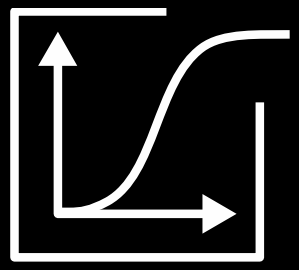


SOURCE DATA

Thomas Lemberger, EMBO

EMBO
excellence in life sciences

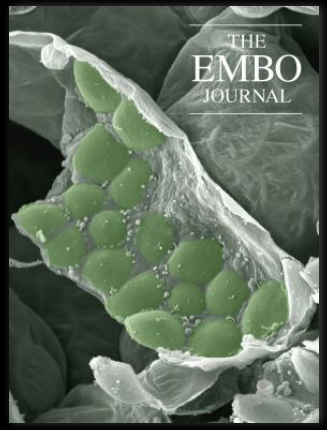
EMBOpress

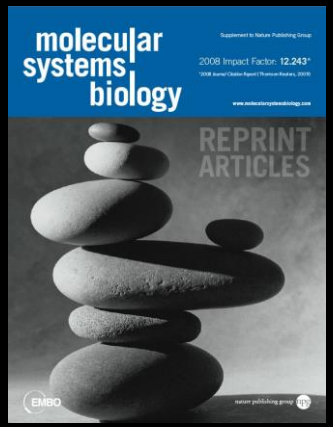
THE EMBO JOURNAL

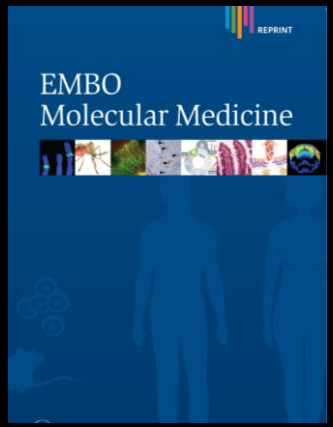
EMBO reports

molecular systems biology
OPEN ACCESS

EMBO Molecular Medicine
OPEN ACCESS







HighWire
Stanford University

