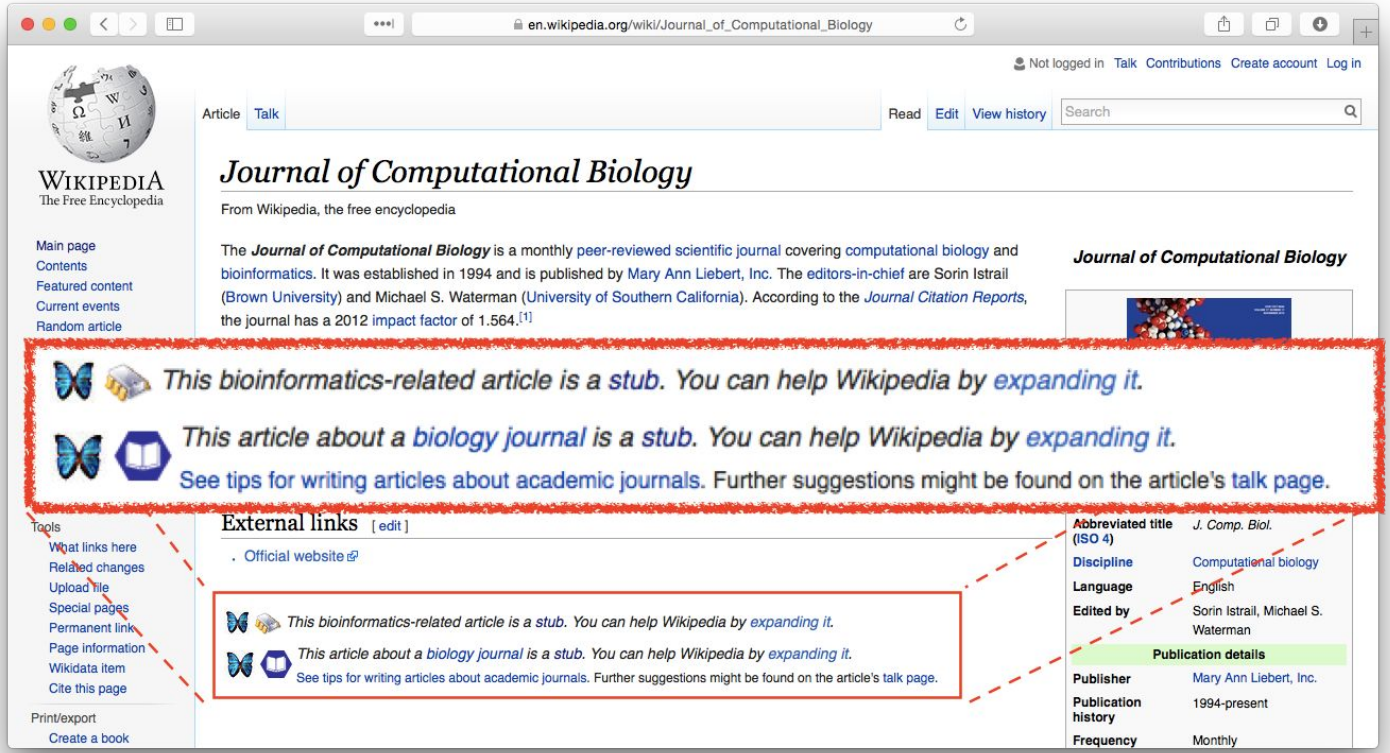
[1]*Peking-Tsinghua Center for Life Sciences at Peking University School of Life Sciences, Beijing 100871, China;*
[2]*Peking University Health Sciences Center, Beijing 100871, China*

Received March 27, 2013; accepted April 5, 2013; published online May 3, 2013

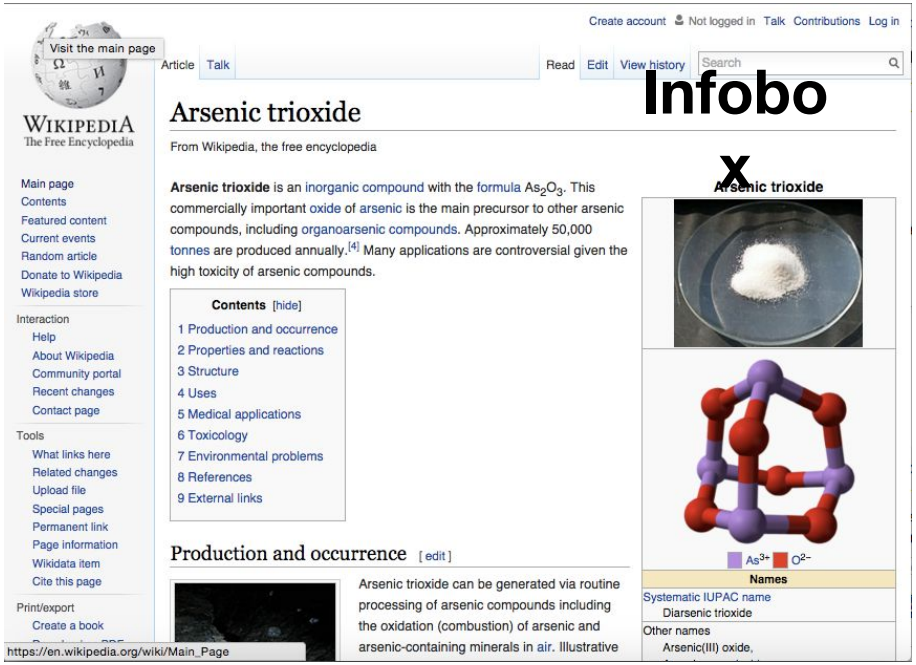
Source: http://jco.ascopubs.org/content/early/2014/09/17/JCO.2014.55.3453

Folk tale of the stone soup

Een Wikipedia artikel wordt net zo bereid als de stenen soep



Multilingual Wikipedia & Infoboxes

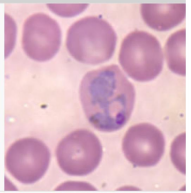


Language independent info boxes

Dutch

nl.m.wikipedia.org

Malaria



Rode bloedcel geïnfecteerd met *P. vivax*

Coderingen

ICD-10	B50
ICD-9	084
OMIM	248310

Greek

el.m.wikipedia.org

Ελονοσία

Ταξινόμηση και εξωτερικές πηγές

Ταξινόμηση ICD-10	B50
Ταξινόμηση ICD-9	084
OMIM	248310
DiseasesDB	7728
MedlinePlus	000621
eMedicine	med/1385 emerg/305 ped/1357

English

en.m.wikipedia.org

Malaria

A *Plasmodium* from the saliva of a female mosquito moving across a mosquito cell

