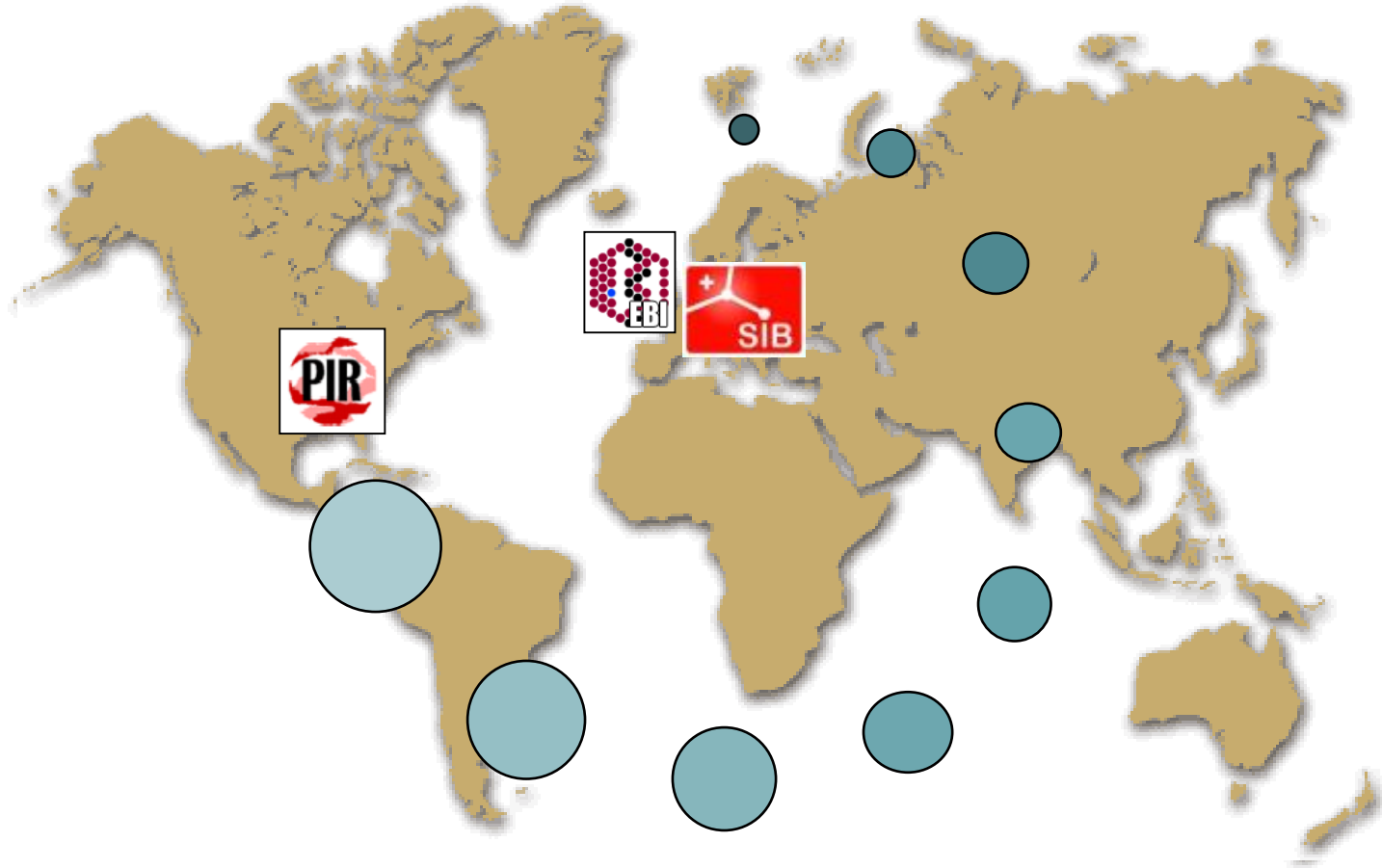


UniProt in 2016

Alex Bateman – EMBL-EBI

Cathy Wu – PIR

Ioannis Xenarios - SIB

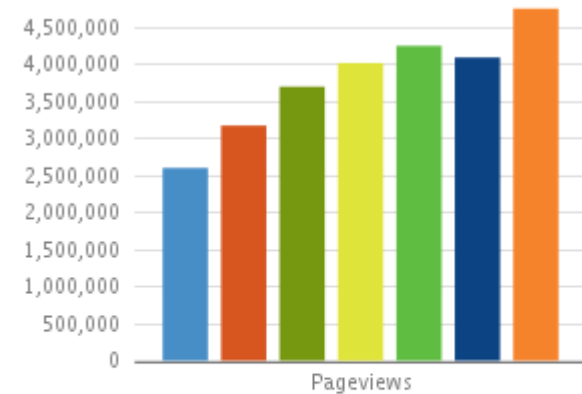
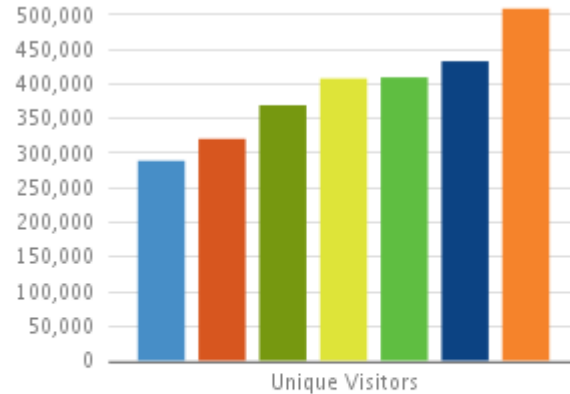
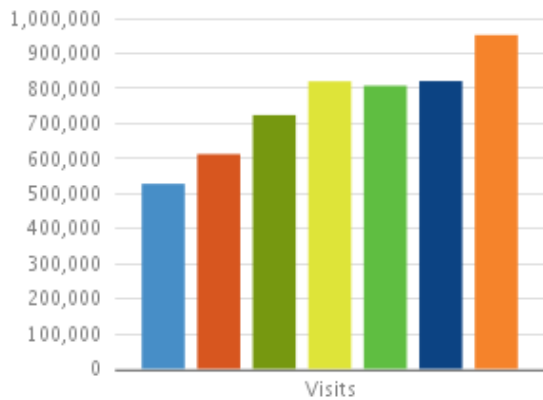


