Medizinische Universität Graz

BIOMEDICAL DATABASES

Stefan Schulz

Institute for Medical Informatics, Statistics and Documentation

stefan.schulz@medunigraz.at

Last revision: July 2024

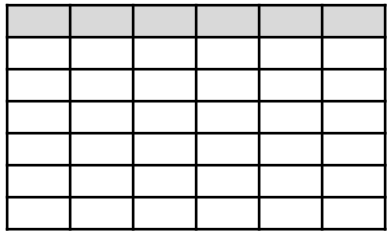
Goals

- Understand basic concepts of databases
  - Database management systems
  - Database access
  - Database curation
  - Database semantics
  - Structured and unstructured database content
  - Controlled vocabularies
- Prototypical databases for biomedical research
  - Databases to support omics research: Uniprot, Ensembl
  - Literature databases: Pubmed / Medline
  - Databases to support clinical research: ClinicalTrials.gov

What are databases?

- IT systems that represent **real world** objects and their dependencies in a **data** model
- DBMS (Database Management System)
  - "container" of a database.
  - It provides a syntax for working with the database.
  - Typical DBMSs: Oracle, MS SQL Server, mySQL, MS Access
- Databases can be found in nearly all IT systems that manage data
  - Banking, booking, eGovernment
  - Health care related information systems
- Interfaces:
  - Database client applications (desktop computers, mobile devices)
  - Web interfaces
  - APIs (application program interfaces)

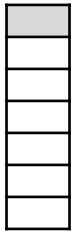
Database Objects – Basic Typology



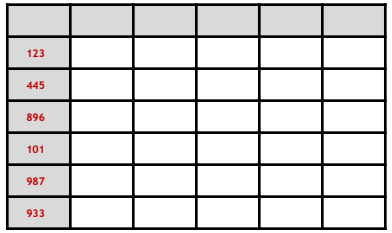
Table

Attributes (column heads) with labels

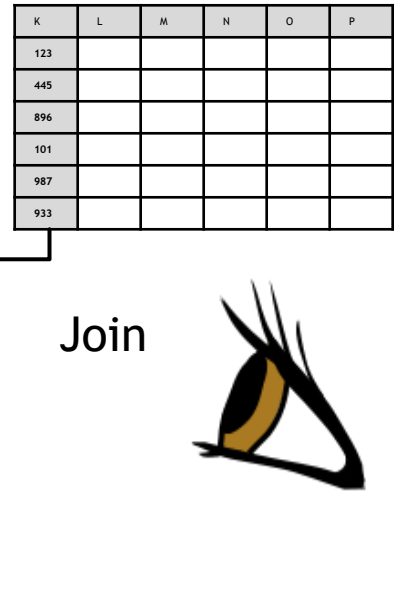
Record (row) with values



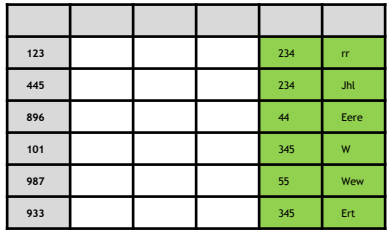
Column



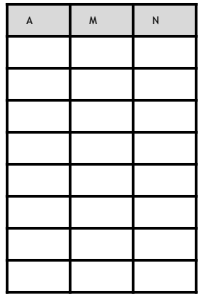
Primary (unique) key



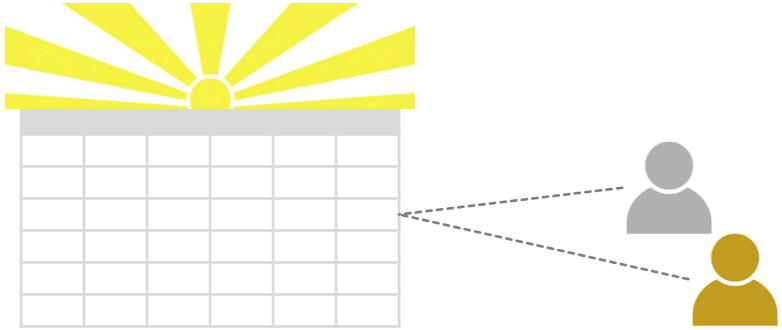
Join



Secondary keys

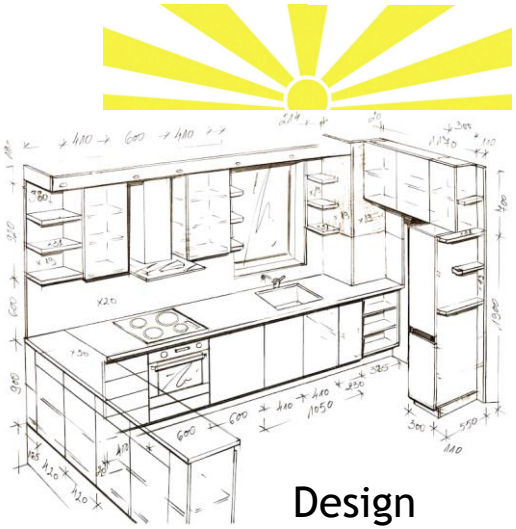


Typical database operations



Devise database architecture

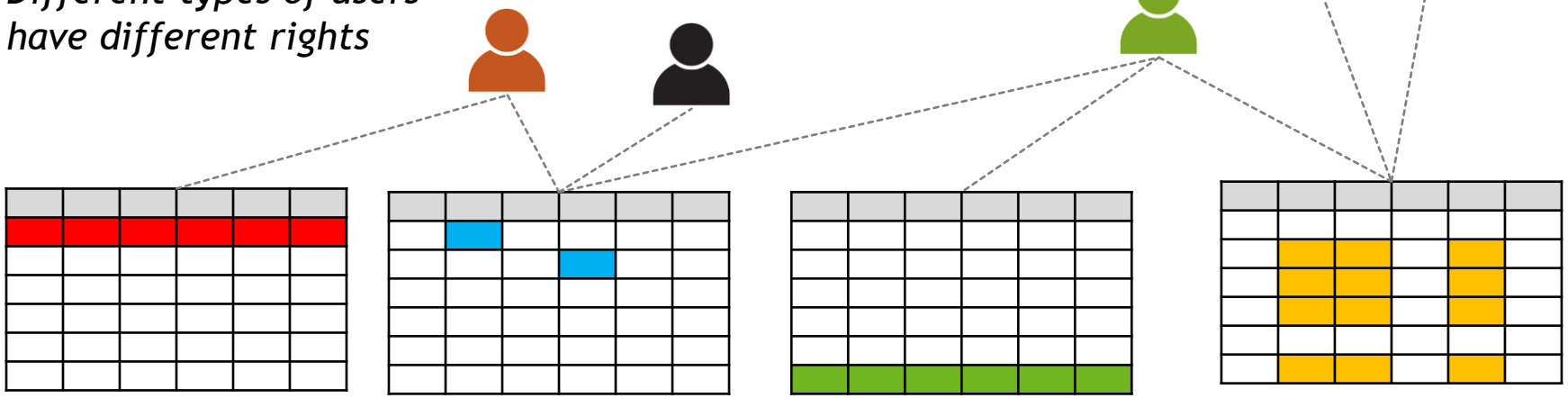
Create new database objects



Design

Use

Different types of users have different rights



Delete data

Change data

Insert new data

Query data

https://www.servus.com/storage/article/haushalt-kuhlschrank-leitfaden-julialammers-SA0009AK.jpg?impolicy=article_short_header

Example: hospital database

PatID	FallID	Vorname	Name	Titel	G	Geburts	SV-Nr	Straße	Ort	PLZ	Land	Aufnahm	Entlassda	Haup	Hauptdiagnose (Te:
234508973	9829993923	Franz	Brunner		m	12.12.1945	7556121245	Kirchweg 26	Gratkorn	8101	Österreich	12.04.2013	22.04.2013	G20.1	Primäres Parkinson-Syndrom mit
549082235	2341221400	Konstantin	Luttenberger		m	31.08.1963	6643310893	Leibnitzer Str 15	Lebring	8403	Österreich	22.04.2013	04.05.2013	I21.1	Akuter transmuraler Myokardinfarkt der
683654353	2000977623	Elke	Schulze	Mag.	w	03.07.1968	9012030768	Anton-Kleinoscheg-Str 29	Graz	8051	Österreich	25.04.2013	30.07.2013	G82.1	Spastische Paraparese und
545454109	2466375743	Jessica	Strohriegl		w	04.10.1988	9970041088	Rudolfstraße 123	Graz	8010	Österreich	08.04.2013	11.04.2013	H66.0	Akute eitrige Otitis media
236519452	7655545877	Jaden	Klötzl		m	05.11.2009	4646051109	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	31.07.2013	S02.1	Schädelbasisfraktur
843656542	4478477543	Margitta	Schwarzenegger-Klötzl	Dr.	w	12.12.1972	5558121272	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	10.06.2013	S72.7	Multiple Frakturen des Femurs
340009212	7887900352	Bettina	Lammer	Mag.	w	14.02.1956	7823140256	Griesplatz 28	Graz	8020	Österreich	01.04.2013	24.07.2013	F30.2	Manie mit psychotischen
202040563	7578888254	Giuseppe	De Tomaso		m	04.10.1952	6542041052	Leitnergasse 11	Graz	8010	Österreich	02.04.2013	04.04.2013	C43.5	Bösartiges Melanom des Rumpfes
421545873	5847364332	Roman	Terbovc		m	05.01.1975	6684050175	Mladinska ulica 29	Šentilj	2212	Slowenien	13.04.2013	16.04.2013	H66.0	Akute eitrige Otitis media
118702653	8559754441	Maria das Neves	Pinheiro da Silva		w	31.12.1969	7546311269	Leitnergasse 23	Graz	8010	Österreich	16.04.2013	30.04.2013	I21.1	Akuter transmuraler Myokardinfarkt der
342444438	3554640992	Kim	Park		m	23.06.1955	4886230675	Wickenburgg 3	Graz	8010	Österreich	05.04.2013	12.04.2013	E11.1	Diabetes mellitus, Typ 2 mit
464346833	9758002454	Maria	Ehrenberger		w	02.02.1946	2399020246	Volksgartenstaße 3	Graz	8020	Österreich	02.04.2013	21.04.2013	C73	Bösartige Neubildung der
203332422	5450025454	Lisbeth	Puntigam		w	24.04.1971	4886240471	Annenstraße 44	Graz	8020	Österreich	17.04.2013	18.05.2013	I26.9	Lungenembolie ohne Angabe eines
210883233	9895576500	Gerhard	Müller		m	01.02.1918	4755010218	Am Arlandgrund 19	Graz	8045	Österreich	22.04.2013	25.04.2013	H25.2	Cataracta senilis, Morgagni-Typ
234094583	9075465823	Adolf	Brunner		m	21.04.1941	6654210441	Preßlgasse 4	Hartberg	8230	Österreich	30.04.2013	23.05.2013	A15.2	Lungentuberkulose, histologisch gesichert
283610234	2002025224	Aise	Devici		w	11.11.1984	5788111184	Hochstadelweg 19	Nußdorf-Debant	9990	Österreich	10.04.2013	13.04.2013	H66.0	Akute eitrige Otitis media
567845242	5564659781	Klaus-Michael	Kohler		m	09.01.1966	7755010966	Müller-Guttenbrunn-Weg 23	Graz-Liebenau	8041	Österreich	21.04.2013	05.05.2013	I21.1	Akuter transmuraler Myokardinfarkt der
457789020	3112143202	Carmen	Meyr		w	03.07.1977	4432030777	Schilfgasse 15	Graz-Straßgang	8054	Österreich	14.04.2013	13.05.2013	S72.7	Multiple Frakturen des Femurs
500545723	5888729364	Muhammad	Üstün		m	06.07.1954	9700060794	Algersdorfer Straße 16	Graz	8020	Österreich	11.04.2013	20.04.2013	E11.1	Diabetes mellitus, Typ 2 mit
687875422	9009736772	Johanna	Hadlic		w	13.08.1929	6734130829	Afritschgasse 36	Graz	8020	Österreich	27.04.2013	03.05.2013	E11.1	Diabetes mellitus, Typ 2 mit
432443502	3434554645	Georg	Moser		m	05.05.1941	7773050541	Freihofanger 2	Graz	8043	Österreich	01.04.2013	30.08.2013	F31.3	Bipolare affektive Störung,
800050524	7676855645	Roberta	Eber		w	03.06.2002	1323030602	Dultstraße 58	Gratkorn	8101	Österreich	05.04.2013	17.06.2013	S02.1	Schädelbasisfraktur
988999523	5000527877	Christiane	Thomüller		w	12.06.1960	2544120690	Göstinger Str 182	Graz	8051	Österreich	12.04.2013	22.04.2013	I21.1	Akuter transmuraler Myokardinfarkt der