Classification and external resources

Specialty	Infectious disease
ICD-10	B50-B54
ICD-9-CM	084

ICD-10 B50

ICD-10 B50-B54

Ταξινόμηση ICD-10 B50

Dutch

nl.m.wikipedia.org

	Papiamento
Hoofdstad	Oranjestad
Regeringsvorm	Constitutionele monarchie
Staatshoofd	Koning Willem-Alexander Fredis Refunjol (gouverneur)
Regeringsleider	Mike Eman (Arubaanse Volkspartij)
Religie	Katholiek 82%, protestant 8%

103,400 [2] (197th)

Greek

el.m.wikipedia.org

Πολίτευμα	Συνταγματική Μοναρχία
Μονάρχης Κυβερνήτης Πρωθυπουργός	Γουλιέλμος-Αλέξανδρος Φρέντις Ρεφουνιόλ Μάικ Έμαν
Πλήρης αυτονομία από το Βασίλειο των Κάτω Χωρών Σύνταγμα	1η Ιανουαρίου 1986
Έκταση	
• Σύνολο Ακτογραμμή	180 km² (213η) 68,5 km
Πληθυσμός	
• Εκτίμηση 2014 • Απογραφή 2000 • Πυκνότητα	107.394[1] (196η) 103.065 556,4 κατ./km² (21
Α.Ε.Π. (PPP) • Ολικό (2005)	2,258 δισ. $[2]

• Εκτίμηση 2014 107.394[1] (196η)

English

pap.m.wikipedia.org

Forma di gobernacion	Democracia p Monarkia con
- Rei	Willem-Alexan
- Gobernador	Fredis Refunjo
- Prome Minister	Mike Eman
Pais den Reino di Hulanda	
Status aparte	1 januari di 19
Area	
- Total	193 km² (n/a)

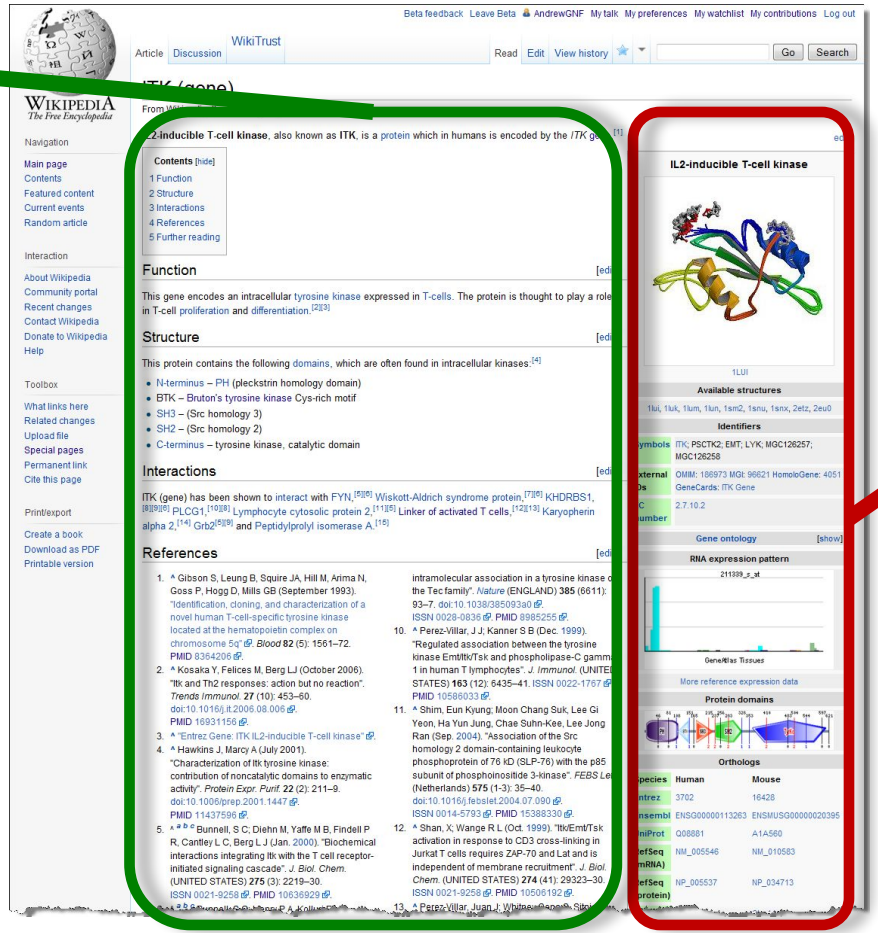
101.484 (2010)[2]

110.663 (2014)[3]

(614,8/km² (2014))

The Gene Wiki project, circa 2008

Summarized knowledge via crowdsourcing



Data imported from structured databases

Huss, PLoS Biol, 2008

Reelin

From Wikipedia, the free encyclopedia

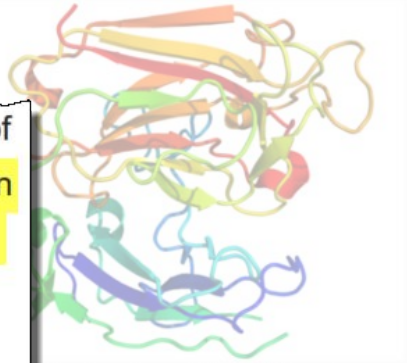
Reelin is a large secreted extracellular matrix glycoprotein that helps regulate processes of neuronal migration and positioning in the developing brain by controlling cell–cell interactions. Besides this important role in early development, reelin continues to work in the adult brain. It modulates synaptic plasticity by [2][3] It also stimulates dendrite[4] migration of neuroblasts generat zones. It is found not only in the tissues.

Reelin has been suggested to be expression of the protein has be bipolar disorder, but the cause of this observation remains uncertain as studies show that psychotropic medication itself affects reelin expression. Moreover, epigenetic hypotheses aimed at explaining the changed levels of reelin expression[6] are controversial.[7][8] Total lack of reelin causes a form of lissencephaly. Reelin may also play a role in Alzheimer's disease, temporal lobe epilepsy and autism.

