

Biocuració i bioinformàtica, o com fer front als nous reptes de la recerca en l'era de les *'big data'*

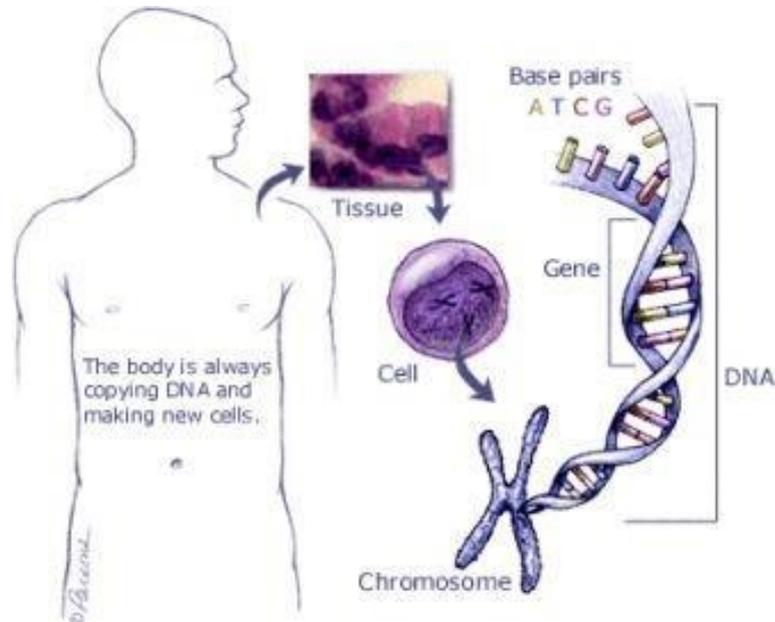
Cristina Casals Casas

10ns Debats de Recerca,
Andorra la Vella, 31^a de maig 2016

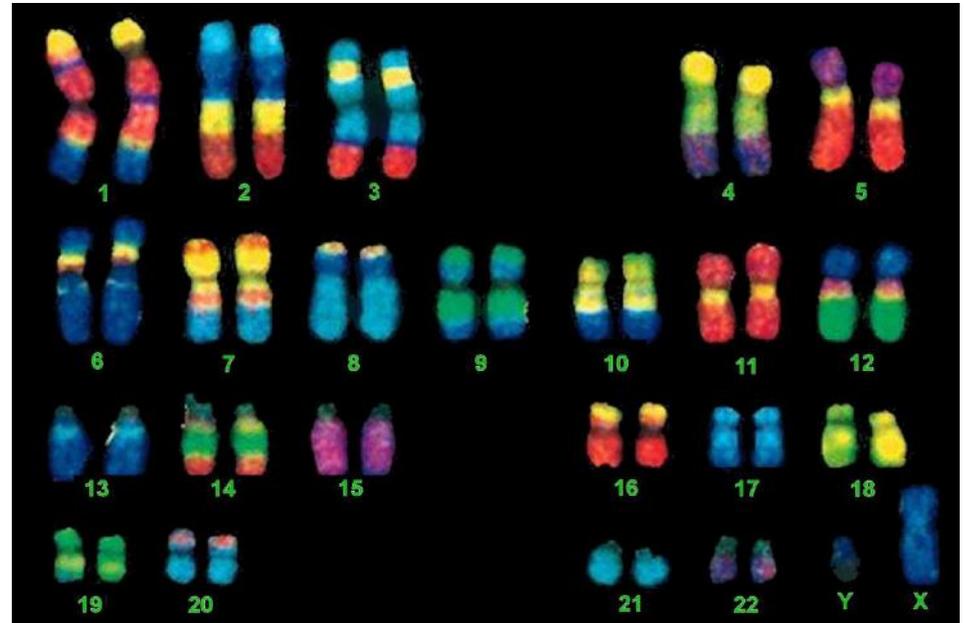


Swiss Institute of
Bioinformatics

Per què la biologia?



www.downscreening.be



Cariotip humà: 23 parells de cromosomes

Formació i recorregut professional

- 1997 Baccalauréat Scientifique, Lycée Comtes de Foix.
- 1997-2001 Maîtrise en Biologie Cellulaire et Moléculaire,
Université Paul Sabatier, Toulouse III.
- 2001-2007 Doctorat en Immunologia,
Institut de Recerca Biomèdica, UB, Barcelona.
- 2007-2012 Postdoctorat, Center for Integrative Genomics,
Université de Lausanne, Suïssa.
- 2013-? Biocuradora per UniprotKB/Swiss-prot,
SIB Swiss Institute of Bioinformatics, Ginebra, Suïssa.



Formació i recorregut professional

1997 Baccalauréat Scientifique, Lycée Comtes de Foix.

1997-2001 Maîtrise en Biologie Cellulaire et Moléculaire,
Université Paul Sabatier, Toulouse III.

Recerca

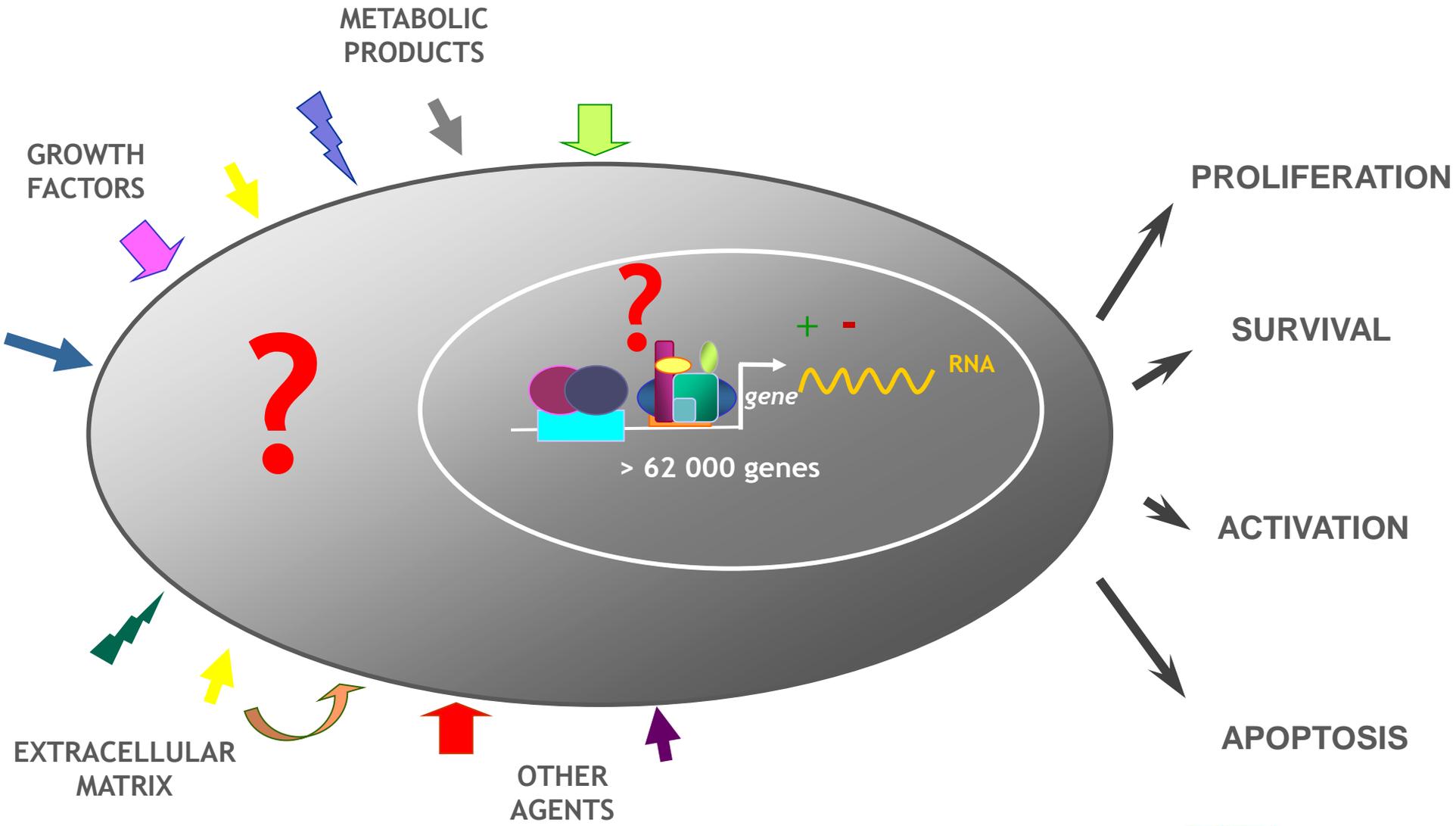
2001-2007 Doctorat en Immunologia,
Institut de Recerca Biomèdica, UB, Barcelona.