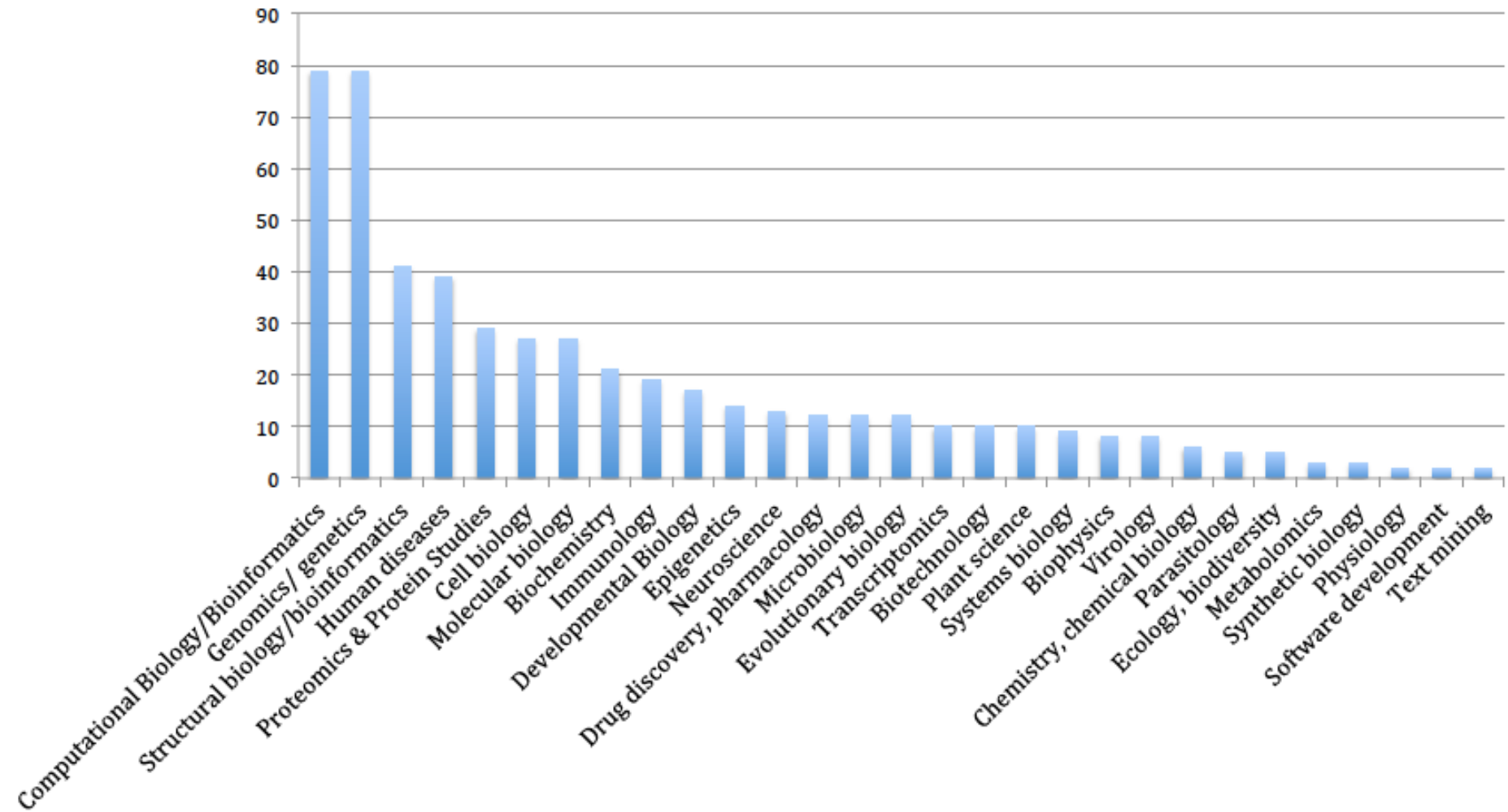
The mission of UniProt is to provide the scientific community with a comprehensive, high-quality and freely accessible resource of **protein sequence** and **functional information**.

UniProt Google Analytics statistics

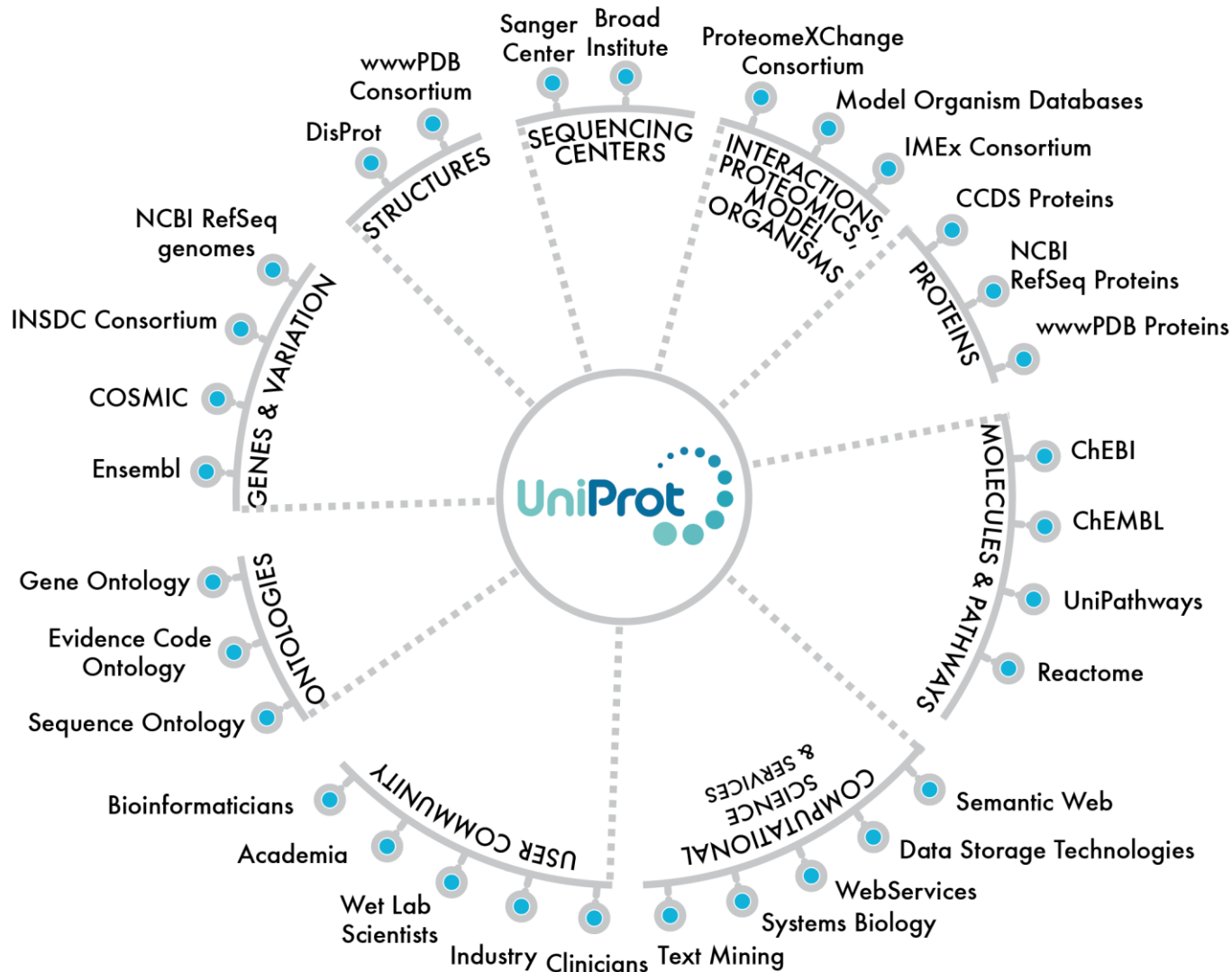
Period / monthly average	Visits	Unique visitors	Pageviews
June 2010 - May 2011	612,905	320,892	3,177,758
June 2011 - May 2012	724,286	369,485	3,703,560
June 2012 - May 2013	820,623	408,244	4,022,786
June 2013 - May 2014	808,135	409,848	4,255,675
March 2014 - Feb. 2015*	821,368	433,136	4,097,871
March 2015 - Feb. 2016	952,837	509,278	4,758,278