Reelin's name comes from the abnormal reeling gait of *reeler* mice,[9] which were later found to have a deficiency of this brain protein and were homozygous for mutation of the RELN gene. The

Reelin has been suggested to be implicated in pathogenesis of several brain diseases. The expression of the protein has been found to be significantly lower in schizophrenia and psychotic bipolar disorder, but the cause of this observation remains uncertain as studies show that psychotropic medication itself

Reelin



hic structure of the third reelin repeat domain.[1]

Available structures	
PDB	Ortholog search: PDBe, RCSB
	List of PDB id codes [show]

Identifiers	
Symbols	RELN ; LIS2; PRO1598; RL
External	OMIM: 600514 MGI: 103022

is to text

is to data

biomedical

> Provide a database of the world's knowledge that anyone can edit
>
> - Denny Vrandečić

Reelin - Wikipedia

Andra

Secure | https://en.wikipedia.org/wiki/Reelin

Andrawaag 21 99+ Talk Sandbox Preferences Beta Watchlist Contributions Log out

Article Talk | Read Edit View history More | Search Wikipedia

WIKIPEDIA
The Free Encyclopedia

Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store

Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page

Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page

Print/export
Create a book
Download as PDF
Printable version

Reelin

From Wikipedia, the free encyclopedia

Reelin (RELN)[5] is a large secreted extracellular matrix glycoprotein that helps regulate processes of neuronal migration and positioning in the developing brain by controlling cell-cell interactions. Besides this important role in early development, reelin continues to work in the adult brain. It modulates synaptic plasticity by enhancing the induction and maintenance of long-term potentiation.[6][7] It also stimulates dendrite[8] and dendritic spine[9] development and regulates the continuing migration of neuroblasts generated in adult neurogenesis sites like subventricular and subgranular zones. It is found not only in the brain, but also in the spinal cord, blood, and other body organs and tissues.[*citation needed*]

Reelin has been suggested to be implicated in pathogenesis of several brain diseases. The expression of the protein has been found to be significantly lower in schizophrenia and psychotic bipolar disorder,[10] but the cause of this observation remains uncertain as studies show that psychotropic medication itself affects reelin expression. Moreover, epigenetic hypotheses aimed at explaining the changed levels of reelin expression[11] are controversial.[12][13] Total lack of reelin causes a form of lissencephaly. Reelin may also play a role in Alzheimer's disease, temporal lobe epilepsy and autism.[*citation needed*]

Reelin's name comes from the abnormal reeling gait of *reeler* mice,[14] which were later found to have a deficiency of this brain protein and were homozygous for mutation of the RELN gene. The primary phenotype associated with loss of reelin function is a failure of neuronal positioning throughout the developing central nervous system (CNS). The mice heterozygous for the reelin gene, while having little neuroanatomical defects, display the endophenotypic traits linked to psychotic disorders.[15]

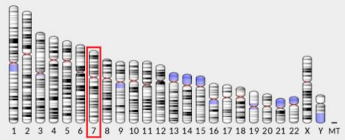
Contents [hide]

RELN

Available structures	
PDB	Ortholog search: PDBe RCSB
	List of PDB id codes [show]

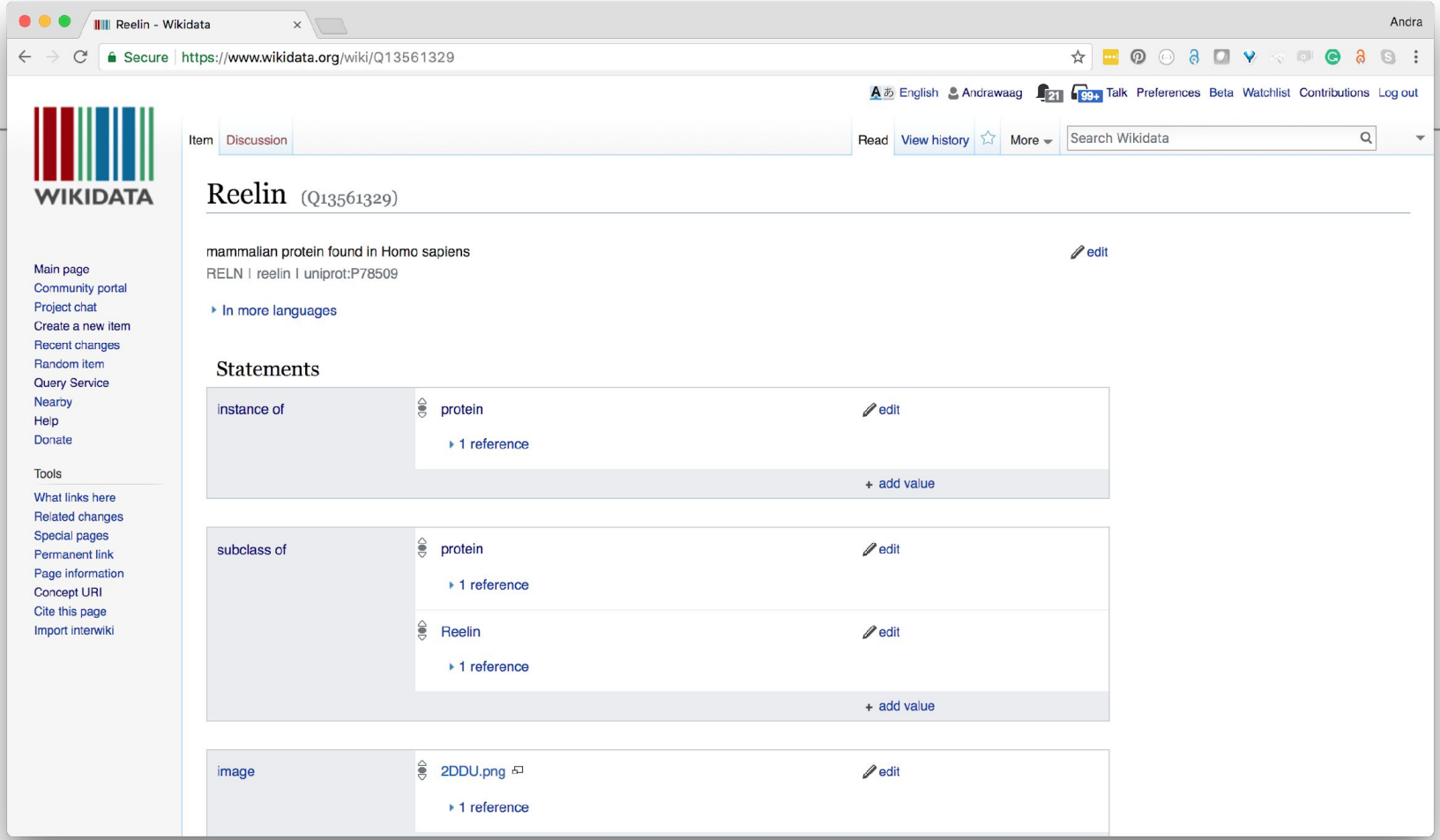
Identifiers	
Aliases	RELN, LIS2, PRO1598, RL, reelin, ETL7
External IDs	OMIM: 600514 MGI: 103022 HomoloGene: 3699 GeneCards: RELN