Med Uni Graz

Names and attributes faked up

PatID	FallID	Vorname	Name	Titel	G	Geburtsd	SV-Nr	Straße	Ort	PLZ	Land	Aufnahm	Entlassda	Haupt	Hauptdiagnose (Te:
234508973	9829993923	Franz	Brunner		m	12.12.1945	7556121245	Kirchweg 26	Gratkorn	8101	Österreich	12.04.2013	22.04.2013	G20.1	Primäres Parkinson-Syndrom mit
549082235	2341221400	Konstantin	Luttenberger		m	31.08.1963	6643310893	Leibnitzer Str 15	Lebring	8403	Österreich	22.04.2013	04.05.2013	I21.1	Akuter transmuraler Myokardinfarkt der
683654353	2000977623	Elke	Schulze	Mag.	w	03.07.1968	9012030768	Anton-Kleinoscheg-Str 29	Graz	8051	Österreich	25.04.2013	30.07.2013	G82.1	Spastische Paraparese und
545454109	2466375743	Jessica	Strohriegl		w	04.10.1988	9970041088	Rudolfstraße 123	Graz	8010	Österreich	08.04.2013	11.04.2013	H66.0	Akute eitrige Otitis media
236519452	7655545877	Jaden	Klötzl		m	05.11.2009	4646051109	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	31.07.2013	S02.1	Schädelbasisfraktur
843656542	4478477543	Margitta	Schwarzenegger-Klötzl	Dr.	w	12.12.197									
340009212	7887900352	Bettina	Lammer	Mag.	w	14.02.19									
202040563	7578888254	Giuseppe	De Tomaso		m	04.10.19									
421545873	5847364332	Roman	Terbovc		m	05.01.19									
118702653	8559754441	Maria das Neves	Pinheiro da Silva		w	31.12.19									
342444438	3554640992	Kim	Park		m	23.06.19									
464346833	9758002454	Maria	Ehrenberger		w	02.02.19									
203332422	5450025454	Lisbeth	Puntigam		w	24.04.19									
210883233	9895576500	Gerhard	Müller		m	01.02.19									
234094583	9075465823	Adolf	Brunner		m	21.04.19									
283610234	2002025224	Aise	Devici		w	11.11.19									
567845242	5564659781	Klaus-Michael	Kohler		m	09.01.19									
457789020	3112143202	Carmen	Meyr		w	03.07.19									
500545723	5888729364	Muhammad	Üstün		m	06.07.19									
															Typ 2 mit
687875422	9009736772	Johanna	Hadlic		w	13.08.1929	6734130829	Afritschgasse 36	Graz	8020	Österreich	27.04.2013	03.05.2013	E11.1	Diabetes mellitus, Typ 2 mit
432443502	3434554645	Georg	Moser		m	05.05.1941	7773050541	Freihofanger 2	Graz	8043	Österreich	01.04.2013	30.08.2013	F31.3	Bipolare affektive Störung,
800050524	7676855645	Roberta	Eber		w	03.06.2002	1323030602	Dultstraße 58	Gratkorn	8101	Österreich	05.04.2013	17.06.2013	S02.1	Schädelbasisfraktur
988999523	5000527877	Christiane	Thomüller		w	12.06.1960	2544120690	Göstinger Str 182	Graz	8051	Österreich	12.04.2013	22.04.2013	I21.1	Akuter transmuraler Myokardinfarkt der

Relational database paradigm

- de-Facto-Standard
- Based on Tables:
  - **Rows** contain database **records** (Tuples)
  - **Columns** contain database **fields** (attributes)
  - **Values** constitute the **content** of a database
- Some other data storage paradigms
  - hierarchical DBs, graph DBs / triple stores, in-memory DBs

PatID	FallID	Vorname	Name	Titel	G	Geburtsd	SV-Nr	Straße	Ort	PLZ	Land	Aufnahm	Entlassda	Haup	Hauptdiagnose (Te:
234508973	9829993923	Franz	Brunner		m	12.12.1945	7556121245	Kirchweg 26	Gratkorn	8101	Österreich	12.04.2013	22.04.2013	G20.1	Primäres Parkinson-Syndrom mit
549082235	2341221400	Konstantin	Luttenberger		m	31.08.1963	6643310893	Leibnitzer Str 15	Lebring	8403	Österreich	22.04.2013	04.05.2013	I21.1	Akuter transmuraler Myokardinfarkt der
683654353	2000977623	Elke	Schulze	Mag.	w	03.07.1968	9012030768	Anton-Kleinoscheg-Str 29	Graz	8051	Österreich	25.04.2013	30.07.2013	G82.1	Spastische Paraparese und
545454109	2466375743	Jessica	Strohriegl		w	04.10.1988	9970041088	Rudolfstraße 123	Graz	8010	Österreich	08.04.2013	11.04.2013	H66.0	Akute eitrige Otitis media
236519452	7655545877	Jaden	Klötzl		m	05.11.2009	4646051109	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	31.07.2013	S02.1	Schädelbasisfraktur
843656542	4478477543	Margitta	Schwarzenegger-Klötzl	Dr.	w	12.12.1972	5558121272	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	10.06.2013	S72.7	Multiple Frakturen des Femurs
340009212	7887900352	Bettina	Lammer	Mag.	w	14.02.1956	7823140256	Griesplatz 28	Graz	8020	Österreich	01.04.2013	24.07.2013	F30.2	Manie mit psychotischen
202040563	7578888254	Giuseppe	De Tomaso		m	04.10.1952	6542041052	Leitnergasse 11	Graz	8010	Österreich	02.04.2013	04.04.2013	C43.5	Bösartiges Melanom des Rumpfes
421545873	5847364332	Roman	Terbovc		m	05.01.1975	6684050								Otitis
118702653	8559754441	Maria das Neves	Pinheiro da Silva		w	31.12.1969	7546311								uraler kt der
342444438	3554640992	Kim	Park		m	23.06.1955	4886230								itus,
464346833	9758002454	Maria	Ehrenberger		w	02.02.1946	2399020								er
203332422	5450025454	Lisbeth	Puntigam		w	24.04.1971	4886240								ie eines
210883233	9895576500	Gerhard	Müller		m	01.02.1918	4755010								lis,
234094583	9075465823	Adolf	Brunner		m	21.04.1941	6654210								ulose, esichert
283610234	2002025224	Aise	Devici		w	11.11.1984	5788111								Otitis
567845242	5564659781	Klaus-Michael	Kohler		m	09.01.1966	7755010								uraler kt der
457789020	3112143202	Carmen	Meyr		w	03.07.1977	4432030								turen
500545723	5888729364	Muhammad	Üstün		m	06.07.1954	9700060								itus,
687875422	9009736772	Johanna	Hadlic		w	13.08.1929	6734130								itus,
432443502	3434554645	Georg	Moser		m	05.05.1941	7773050								tive
800050524	7676855645	Roberta	Eber		w	03.06.2002	1323030								raktur
988999523	5000527877	Christiane	Thomüller		w	12.06.1960	2544120690	Göstinger Str 182	Graz	8051	Österreich	12.04.2013	22.04.2013	I21.1	Akuter transmuraler Myokardinfarkt der

Database queries

Selection by rows

Projection by columns

SELECT **Name, Vorname** from Patienten
WHERE **Ort = „Hausmannstätten“**

PatID	FallID	Vorname	Name	Titel	G	Geburtsd	SV-Nr	Straße	Ort	PLZ	Land	Aufnahm	Entlassda	Haupt	Hauptdiagnose (Te:
234508973	9829993923	Franz	Brunner		m	12.12.1945	7556121245	Kirchweg 26	Gratkorn	8101	Österreich	12.04.2013	22.04.2013	G20.1	Primäres Parkinson-Syndrom mit
549082235	2341221400	Konstantin	Luttenberger		m	31.08.1963	6643310893	Leibnitzer Str 15	Lebring	8403	Österreich	22.04.2013	04.05.2013	I21.1	Akuter transmuraler Myokardinfarkt der
683654353	2000977623	Elke	Schulze	Mag.	w	03.07.1968	9012030768	Anton-Kleinoscheg-Str 29	Graz	8051	Österreich	25.04.2013	30.07.2013	G82.1	Spastische Paraparese und
545454109	2466375743	Jessica	Strohriegl		w	04.10.1988	9970041088	Rudolfstraße 123	Graz	8010	Österreich	08.04.2013	11.04.2013	H66.0	Akute eitrige Otitis media
236519452	7655545877	Jaden	Klötzl		m	05.11.2009	4646051109	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	31.07.2013	S02.1	Schädelbasisfraktur
843656542	4478477543	Margitta	Schwarzenegger-Klötzl	Dr.	w	12.12									
340009212	7887900352	Bettina	Lammer	Mag.	w	14.02									
202040563	7578888254	Giuseppe	De Tomaso		m	04.10									
421545873	5847364332	Roman	Terbovc		m	05.01									
118702653	8559754441	Maria das Neves	Pinheiro da Silva		w	31.12									
342444438	3554640992	Kim	Park		m	23.06									
464346833	9758002454	Maria	Ehrenberger		w	02.02									
203332422	5450025454	Lisbeth	Puntigam		w	24.04									
210883233	9895576500	Gerhard	Müller		m	01.02									
234094583	9075465823	Adolf	Brunner		m	21.04									
283610234	2002025224	Aise	Devici		w	11.11									
567845242	5564659781	Klaus-Michael	Kohler		m	09.01									
457789020	3112143202	Carmen	Meyr		w	03.07									
500545723	5888729364	Muhammad	Üstün		m	06.07									
687875422	9009736772	Johanna	Hadlic		w	13.08.1929	6734130829	Afritschgasse 36	Graz	8020	Österreich	27.04.2013	03.05.2013	E11.1	Typ 2 mit Diabetes mellitus, Typ 2 mit
432443502	3434554645	Georg	Moser		m	05.05.1941	7773050541	Freihofanger 2	Graz	8043	Österreich	01.04.2013	30.08.2013	F31.3	Bipolare affektive Störung,
800050524	7676855645	Roberta	Eber		w	03.06.2002	1323030602	Dultstraße 58	Gratkorn	8101	Österreich	05.04.2013	17.06.2013	S02.1	Schädelbasisfraktur
988999523	5000527877	Christiane	Thomüller		w	12.06.1960	2544120690	Göstinger Str 182	Graz	8051	Österreich	12.04.2013	22.04.2013	I21.1	Akuter transmuraler Myokardinfarkt der

Database keys

- Cover one or more fields
- Speed up ordering and retrieval
- Primary keys are univocal keys that precisely identify exactly one database record

(Many primary keys are part of our daily life: Social security number, IBAN, customer ID...)

Table “Patients”

SV-Nr	Straße	Ort	PLZ	Land	Aufnahm	Entlassda	Haup
556121245	Kirchweg 26	Gratkorn	8101	Österreich	12.04.2013	22.04.2013	G20.1
643310893	Leibnitzer Str 15	Lebring	8403	Österreich	22.04.2013	04.05.2013	I21.1
012030768	Anton-Kleinoscheg-Str 29	Graz	8051	Österreich	25.04.2013	30.07.2013	G82.1
970041088	Rudolfstraße 123	Graz	8010	Österreich	08.04.2013	11.04.2013	H66.0
646051109	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	31.07.2013	S02.1
558121272	Hofstr 1	Hausmannstätten	8071	Österreich	05.04.2013	10.06.2013	S72.7
823140256	Griesplatz 28	Graz	8020	Österreich	01.04.2013	24.07.2013	F30.2
542041052	Leitnergasse 11	Graz	8010	Österreich	02.04.2013	04.04.2013	C43.5
684050175	Mladinska ulica 29	Šentilj	2212	Slowenien	13.04.2013	16.04.2013	H66.0
546311269	Leitnergasse 23	Graz	8010	Österreich	16.04.2013	30.04.2013	I21.1
886230675	Wickenburgg 3	Graz	8010	Österreich	05.04.2013	12.04.2013	E11.1
399020246	Volksgartenstaße 3	Graz	8020	Österreich	02.04.2013	21.04.2013	C73
886240471	Annenstraße 44	Graz	8020	Österreich	17.04.2013	18.05.2013	I26.9
755010218	Am Arlandgrund 19	Graz	8045	Österreich	22.04.2013	25.04.2013	H25.2
654210441	Preßlgasse 4	Hartberg	8230	Österreich	30.04.2013	23.05.2013	A15.2
788111184	Hochstadelweg 19	Nußdorf-Debant	9990	Österreich	10.04.2013	13.04.2013	H66.0
755010966	Müller-Guttenbrunn-Weg 23	Graz-Liebenau	8041	Österreich	21.04.2013	05.05.2013	I21.1
432030777	Schilfgasse 15	Graz-Straßgang	8054	Österreich	14.04.2013	13.05.2013	S72.7
700060794	Algersdorfer Straße 16	Graz	8020	Österreich	11.04.2013	20.04.2013	E11.1
734130829	Afritschgasse 36	Graz	8020	Österreich	27.04.2013	03.05.2013	E11.1
773050541	Freihofanger 2	Graz	8043	Österreich	01.04.2013	30.08.2013	F31.3
323030602	Dultstraße 58	Gratkorn	8101	Österreich	05.04.2013	17.06.2013	S02.1
544120690	Göstinger Str 182	Graz	8051	Österreich	12.04.2013	22.04.2013	I21.1

Table “ICD-10”