* Different period due to new NIH grant period; these numbers were reported to the NIH





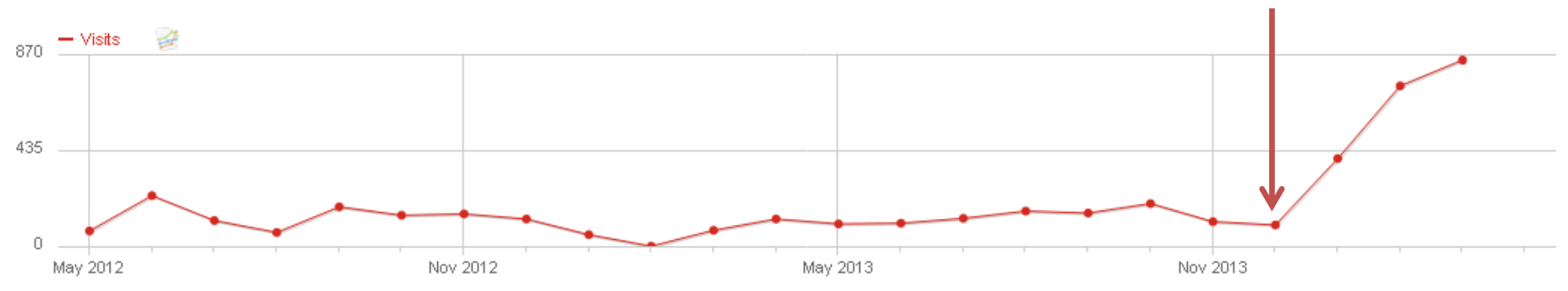
UniProt collaborations



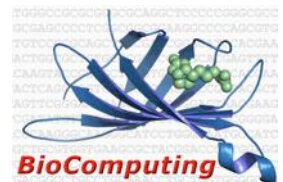
Impact of Linking from UniProt