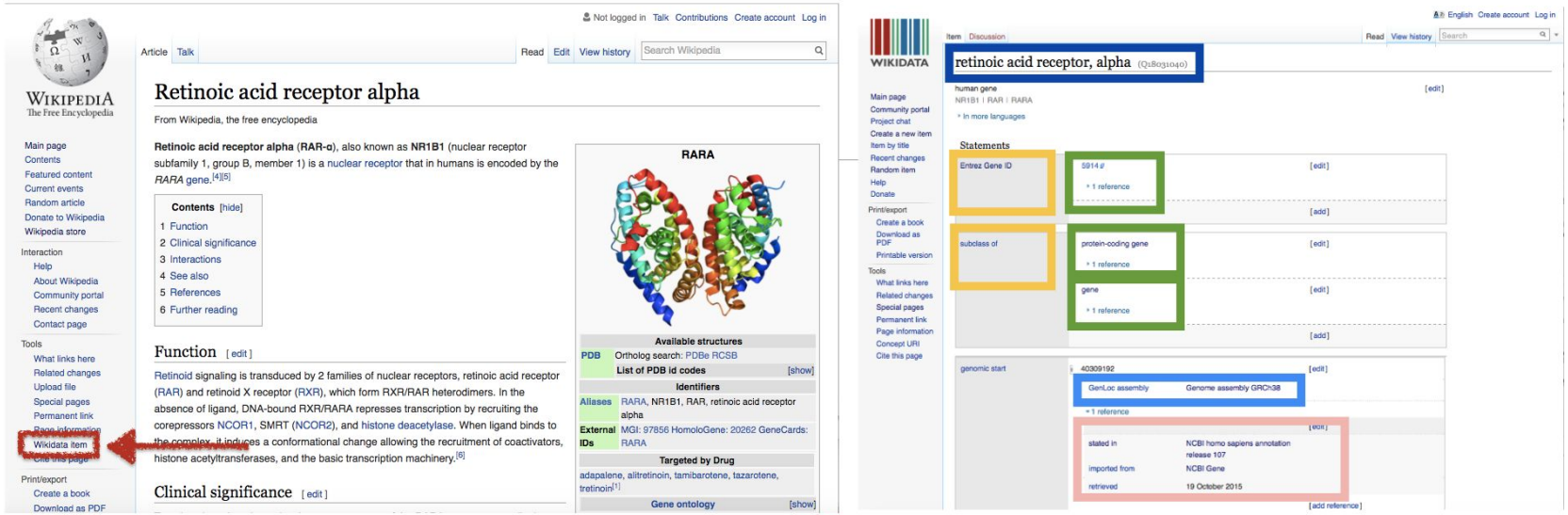
Gene location (Human) [hide]

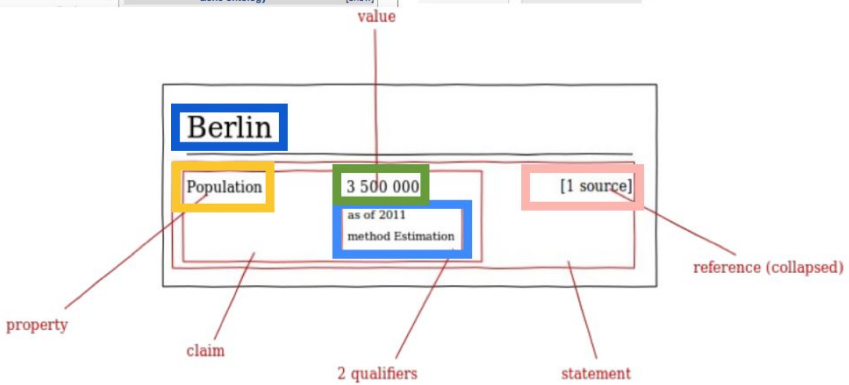


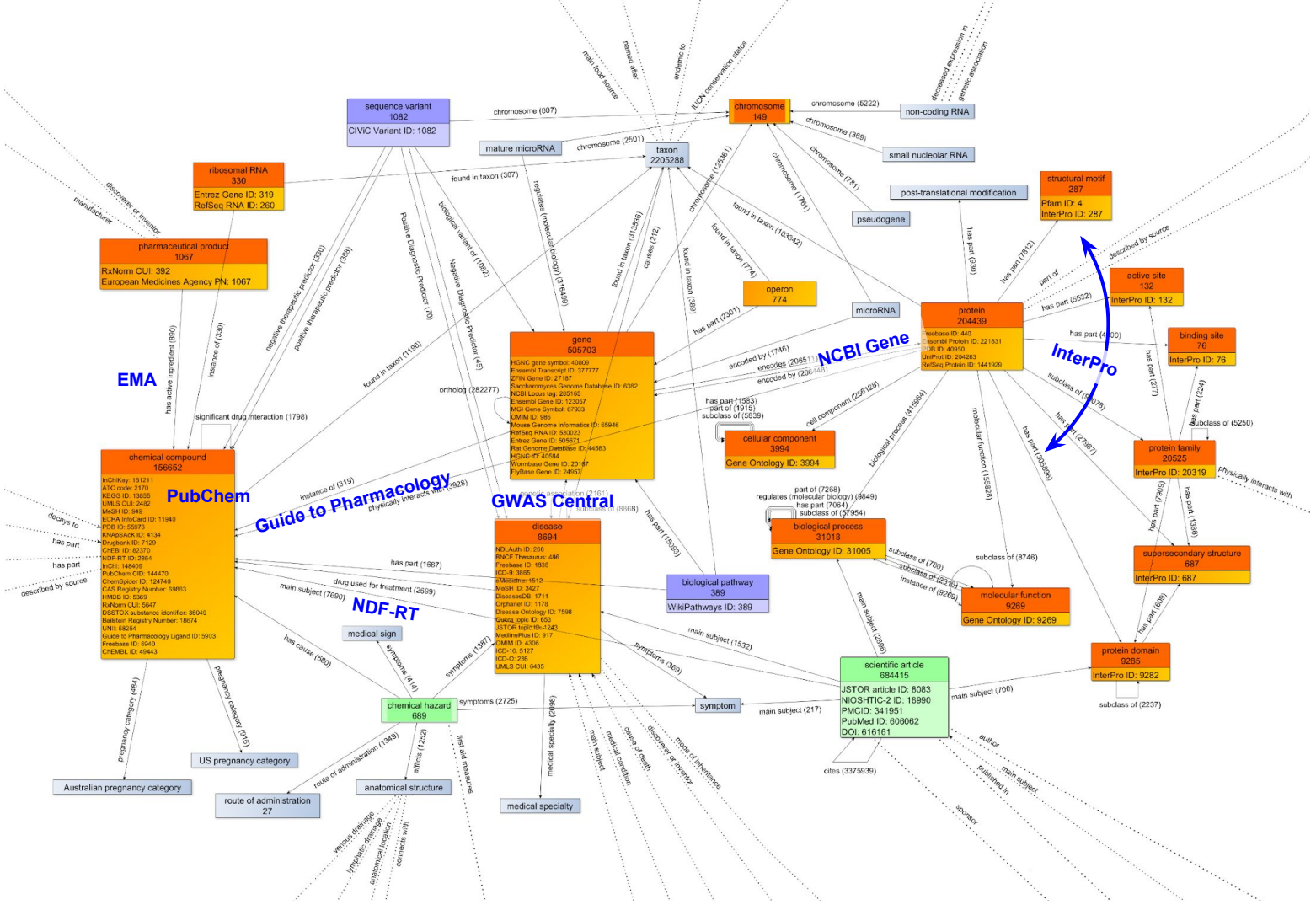
Chr. Chromosome 7 (human)[1]