Code	Text
I21.1	Akuter transmuraler Myokardinfarkt der Hinterwand
C43.5	Bösartiges Melanom des Rumpfes
I26.9	Lungenembolie ohne Angabe eines akuten Cor pulmonale
H25.2	Cataracta senilis, Morgagni-Typ
A15.2	Lungentuberkulose, histologisch gesichert
G20.1	Primäres Parkinson-Syndrom mit mäßiger bis schwerer Beeinträch
S02.1	Schädelbasisfraktur
H66.0	Akute eitrige Otitis media
F30.2	Manie mit psychotischen Symptomen
E11.1	Diabetes mellitus, Typ 2 mit Ketoazidose
F31.3	Bipolare affektive Störung, gegenwärtig leichte oder mittelgradige
S72.7	Multiple Frakturen des Femurs
C73	Bösartige Neubildung der Schilddrüse
G82.1	Spastische Paraparese und Paraplegie

Med Uni Graz

„Normalisation“

- Removal of redundant information and thus sources of error
- More compact representation of content
- Primary key of detail table is foreign key of main table

Database semantics

- Semantics
  - The meaning behind names, identifiers, values in a database
  - The way how they are related to the real world
  - Database content denotes (types of) entities in the real world
- Example
  - Field *AdmDia* Value: T1DM
  - Field *Hb* Value: 13,3
  - Field *Gender* Value: 1
  - Field DOB Value: 11/12/23
  - Field *Inclusion criteria* Value: Over 18 years old. ASA classification less than or equal to 3. Patients who live accompanied in a home at a maximum distance of 30 minutes from the hospital and an adequate cognitive capacity.
- Datatypes
  - Numeric, Date, Text, Boolean (True / False)
  - "Controlled vocabularies" (CV): **Coding systems / thesauri / ontologies: One meaning – one controlled term (+ ID / code)**

Problems
- Number/date formats
- Unclear encodings
- Acronyms
- Underspecified field labels
- Free text

Examples for controlled vocabularies

ICD-10

Chapter XII
Diseases of the skin and subcutaneous tissue
(L00-L99)

Urticaria and erythema (L50-L54)

Excl.: Lyme disease (A69.2)
rosacea (L71.-)

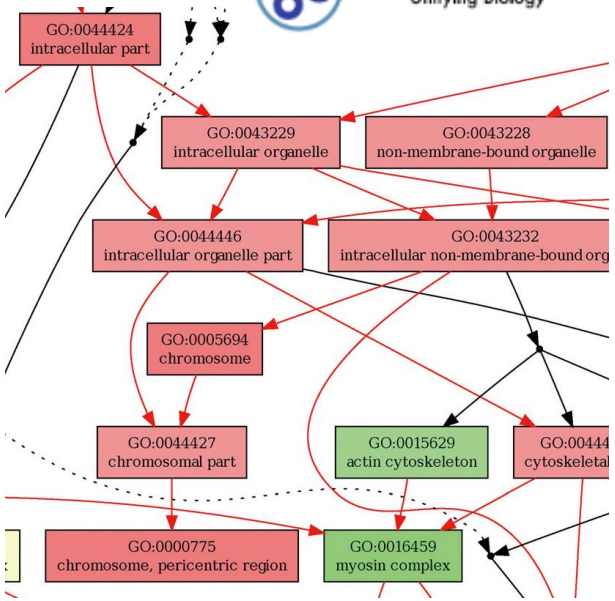
L50 Urticaria

Excl.: allergic contact dermatitis (L23.-)
angioneurotic oedema (T78.3)
hereditary angio-oedema (D84.1)
Quincke oedema (T78.3)
urticaria:
- giant (T78.3)
- neonatorum (P83.8)
- papulosa (L28.2)
- pigmentosa (Q82.2)
- serum (T80.6)
- solar (L56.3)

L50.0 Allergic urticaria
L50.1 Idiopathic urticaria
L50.2 Urticaria due to cold and heat
L50.3 Dermatographic urticaria
L50.4 Vibratory urticaria
L50.5 Cholinergic urticaria
L50.6 Contact urticaria
L50.8 Other urticaria
Urticaria:
- chronic
- recurrent periodic

L50.9 Urticaria, unspecified

GENEONTOLOGY



Medical Subject Headings

Pathological Conditions, Signs and Symptoms
Pathologic Processes
Hemorrhage

Hemorrhage
Bleeding or escape of blood

MeSH Unique ID: D006
Entry Terms:
- Hemorrhages
- Bleeding

Blood Loss, Surgical
Ecchymosis
Epistaxis
Exsanguination
Eye Hemorrhage
Choroid Hemorrhage
Hyphema
Retinal Hemorrhage
Vitreous Hemorrhage

SNOMED CT

The global language of healthcare

Parents
- Ischemic heart disease (disorder)
- Myocardial disease (disorder)
- Myocardial necrosis (finding)
- Necrosis of anatomical site (disorder)

Myocardial infarction (disorder)
SCTID: 22298006
22298006 | Myocardial infarction (disorder) |
en Myocardial infarction
en Infarction of heart
en Cardiac infarction
en Heart attack
en Myocardial infarction (disorder)
en MI - Myocardial infarction
en Myocardial infarct

Associated morphology → Infarct
Finding site → Myocardium structure

Children
- Acute myocardial infarction (disorder)
- First myocardial infarction (disorder)
- Microinfarct of heart (disorder)

Controlled vocabularies (CVs) and Databases

Med Uni Graz

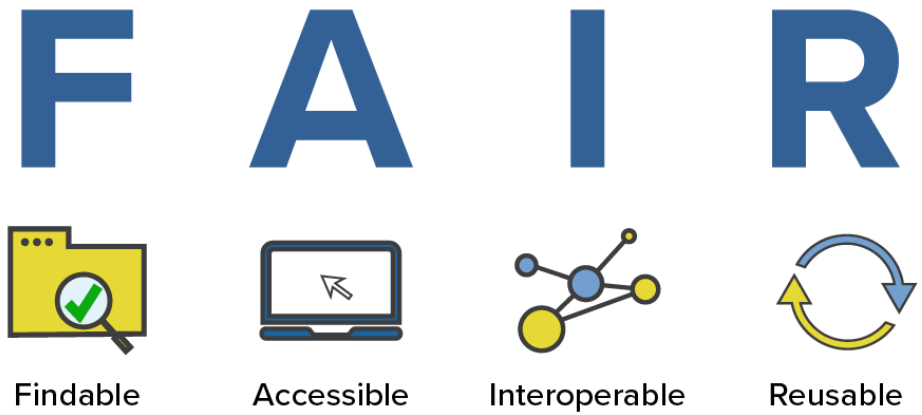
- CVs provide standardised semantic identifiers
  - AKA terminologies, thesauri (additional relations), ontologies (based on logic)
- CVs constitute the semantic building blocks for databases (like words that constitute a text). They provide standardised meaning for
  - Database tables
  - Database fields
  - Database cells
- Without CVs users need to interpret free text entries
  - Variety of language (“bleeding”, “haemorrhage”, “haemorrhage”, “hemorrhagic”, “haemorrhagic”)
  - Hierarchical relations (“heart disease”, “ischemic heart disease”, “MI”, “STEMI”) (“Africa”, “Western Africa”, “Liberia”)
  - Free text, to different extent, exists in all biomedical databases

Database annotations

- The addition of identifiers IDs from CVs is known as annotation
- Annotations are normally done by domain experts, AKA curators
- Tools that do automated processing of natural language using AI technology increasingly assist curators and accelerate their work
- Limitations:
  - Fully automated annotations (without human review) raise quality issues, the less training data are available (AI, large language models,…)
  - Manual annotations can also be faulty and incomplete
  - No CV is ideal – often there are several options of expressing one entity of meaning
- Efficient use of biomedical databases requires considerable familiarity both with CVs and domain language

Criteria to describe biomedical databases

- Access (free or subscription-based)
- Availability of database content (downloadable)
- Kind of interfaces (User, API)
- Transparency of used algorithms
- Human annotation effort
- Connection with other databases
- Structuredness (unstructured, semistructured, structured, structured + coded)
- Use of standards (CVs, information models)
- Sophistication / community involvement



Wilkinson, M. D. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, 3, 160018. doi:10.1038/sdata.2016.18

Note: Search engines such as Google Scholar are not databases

Biological Databases

Biological Databases

- Increasing amount of partly overlapping databases
- Huge amount of data
  - Sequences
  - Annotations (Gene ontology, organisms, MeSH descriptors)
  - Bibliographic information
- In-built visualization tools
- Plug-ins, e.g. for sequence alignment
- Heavy curation effort by the involved communities
- Heavily interlinked (use of IDs of other databases and CVs as foreign keys)
- Linked with original literature sources (PMID)
- Many only exist due to public funding (EU, US)

Wheeler, David L., et al. "Database resources of the national center for biotechnology information." *Nucleic acids research*35.suppl_1 (2006): D5-D12.

Toomula, Nishant, et al. "Biological databases-integration of life science data." *J. Comput. Sci. Syst. Biol* 4 (2012): 87-92.

Most popular biomedical databases

Med Uni Graz

UniProt

Resource for protein sequence and functionwith data on protein names, functions, and interactions

DrugBank

Drug related chemical, pharmacologic, and pharmaceutical information combined with drug target information

EMBASE

Large biomedical and pharmacological database. Extensive drug and disease information.

Ensembl

Browser for vertebrate genomes with annotation and data from multiple sources. Used for comparative genomics

PubMed / MEDLINE

Over 30 million biomedical biblio-graphic records. 90% with human annotations (MEDLINE)

ClinicalTrials.gov

Database of clinical studies (around 400k). Key resource for clinical trial information

OMIM

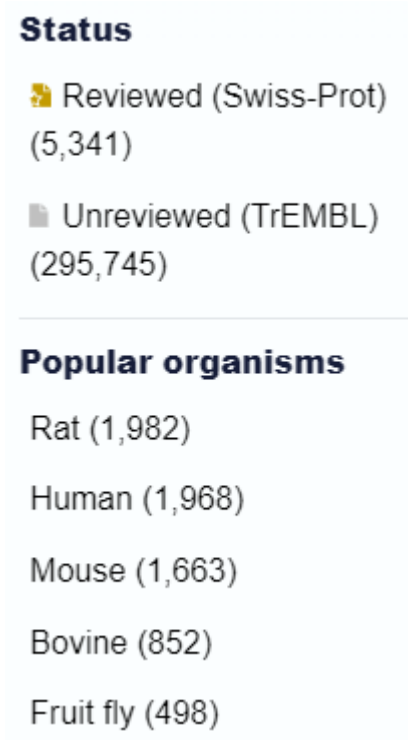
Contains human genes and genetic disorders. Essential for genetic research and diagnosis

Cochrane Library

Focuses on evidence-based medicine. Includes systematic reviews and meta-analyses.

Uniprot

- Huge database protein sequence and functional information for organisms including viruses
  - TrEMBL: computationally analysed records + automatic annotations
  - Swiss-Prot: manually reviewed annotations about all known relevant information about a protein from literature and sequence data.
  - One database record per gene and species
    - Location, biological processes, catalytic activity
    - Protein-protein interactions
    - Domains, binding sites
    - Expression patterns
    - Variant forms



https://www.uniprot.org/

UniProt: Functional annotation with Gene Ontology

Med Uni Graz

Q9BZA8 · PC11Y_HUMAN

Protein	Protocadherin-11 Y-linked	Amino acids	1340 (go to sequence)
Gene	PCDH11Y	Protein existence	Evidence at protein le
Status	UniProtKB reviewed (Swiss-Prot)	Annotation score	5/5
Organism	Homo sapiens (Human)		

Entry | Variant viewer 472 | Feature viewer | Genomic coordinates | Publications | External links | History

BLAST | Align | Download | Add | Add a publication | Entry feedback

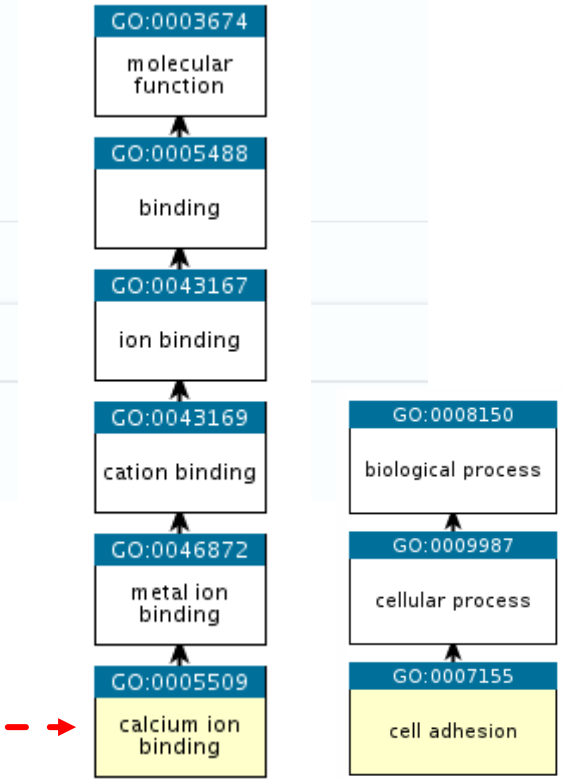
Function

Potential calcium-dependent cell-adhesion protein.