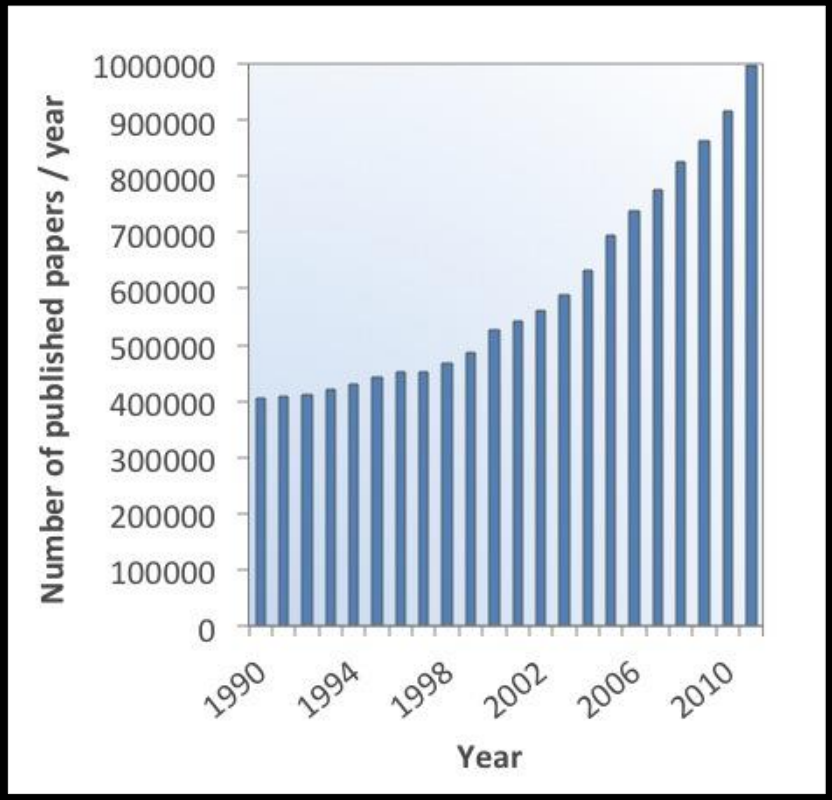
EMBO
excellence in life sciences

WILEY

Scientific publishing

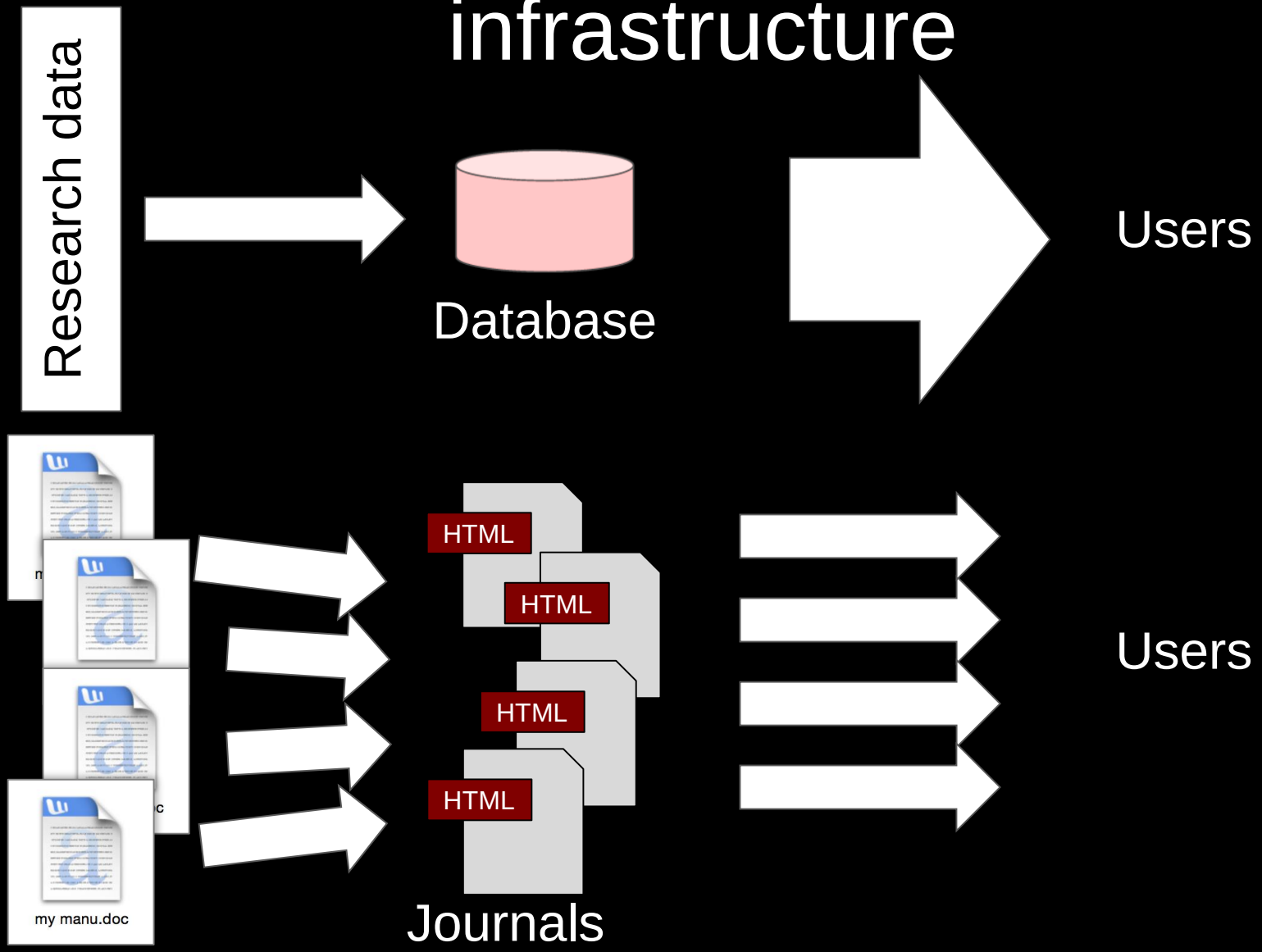
- Dominant channel for the dissemination of peer-reviewed data.
- Journals function as a proxy for quality in research assessment

- The rate of publishing keeps increasing.
- Papers are human-readable but poorly machine-readable.