https://en.wikipedia.org/wiki/Reelin#Psychotropic_medication


Reelin - Wikidata
Andra
Secure https://www.wikidata.org/wiki/Q13561329
English Andrawaag 21 99+ Talk Preferences Beta Watchlist Contributions Log out
WIKIDATA
Item Discussion
Read View history More
Search Wikidata
Reelin (Q13561329)
mammalian protein found in Homo sapiens
edit
RELN | reelin | uniprot:P78509
In more languages
Main page
Community portal
Project chat
Create a new item
Recent changes
Random item
Query Service
Nearby
Help
Donate
Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page
Import interwiki
Statements
instance of
protein
edit
1 reference
add value
subclass of
protein
edit
1 reference
Reelin
edit
1 reference
add value
image
2DDU.png
edit
1 reference

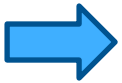

Not logged in Talk Contributions Create account Log in
Article Talk
Read Edit View history
Search Wikipedia
Retinoic acid receptor alpha
From Wikipedia, the free encyclopedia
Retinoic acid receptor alpha (RAR-α), also known as NR1B1 (nuclear receptor subfamily 1, group B, member 1) is a nuclear receptor that in humans is encoded by the RARA gene.[4][5]
Contents [hide]
1 Function
2 Clinical significance
3 Interactions
4 See also
5 References
6 Further reading
Function [edit]
Retinoid signaling is transduced by 2 families of nuclear receptors, retinoic acid receptor (RAR) and retinoid X receptor (RXR), which form RXR/RAR heterodimers. In the absence of ligand, DNA-bound RXR/RARA represses transcription by recruiting the corepressors NCOR1, SMRT (NCOR2), and histone deacetylase. When ligand binds to the complex, it induces a conformational change allowing the recruitment of coactivators, histone acetyltransferases, and the basic transcription machinery.[6]
Clinical significance [edit]
RARA
Available structures
PDB Ortholog search: PDBe RCSB
List of PDB id codes [show]
Identifiers
Aliases RARA, NR1B1, RAR, retinoic acid receptor alpha
External IDs MGI: 97856 HomoloGene: 20262 GeneCards: RARA
Targeted by Drug
adapalene, alitretinoin, tamibarotene, tazarotene, tretinoin[1]
Gene ontology [show]
Main page
Contents
Featured content
Current events
Random article
Donate to Wikipedia
Wikipedia store
Interaction
Help
About Wikipedia
Community portal
Recent changes
Contact page
Tools
What links here
Related changes
Upload file
Special pages
Permanent link
Page information
Wikidata item
Cite this page
Print/export
Create a book
Download as PDF
WIKIDATA
Item Discussion
Read View history Search
retinoic acid receptor, alpha (Q18031040)
human gene
NR1B1 | RAR | RARA
In more languages
Statements
Entrez Gene ID
5914 #
1 reference
subclass of
protein-coding gene
1 reference
gene
1 reference
genomic start
40309192
GenLoc assembly
Genome assembly GRCh38
1 reference
stated in
NCBI homo sapiens annotation release 107
imported from
NCBI Gene
retrieved
19 October 2015
[add reference]
Main page
Community portal
Project chat
Create a new item
Item by title
Recent changes
Random item
Help
Donate
Print/export
Create a book
Download as PDF
Printable version
Tools
What links here
Related changes
Special pages
Permanent link
Page information
Concept URI
Cite this page


value
Berlin
Population
3 500 000
as of 2011
method Estimation
[1 source]
property
claim
2 qualifiers
statement
reference (collapsed)