2007-2012 Postdoctorat, Center for Integrative Genomics,
Université de Lausanne, Suïssa.

2012-? Biocuradora per UniprotKB/Swiss-prot,
SIB Swiss Institute of Bioinformatics, Ginebra, Suïssa.

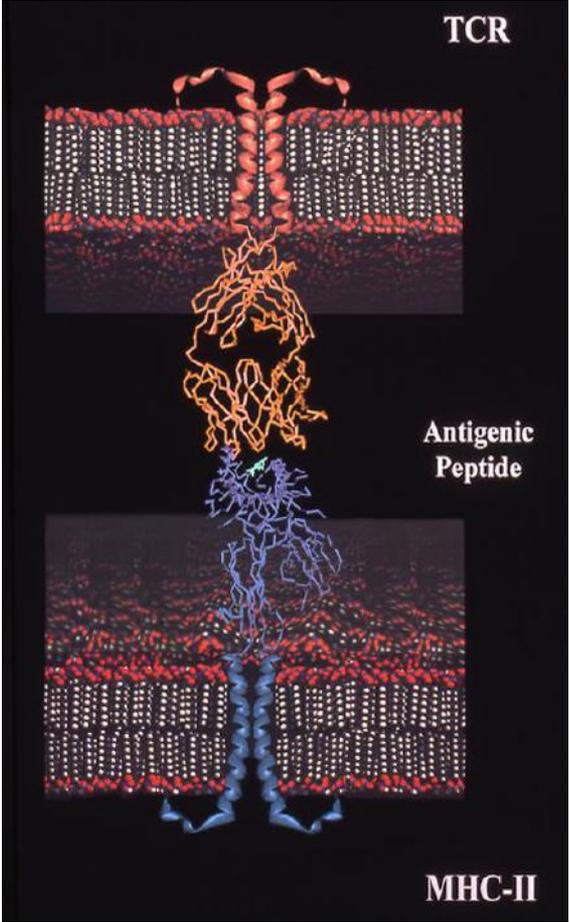
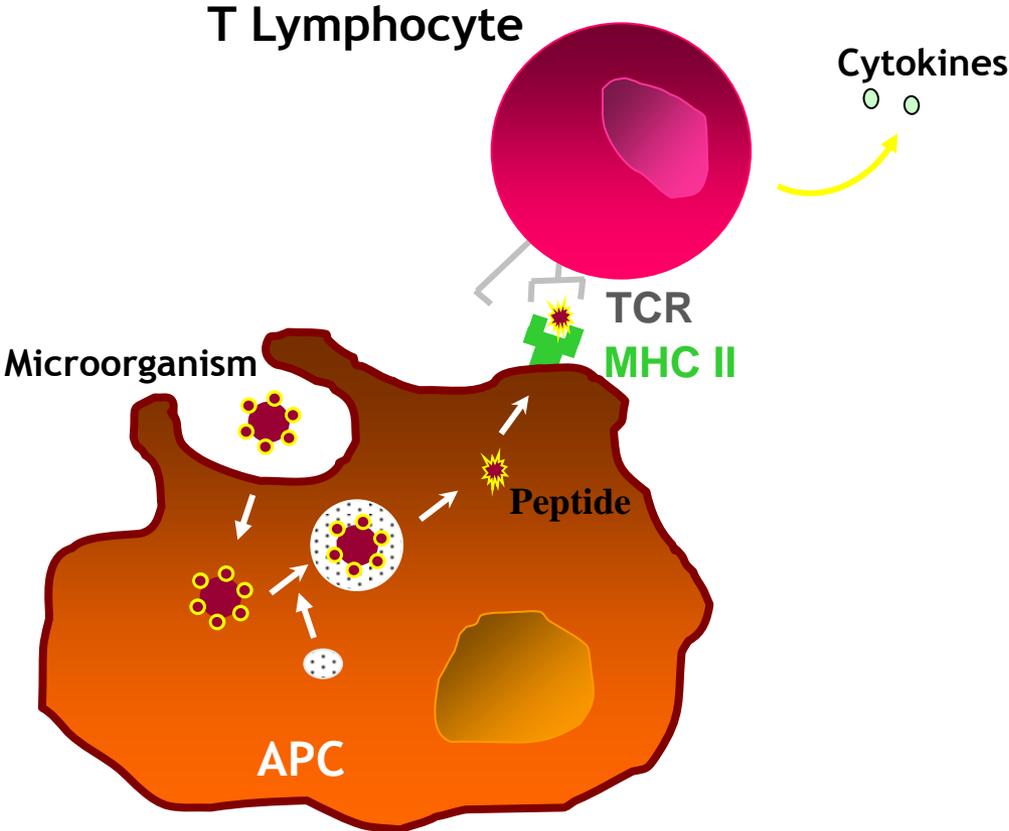


Recerca: regulació de l'expressió gènica



Role of AP-1 in the regulation of MHC-induced expression by LPS in Dendritic Cells and B lymphocytes