GO annotations

ASPECT	TERM	
Cellular Component	plasma membrane	Source:UniProtKB-SubCell
Molecular Function	calcium ion binding	Source:InterPro
Biological Process	cell adhesion	Source:GO_Central

GENEONTOLOGY Unifying Biology



UniProt: Link to literature

Q9BZA8 · PC11Y_HUMAN

Protein	Protocadherin-11 Y-linked	Amino acids	1340 (go to sequence)
Gene	PCDH11Y	Protein existence	Evidence at protein level
Status	UniProtKB reviewed (Swiss-Prot)	Annotation score	5/5
Organism	Homo sapiens (Human)		

Entry | Variant viewer 472 | Feature viewer | Genomic coordinates | Publications | External links | History

BLAST | Align | Download | Add | Add a publication | Entry feedback

Publications for Q9BZA8

Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain.

Blanco P., Sargent C.A., Boucher C., Mitchell M., Affara N.

View abstract

Cited for NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 3), TISSUE SPECIFICITY
Tissue Brain
Categories Sequences, Expression
Source UniProtKB reviewed (Swiss-Prot)

PubMed
Europe PMC
Mamm. Genome 11:906-914 (2000)

Cited in 2

The emergence of protocadherin-PC expression during the acquisition of apoptosis-resistance by prostate cancer cells.

Chen M.-W., Vacherot F., De La Taille A., Gil-Diez-De-Medina S., Shen R., Friedman R.A., Burchardt M., Chopin D.K., Buttyan R.

View abstract

Cited for NUCLEOTIDE SEQUENCE [MRNA] (ISOFORM 3), INTERACTION WITH CTNNB1, SUBCELLULAR LOCATION, TISSUE SPECIFICITY
Tissue Prostatic carcinoma

PubMed
Europe PMC
Oncogene 21:7861-7871 (2002)

Cited in | Mapped to

Mammalian Genome 11, 906–914 (2000).
DOI: 10.1007/s003350010177

Mammalian Genome

Conservation of PCDHX in mammals; expression of human X/Y genes predominantly in brain

Patricia Blanco,[1] Carole A. Sargent,[1] Catherine A. Boucher,[1] Michael Mitchell,[2] Nabeel A. Affara[1]

[1]Human Molecular Genetics Group, Division of Cellular and Molecular Pathology, University of Cambridge, Department of Pathology, Tennis Court Road, Cambridge CB2 1QP, England, UK
[2]INSERM UNITE 491, Unite de Genetique Medicale et Developpment, Faculte de Medecine, 27 Boulevard Jean Moulin, 13385 Marseille, Cedex 05, France

Received: 27 March 2000 / Accepted: 2 June 2000

Abstract. Protocadherins are members of the cadherin superfamily involved in cell-cell interactions critical in the development of the central nervous system. This paper describes the isolation, sequence, and expression analysis of two novel protocadherin genes from the hominid specific Yp11.2/Xq21.3 block of homology between the sex chromosomes. The X-(PCDHX) and Y-linked (PCDHY) genes share 98.1% nucleotide and 98.3% amino acid identity and have an identical gene structure of six exons. The open reading frames of PCDHX and PCDHY encode proteins of 1025 and 1037 amino acids respectively and specify seven extracellular cadherin domains. Small differences in amino acid sequence affect regions that potentially have a large impact on function: thus, the X and Y genes may be differentiated in this respect. Sequence analysis of cDNA clones shows that both the X and Y loci are transcribed. RT-PCR expression analysis of mRNA from a variety of tissues and cell lines has demonstrated that both transcripts are expressed predominantly in the brain, with differential regional expression. From studies in the NTERA pluripotential cell line (which differentiates along neuronal and spermatogenic pathways in response to retinoic acid), it emerges that the X and Y-linked genes are regulated differently. This indicates that PCDHX and PCDHY possess different promoter regions. These findings suggest a role for PCDHX and PCDHY in the brain, consistent with the involvement of protocadherins in segmental brain morphogenesis and function. The implications of Y-linked genes expressed predominantly in tissues and organs other than the testis are considered within the context of the concept of sexual selection.

Introduction

It is generally accepted that the mammalian sex chromosomes emerged from an ancestral pair of autosomes and that sexually dimorphic chromosomes evolved as a chromosomal basis of sex determination and dosage compensation were established (Ohno 1967; Graves 1995; Charlesworth 1991). The suppression of recombination between the X and Y chromosomes has led to the degeneration of most genes located in the genetically isolated, non-recombining region of the Y, and it has been proposed that male-benefit genes assisting male reproduction and testicular function will accumulate in this chromosomal segment (Fisher 1931). Few other functional genetic loci are expected on the differential region of the Y Chromosome (Chr) that are not directly related to sex determination and spermatogenesis (or indirectly aid male reproduction through sexual selection) unless they encode genes that are required in diploid dose in both males and females. It has been suggested by Ferguson-Smith (1965) that the latter category will include X-Y homologous genes that are likely to underpin the gonadal and somatic stigmata characteristic of Turner syndrome.

In addition to the two pseudoautosomal regions, the present-day sex chromosomes retain several homologous loci and chromosomal intervals mapping to the non-recombining part of the Y Chr, some of which date from the ancestral homologs, and others that have arisen more recently during mammalian radiation (Lambson et al. 1992; Lahn and Page 1999; Vogt et al. 1997). Thus, functional loci shared between the sex chromosomes have been recruited to the Y at different points in evolution and provide a potential source of genes on the Y that can diverge and degenerate or be selected if they provide a benefit to the male. To date, no genes have been described that map in the major Yp11.2/Xq21.3 block of homology specific to hominids which is believed to have arisen as a result of a transposition from Xq after the divergence of chimpanzees and humans (Lambson et al. 1992; Page et al. 1984). Deletion studies have indicated that one or more genes responsible for the lymphoedemic anomalies of Turner syndrome lie in this block on Yp (Ogata et al. 1993). The homologous region on the X has been associated with premature ovarian failure (Sala et al. 1997) and epilepsy (Ryan et al. 1997).

This paper reports the analysis of the first functional X-Y homologous gene located within the hominid specific Yp11.2/Xq21.3 block of homology. From the sequence analysis of both the X and Y copies, it is clear that they form part of the cadherin gene superfamily and, more specifically, the protocadherin subfamily. Following the convention for protocadherin genes, the two loci have been termed PCDHX and PCDHY. Very recently, the partial sequence and structure of the X-linked gene was reported (Yoshida and Sugano 1999). Here, it is shown that the X and Y homologs have 98.1% nucleotide and 98.3% amino acid identity, reflecting the recent arrival of this gene on the Y Chr and, hence, the short period of time the homologs have been diverging. Surprisingly for a Y-linked transcript, the genes are expressed primarily in the brain, consistent with the finding that protocadherins are predominantly expressed in the central nervous system (Sano et al. 1993) and are believed to have an important role in the segmental development and function of the brain (Gumbiner 1996; Redies and Takeichi 1996).

Materials and Methods

PCR amplification. PCRs were performed typically in 20-µl volume containing 10 mM Tris-HCl, pH 9.0, 1.5 mM Mg^{2+}, 50 mM KCl, 0.1% Triton X-100, 0.01% (wt/vol) gelatin, 0.2 U *Taq* polymerase, 0.125 mM of each dNTP, 0.6 µM of each primer, 50–100 ng of template on a Techne

Correspondence to: N.A. Affara; E-mail: na@mole.bio.cam.ac.uk

The nucleotide sequence data reported in this paper have been submitted to EMBL and have been assigned the accession numbers AJ276805 (PCDHY) and AJ276804 (PCDHX).

Ensembl

Med Uni Graz

- Genome database for selected species (Homo sapiens and key model organisms)
- Important features
  - Graphical views
  - Gene Tree
  - Orthologues
  - Gene Variants
- Annotations
  - Gene Ontology: Biological Process, Molecular Function, Cellular Component
  - Phenotypes
  - Sources (PMIDs)

https://www.ensembl.org

GENEONTOLOGY Unifying Biology

Show/hide columns (1 hidden)

Accession	Term
GO:0002020	protease binding
GO:0003677	DNA binding
GO:0003697	single-stranded DNA binding
GO:0005515	protein binding

Ensemble: Link to controlled vocabularies and other databases

Med Uni Graz

Human (GRCh38.p14) ▾

Location: 13:32,315,086-32,400,268 | Gene: BRCA2

Gene-based displays

- Summary
  - Splice variants
  - Transcript comparison
  - Gene alleles
- Sequence
  - Secondary Structure
- Comparative Genomics
  - Genomic alignments
  - Gene tree
  - Gene gain/loss tree
  - Orthologues
  - Paralogues
- Ontologies
  - GO: Biological process
  - GO: Cellular component
  - GO: Molecular function
- Phenotypes
- Genetic Variation
  - Variant table
  - Variant image
  - Structural variants
- Gene expression
- Pathway
- Molecular interactions
- Regulation
- External references
- Supporting evidence
- ID History
  - Gene history

Gene: BRCA2 ENSG00000139618

Description	BRCA2 DNA repair associated [Source:HGNC Symbol;Acc:HGNC:1101]
Gene Synonyms	BRCC2, FACD, FAD, FAD1, FANCD, FANCD1, XRCC11
Location	Chromosome 13: 32,315,086-32,400,268 forward strand. GRCh38:CM000675.2
About this gene	This gene has 19 transcripts (splice variants), 173 orthologues and is associated with 123 phenotypes.
Transcripts	Show transcript table

HGNC

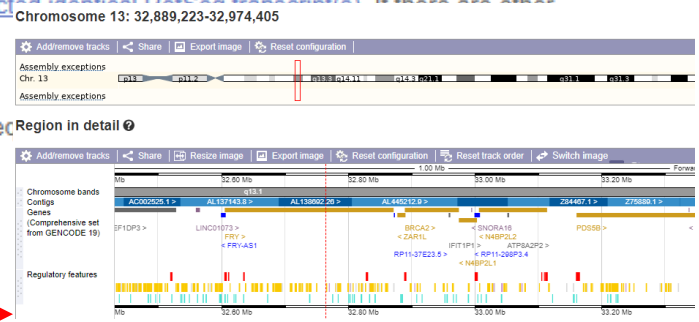
Gene data ▾ | Tools ▾ | Downloads ▾ | VGNC ▾ | Contact us ▾ | More ▾

Symbol report for BRCA2 Stable symbol

Summary

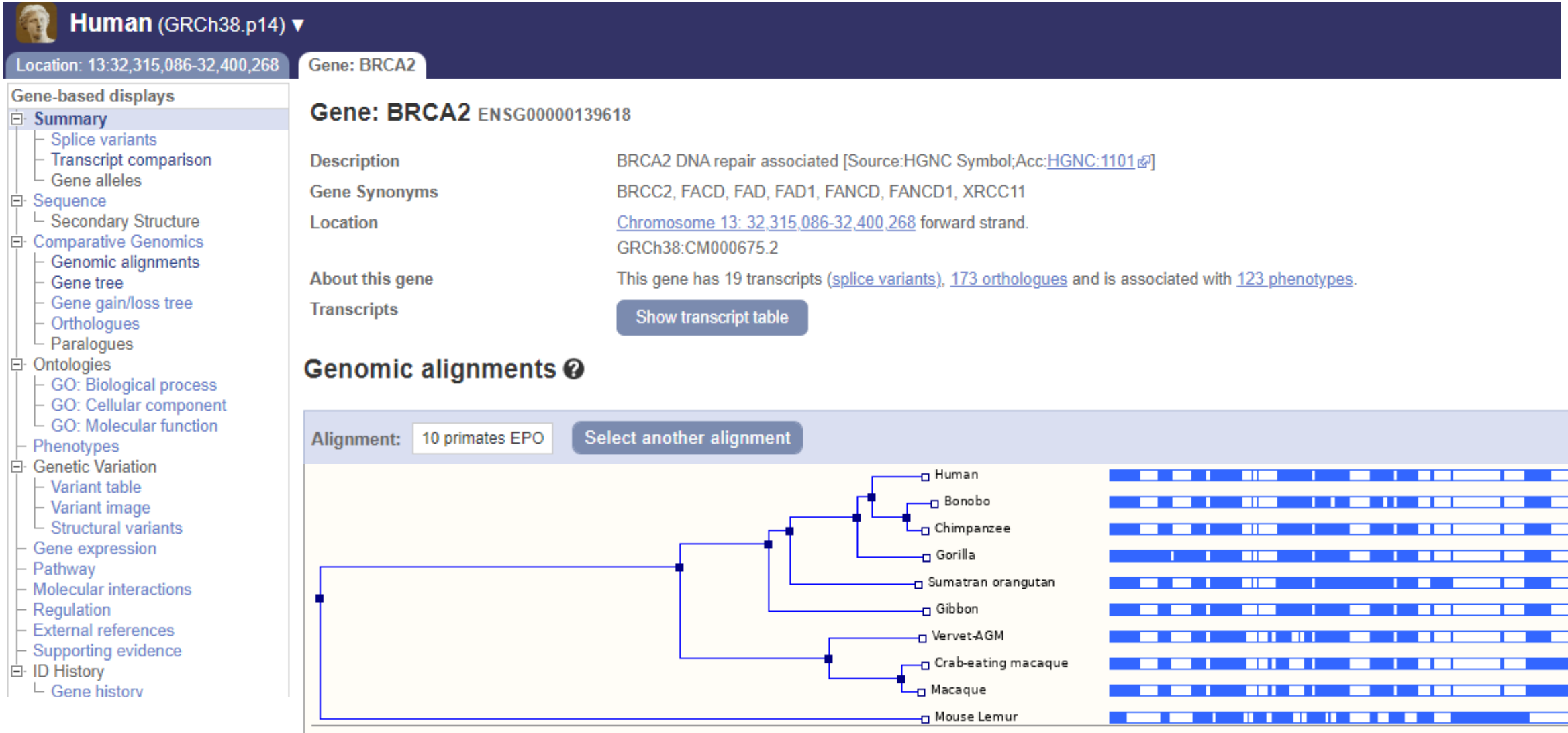
Name	BRCA2 (HGNC Symbol)
MANE	This gene contains MANE Select ENST00000380152, ENSP00000369497
UniProtKB	This gene has proteins that correspond to the following UniProtKB identifiers: P51587
RefSeq	This Ensembl/Gencode gene contains transcript(s) for which we have selected ... RefSeq transcripts available they will be in the External references table
CCDS	This gene is a member of the Human CCDS set: CCDS9344.1
LRG	LRG_293 provides a stable genomic reference framework for describing sec
Ensembl version	ENSG00000139618.19
Other assemblies	This gene maps to 32,889,223-32,974,405 in GRCh37 coordinates. View this locus in the GRCh37 archive: ENSG00000139618