sequence variant
1082
CIViC Variant ID: 1082
chromosome
149
chromosome (5222)
non-coding RNA
chromosome (807)
chromosome (369)
small nucleolar RNA
mature microRNA
chromosome (2501)
taxon
2205288
chromosome (781)
pseudogene
found in taxon (307)
ribosomal RNA
330
Entrez Gene ID: 319
RefSeq RNA ID: 260
chromosome (1761)
post-translational modification
found in taxon (103342)
structural motif
287
Pfam ID: 4
InterPro ID: 287
pharmaceutical product
1067
RxNorm CUI: 392
European Medicines Agency PN: 1067
found in taxon (714)
operon
774
has part (930)
has part (7812)
active site
132
InterPro ID: 132
has part (5532)
microRNA
protein
204439
Refbase ID: 440
Ensembl Protein ID: 221831
UCJB ID: 40950
UniProt ID: 294263
RefSeq Protein ID: 1441929
has part (4430)
binding site
76
InterPro ID: 76
has part (277)
has part (224)
gene
505703
HGNC gene symbol: 40809
Ensembl Transcript ID: 377777
ZFIN Gene ID: 27187
Saccharomyces Genome Database ID: 6382
NCBI Locus tag: 285165
Ensembl Gene ID: 123057
MGI Gene Symbol: 67933
OMIM ID: 986
Mouse Genome Informatics ID: 65946
RefSeq RNA ID: 830023
Entrez Gene ID: 505671
Rat Genome Database ID: 44583
HGNC ID: 40594
Wormbase Gene ID: 20167
FlyBase Gene ID: 24957
encoded by (1746)
encodes (20651)
encoded by (20645)
has part (1583)
part of (1915)
subclass of (5839)
cellular component
3994
Gene Ontology ID: 3994
cell component (256128)
biological process (159864)
molecular function (153928)
has part (27987)
has part (305686)
subclass of (5250)
protein family
20525
InterPro ID: 20319
physically interacts with
chemical compound
156652
InChIKey: 151211
ATC code: 2170
KEGG ID: 13855
UMLS CUI: 2482
MeSH ID: 949
ECHA InfoCard ID: 11940
PDB ID: 55673
KNApSAcK ID: 4134
Drugbank ID: 7129
ChEBI ID: 82370
NDF-RT ID: 2864
InChI: 146409
PubChem CID: 144470
ChemSpider ID: 124740
CAS Registry Number: 69883
HMDB ID: 5369
RxNorm CUI: 5647
DSSTOX substance identifier: 36049
Beilstein Registry Number: 18674
UNII: 58254
Guide to Pharmacology Ligand ID: 5903
Freebase ID: 6940
ChEMBL ID: 49443
PubChem
Guide to Pharmacology
NCBI Gene
InterPro
EMA
GWAS Central
NDF-RT
part of (7268)
regulates (molecular biology) (8649)
has part (7064)
subclass of (57954)
biological process
31018
Gene Ontology ID: 31005
subclass of (780)
subclass of (8746)
instance of (2310)
molecular function
9269
Gene Ontology ID: 9269
supersecondary structure
687
InterPro ID: 687
has part (7409)
has part (1365)
has part (605)
disease
8694
NDLAuth ID: 286
BNCF Thesaurus: 496
Freebase ID: 1936
ICD-9: 3855
eMedicine: 1512
MeSH ID: 3427
DiseasesDB: 1711
Orphanet ID: 1178
Disease Ontology ID: 7598
Gwoza topic ID: 653
JSTOR topic ID: 1243
MedlinePlus ID: 917
OMIM ID: 4306
ICD-10: 5127
ICD-O: 238
UMLS CUI: 6435
biological pathway
389
WikiPathways ID: 389
protein domain
9285
InterPro ID: 9282
subclass of (2237)
medical sign
scientific article
684415
JSTOR article ID: 8083
NIOSHTIC-2 ID: 18990
PMCID: 341951
PubMed ID: 606062
DOI: 616161
chemical hazard
689
symptom
cites (3375939)
US pregnancy category
Australian pregnancy category
route of administration
27
anatomical structure
medical specialty

Simple data retrieval

"Retrieve genes with GWAS association with asthma"



39 genes

gene	geneLabel		gene	geneLabel		gene	geneLabel		gene	geneLabel
Q5013317	COL22A1		Q18027370	IGSF3		Q18053559	CDHR3		Q14903974	SMAD3
Q14912759	SLC22A5		Q18045382	HPSE2		Q18045669	ATG3		Q18033889	IL1RL1
Q14914243	PSAP		Q18048437	IL33		Q18035037	RAD50		Q17917202	ERBB4
Q14907990	SLC30A8		Q18051900	PYHIN1		Q18036984	FBXL7		Q18027836	IL6R
Q18025002	GAB1		Q17709208	ACO1		Q18033919	XPR1		Q18030185	NOTCH4
Q18035589	C6orf10		Q18027822	IL2RB		Q15326496	RORA		Q18030409	PDE4D
Q18054256	GSDMA		Q18030364	PBX2		Q18042132	GSDMB		Q18045645	IKZF4
Q18058487	C5orf56		Q18037773	ABI3BP		Q18029145	MKLN1		Q18039979	KLHL5
Q18030785	PRKG1		Q18039623	CTNNA3		Q18036729	RAP1GAP2		Q18026947	HLA-DQA1
Q18033424	IL18R1		Q18046350	ZNF665		Q14878303	IL13			

```
SELECT DISTINCT ?gene ?geneLabel where {
  ?gene wdt:P2293 wd:Q35869 .  # gene has genetic association to "asthma"
  ?gene wdt:P31 wd:Q7187 .     # gene is subclass of "gene"
  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
```

http://bit.ly/bosc2017_wikidata

Data integration

"Retrieve genes with GWAS association with asthma and gene product is localized to membrane"

22 genes

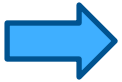
gene	geneLabel	gene	geneLabel	gene	geneLabel	gene	geneLabel
Q14912759	SLC22A5	Q18027370	IGSF3	Q18035037	RAD50	Q18027836	IL6R
Q14914243	PSAP	Q18033424	IL18R1	Q18033919	XPR1	Q18030409	PDE4D
Q14907990	SLC30A8	Q18045382	HPSE2	Q18042132	GSDMB	Q18030185	NOTCH4
Q18035589	C6orf10	Q18027822	IL2RB	Q18036729	RAP1GAP2	Q18026947	HLA-DQA1

```
SELECT DISTINCT ?gene ?geneLabel where {
  ?gene wdt:P2293 wd:Q35869 .  # gene has genetic association to "asthma"

  ?gene wdt:P31 wd:Q7187 .     # gene is subclass of "gene"

  ?gene wdt:P688 ?protein .                # gene encodes a protein
  ?protein wdt:P681 ?cc .                  # protein has a cellular component
  ?cc wdt:P279*|wdt:P361* wd:Q14349455 .   # cell component is 'part of' or 'subclass of' membrane

  SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
}
```

http://bit.ly/bosc2017_wikidata

Computing on provenance

"Retrieve genes with GWAS association with asthma and gene product is localized to membrane (non-IEA)"



15 genes

gene	geneLabel		gene	geneLabel		gene	geneLabel
Q14912759	SLC22A5		Q18045382	HPSE2		Q17917202	ERBB4
Q14914243	PSAP		Q18027822	IL2RB		Q18027836	IL6R
Q14907990	SLC30A8		Q14903974	SMAD3		Q18030409	PDE4D
Q18027370	IGSF3		Q18035037	RAD50		Q18030185	NOTCH4
Q18033424	IL18R1		Q18036729	RAP1GAP2		Q18026947	HLA-DQA1

```
6   ?gene wdt:P31 wd:Q7187 ;      # gene is subclass of "gene"
7         wdt:P688 ?protein ;     # gene encodes a protein
8         rdfs:label ?geneLabel .
9   FILTER (lang(?geneLabel) = "en")
10  ?protein p:P681 ?s .                          # protein's cell component statement
11    ?s ps:P681 ?cp .                            # get statement value
12    FILTER NOT EXISTS {?s pq:P459 wd:Q23190881 .}  # determination method is not IEA
13    ?cp wdt:P279*|wdt:P361* wd:Q14349455 .       # statement value is 'part of' or 'subclass of' membrane
14
```

http://bit.ly/bosc2017_wikidata

Opportunistic integration

"Retrieve genes with GWAS association with any respiratory disease and gene product is localized to membrane (non-IEA) and show causative chemical hazards"