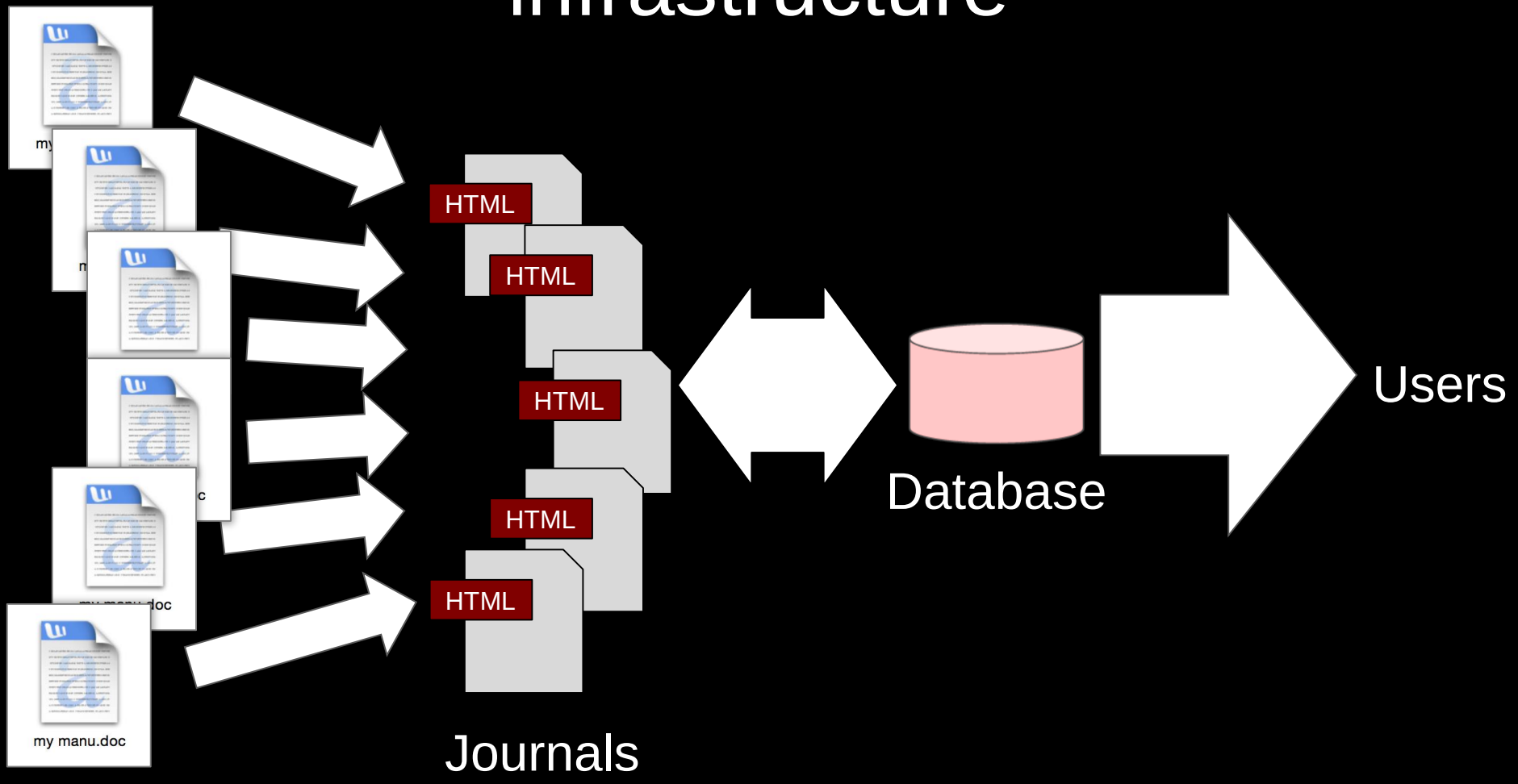



search

Publishing as a distributed