UniProt



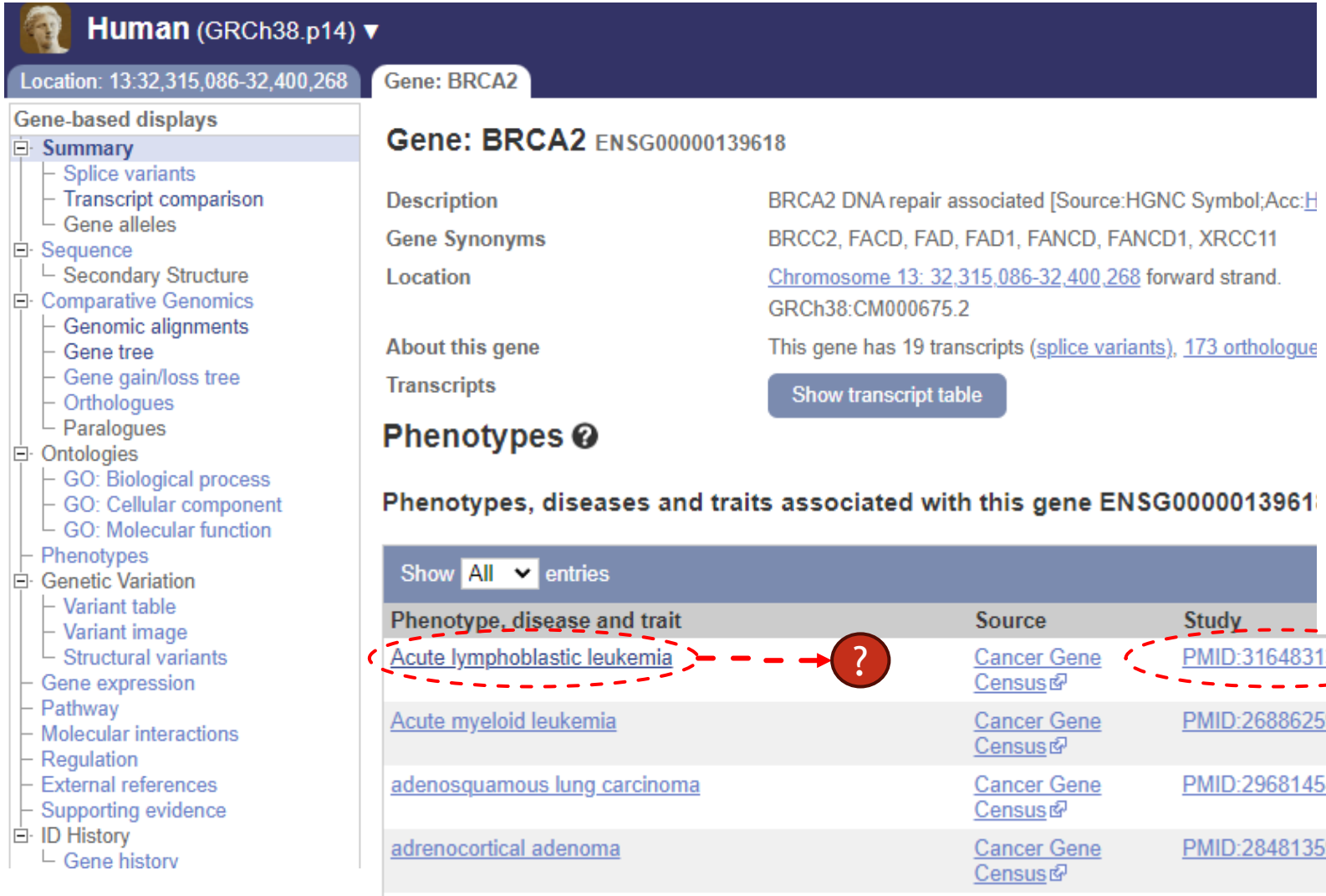
Ensemble: Genomic alignments

Med Uni Graz



Ensemble: Phenotype annotations

Med Uni



Pediatric ALL relapses after allo-SCT show high individuality, clonal dynamics, selective pressure, and druggable targets

Jessica I. Hoell,[1-3,*] Sebastian Ginzel,[1,2,4,5,*] Michaela Kuhlen,[1,2,*] Andreas Kloetgen,[1,2,6] Michael Gombert,[1,2] Ute Fischer,[1,2] Daniel Hein,[1,2] Salih Demir,[7,8] Martin Stanulla,[9] Martin Schrappe,[10] Udo zur Stadt,[11] Peter Bader,[12] Florian Babor,[1,2] Friedhelm Schuster,[1,2] Brigitte Strahm,[13] Julia Alten,[10] Anja Moericke,[10] Gabriele Escherich,[14] Arend von Stackelberg,[15] Ralf Thiele,[4] Alice C. McHardy,[6] Christina Peters,[16] Beat Bornhauser,[17] Jean-Pierre Bourquin,[17] Stefan Krause,[18] Juergen Enczmann,[18] Lüder Hinrich Meyer,[7] Cornelia Eckert,[2,15,19] Arndt Borkhardt,[1,2] and Roland Meisel[1,2]

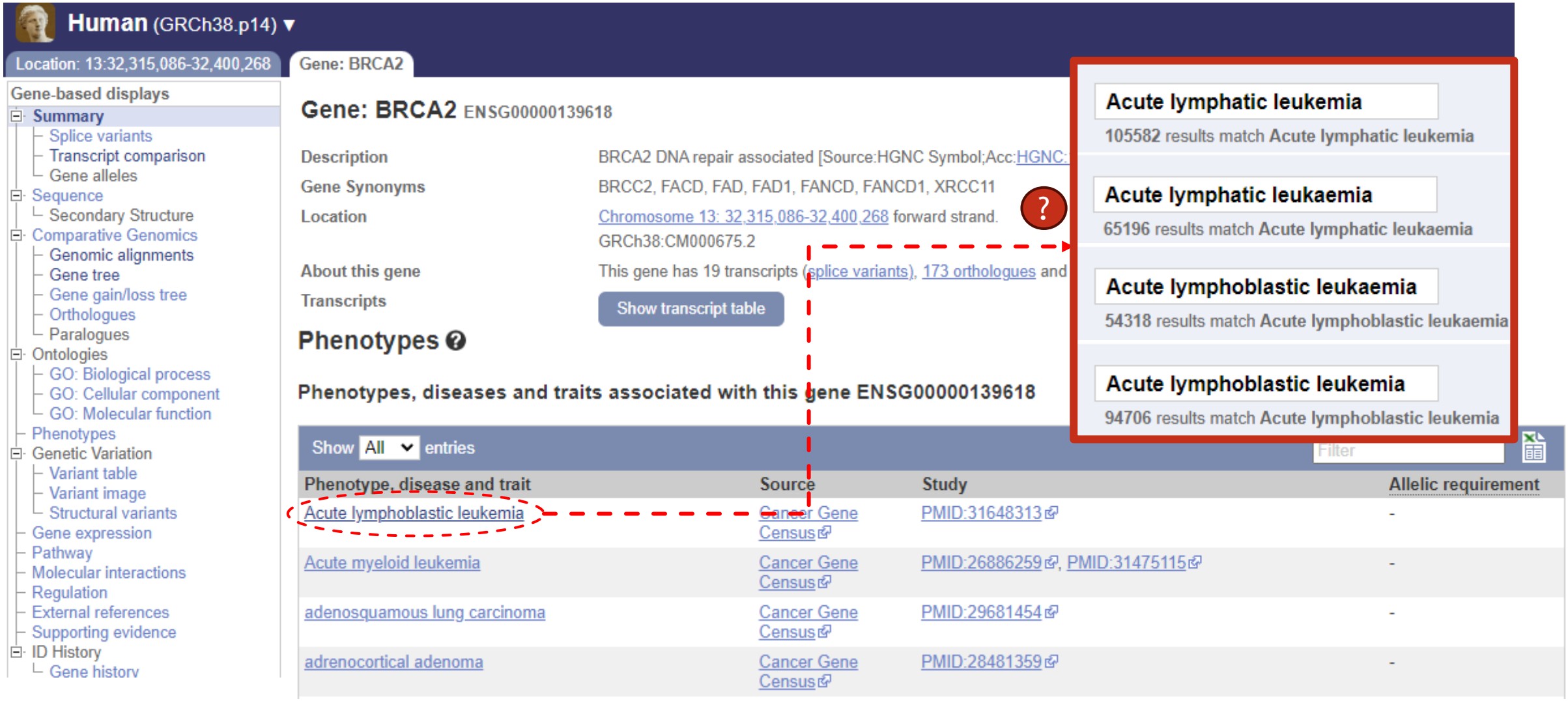
[1]Department of Pediatric Oncology, Hematology and Clinical Immunology, University Children's Hospital, Medical Faculty, Heinrich-Heine University, Düsseldorf, Germany; [2]German Consortium for Translational Cancer Research (DKTK), Partner site Essen/Düsseldorf, Heidelberg, Germany; [3]Department of Pediatric Hematology and Oncology, Martin Luther University Halle-Wittenberg, Halle, Germany; [4]Department of Computer Science, Bonn-Rhine-Sieg University of Applied Sciences, Sankt-Augustin, Germany; [5]Fraunhofer Institut für Intelligente Analyse und Informationssysteme, Schloss Birlinghoven, St. Augustin, Germany; [6]Computational Biology of Infection Research, Helmholtz Center for Infection Research, Braunschweig, Germany; [7]Department of Pediatrics and Adolescent Medicine, Ulm University Medical Center, Ulm, Germany; [8]International Graduate School of Molecular Medicine, Ulm University, Ulm, Germany. [9]Pediatric Hematology and Oncology, Hannover Medical School, Hannover, Germany; [10]Department of Pediatrics, University Medical Center Schleswig-Holstein, Kiel Campus, Kiel, Germany; [11]Center for Diagnostics, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [12]University Hospital for Children and Adolescents, Frankfurt am Main, Germany; [13]Division of Pediatric Hematology and Oncology, Department of Pediatrics and Adolescent Medicine, Medical Center–University of Freiburg, Faculty of Medicine, University of Freiburg, Freiburg, Germany; [14]Clinic of Pediatric Hematology and Oncology, University Medical Center Hamburg-Eppendorf, Hamburg, Germany; [15]Pediatric Hematology and Oncology, Charité University Hospital, Berlin, Germany; [16]St. Anna Children's Hospital, Vienna, Austria; [17]Department of Pediatric Oncology and Children's Research Centre, University Children's Hospital Zürich, Zürich, Switzerland; [18]Institute for Transplantation Diagnostics and Cell Therapeutics, University Hospital of Düsseldorf, Düsseldorf, Germany; and [19]German Cancer Research Center, Heidelberg, Germany

Key Points

- Pediatric ALL relapses after allogeneic stem cell transplantation display highly diverse, dynamic and patient-individual genetic lesions.
- Nine of 10 relapsing pediatric transplant recipients present with genetic alterations for which novel targeted therapies are available.

Survival of patients with pediatric acute lymphoblastic leukemia (ALL) after allogeneic hematopoietic stem cell transplantation (allo-SCT) is mainly compromised by leukemia relapse, carrying dismal prognosis. As novel individualized therapeutic approaches are urgently needed, we performed whole-exome sequencing of leukemic blasts of 10 children with post–allo-SCT relapses with the aim of thoroughly characterizing the mutational landscape and identifying druggable mutations. We found that post–allo-SCT ALL relapses display highly diverse and mostly patient-individual genetic lesions. Moreover, mutational cluster analysis showed substantial clonal dynamics during leukemia progression from initial diagnosis to relapse after allo-SCT. Only very few alterations stayed constant over time. This dynamic clonality was exemplified by the detection of thiopurine resistance-mediating mutations in the nucleotidase *NT5C2* in 3 patients' first relapses, which disappeared in the post–allo-SCT relapses on relief of selective pressure of maintenance chemotherapy. Moreover, we identified *TP53* mutations in 4 of 10 patients after allo-SCT, reflecting acquired chemoresistance associated with selective pressure of prior antineoplastic treatment. Finally, in 9 of 10 children's post–allo-SCT relapse, we found alterations in genes for which targeted therapies with novel agents are readily available. We could show efficient targeting of leukemic blasts by APR-246 in 2 patients carrying TP53 mutations. Our findings shed light on the genetic basis of post–allo-SCT relapse and may pave the way for unraveling novel therapeutic strategies in this challenging situation.