Casals C. et al. *J. Immunol.* 2007

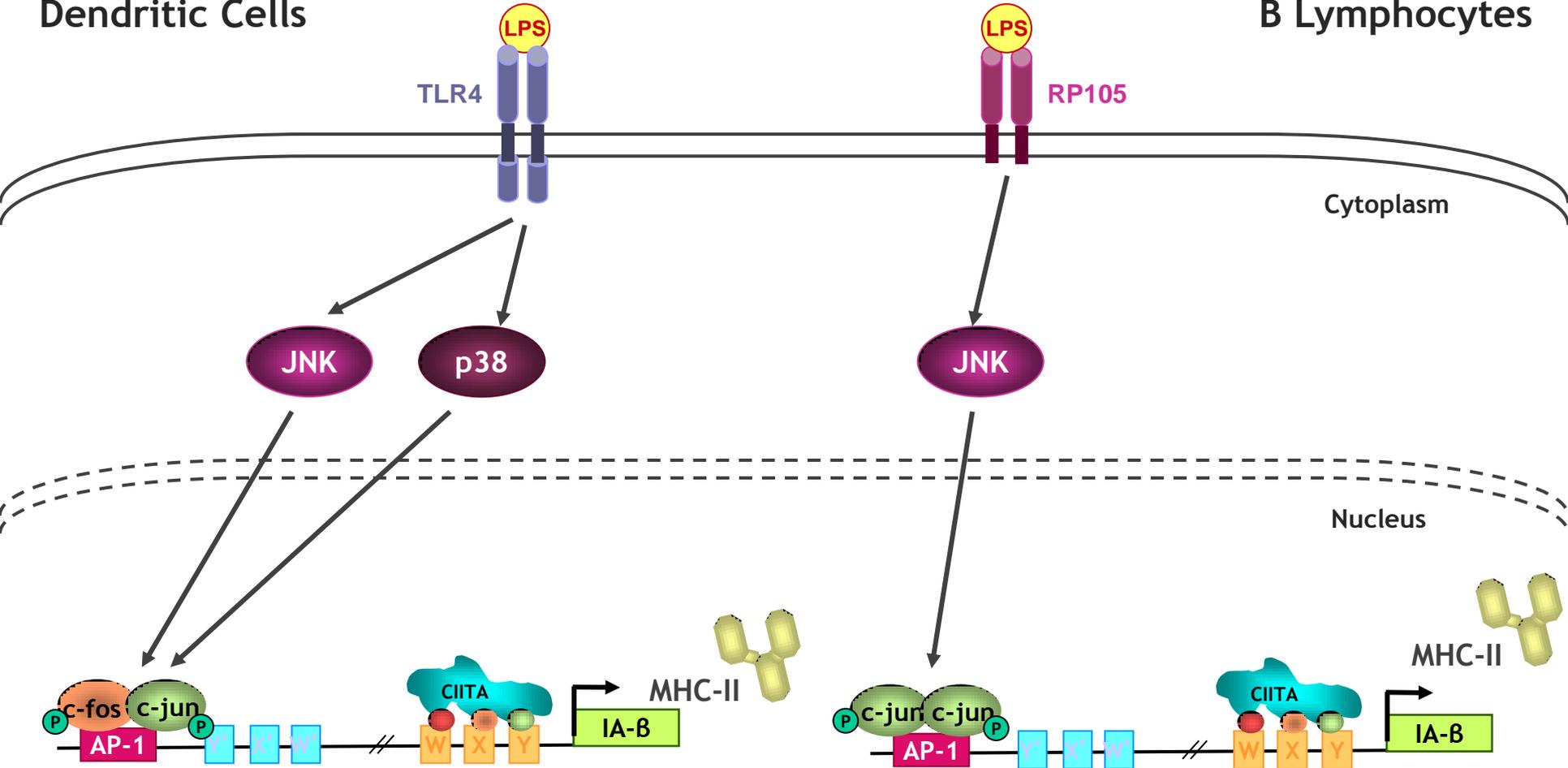


Role of AP-1 in the regulation of MHC-induced expression by LPS in Dendritic Cells and B lymphocytes

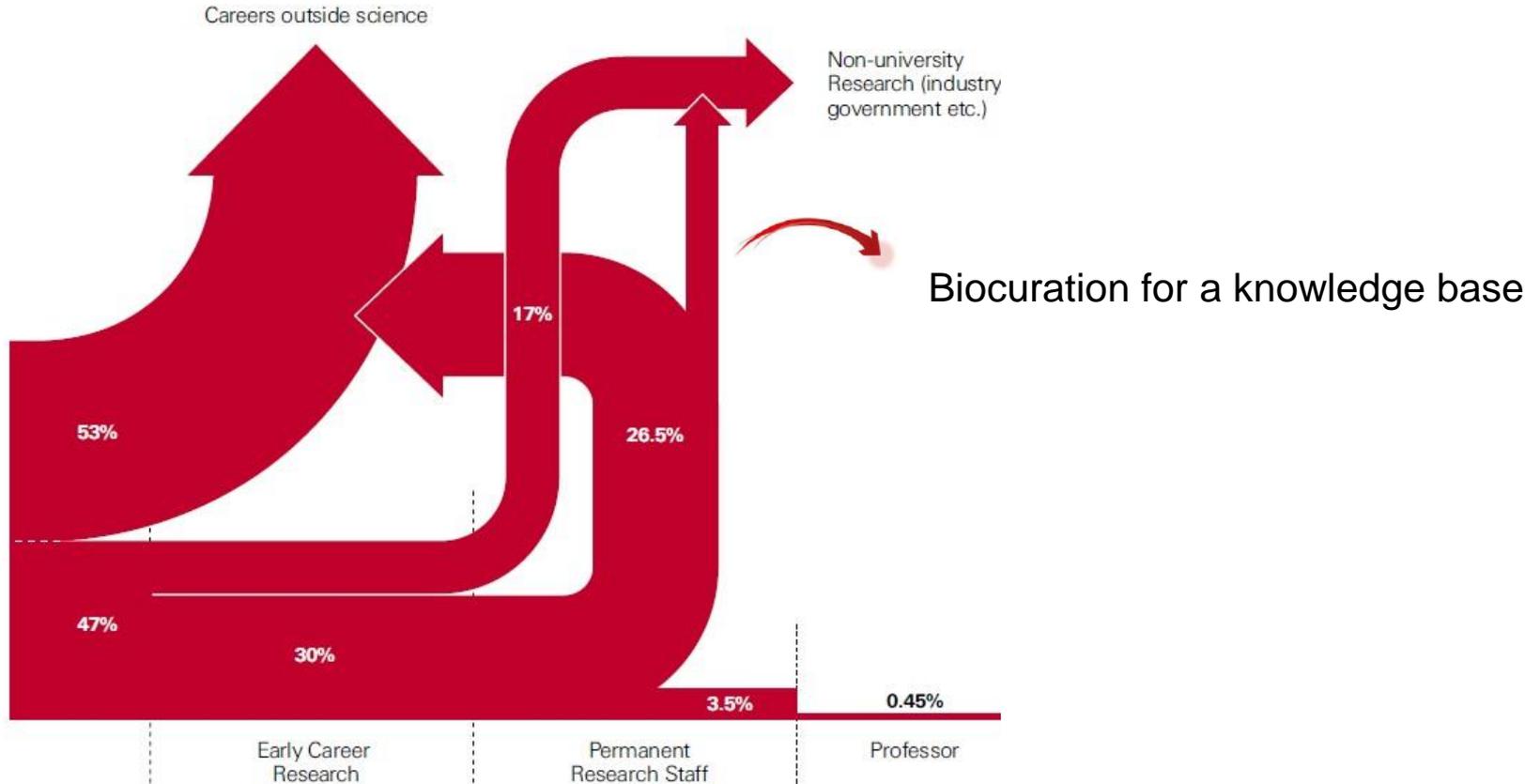
Casals C. et al. *J. Immunol.* 2007

Dendritic Cells

B Lymphocytes



... i després d'11 anys en la recerca pública?