infrastructure

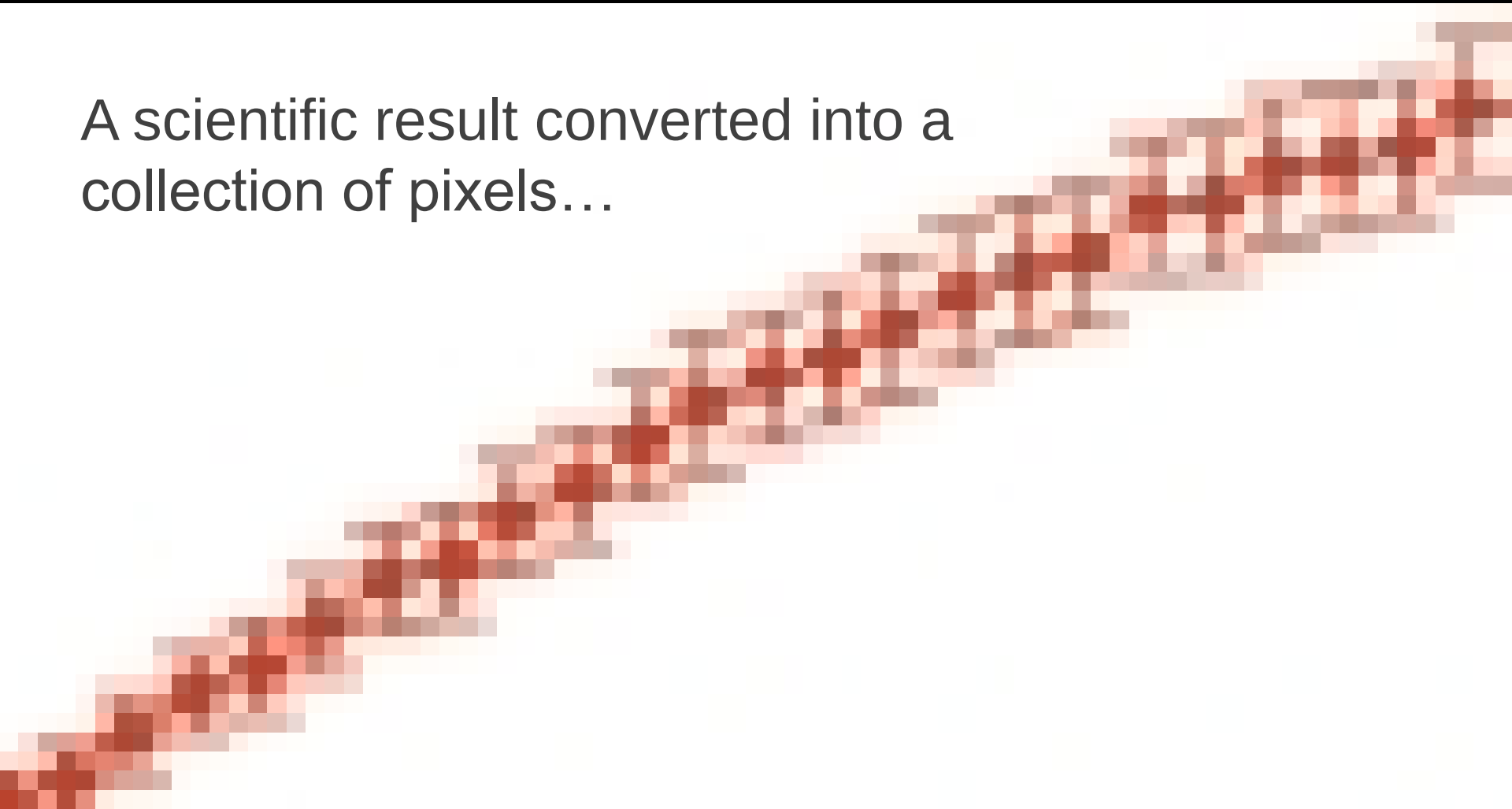


Publishing as a distributed