Ensemble: Phenotype annotations



NCBI Databases

- NCBI databases ("Entrez")
  - Using platform known from Pubmed
  - Interlinked
- Important domains
  - Protein sequences
  - Genes
  - Gene expression maps
  - Complete genomes
  - Human genetic disorders (OMIM)
  - Chemicals (substance, Compound, BioAssay)

https://www.ncbi.nlm.nih.gov/search/

Medline & PubMed

- MEDLINE is the database, PubMed the search interface
  - July 2024: 31,521,180 million + 5,895,750 “in process” (not indexed)
  - 90% English-language articles
  - approx. 6,000 Publication organs (Journals, Proceedings, Books)
- Beyond MEDLINE
  - IN-PROCESS (publications in "waiting position")
  - MeSH (Medical Subject Headings) : Keyword Thesaurus
- Manual indexing by NLM experts (manual)
  - MeSH headings / subheadings
  - Publication type
  - Substances, enzymes, organisms

National Library of Medicine
National Center for Biotechnology Information

Log in

PubMed®

Search

Advanced

PubMed® comprises more than 37 million citations for biomedical literature from MEDLINE, life science journals, and online books. Citations may include links to full text content from PubMed Central and publisher web sites.

Learn

About PubMed
FAQs & User Guide
Finding Full Text

Find

Advanced Search
Clinical Queries
Single Citation Matcher

Download

E-utilities API
FTP
Batch Citation Matcher

Explore

MeSH Database
Journals

National Library of Medicine
National Center for Biotechnology Information

Log in

PubMed®

PubMed Advanced Search Builder

PubMed®

User Guide

Add terms to the query box

All Fields

Enter a search term

ADD

Show Index

Query box

Enter / edit your search query here

Search

History and Search Details

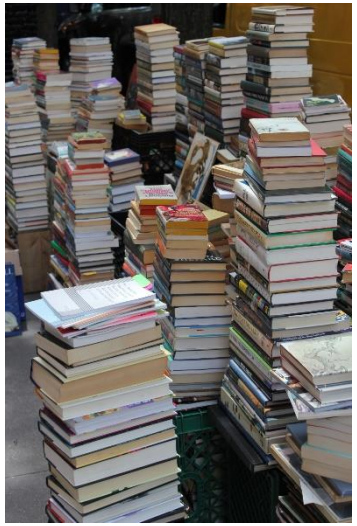
Your history is currently empty! As you use PubMed your recent searches will appear here.

Pubmed vs. Medline vs. MeSH

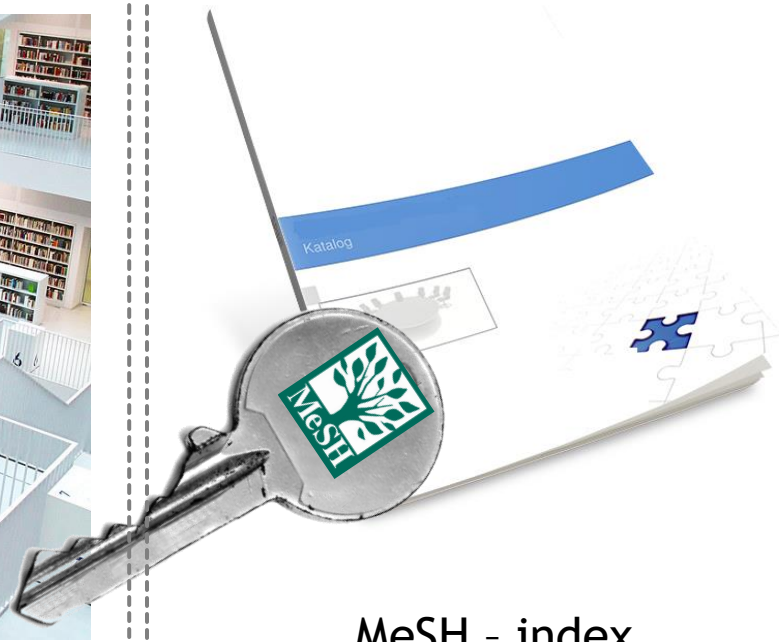
Med Uni Graz

PubMed

National Library of Medicine NLM



MEDLINE



MeSH – index

National Library of Medicine
National Center for Biotechnology Information

Log in

MeSH

MeSH

Search

Limits Advanced

Help

MeSH

MeSH (Medical Subject Headings) is the NLM controlled vocabulary thesaurus used for indexing articles for PubMed.

Using MeSH

Help

Tutorials

More Resources

E-Utilities

NLM MeSH Homepage

Medical Subject Headings (MeSH)

- One of the most important CVs in biomedicine: “Key” to MEDLINE
- 20,000 keywords, hierarchically structured:
  - Documents annotated with more specific keywords will also be found using more general keywords.
- Every MeSH term
  - Has a preferred term (“Hemorrhage")
  - Has synonyms and hyponyms (Entry Terms): “Bleeding“, “Haemorrhage”.
  - Is in one or more "trees": "Tuberculosis, Pulmonary" both under "Lung Diseases" and "Bacterial Infections"
  - Can be further specified by "subheadings", e.g. Tuberculosis, Pulmonary / * drug therapy.

https://www.ncbi.nlm.nih.gov/mesh/

Fields | Values

```
PMID- 25643895
STAT- MEDLINE
TI  - Lower hazard ratio for death in women with cerebral hemorrhage.
PG  - 59-64
LID - 10.1111/ane.12359 [doi]
AB  - OBJECTIVES: The aim of the study was to clarify the hazard ratio for death within
      30 days after stroke comparing women to men. MATERIAL AND METHODS: We reviewed
      all stroke patients registered in the Kyoto Stroke Registry (from January 1999 to
      December 2009) in Japan. Hazard ratio (HR) for death and 95% confidence interval
      were calculated by the Cox regression in stroke and in each stroke subtype:
      cerebral infarction(…)
CI  - (c) 2015 John Wiley & Sons A/S. Published by John Wiley & Sons Ltd.
FAU - Shigematsu, K
AU  - Shigematsu K
AUID- ORCID: http://orcid.org/0000-0003-3747-8115
AD  - Department of Neurology, National Hospital Organization, Minami Kyoto Hospital,
      Kyoto, Japan.
FAU - Watanabe, Y
AU  - Watanabe Y
AD  - Department of Epidemiology for Community Health and Medicine, Graduate School of
      Medical Science, Kyoto Prefectural University of Medicine, Kyoto, Japan.
FAU - Nakano, H
AD  - Department of Neurosurgery, Kyoto Kidugawa Hospital, Kyoto, Japan.
CN  - Kyoto Stroke Registry Committee
LA  - eng
PT  - Journal Article
TA  - Acta Neurol Scand
SB  - IM
MH  - Adult
MH  - Aged
MH  - Cerebral Hemorrhage/etiology/*mortality
MH  - Female
MH  - *Sex Characteristics
MH  - Stroke/complications/*mortality
MH  - Subarachnoid Hemorrhage/etiology/mortality
OT  - cerebrovascular diseases
OT  - strokes
EDAT- 2015/02/04 06:00
MHDA- 2015/11/11 06:00
CRDT- 2015/02/04 06:00
PHST- 2014/11/11 00:00 [accepted]
PHST- 2015/02/04 06:00 [pubmed]
PHST- 2015/11/11 06:00 [medline]
SO  - Acta Neurol Scand. 2015 Jul;132(1):59-64. doi: 10.1111/ane.12359. Epub 2015 Feb
```

MeSH

Title

Abstract

Authors

Affiliation

Pub Type

MeSH

Dates

Source

Med Uni Graz

Tuberculosis, Pulmonary

MYCOBACTERIUM infections of the lung.
Year introduced: TUBERCULOSIS IN CHILDHOOD was heading 1963-1989

PubMed search builder options
Subheadings:

- analysis
- anatomy and histology
- blood
- blood supply
- cerebrospinal fluid
- chemically induced
- chemistry
- classification
- complications
- congenital
- cytology
- diagnosis
- diet therapy
- drug therapy
- economics
- embryology
- enzymology
- epidemiology
- ethnology
- etiology
- genetics
- history
- immunology
- legislation and jurisprudence
- metabolism
- microbiology
- mortality
- nursing
- organization and administration
- parasitology
- pathology
- pharmacology
- physiology
- physiopathology
- prevention and control
- psychology
- radiography
- radionuclide imaging
- radiotherapy
- rehabilitation
- statistics and numerical data
- surgery
- therapeutic use
- therapy
- transmission
- ultrasonography
- urine
- veterinary
- virology

MeSH-Subheadings

MeSH-Entry Terms

MeSH Unique ID: D014397
Entry Terms:

- Tuberculoses, Pulmonary
- Pulmonary Tuberculoses
- Pulmonary Tuberculosis
- Pulmonary Consumption
- Consumption, Pulmonary
- Consumptions, Pulmonary
- Pulmonary Consumptions
- Pulmonary Phthisis
- Phthises, Pulmonary
- Phthisis, Pulmonary
- Pulmonary Phthises

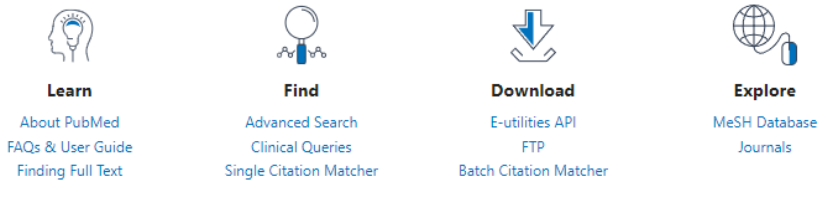
MeSH-Trees

All MeSH Categories
Diseases Category
Bacterial Infections and Mycoses
Bacterial Infections
Gram-Positive Bacterial Infections
Actinomycetales Infections
Mycobacterium Infections
Tuberculosis
Tuberculosis, Pulmonary
Silicotuberculosis

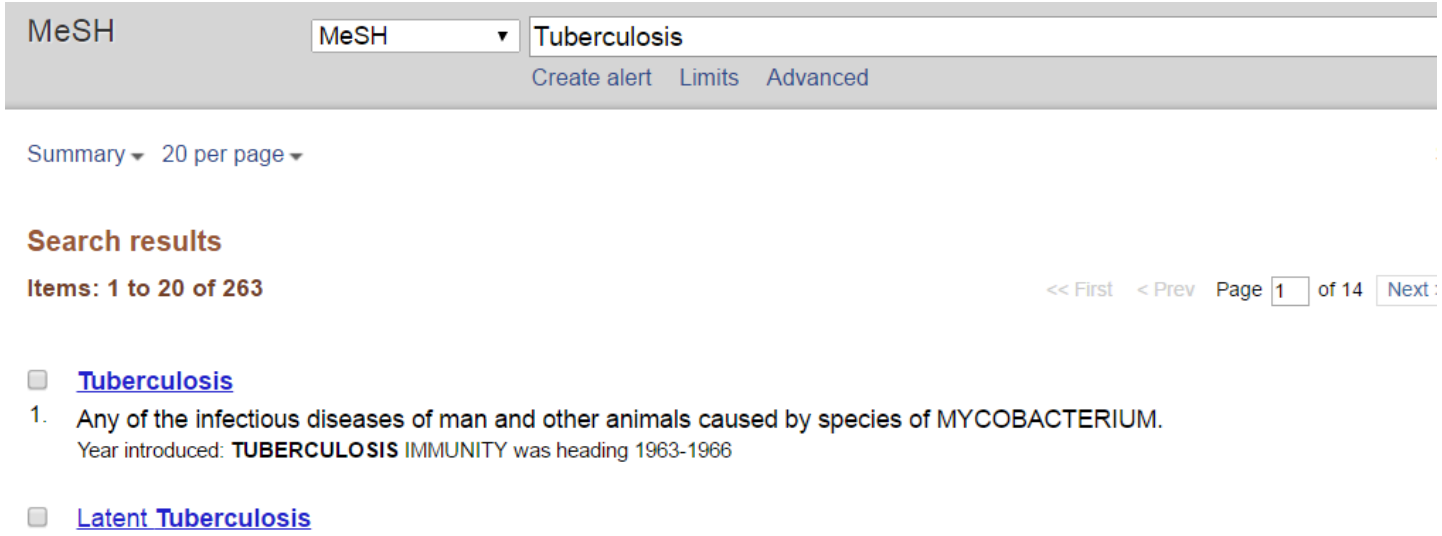
All MeSH Categories
Diseases Category
Respiratory Tract Diseases
Lung Diseases
Tuberculosis, Pulmonary
Silicotuberculosis