SIB Swiss Institute of Bioinformatics (since 1998)

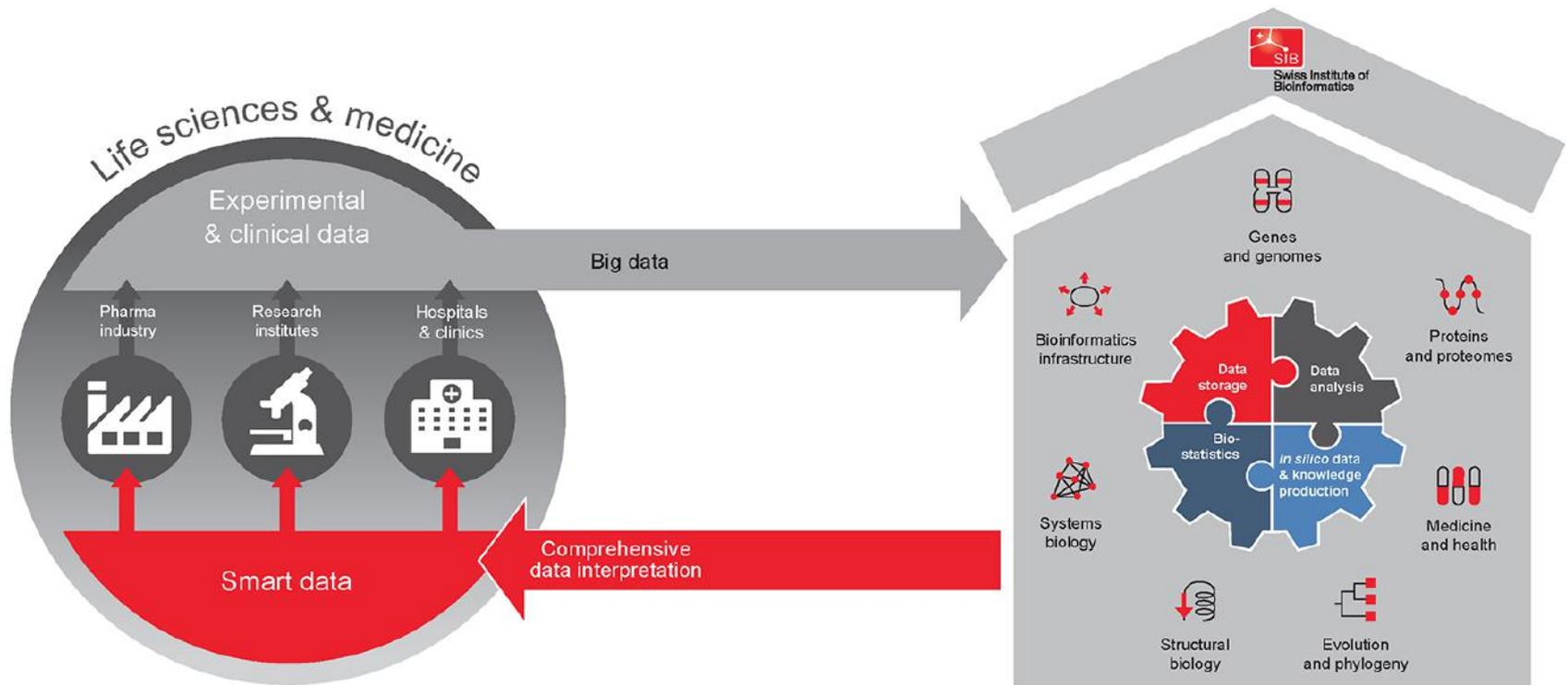
- 65 groups
- 750 collaborators: mathematicians, biologists, clinicians, chemists, physicists, computer scientists,
- Grants: US, EU and CH
- Common point: bioinformatics

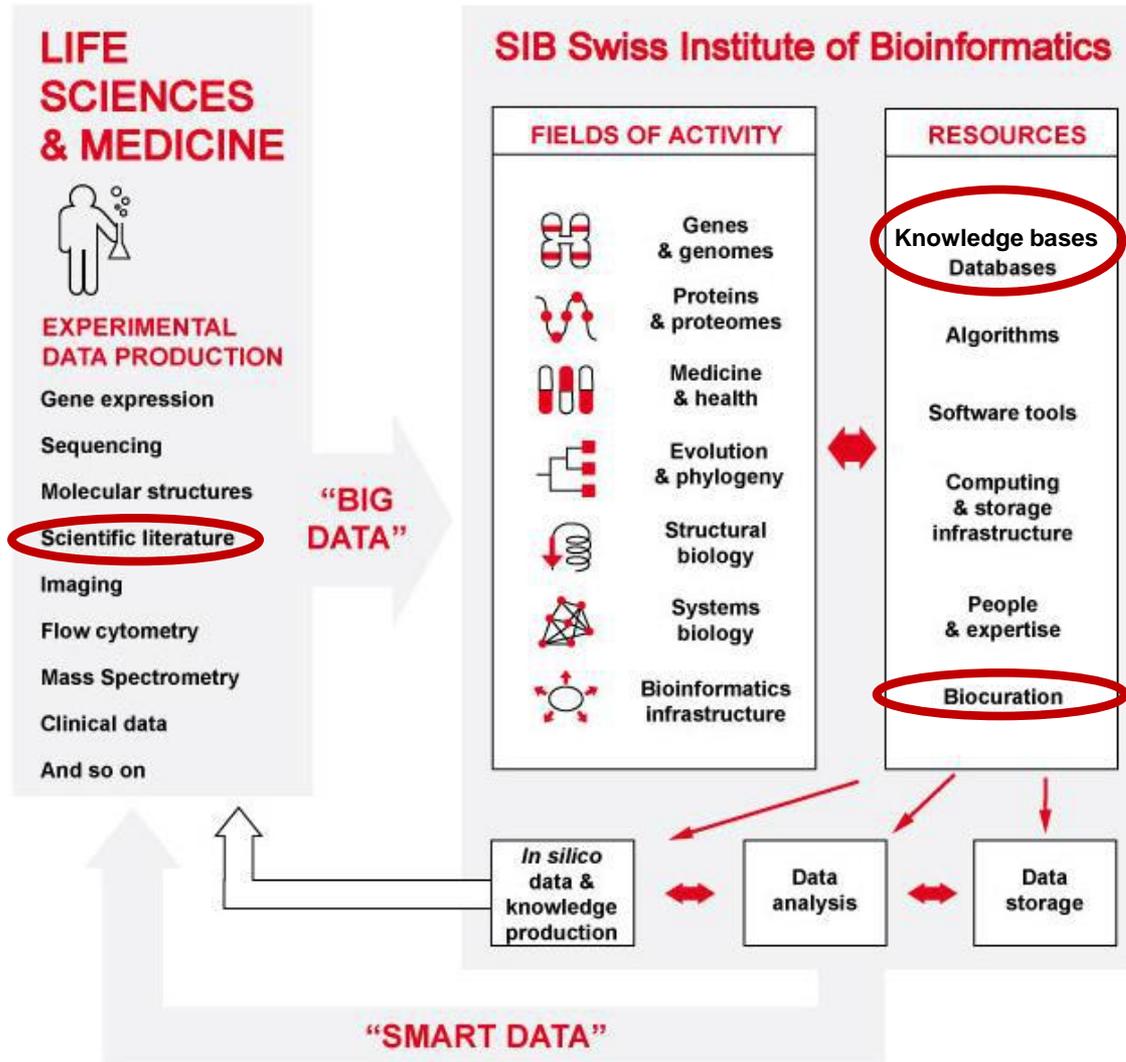
www.sib.swiss



Swiss Institute of
Bioinformatics

Bioinformatics is the application of computer technology to the understanding and effective use of biological data





Knowledge Base



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](#)

Not logged in [Talk](#) [Contributions](#) [Create account](#) [Log in](#)

Article [Talk](#)

[Read](#) [Edit](#) [View history](#)

Search

Andorra

From Wikipedia, the free encyclopedia

Coordinates: 42°30′N 1°30′E

Not to be confused with Andora.

For other uses, see Andorra (disambiguation).