infrastructure

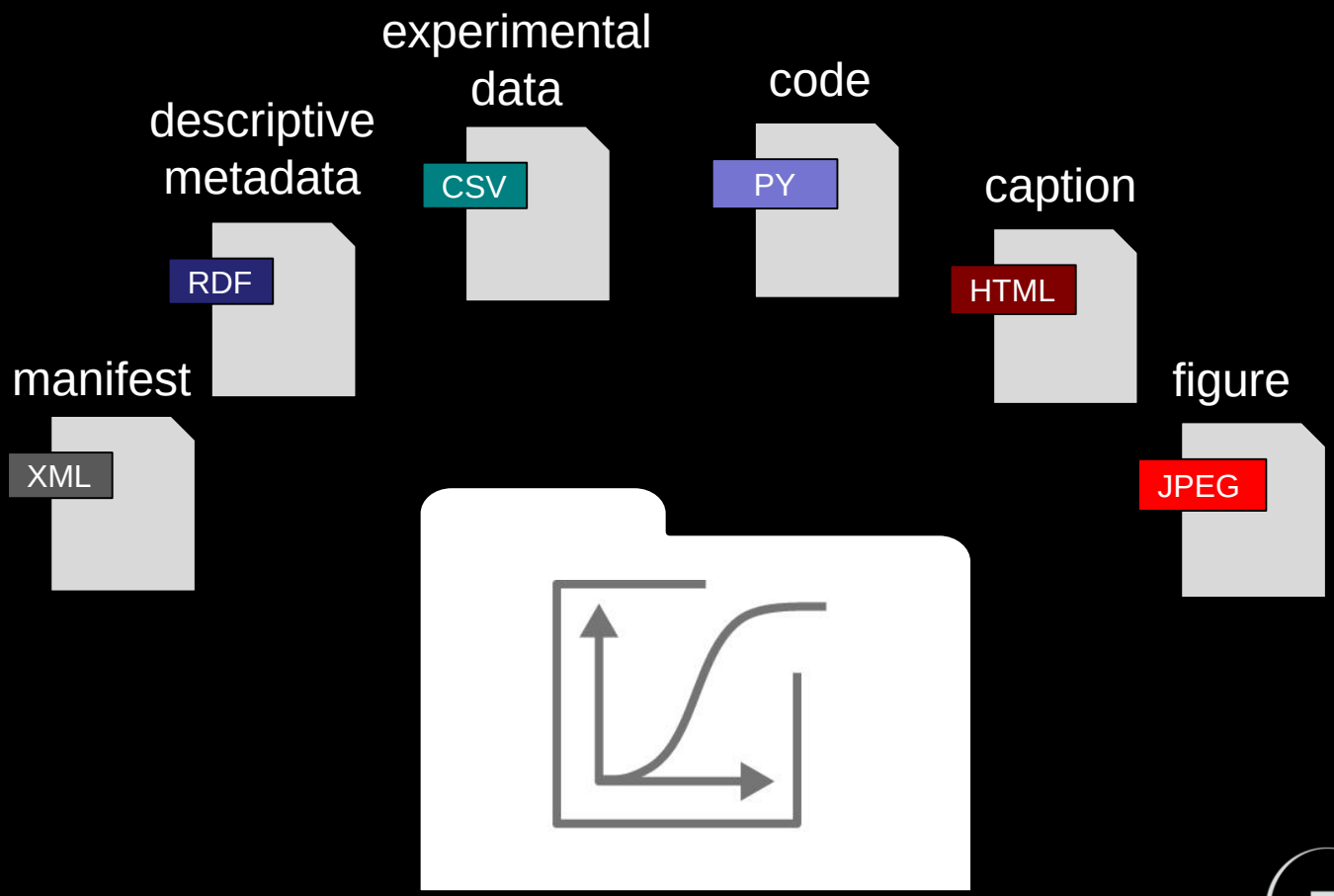


What is a figure?