MeSH in PubMed: Querying

- Jump from Pubmed to MeSH

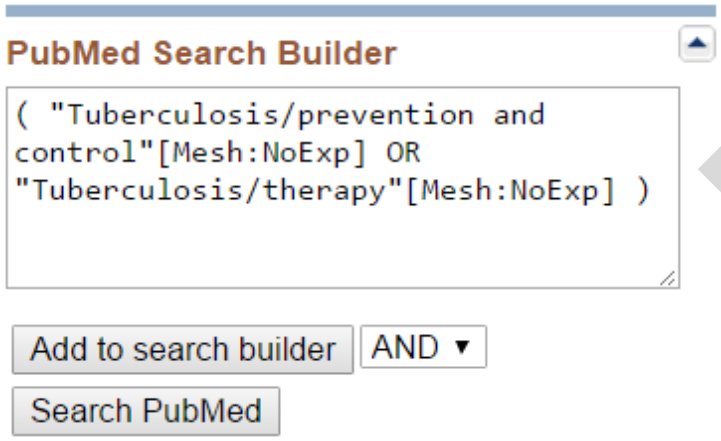


- Complex Search: build from individual queries :



MeSH in PubMed: refine query

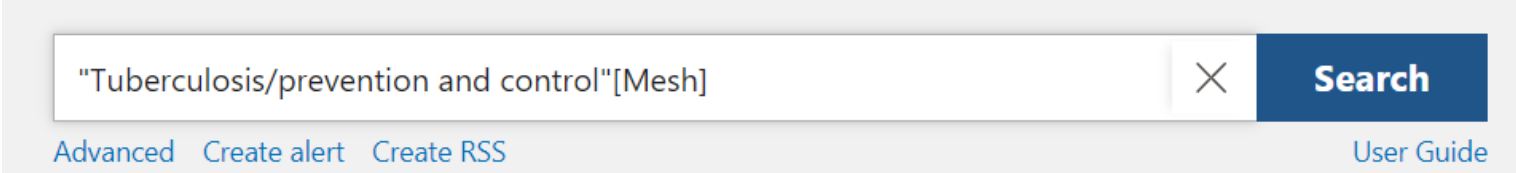
- If multiple hits: click correct one
- Restrict search:
  - By MeSH Subheadings
  - By "MeSH Major Topic“ (*)
- Add to search builder



MeSH-based query using PubMed query syntax

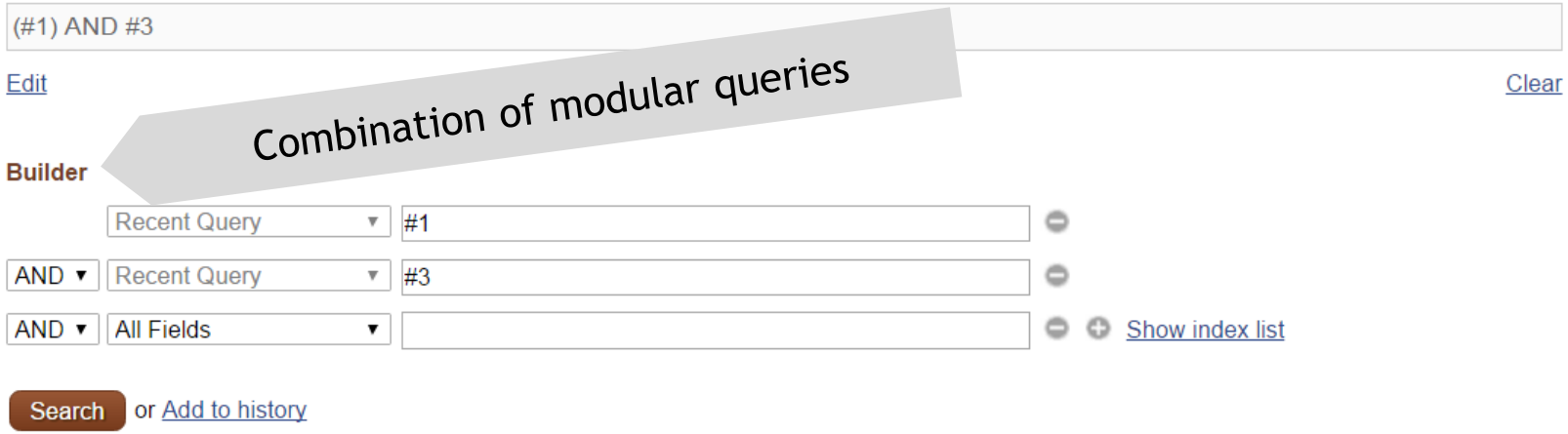
MeSH in PubMed: Complex Queries

- Division into individual search steps, each of which generated with the SearchBuilder. Typically: PICO
- Use „Advanced“ for modularising complex PubMed queries

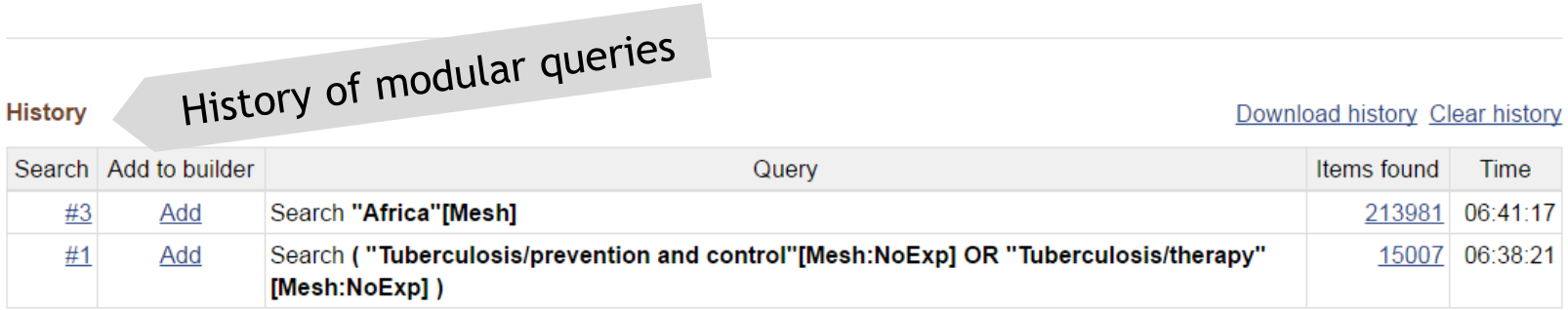


- In "History" each individual query is numbered for the creation of combined queries
- Complex searches may include all fields of a MEDLINE record, e.g. Authors, journals, time periods, etc. These are also selected in the "Builder".

MeSH in PubMed: Use of "Builder"



- Division into individual search steps, which one generates using SearchBuilder.
- Using the logical operators "OR“ (disjunction, union), "AND“ (conjunction, intersection), "NOT" (complement)



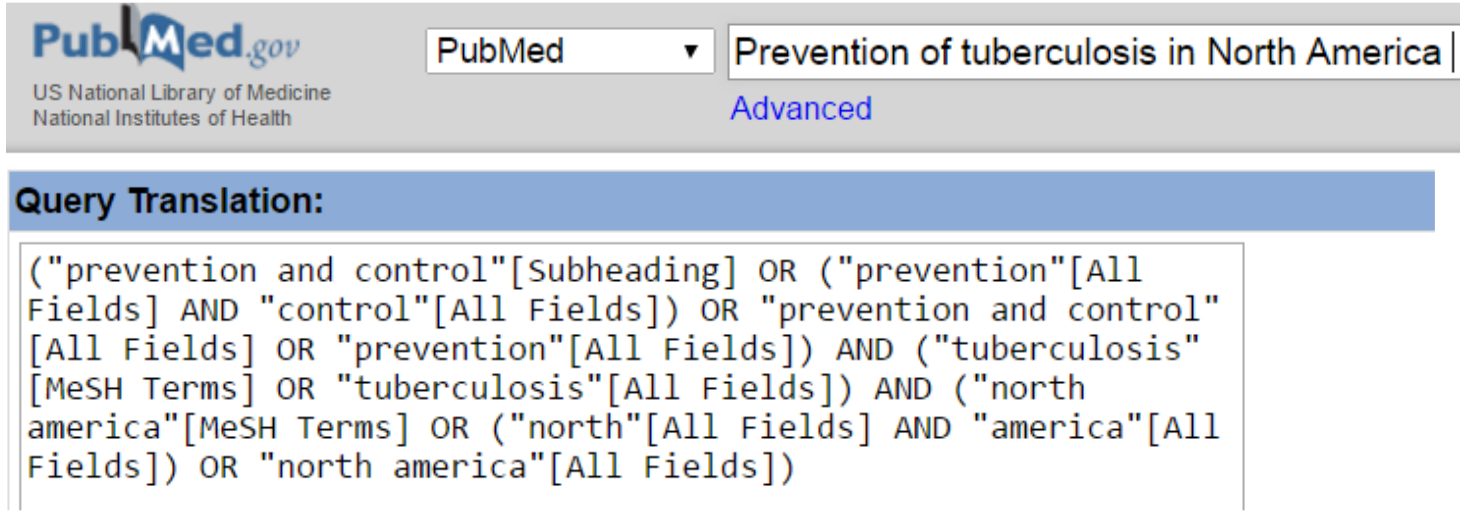
Search	Add to builder	Query	Items found	Time
#3	Add	Search **"Africa"[Mesh]**	213981	06:41:17
#1	Add	Search **("Tuberculosis/prevention and control"[Mesh:NoExp] OR "Tuberculosis/therapy" [Mesh:NoExp])**	15007	06:38:21

PubMed: Free-text search

- Free text search as alternative / supplement:
  - Include articles not yet been indexed in MEDLINE
  - Not sufficiently accurate MeSH terms
  - Search in foreign language titles
  - Doubts about the completeness of MeSH index

2023: 63 % are MeSH indexed
2003: 94 % are MeSH indexed
1983: 96 % are MeSH indexed

- Automatic term mapping: produces a combination of free text and MeSH search
  - Usually suboptimal result, but good start, particularly for beginners



PubMed: Free text search principles

Med Uni Graz

- Specify using "Field tags" Text Word [tw]
- Synonyms und hypernyms and all variants have to be added manually!
- Truncation operator (wildcard) "*":
  - **`cholangio*`** retrieves **`cholangiohepatography`**, **`cholangiovenous`**, …
- Phrase search (double quotes):
  - **`CD8 T cell memory`** more but less specific hits compared to **`"CD8 T cell memory"`**.
  - Everything between an operator and a tag is interpreted as a phrase:
    `AND CD8 T cell memory [tw]`
- Challenges of free text search
  - Synonyms and hyponyms have to be considered and entered (**`OR`** operator).
  - Spelling variants must be considered: **`esophagus`** (American) and **`oesophagus`** (British)
  - Ambiguous phrases: low precision - Search only in the title leads: low recall.

PubMed / MEDLINE

Med Uni Graz

- Before using it seriously: watch Tutorials
- Automatic Term mapping: suboptimal results but good for a first try
- MeSH search: captures only manually indexed content. Recent publications are not yet indexed!
- Free text query: captures also non-indexed content
  - Require manual addition of synonyms, hypernyms
  - Short free text searches are normally bad
- Both MeSH and free-text search require
  - Boolean operators **`AND OR NOT`**
  - Parentheses (of more than one operator)
  - Field tags
- Publication types only cover indexed content!
- Modularize your query – assess plausibility of partial searches (e.g. is a known relevant article retrieved)?
- Minor mistakes may completely destroy your search!
- Don't take PubMed querying easy !

Usually, a free-text search is either short or good!

Clinical databases

- ClinicalTrials.gov as one example for clinical trials
- UpToDate as an example for database support at the doctor's workplace
- Just to mention:
  - also clinical or epidemiological registries are databases
  - and hospital information systems heavily rely on database technology

ClinicalTrials.gov

NIH U.S. National Library of Medicine

ClinicalTrials.gov

- Database of clinical studies
- Maintained by the U.S. National Library of Medicine
- Nearly 300,000 records in 2019
- No use of index terms
- Synonym matching in the background

- Important fields:
  - Dates, locations
  - Primary outcome (e.g. success of treatment)
  - Secondary outcome (e.g. costs, complication, morbidity)
  - Text description
  - Study type and design (e.g. RCT)
  - Population, inclusion, exclusion criteria
  - Intervention (drug, surgery, …)
  - Publications
  - Recruitment information
  - Sponsor

ClinicalTrials.gov is a place to learn about clinical studies from around the world.

The U.S. government does not review or approve the safety and science of all studies listed on this website.

Read our full disclaimer for details.