4 diseases / 6 chemical hazards

diseaseGALabel	exposureLabel
lung cancer	arsenic pentoxide exposure
lung cancer	HN1 exposure
lung cancer	mechlorethamine exposure
lung cancer	HN3 exposure
asthma	Phenacyl chloride exposure
pulmonary emphysema	phosgene exposure

```
12
13    ?exposure wdt:P1542 ?diseaseGA .  # something causes disease
14    ?exposure wdt:P279 wd:Q21167512 . # and that something is a chemical hazard
15
16    SERVICE wikibase:label { bd:serviceParam wikibase:language "en". }
17  }
```

http://bit.ly/bosc2017_wikidata

Leveraging the Disease Ontology structure

"Retrieve genes with GWAS association with any respiratory disease and gene product is localized to membrane (non-IEA)"

31 genes / 8 diseases

diseaseGALabel	gene_counts	geneList
asthma	15	SMAD3, RAP1GAP2, IL18R1, HPSE2, SLC30A8, SLC22A5, PSAP, ERBB4, HLA-DQA1, IGSF3, IL2RB, IL6R, NOTCH4, PDE4D, RAD50
chronic obstructive pulmonary disease	5	HLA-C, SFTPD, ANXA5, ANXA11, ATP2C2
lung cancer	3	TGM5, VTI1A, PHACTR2
interstitial lung disease	2	DSP, ATP11A
non-small-cell lung carcinoma	2	NALCN, DLST
nasopharynx carcinoma	2	ITGA9, TNFRSF19
adenocarcinoma of the lung	1	BTNL2
pulmonary emphysema	1	BICD1

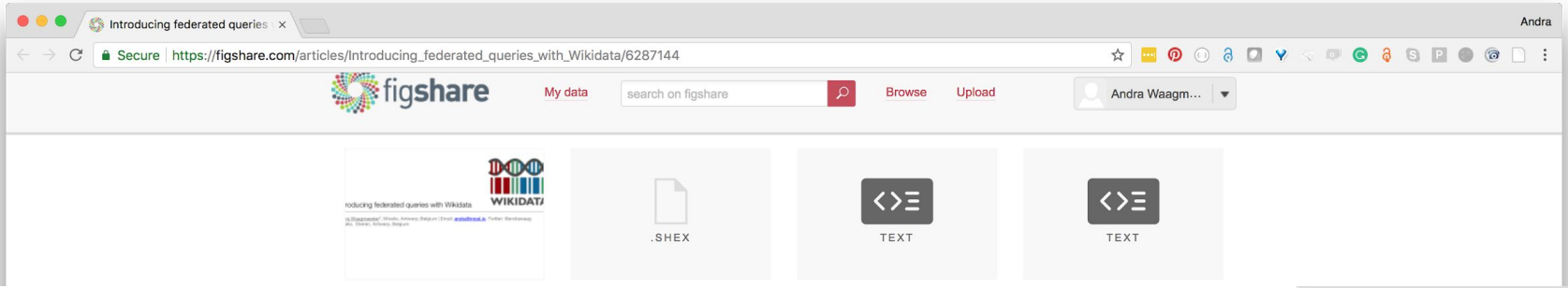
```
1  SELECT ?diseaseGALabel (count (DISTINCT ?ger
2  (group_concat(DISTINCT ?geneLabel; separator
3    ?gene wdt:P2293 ?diseaseGA .          # gene
4    ?diseaseGA wdt:P279* wd:Q3286546 .    # to a
5
6    ?gene wdt:P31 wd:Q7187 ; wdt:P688 ?protein ;    # gene is subclass of "gene" and encodes protein
7          rdfs:label ?geneLabel .
8    FILTER (lang(?geneLabel) = "en")
9    ?protein p:P681 ?s .                            # protein's cell component statement
10     ?s ps:P681 ?cp .                              # get statement value
```

http://bit.ly/bosc2017_wikidata

Where Figshare meets Wikidata

Source: https://doi.org/10.6084/m9.figshare.97662.v1

A deposit for data, posters and slides



Waagmeester, Andra (2018): Introducing federated queries with Wikidata. figshare. Fileset.

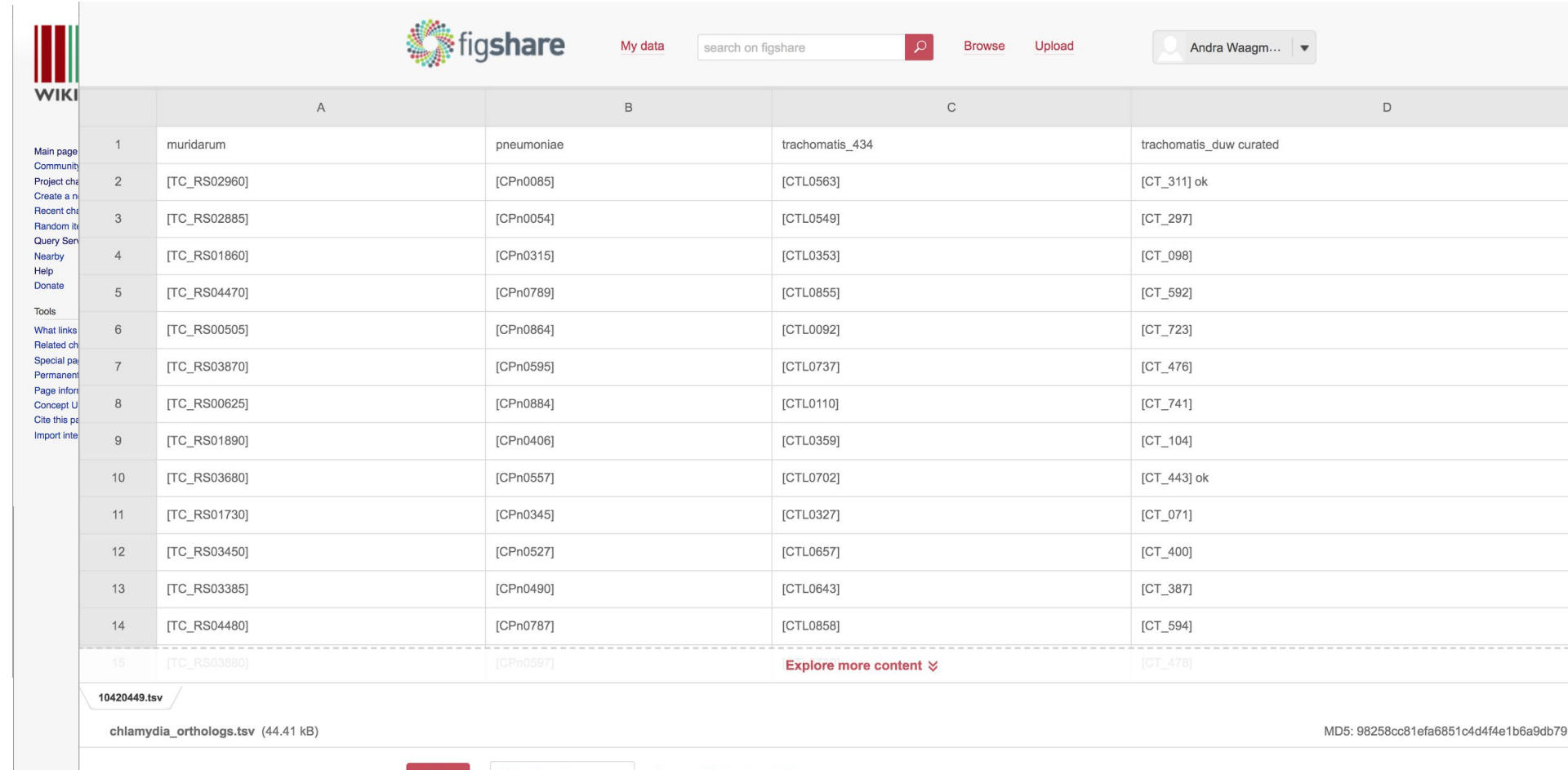
https://doi.org/10.6084/m9.figshare.6287144.v1