MobiDB

Database of protein disorder

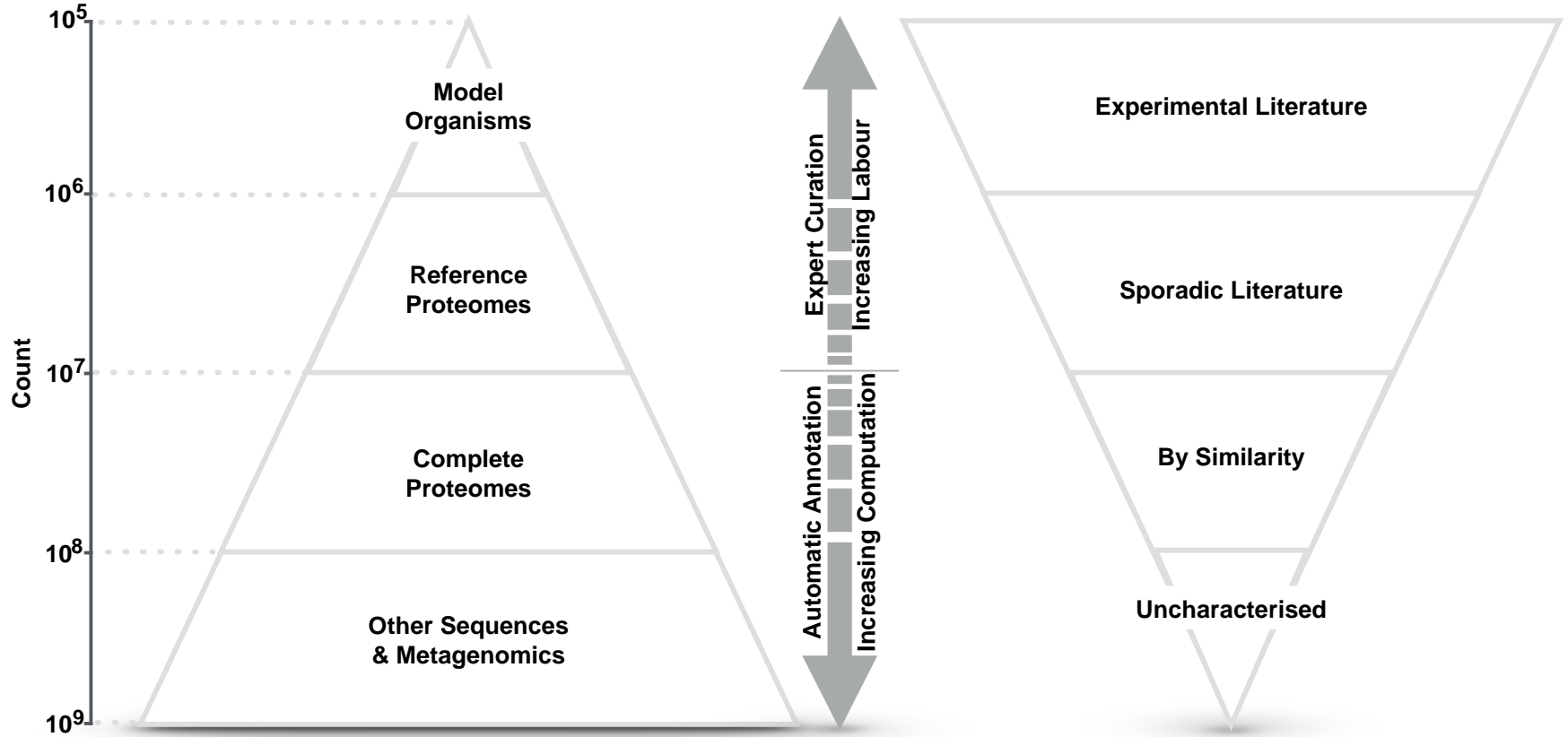


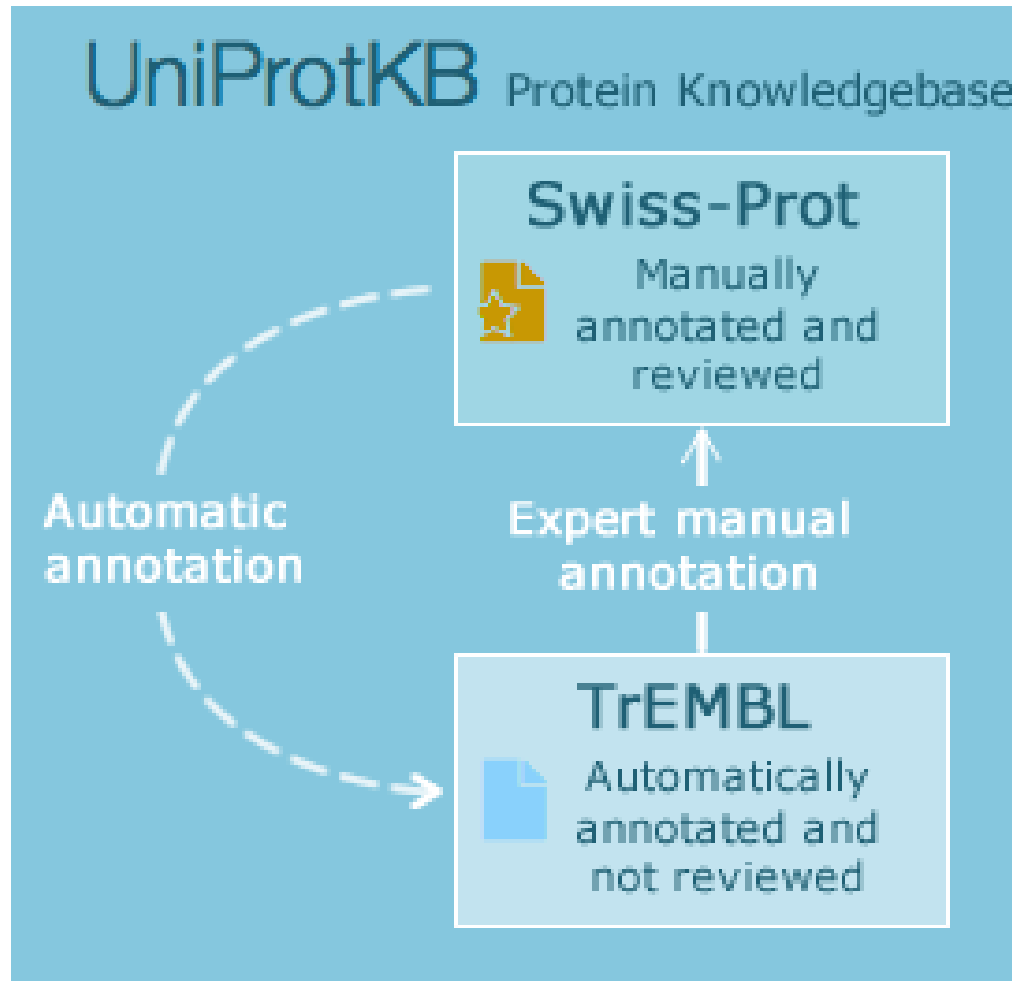
MobiDB:	visits	pageviews
2013 (12 months)	1,287	5,619
2014 (Jan through Mar)	2,010	6,474