A scientific result converted into a collection of pixels…



Figures as data packages

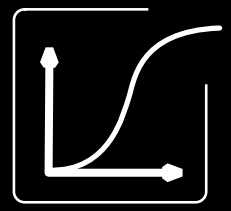


EMBO
excellence in life sciences

SourceData

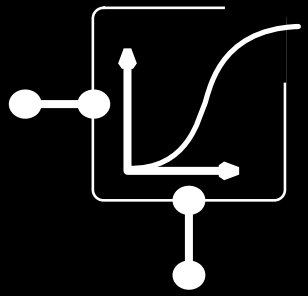
- Issue 1: data availability for re-analysis

  ⇒ Source data behind the figures



- Issue 2: structure

  ⇒ Structured descriptive metadata



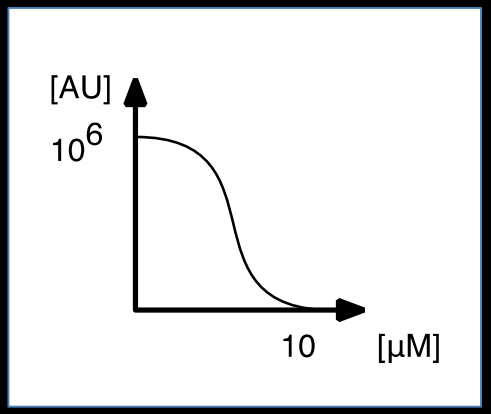
- Issue 3: discoverability

  ⇒ Data-oriented search

EMBO
excellence in life sciences

Data / Metadata

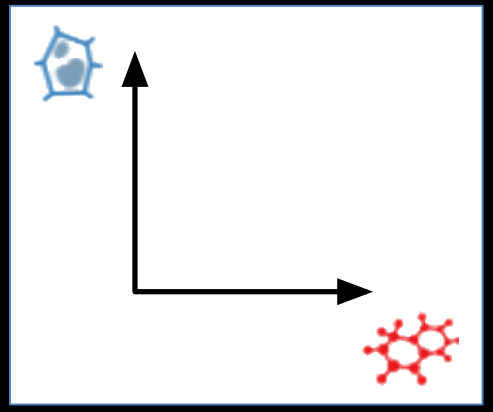
Measurement



Data

+

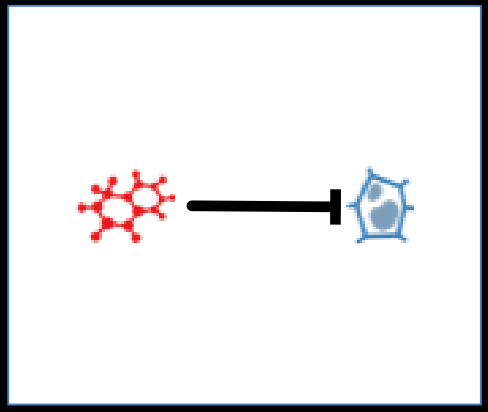
Design



Descriptive metadata

=

Interpretation

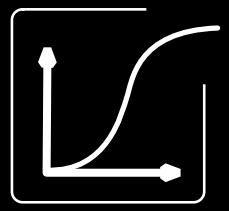


Statement

SourceData

- Issue 1: data availability for re-analysis

  ⇒ Source data behind the figures



- Issue 2: structure

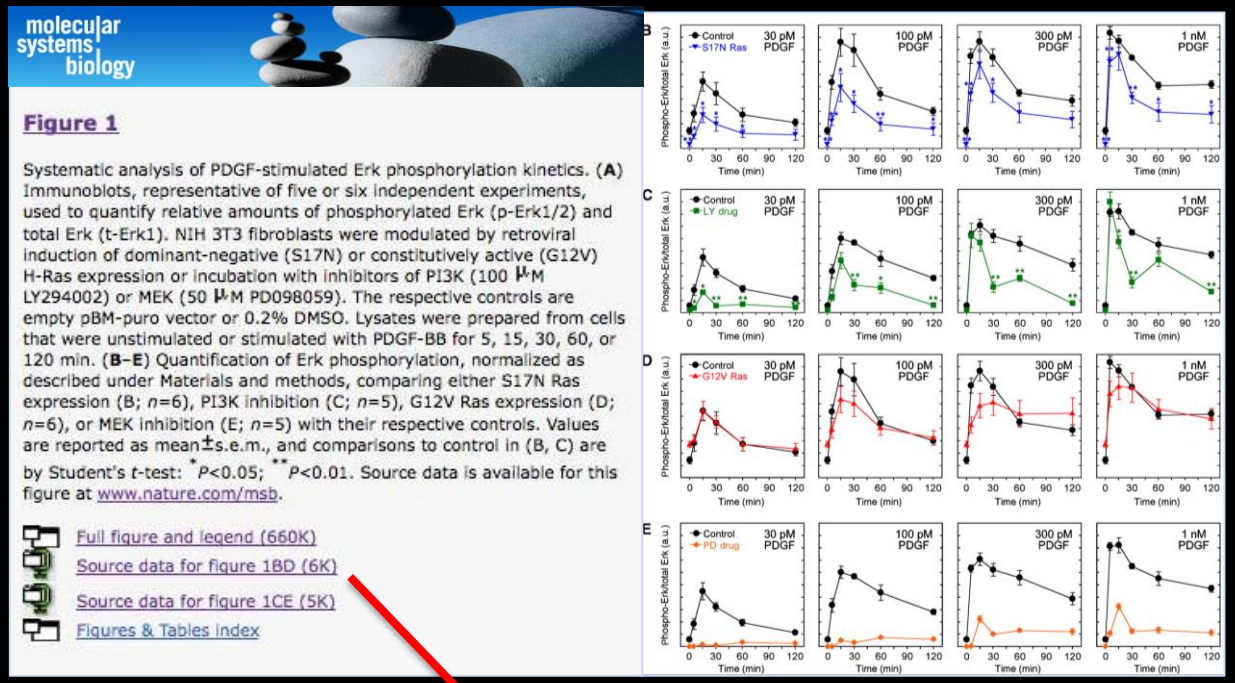
  ⇒ Structured descriptive metadata

- Issue 3: discoverability

  ⇒ Data-oriented search

EMBO
excellence in life sciences

Figure source data



molecular systems biology

Figure 1