Andorra (ⁱˈænˈdɔːrə; Catalan: [ənˈdɔɾə], locally: [anˈdɔɾa]), officially the **Principality of Andorra** (Catalan: *Principat d'Andorra*), also called the **Principality of the Valleys of Andorra**^[5] (Catalan: *Principat de les Valls d'Andorra*), is a sovereign landlocked microstate in Southwestern Europe, located in the eastern Pyrenees mountains and bordered by Spain and France. Created under a charter in A.D. 988, the present Principality was formed in A.D. 1278. It is known as a *principality* as it is a *monarchy* headed by two *Co-Princes* – the Spanish/Roman Catholic *Bishop of Urgell* and the *President of France*.

Andorra is the *sixth-smallest nation in Europe*, having an area of 468 km² (181 sq mi) and a population of approximately 85,000.^[1] Its capital *Andorra la Vella* is the highest capital city in Europe, at an elevation of 1,023 metres (3,356 ft) *above sea level*.^[6] The official language is *Catalan*, although Spanish, Portuguese, and French are also commonly spoken.^{[1][7]}

Andorra's *tourism* services an estimated 10.2 million visitors annually.^[8] It is not a member of the *European Union*, but the euro is the official currency. It has been a member of the *United Nations* since 1993.^[9] In 2013, the people of Andorra had the highest life expectancy in the world at 81 years, according to *The Lancet*.^[10]

Contents [hide]

- Etymology
- History
 - 17th to 19th centuries
 - 20th century
- Politics
- Law and criminal justice
- Foreign relations, defence, and security
 - Military
 - GIPA
 - Police Corps

Principality of Andorra

Principat d'Andorra (Catalan)



Flag



Coat of arms

Motto: *Virtus Unita Fortior*

United virtue is stronger

Anthem: *El Gran Carlemany*

The Great Charlemagne



Swiss Institute of
Bioinformatics

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.

UniProtKB
UniProt Knowledgebase

Swiss-Prot (550,960)
Manually annotated and reviewed.

TrEMBL (63,686,057)
Automatically annotated and not reviewed.

UniRef
Sequence clusters

UniParc
Sequence archive

Proteomes

Supporting data

Literature citations 	Taxonomy 	Subcellular locations
Cross-ref. databases 	Diseases 	Keywords

News

[Forthcoming changes](#)
Planned changes for UniProt

[UniProt release 2016_04](#)
Small changes, big effects | New UniProt JAPI | Change of the UniProt RDF files distribution

[UniProt release 2016_03](#)
From the Zika forest to the Amazon, news from a viral wanderer | Cross-

[News archive](#)

Getting started

Text search

Our basic text search allows you to search all the resources available

BLAST

Find regions of similarity between your sequences

Sequence alignments

Align two or more protein sequences using the



UniProt data

Download latest release

Get the UniProt data

Statistics

View Swiss-Prot and TrEMBL statistics

How to cite us

The UniProt Consortium

Submit your data

Protein spotlight



A Tighter Mesh

April 2016

A fabric whose weave is slack is less resistant to strain than one whose knit is close. These physical principles of toughness are followed by Nature all the way down the molecular scale. At the cellular level, protrusions known as microvilli grow at the apical end of the epithelial cells that line the intestinal tract. Each protrusion is like a finger

Knowledge Base

On the Wiki
Episodes
Characters
Culture&Society

Wiki Activity
Random page
Videos
Images
Chat

Tyrion Lannister

3,203 PAGES ON THIS WIKI

View source
Talk: 27

Tyrion Lannister: *"I'm guilty of a far more monstrous crime: I'm guilty of being a dwarf!"*

Tywin Lannister: *"You are not on trial for being a dwarf."*

Tyrion Lannister: *"Oh, yes I am! I've been on trial for that my entire life!"*

— Tyrion at his trial, falsely accused of killing his nephew King Joffrey^[src]

Tyrion Lannister is a major character in the first, second, third, fourth, fifth, and sixth seasons. He is played by starring cast member Peter Dinklage and debuts in the series premiere. Tyrion Lannister is the youngest son of Lord Tywin Lannister and younger brother of Cersei and Jaime Lannister. A dwarf, he uses his wit and intellect to overcome the prejudice he faces.

Tyrion Lannister

Tyrion has been born a commoner, he'd have simply been left out in the woods to die. However, Tyrion was born into a powerful noble House, and was therefore spared. Even though his father doesn't think much of him, he has had the benefits of being raised with wealth and education, and is expected to lead his life as a credit to the Lannister name.^[3] He is committed to the good of both his House and, since Cersei married King Robert Baratheon, his

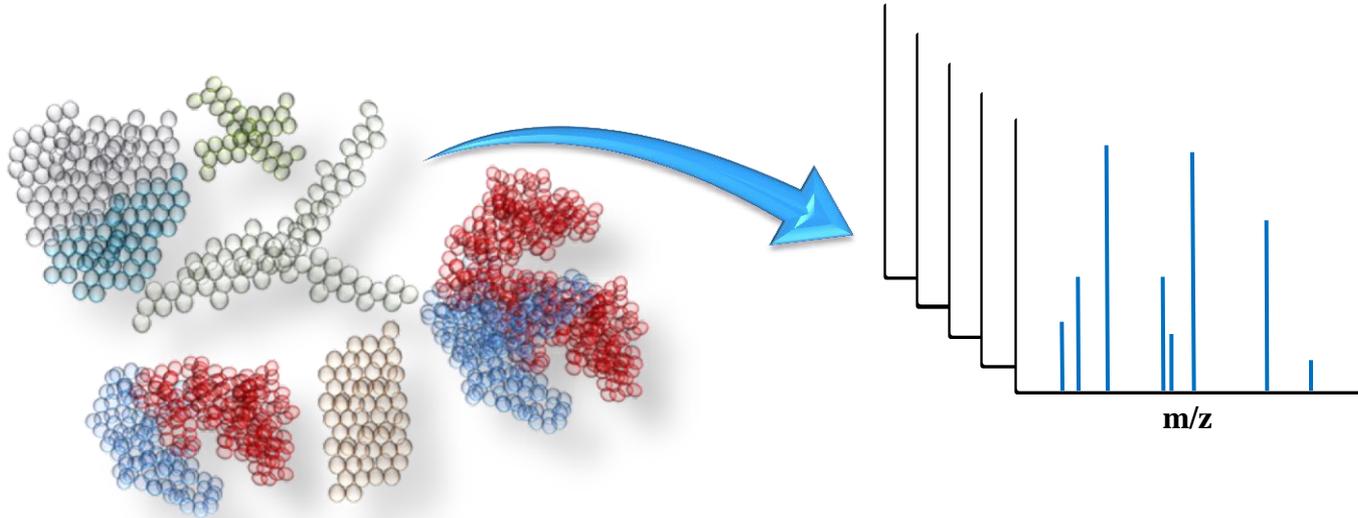
r, the richest the
tuent regions
erly Rock is
ger brother of
ister died
ntempt as