Sequences

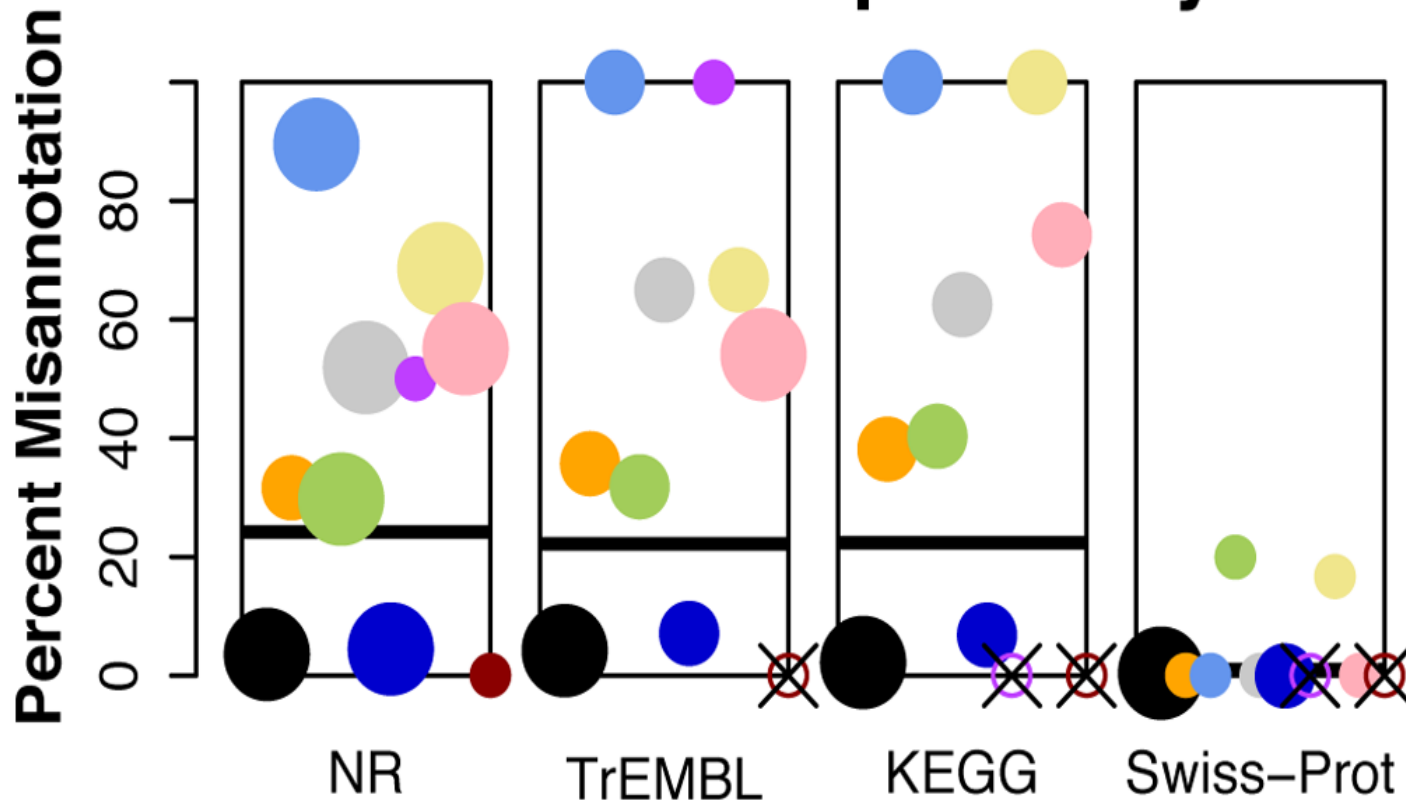
Information





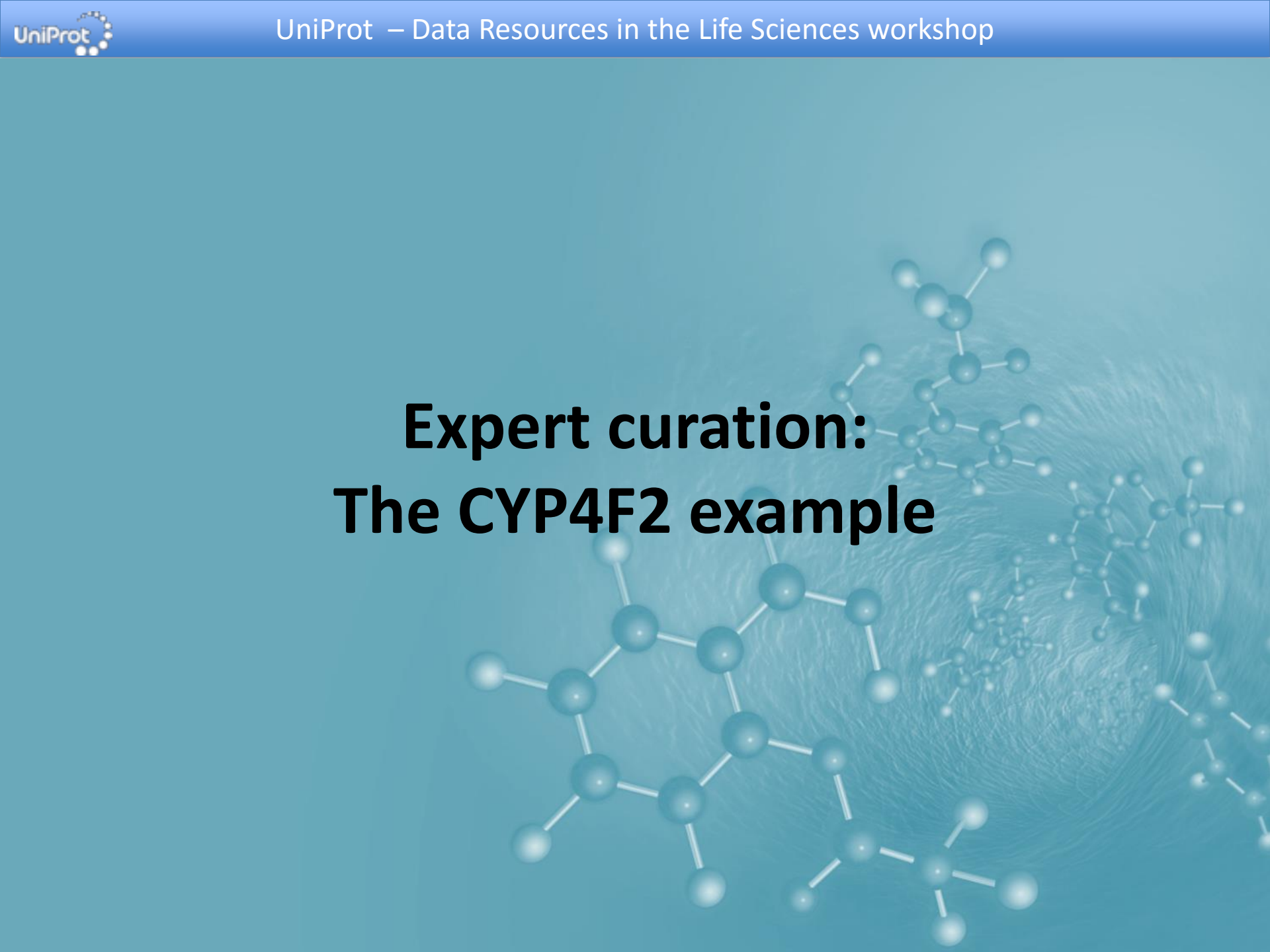
Biocuration is essential

Enolase Superfamily



Schnoes AM, Brown SD, Dodevski I, Babbitt PC (2009) Annotation Error in Public Databases: Misannotation of Molecular Function in Enzyme Superfamilies. *PLoS Comput Biol* 5(12): e1000605.