Wikidata:Notability

1. It contains at least one valid sitelink to a page on Wikipedia, Wikivoyage, Wikisource, Wikiquote, Wikinews, Wikibooks, Wikidata, Wikispecies, Wikiversity, or Wikimedia Commons.
2. It refers to an instance of a clearly identifiable conceptual or material entity. The entity must be notable, in the sense that it can be described using serious and publicly available references.
3. It fulfills some structural need

source: https://www.wikidata.org/wiki/Wikidata:Notability

Figshare as host for serious and publicly available references



figshare My data search on figshare Browse Upload Andra Waagm…

WIKI

Main page
Community
Project cha
Create a n
Recent cha
Random ite
Query Serv
Nearby
Help
Donate

Tools

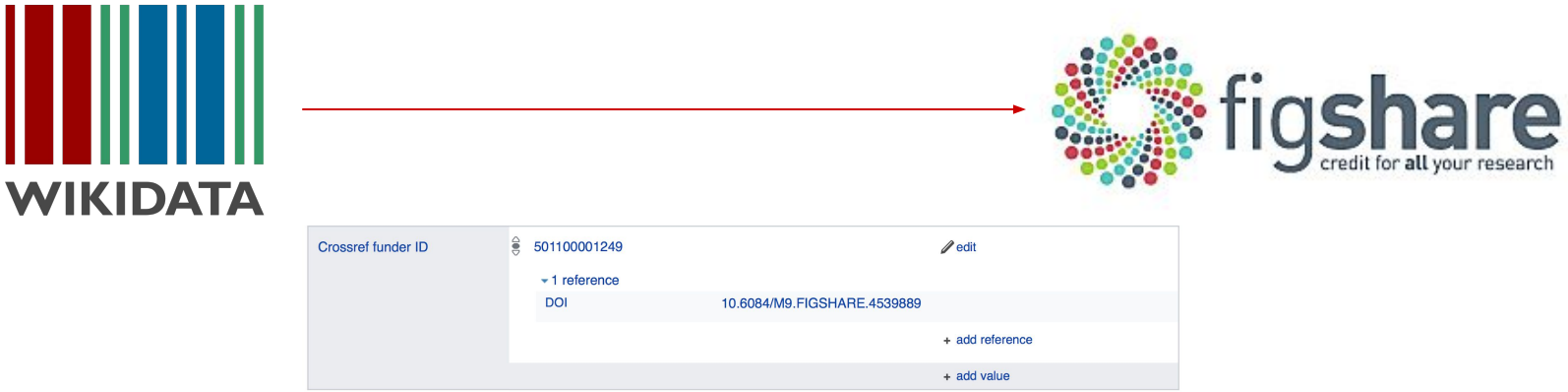
What links
Related ch
Special pa
Permanent
Page infor
Concept U
Cite this pa
Import inte

	A	B	C	D
1	muridarum	pneumoniae	trachomatis_434	trachomatis_duw curated
2	[TC_RS02960]	[CPn0085]	[CTL0563]	[CT_311] ok
3	[TC_RS02885]	[CPn0054]	[CTL0549]	[CT_297]
4	[TC_RS01860]	[CPn0315]	[CTL0353]	[CT_098]
5	[TC_RS04470]	[CPn0789]	[CTL0855]	[CT_592]
6	[TC_RS00505]	[CPn0864]	[CTL0092]	[CT_723]
7	[TC_RS03870]	[CPn0595]	[CTL0737]	[CT_476]
8	[TC_RS00625]	[CPn0884]	[CTL0110]	[CT_741]
9	[TC_RS01890]	[CPn0406]	[CTL0359]	[CT_104]
10	[TC_RS03680]	[CPn0557]	[CTL0702]	[CT_443] ok
11	[TC_RS01730]	[CPn0345]	[CTL0327]	[CT_071]
12	[TC_RS03450]	[CPn0527]	[CTL0657]	[CT_400]
13	[TC_RS03385]	[CPn0490]	[CTL0643]	[CT_387]
14	[TC_RS04480]	[CPn0787]	[CTL0858]	[CT_594]
15	[TC_RS03880]	[CPn0597]	Explore more content	[CT_478]

10420449.tsv

chlamydia_orthologs.tsv (44.41 kB) MD5: 98258cc81efa6851c4d4f4e1b6a9db79

Citing Figshare



Citations for the sum of all human knowledge

Source: https://doi.org/10.6084/m9.figshare.5648233.v3

Availability

Platforms

- www.figshare.com - Figshare main
- www.wikidata.org – Wikidata main
- Query.wikidata.org - Wikidata Query Service

github.com/SuLab/GeneWikiCentral

- github.com/SuLab/**wikidataintegrator** – python module for Wikidata
- github.com/SuLab/**scheduled-bots** – bot automation framework
- github.com/SuLab/**Genewiki-ShEx** – data models

Acknowledgments

Gene Wiki:

- Andrew Su - The Scripps Research Institute
- Greg Stupp - The Scripps Research Institute
- Lynn Schriml - University of Maryland

Wikicite

- Finn Årup Nielsen - Technical University of Denmark
- Dario Taraborelli - Wikimedia Foundation
- Egon Wilighagen - Maastricht University

Funding:

- This work was supported by the National Institutes of Health under grant GM089820

If time permits …..

DEMO