His size has
ies, such as
intellect and
rism like

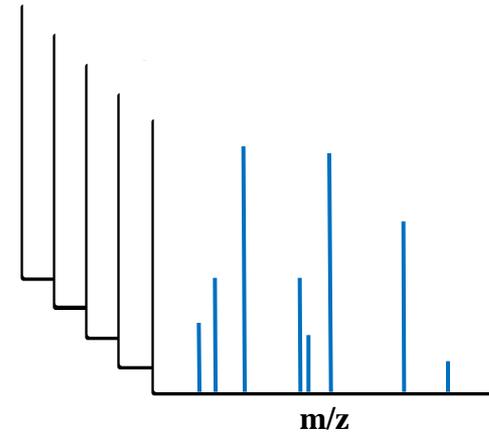
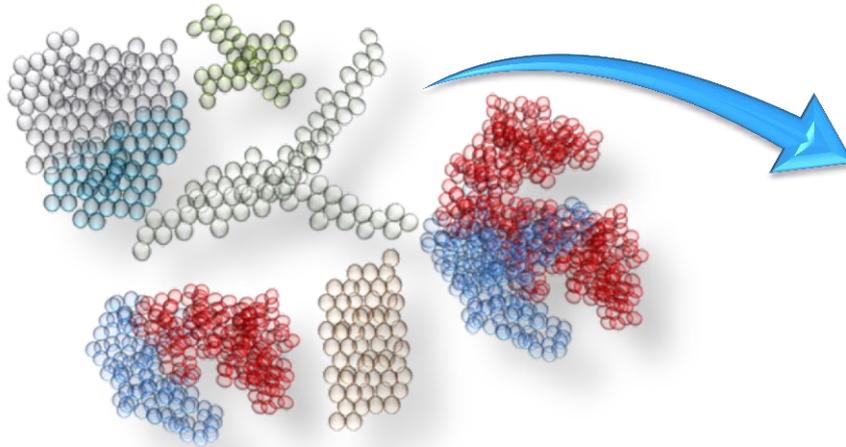
Season(s)	1, 2, 3, 4, 5, 6
First seen	"Winter is Coming"
Appeared in	50 episodes (see below)
Titles	Master of Coin (formerly) Acting Hand of the King (formerly)
Also known as	The Imp Halfman The Little Lion Demon Monkey
Status	Alive
Age	Approx. 38 in Season 6 ^[1]
Origin	Casterly Rock
Allegiance	House Lannister (formerly) House Targaryen
Culture	Andal
Religion	Faith of the Seven
Family	{Tywin Lannister} - father/first cousin once removed {Joanna Lannister} - mother/first cousin once removed {Tytos Lannister} - paternal grand- father/great-uncle Cersei Lannister - sister/second



Interpretation of [prote]omics



Interpretation of [prote]omics



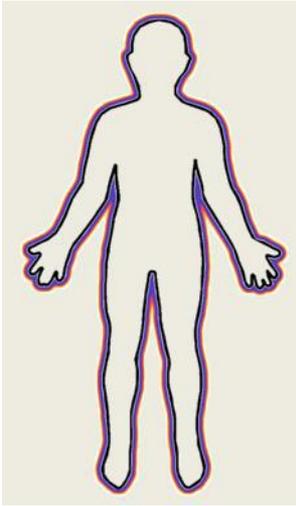
Which proteins?

Which biological processes?



The screenshot shows the UniProt website interface. At the top, there is a search bar with 'UniProtKB' entered. Below the search bar, there are navigation links for 'BLAST', 'Align', and 'Retrieve/ID mapping'. The main content area is divided into several sections: 'UniProtKB' (Swiss-Prot 550,740 and TrEMBL 63,039,659), 'UniRef' (UniProt Reference Clusters), 'UniParc' (comprehensive and non-redundant database), and 'Proteomes' (set of proteins thought to be expressed by an organism). Below these sections is a 'Supporting data' section with links for 'Literature citations', 'Cross-ref. databases', 'Taxonomy', 'Diseases', 'Subcellular locations', and 'Keywords'.

Variation and protein biology

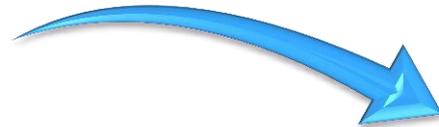
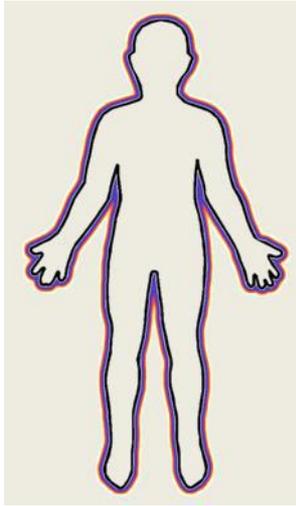


ATTAGTTAGTCCAATGCTC

ATTAGTT**C**GTCCAATGCTC

p.Leu246Phe

Variation and protein biology



ATTAGTTAGTCCAATGCTC
ATTAGTT**C**GTCCAATGCTC
p.Leu246Phe



*Effect on
protein activity?
In a conserved
functional domain?*

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.

UniProtKB UniProt Knowledgebase Swiss-Prot (550,740) Manually annotated and reviewed. Records with information extracted from literature and curator-evaluated computational analysis. TrEMBL (63,039,659) Automatically annotated and not reviewed. Records that await full manual annotation.	UniRef The UniProt Reference Clusters (UniRef) provide clustered sets of sequences from the UniProt Knowledgebase (including isoforms) and selected UniParc records.	UniParc UniParc is a comprehensive and non-redundant database that contains most of the publicly available protein sequences in the world.	Proteomes A proteome is the set of proteins thought to be expressed by an organism. UniProt provides proteomes for species with completely sequenced genomes.
---	--	--	---

Supporting data

- Literature citations
- Taxonomy
- Subcellular locations
- Cross-ref. databases
- Diseases
- Keywords

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.

UniProtKB

UniProt Knowledgebase

Swiss-Prot
(550,960)

Manually annotated and reviewed.

TrEMBL
(63,686,057)

Automatically annotated and not reviewed.

UniRef

Sequence clusters



UniParc

Sequence archive



Proteomes



Supporting data

Literature citations



Taxonomy



Subcellular locations



Cross-ref. databases



Diseases



Keywords



News



Forthcoming changes

Planned changes for UniProt

UniProt release 2016_04

Small changes, big effects | New UniProt JAPI | Change of the UniProt RDF files distribution

UniProt release 2016_03

From the Zika forest to the Amazon, news from a viral wanderer | Cross-

News archive

Getting started

Text search

Our basic text search allows you to search all the resources available

BLAST

Find regions of similarity between your sequences

Sequence alignments

Align two or more protein sequences using the



UniProt data

Download latest release

Get the UniProt data

Statistics

View Swiss-Prot and TrEMBL statistics

How to cite us

The UniProt Consortium

Submit your data

Protein spotlight



A Tighter Mesh

April 2016

A fabric whose weave is slack is less resistant to strain than one whose knit is close. These physical principles of toughness are followed by Nature all the way down the molecular scale. At the cellular level, protrusions known as microvilli grow at the apical end of the epithelial cells that line the intestinal tract. Each protrusion is like a finger

Biocurator / Data Scientist



WIKIPEDIA
The Free Encyclopedia

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

- Interaction
- Help

Article [Talk](#)

Read [Edit](#) [View history](#)

Search

Biocurator

From Wikipedia, the free encyclopedia

A **biocurator** is a professional **scientist** who **curates**, collects, annotates, and validates information that is disseminated by **biological** and **model organism databases**.^[1] The role of a biocurator encompasses quality control of primary biological research data intended for publication, extracting and organizing data from original scientific literature, and describing the data with standard **annotation** protocols and vocabularies that enable powerful queries and **biological database** inter-operability. Biocurators communicate with researchers to ensure the accuracy of curated information and to foster data exchanges with research laboratories.^[2]

Biocurators (also called **scientific curators**, **data curators** or **annotators**) have been recognized as the "museum catalogers of the Internet age".^{[3][4]}



They read publications and extract the KNOWLEDGE that will help explain the function of a protein.

enabling life scientists to have an accurate, comprehensive and computer friendly representation of biological knowledge