Expert curation: The CYP4F2 example



The CYP4F2 example

- CYP4F2 is a member of the cytochrome P450 family
- Oxidoreductase that oxidizes a variety of structurally unrelated compounds *in vitro* such as steroids, fatty acids and xenobiotics

FEBS 14201

Cloning and expression of a novel form of leukotriene B₄ ω-hydroxylase from human liver

Yasushi Kikuta^a, Emi Kusunose^b, Tetsuaki Kondo^a, Satoru Yamamoto^a, Hiroaki Kinoshita^c, Masamichi Kusunose^{a,*}

^a Department of Food Science and Technology, Faculty of Engineering, Fukuyama University, Fukuyama, Hiroshima 729-02, Japan

^b Toneyama Institute for Tuberculosis Research, Osaka City University Medical School, Toyonaka, Osaka 560, Japan

^c 2nd Department of Surgery, Osaka City University Medical School, Abeno, Osaka 545, Japan

Received 6 May 1994; revised version received 26 May 1994

Formation of 20-Hydroxyeicosatetraenoic Acid, a Vasoactive and Natriuretic Eicosanoid, in Human Kidney

ROLE OF CYP4F2 AND CYP4A11*

(Received for publication, September 2, 1999, and in revised form, October 16, 1999)

Jerome M. Lasker‡, W. Bill Chen, Imre Wolf, Barbara P. Bloswick§, Patricia D. Wilson§, and Pnina K. Powell

The CYP4F2 example

- CYP4F2 oxidizes vitamin K, an essential vitamin for blood coagulation
- It is involved in vitamin K catabolism by mediating omega-hydroxylation of vitamin K, a new enzymatic reaction

BIOCHEMISTRY
including biophysical chemistry & molecular biology

Article

pubs.acs.org/biochemistry

Cytochrome P450-Dependent Catabolism of Vitamin K: ω -Hydroxylation Catalyzed by Human CYP4F2 and CYP4F11



Katheryne Z. Edson,[†] Bhagwat Prasad,[‡] Jashvant D. Unadkat,[‡] Yoshitomo Suhara,[§] Toshio Okano,^{||}
F. Peter Guengerich,[⊥] and Allan E. Rettie^{*,†}

How to summarize and represent such information in UniProt?



CYP4F2 annotation



P78329 - CP4F2_HUMAN

Functionⁱ



Omega-hydroxylase that oxidizes a variety of structurally unrelated compounds, including steroids, fatty acids and xenobiotics. Plays a key role in vitamin K catabolism by mediating omega-hydroxylation of vitamin K1 (phylloquinone), and menaquinone-4 (MK-4), a form of vitamin K2. Hydroxylation of phylloquinone and MK-4 probably regulates blood coagulation (PubMed:19297519, PubMed:24138531). Also shows arachidonic acid omega-hydroxylase activity in kidney, by mediating conversion of arachidonic acid to 20-hydroxyeicosatetraenoic acid (20-HETE), possibly influencing blood pressure control (PubMed:10660572, PubMed:17341693, PubMed:18574070). Also acts as a leukotriene-B₄ omega-hydroxylase by mediating conversion of leukotriene-B₄ (LTB₄) to its omega-hydroxylated metabolite 20-hydroxyleukotriene-B₄ (20-OH LTB₄) (PubMed:8026587, PubMed:9799565).  7 Publications 

Catalytic activityⁱ



(6Z,8E,10E,14Z)-(5S,12R)-5,12-dihydroxyicosa-6,8,10,14-tetraenoate + NADPH + O₂ = (6Z,8E,10E,14Z)-(5S,12R)-5,12,20-trihydroxyicosa-6,8,10,14-tetraenoate + NADP⁺ + H₂O.  2 Publications 

Phylloquinone + NADPH + O₂ = omega-hydroxyphylloquinone + NADP⁺ + H₂O.  1 Publication 



(5Z,8Z,11Z,14Z)-icosatetraenoate + NADPH + O₂ = (5Z,8Z,11Z,14Z)-20-hydroxyicosa-5,8,11,14-tetraenoate + NADP⁺ + H₂O.

 1 Publication 



Cofactorⁱ



heme  By similarity 

Kineticsⁱ

k_{cat} is 0.067 min⁻¹ with menaquinone-4 (MK-4) as substrate.  1 Publication 

K_M=74.8 μM for leukotriene-B₄  1 Publication 

K_M=1.7 μM for menaquinone-4 (MK-4)  1 Publication 





V_{max}=2.42 nmol/min/mg enzyme  1 Publication 




CYP4F2 annotation

P78329 - CP4F2_HUMAN

Involvement in disease¹

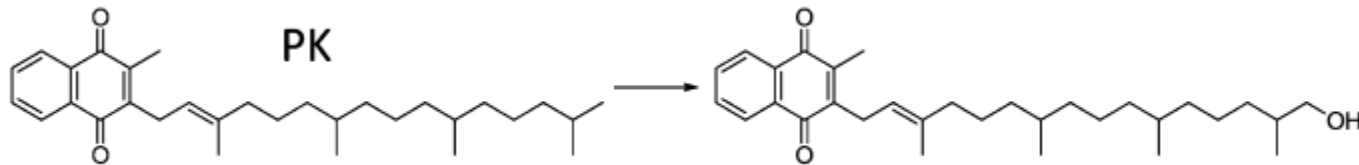
Coumarin resistance (CMRES) [MIM:122700]: A condition characterized by partial or complete resistance to warfarin or other 4-hydroxycoumarin derivatives. These drugs are used as anti-coagulants for the prevention of thromboembolic diseases in subjects with deep vein thrombosis, atrial fibrillation, or mechanical heart valve replacement.

Note: Disease susceptibility may be associated with variations affecting the gene represented in this entry. The variant Met-433 is associated with coumarin (the brand name of warfarin) resistance by increasing coumarin maintenance dose in patients on this anti-coagulant therapy. This is probably due to decreased activity of the phyloquinone omega-hydroxylase activity, leading to an increase in hepatic vitamin K levels that warfarin must antagonize ( 1 Publication ),  7 Publications ).

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier	Actions
Natural variant ¹	433 – 433	1	<p>V → M Polymorphism probably associated with CMRES; increases warfarin maintenance dose in patients on warfarin anti-coagulant therapy, possibly due to increased hepatic vitamin K levels that warfarin must antagonize. Decreased phyloquinone omega-hydroxylase activity. Decreased production of 20-hydroxyeicosatetraenoic acid (20-HETE).</p> <p> 10 Publications </p>		VAR_013119	

Collaboration with other resources

- Catalytic reactions follow recommendations from IUBMB



- Manual curation of GO terms from the literature is part of the UniProt curation process