Focus Your Search (all filters optional)

Condition/disease

Other terms

Intervention/treatment

Location

Search by address, city, state, or country and select from the dropdown list

Study Status

- (●) All studies
- ( ) Recruiting and not yet recruiting studies

More Filters +

Search

ClinicalTrials.gov
National Library of Medicine
8600 Rockville Pike, Bethesda, MD
20894

Descriptive Information	
Brief Title ICMJE	Appendicectomy Versus Antibiotics in the Treatment of **Acute** Uncomplicated **Appendicitis**
Official Title ICMJE	Study of Surgical Treatment (Open Appendicectomy) Versus Antibiotic Treatment (Ertapenem) in the Treatment of **Acute** Uncomplicated **Appendicitis**
Brief Summary	Appendicectomy has been the treatment of acute appendicitis for over a hundred years. Appendicectomy, however, includes operative and postoperative risks despite being a "routine" operation. At the same time other similar intra-abdominal infections, such as diverticulitis, are treated with antibiotics. There have been some encouraging reports on successful treatment of appendicitis with antibiotics and it has been estimated that operative treatment might be necessary for only 15 - 20 % of patients with acute appendicitis. The aim of this randomized prospective study is to compare operative treatment (open appendicectomy) with conservative treatment with antibiotics (ertapenem, Invanz). Before randomization acute uncomplicated appendicitis is diagnosed with a CT scan.The hypothesis of the study is that the majority of patients with uncomplicated acute appendicitis can be treated successfully with antibiotics and unnecessary appendicectomies can be avoided.
Detailed Description	*Not Provided*
Study Type ICMJE	Interventional
Study Phase	Not Applicable
Study Design ICMJE	Allocation: Randomized Intervention Model: Parallel Assignment Masking: None (Open Label) Primary Purpose: Treatment
Condition ICMJE	**Acute Appendicitis**
Intervention ICMJE	• Procedure: Appendicectomy Standard appendicectomy • Drug: Ertapenem ertapenem 1g x 1 i.v.for three days + after discharge levofloxacin 500 mg 1 x 1 + metronidazole 500 mg 1x3 for 7 days p.o.
Study Arms	• Active Comparator: Operative treatment Regular open appendicectomy Intervention: Procedure: Appendicectomy • Active Comparator: Antibiotic treatment Ertapenem 1 g i.v. x 1 three days Intervention: Drug: Ertapenem

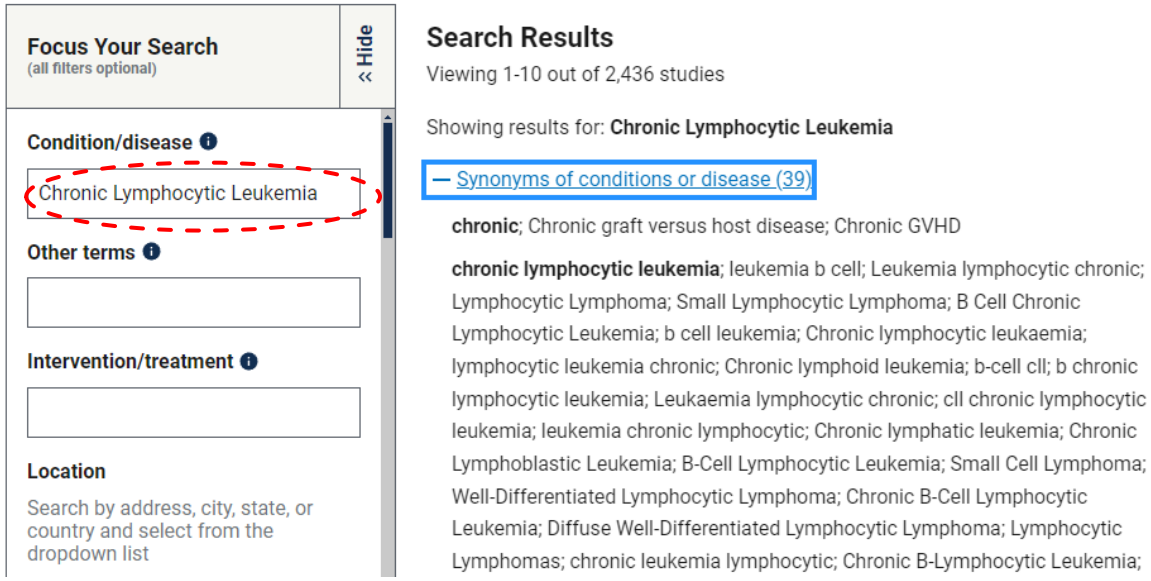
Med Uni Graz

Tracking Information	
First Submitted Date ICMJE	November 30, 2009
First Posted Date ICMJE	December 1, 2009
Results First Submitted Date	August 21, 2016
Results First Posted Date	February 23, 2017
Last Update Posted Date	June 28, 2018
Study Start Date ICMJE	November 2009
Actual Primary Completion Date	June 2012 (Final data collection date for primary outcome measure)
Current Primary Outcome Measures ICMJE (submitted: January 2, 2017)	The Success of Antibiotic and Surgical Treatment in the Treatment of Acute Uncomplicated Appendicitis [Time Frame: Up to 10 years] A successful treatment is determined by resolution of the appendicitis by means of the assigned treatment.
Original Primary Outcome Measures ICMJE (submitted: November 30, 2009)	The success of antibiotic treatment in patients with acute uncomplicated appendicitis [Time Frame: 1.-3. days, 1 week, 2 months, 1 year, 3,5,10 y]
Change History	Complete list of historical versions of study NCT01022567 on ClinicalTrials.gov Archive Site
Current Secondary Outcome Measures ICMJE (submitted: January 2, 2017)	• The Possible Complications, Morbidity and Mortality of Operative and Conservative Treatment [Time Frame: 1 year] • The Direct and Indirect Costs of Both Treatment Arms [Time Frame: 1 year] • The Recurrence of Conservatively Treated Appendicitis [Time Frame: up to 10 years]
Original Secondary Outcome Measures ICMJE (submitted: November 30, 2009)	• The Possible Complications, Morbidity and Mortality of Operative and Conservative Treatment [Time Frame: Same as primary outcome measure] • The Direct and Indirect Costs of Both Treatment Arms [Time Frame: Same as primary outcome measure] • The Recurrence of Conservatively Treated Appendicitis [Time Frame: Same as primary outcome measure]
Current Other Outcome Measures ICMJE	*Not Provided*
Original Other Outcome Measures ICMJE	*Not Provided*

Recruitment Information	
Recruitment Status ICMJE	Active, not recruiting
Actual Enrollment ICMJE (submitted: December 1, 2014)	530
Original Actual Enrollment ICMJE (submitted: November 30, 2009)	600
Estimated Study Completion Date	December 2025
Actual Primary Completion Date	June 2012 (Final data collection date for primary outcome measure)
Eligibility Criteria ICMJE	Inclusion Criteria: • Age range from 18 to 60 years • CT scan diagnosed uncomplicated acute appendicitis Exclusion Criteria: • Age under 18 years or age over 60 years • Pregnancy or breast-feeding • Allergy to contrast media or iodine • Renal insufficiency • metformin medication (DM) • Peritonitis (a perforated appendix) • Lack of co-operation (unable to give consent) • A severe other medical condition • CT-scan: other diagnosis, fecal lithiasis in appendix, perforation, abscess, suspicion of a tumour
Sex/Gender	Sexes Eligible for Study: All
Ages	18 Years to 60 Years (Adult)
Accepts Healthy Volunteers	Yes
Contacts ICMJE	*Contact information is only displayed when the study is recruiting subjects*
Listed Location Countries ICMJE	Finland

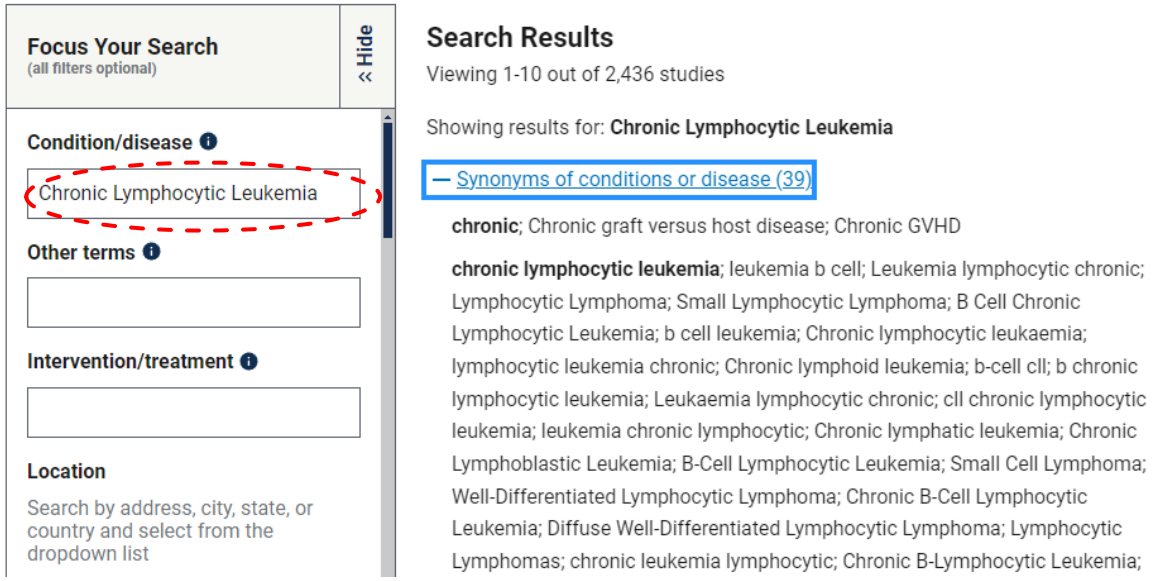
Med Uni Graz

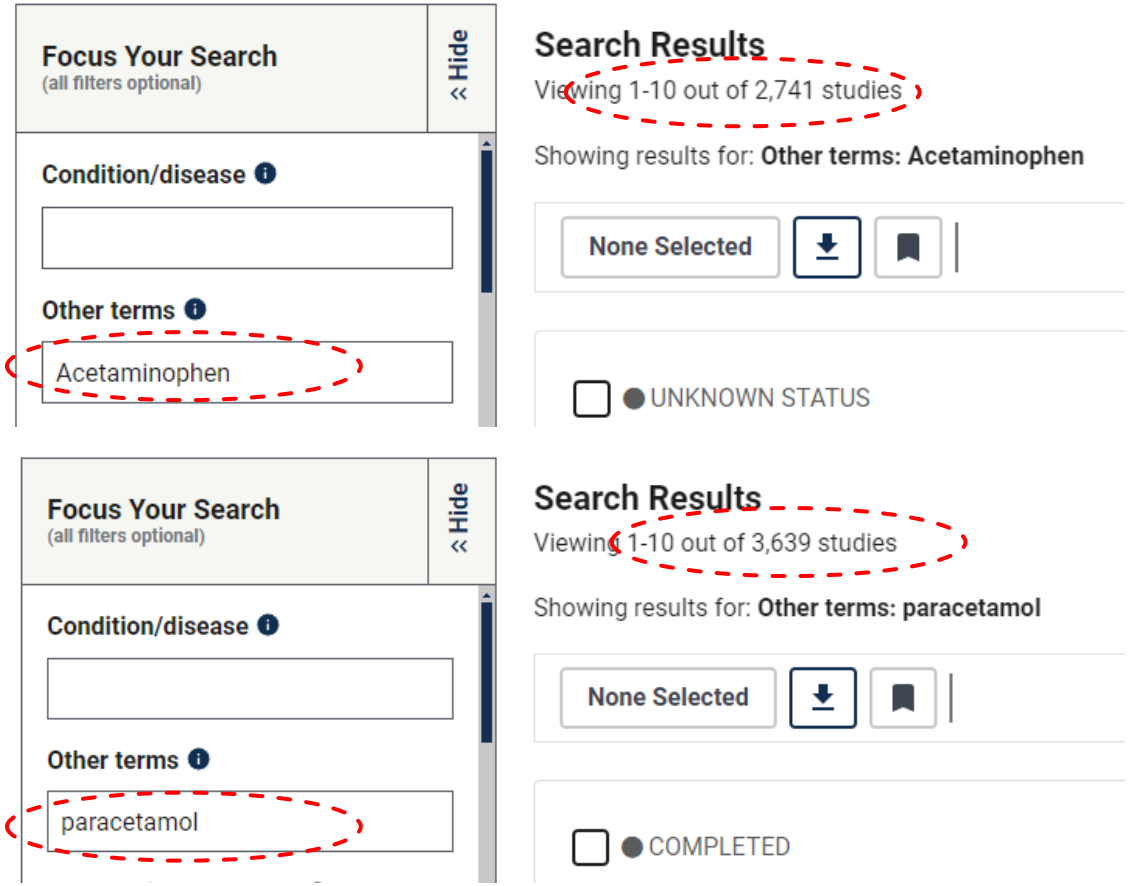
ClinicalTrials.gov - Query



- All queries are free text
- Queries for conditions / diseases:
  Synonym matching in the background
- No other semantic processing
- Responsibility for comprehensive free text mapping up to the user !

ClinicalTrials.gov - Query





- All queries are free text
- Queries for conditions / diseases:
  Synonym matching in the background
- No other semantic processing
- Responsibility for comprehensive free text mapping up to the user !

Take home messages

- Biomedical databases are great tools for life science and clinical research
- Several challenges:
  - Understand which databases you need for your research
  - Get familiar with the internal structure of the database as well as the functionality of the frontend
  - Be aware that CVs are used by humans and humans err !
  - Understand exactly the structure of a CV
  - Get acquainted with the query syntax, particularly in Pubmed
  - Good free text queries require excellent understanding of the domain and its (English) terminology. Independent of its use in MEDLINE, MeSH can also be used as a dictionary to retrieve synonyms and hyponyms
  - Database systems have no inbuilt intelligence – they mostly do not recognize even the most obvious type or spelling mistakes
  - Use tutorials ad videos and team up with your colleagues