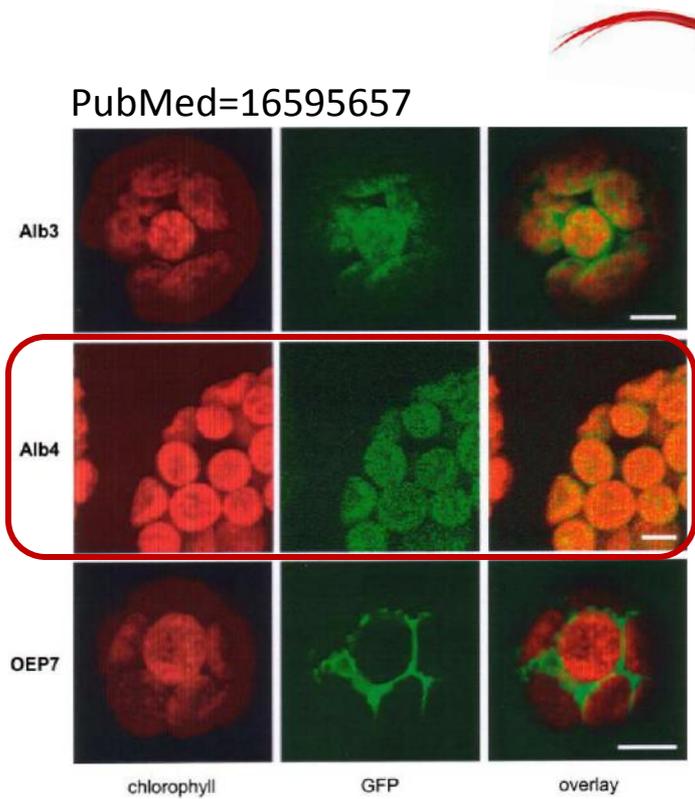


FIGURE 3. Subcellular localization of Alb3-GFP and Alb4-GFP fusion proteins. *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3- or Alb4-GFP. Maximum intensity signals from confocal images are shown for chlorophyll autofluorescence, GFP fluorescence, and an overlay of both. OEP7-GFP is included as a

corresponds to the predicted length of the Alb4 mRNA. Even after prolonged exposure of the blots treated with the Alb4 probe, no signal could be found at ~3.5 kb, the predicted size of the Artemis transcript.

Alb4 Is a Thylakoid Membrane Protein—Alb4 is predicted to have a chloroplast transit peptide with a processing site after amino acid resi-

UniProtKB Q9FYL3

Subcellular location¹

- Plastid > chloroplast thylakoid membrane 1 Publication
- Multi-pass membrane protein 1 Publication

Topology

Feature key	Position(s)	Length	Description	Graphical view
Transmembrane ¹	115 - 135	21	Helical	
Transmembrane ¹	184 - 204	21	Helical	
Transmembrane ¹	263 - 283	21	Helical	
Transmembrane ¹	302 - 322	21	Helical	

GO - Cellular component²

- chloroplast
- chloroplast thylakoid membrane
- integral component of membrane
- thylakoid

Controlled vocabulary

2. "A second thylakoid membrane-localized Alb3/OxaI/YidC homologue is involved in proper chloroplast biogenesis in *Arabidopsis thaliana*."
 Gerdes L., Bals T., Klostermann E., Karl M., Philipp K., Huenken M., Soll J., Schuenemann D.
 J. Biol. Chem. 281:16632-16642(2006) [PubMed] [Europe PMC] [Abstract]
Cited for: SEQUENCE REVISION, TISSUE SPECIFICITY, SUBCELLULAR LOCATION.

due 45 based on the ChloroP prediction program (33). To test this prediction, *Arabidopsis* mesophyll protoplasts were transiently transformed with constructs for Alb3-GFP or Alb4-GFP. Merging of the GFP and autofluorescence images indicated a thylakoid localization of Alb4. The GFP distribution for Alb4 is similar to that of Alb3 and not to that of outer envelope protein AtOEP7 (Fig. 3). To test this assumption, *in vitro* translated radiolabeled Alb4 was imported into isolated pea

Function annotation for humans

Function¹

Golgi serine/threonine protein kinase that phosphorylates secretory pathway proteins within Ser-x-Glu/pSer motifs and plays a key role in biomineralization of bones and teeth (PubMed:22582013, PubMed:23754375, PubMed:25789606). Constitutes the main protein kinase for extracellular proteins, generating the majority of the extracellular phosphoproteome (PubMed:26091039). Mainly phosphorylates proteins within the Ser-x-Glu/pSer motif, but also displays a broader substrate specificity (PubMed:26091039). Phosphorylates casein as well as a number of proteins involved in biomineralization such as AMELX, AMTN, ENAM and SPP1 (PubMed:22582013, PubMed:25789606). In addition to its role in biomineralization, also plays a role in lipid homeostasis, wound healing and cell migration and adhesion (PubMed:26091039).  4 Publications ▾

Manual assertion based on experiment in¹

"A single kinase generates the majority of the secreted phosphoproteome."

Tagliabracci V.S., Wiley S.E., Guo X., Kinch L.N., Durrant E., Wen J., Xiao J., Cui J., Nguyen K.B., Engel J.L., Coon J.J., Grishin N., Pinna L.A., Pagliarini D.J., Dixon J.E.

Cell 161:1619-1632(2015) [PubMed] [Europe PMC] [Abstract]

Cited for: PROTEIN SEQUENCE OF N-TERMINUS, FUNCTION, SUBCELLULAR LOCATION, AUTOPHOSPHORYLATION, GLYCOSYLATION



Function annotation for computers

- `<comment type="function">`
 - `<text evidence="10 48 50 66 69 74">`

Isoform Beta: Acts as a dominant negative inhibitor of isoform Alpha (PubMed:7769088, PubMed:8621628, PubMed:20484466). Has intrinsic transcriptional activity independent of isoform Alpha when both isoforms are coexpressed (PubMed:19248771, PubMed:26711253). Loses this transcription modulator function on its own (PubMed:20484466). Has no hormone-binding activity (PubMed:8621628). May play a role in controlling glucose metabolism by maintaining insulin sensitivity (By similarity). Reduces hepatic gluconeogenesis through down-regulation of PEPCK in an isoform Alpha-dependent manner (PubMed:26711253). Directly regulates STAT1 expression in isoform Alpha-independent manner (PubMed:26711253).

`</text>`
 - `</comment>`
- `<comment type="function">`
 - `<text>`

Isoform Alpha-2: Has lower transcriptional activation activity than isoform Alpha. Exerts a dominant negative effect on isoform Alpha trans-repression mechanism (PubMed:20484466).

`</text>`
 - `</comment>`
- `<comment type="function">`
 - `<text evidence="20">`Isoform GR-P: Increases activity of isoform Alpha.`</text>`
 - `</comment>`
- `<comment type="function">`
 - `<text evidence="22 39">`

Isoform Alpha-B: More effective than isoform Alpha in transcriptional activation, but not repression activity.

`</text>`



Proteins and diseases

Pathology & Biotechⁱ

Involvement in diseaseⁱ

Raine syndrome (RNS)  3 Publications ▼

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

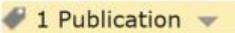
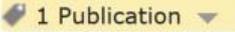
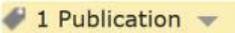
Disease description: Autosomal recessive osteosclerotic bone dysplasia with neonatal lethal outcome. Clinical features include generalized osteosclerosis, craniofacial dysplasia and microcephaly.