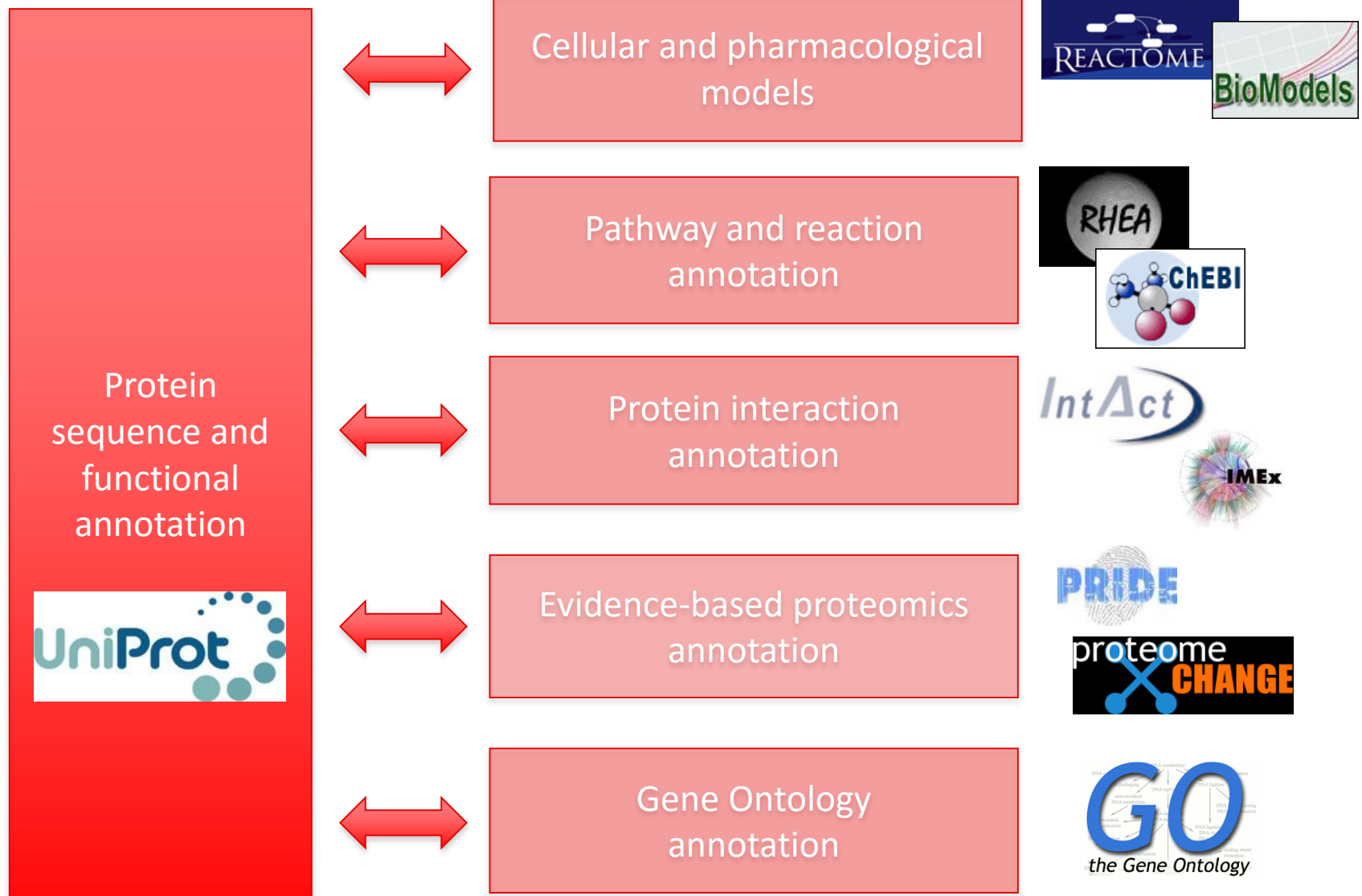
P78329	CYP4F2	GO:0007596	blood coagulation	TAS	ECO:0000304	PMID:24138531
P78329	CYP4F2	GO:0007596	blood coagulation	TAS	ECO:0000304	PMID:19297519
P78329	CYP4F2	GO:0042376	phylloquinone catabolic process	IDA	ECO:0000314	PMID:24138531
P78329	CYP4F2	GO:0042377	vitamin K catabolic process	IDA	ECO:0000314	PMID:24138531

- When possible, disease nomenclature is associated with OMIM

Involvement in diseaseⁱ

Coumarin resistance (CMRES) [MIM: 122700]: A condition characterized by partial or complete resistance to warfarin or other 4-hydroxycoumarin derivatives. These drugs are used as anti-coagulants for the prevention of thromboembolic diseases in subjects with deep vein thrombosis, atrial fibrillation, or mechanical heart valve replacement.

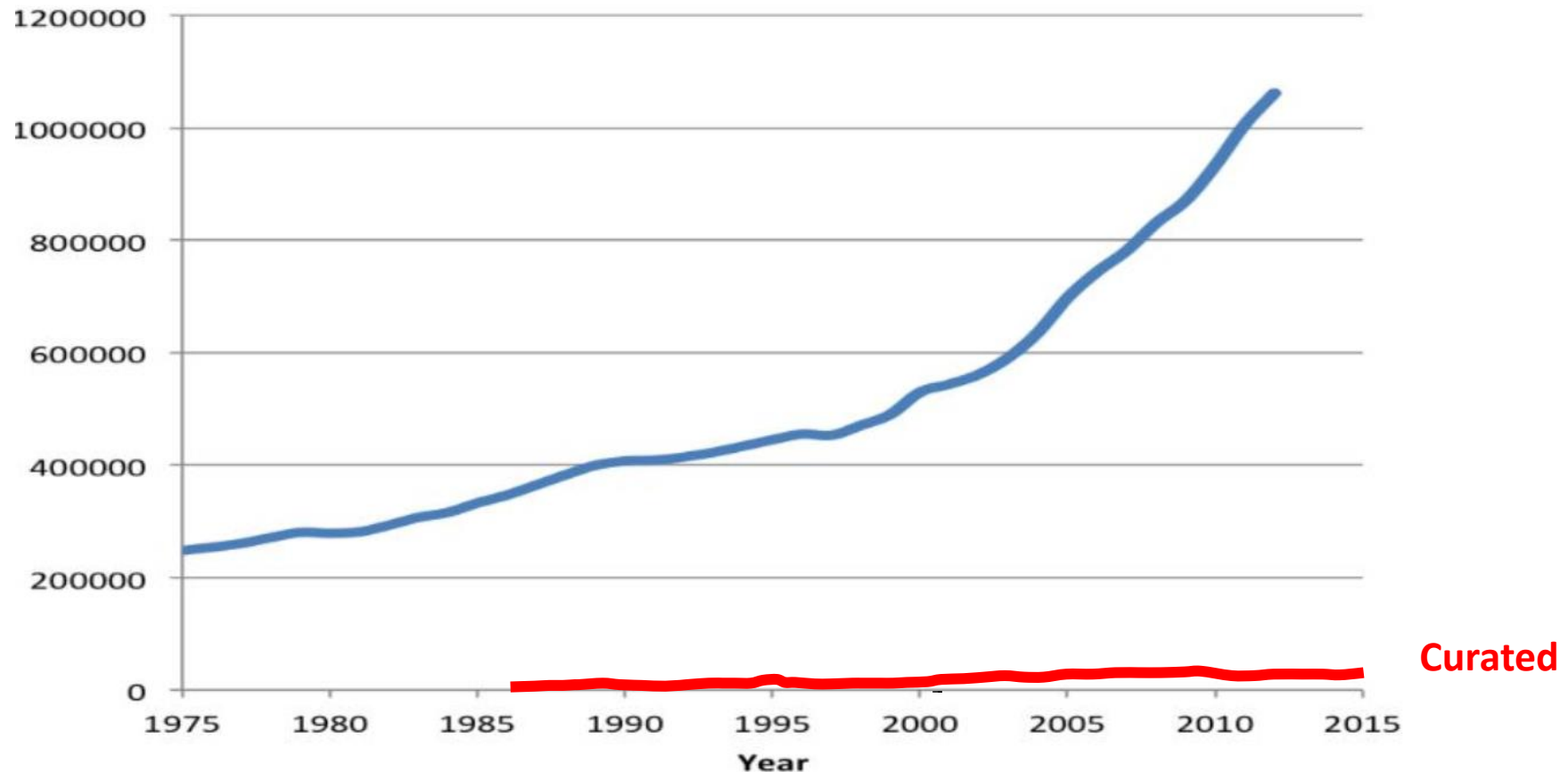
Biocurating in the best resource



Literature curation

- Of **253** papers indexed in PubMed for CYP4F2, only **15** have been used for annotation
- Most information is taken from 4 papers
- Other publications were either not relevant for annotation in UniProtKB or presented redundant or weaker evidence

The growth of biomedical literature




Is expert curation sustainable?

Literature Triage: Curation workflow

- 4 curators from different annotation programs
- Run tests over 8 months
- Use PubTator to select publications

12 [A general strategy to construct small molecule biosensors in eukaryotes.](#)

- Last updated at Mon Jan 4 07:01:45 2016 [\[Download Annotation\]](#) 

Feng J, Jester BW, Tinberg CE, [...], Baker D


eLife; 2015 Dec 29 ; 4 [\[Full text links\]](#)

PMID:26714111 - [Related citations](#) Curatable Not Priority Not Curatable

ABSTRACT

Out of scope	▼
- No comment -	
Out of scope	
Redundant	
High-throughput	
Weak Evidence	
Review/comment	

13 [Mapping residual transmission for malaria elimination.](#)

- Last updated at Mon Jan 4 07:01:45 2016 [\[Download Annotation\]](#) 

Reiner RC. Le Manach A. Kunene S. [...]. Smith DL

- Tag curatable papers
- Tag non-curatable papers and describe why

Sustainability of Literature curation

- Curators evaluate ~50-60,000 papers per year
- ~10K are curated and added to UniProt
- ~10K are redundant to existing information
- ~10K low priority
- ~10K are not well supported
- ~20K out of scope
- Sampling shows 90% PubMed out of scope

Challenges to sustainability

- Short term duration of grants
- Funding to create new resources
- Competition against research proposals
- Crowdsourcing over optimism
- Lack of understanding of importance of expert curation

The Team

PIs: Alex Bateman, Cathy Wu, Ioannis Xenarios

Key staff: Cecilia Arighi (Curation), Lydie Bougueleret (Co-Direction), Alan Bridge (Content), Hongzhan Huang (Development), Michele Magrane (Curation), Maria Martin (Development), Peter McGarvey (Content), Darren Natale (Content), Claire O'Donovan (Content), Sylvain Poux (Curation), Manuela Pruess (Coordination), Nicole Redaschi (Development)

Content/Curation: Lucila Aimo, Ghislaine Argoud-Puy, Andrea Auchincloss, Kristian Axelsen, Sara Benmohamed, Brigitte Boeckmann, Emmanuel Boutet, Lionel Breuza, Ramona Britto, Hema Bye-A-Jee, Cristina Casals Casas, Elisabeth Coudert, Melanie Courtot, Anne Estreicher, Livia Famiglietti, Marc Feuermann, John S. Garavelli, Penelope Garmiri, Daniel Gonzalez, Arnaud Gos, Nadine Gruaz, Emma Hatton-Ellis, Ursula Hinz, Chantal Hulo, Nevila Hyka-Nouspikel, Florence Jungo, Guillaume Keller, Kati Laiho, Philippe Lemercier, Damien Lieberherr, Alistair MacDougall, Patrick Masson, Anne Morgat, Barbara Palka, Ivo Pedruzzi, Klemens Pichler, Sandrine Pilbout, Catherine Rivoire, Bernd Roechert, Karen Ross, Michel Schneider, Aleksandra Shypitsyna, Christian Sigrist, Elena Speretta, Andre Stutz, Shyamala Sundaram, Michael Tognolli, Nidhi Tyagi, C. R. Vinayaka, Qinghua Wang, Kate Warner, Lai-Su Yeh, Rosanna Zaru

Development: Emanuele Alpi, Ricardo Antunes, Leslie Arminski, Parit Bansal, Delphine Baratin, Teresa Batista Neto, Benoit Bely, Mark Bingley, Jerven Bolleman, Borisas Bursteinas, Chuming Chen, Yongxing Chen, Beatrice Cuche, Alan Da Silva, Edouard De Castro, Maurizio De Giorgi, Tunca Dogan, Leyla Garcia Castro, Elisabeth Gasteiger, Sebastien Gehant, Leonardo Gonzales, Arnaud Kerhornou, Vicente Lara, Wudong Liu, Thierry Lombardot, Jie Luo, Xavier Martin, Andrew Nightingale, Joseph Onwubiko, Monica Pozzato, Sangya Pundir, Guoying Qi, Alexandre Renaux, Steven Rosanoff, Rabie Saidi, Tony Sawford, Edward Turner, Vladimir Volynkin, Yuqi Wang, Tony Wardell, Xavier Watkins, Hermann Zellner, Jian Zhang

European Bioinformatics Institute (EMBL-EBI), Hinxton, Cambridge, UK
Protein Information Resource (PIR), Washington DC and Delaware, USA
SIB Swiss Institute of Bioinformatics (SIB), Geneva, Switzerland