[See also OMIM:259775](#)



Natural variants

Natural variant

Feature key	Position(s)	Length	Description	Graphical view	Feature identifier
Natural variant ⁱ	258 - 258	1	I → N in RNS. 		VAR_073660
Natural variant ⁱ	268 - 268	1	T → M in RNS; mild non-lethal form; decreased protein kinase activity. 		VAR_073661
Natural variant ⁱ	280 - 280	1	G → R in RNS. 		VAR_073662
Natural variant ⁱ	328 - 328	1	P → S in RNS; mild non-lethal form; decreased protein		VAR_073663

Updated annotations




... altres tasques i col·laboracions

Projecte Europeu **SYSVASC** 



- Detectar marcadors primerencs de malalties cardiovasculars
- 17 grups de 7 països diferents.
- Col·laboració interdisciplinària:
Metges, Biòlegs, Bioinformàtics, Químics i Biocuradors.

Curation: example from SysVasc PKN



Cell Reports
Article

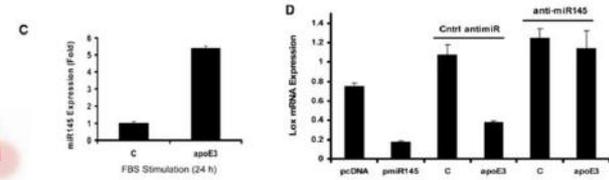


Cardiovascular Protection by ApoE and ApoE-HDL Linked to Suppression of ECM Gene Expression and Arterial Stiffening

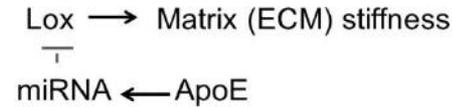
Devashish Kothapalli,^{1,4} Shu-Lin Liu,^{1,4} Yong Ho Bae,¹ James Monslow,² Tina Xu,¹ Elizabeth A. Hawthorne,¹ Fitzroy J. Byfield,¹ Paola Castagnino,¹ Shilpa Rao,¹ Daniel J. Rader,¹ Ellen Pure,² Michael C. Phillips,³ Sissei Lund-Katz,³ Paul A. Janney,¹ and Richard K. Asselant^{1,*}
¹Institute for Translational Medicine and Therapeutics, Departments of Pharmacology, Medicine, and Physiology, and the Molecular Profiling Facility, University of Pennsylvania, Philadelphia, PA 19104, USA
²Wistar Institute, Philadelphia, Philadelphia, PA 19104, USA
³The Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA
⁴These authors contributed equally to this work
 *Correspondence: assolan@mail.med.upenn.edu
<http://dx.doi.org/10.1016/j.celrep.2012.09.018>

ApoE maintains arterial elasticity by suppressing ECM gene expression.

Curation: example from SysVasc PKN



...enhanced crosslinking by elevated Lox expression has the potential to increase tissue stiffness...

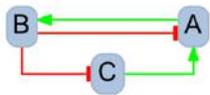


Network dynamics

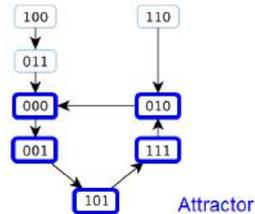
Synchronous update of all nodes:

Example: A,B,C = 1,0,0

A,B,C = 0,1,1



State transition graph:



A' = C AND NOT B

B' = A

C' = NOT B

Terminal strongly connected components of the state transition graph ~ biological phenotypes

Curation: example from SysVasc PKN



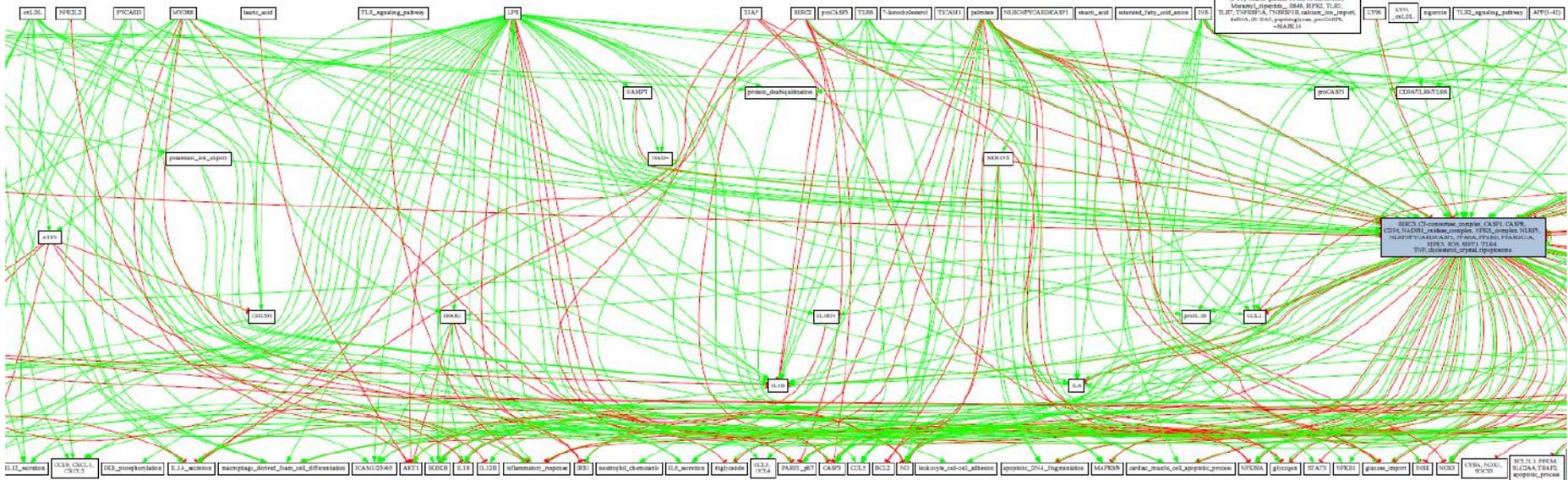
Source name	Source ID	Target name	Target ID	PMID	Source text	Evidence code def.	Evidence code ID	Expt. model ID
ApoE	P08226	miR145	M10000169	23103162	...	Exptl	ECO:0000006	CL:0000359
miR145	M10000169	Lox	P28301	23103162	...	Exptl	ECO:0000006	CL:0000359
Lox	P28301	Matrix stiffness	None	23103162	...	Author statement	ECO:0000033	None

ApoE → miR145 ⊥ Lox → Matrix stiffness

Projecte SysVasc

Prior Knowledge Network (PKN):

- > 200 publicacions sobre la formació de placa ateroscleròtica
- > 600 'nodes': >400 proteïnes, >200 d'altres entitats (metabòlits, processos...)
- ... finalització al desembre 2018



Divulgació científica



Swiss Institute of
Bioinformatics

Services & infrastructure ▾ Research ▾ Training ▾ Events **Bioinformatics for all** ▾ About us ▾



www.sib.swiss



Swiss Institute of
Bioinformatics

Agraïments



Prof. A. Celada group

Jorge Lloberas, PhD
Monica Comalada, PhD
Annabel F. Valledor, PhD
Eva Alvarez, PhD
Maria Serra, PhD
Carlos Sebastian, PhD
Ester Sanchez, PhD
Neus Serrat, PhD
Consol Farrera, PhD



Prof. B. Desvergne group

Aurélien Naldi, PhD
Christophe Héligon, PhD
Jérôme Feige, PhD
Federica Gilardi, PhD
Carine Winkler
Michael Baruchet



Prof. Ioannis Xenarios group

Lydie Bougueleret, PhD
Sylvain Poux, PhD
Alan Bridge, PhD
Lionel Breuza, PhD
Marie-Claude Blatter, PhD
Anne Estreicher, PhD



Per què la biologia?

Síndrome de Down