Systematic analysis of PDGF-stimulated Erk phosphorylation kinetics. (**A**) Immunoblots, representative of five or six independent experiments, used to quantify relative amounts of phosphorylated Erk (p-Erk1/2) and total Erk (t-Erk1). NIH 3T3 fibroblasts were modulated by retroviral induction of dominant-negative (S17N) or constitutively active (G12V) H-Ras expression or incubation with inhibitors of PI3K (100 μM LY294002) or MEK (50 μM PD098059). The respective controls are empty pBM-puro vector or 0.2% DMSO. Lysates were prepared from cells that were unstimulated or stimulated with PDGF-BB for 5, 15, 30, 60, or 120 min. (**B–E**) Quantification of Erk phosphorylation, normalized as described under Materials and methods, comparing either S17N Ras expression (B; *n*=6), PI3K inhibition (C; *n*=5), G12V Ras expression (D; *n*=6), or MEK inhibition (E; *n*=5) with their respective controls. Values are reported as mean±s.e.m., and comparisons to control in (B, C) are by Student's *t*-test: $^{*}P<0.05$; $^{**}P<0.01$. Source data is available for this figure at www.nature.com/msb.

Full figure and legend (660K)
Source data for figure 1BD (6K)
Source data for figure 1CE (5K)
Figures & Tables index

Fig1b&d_raw.txt

	A	B	C	D
1	Fig1b&d-Column1	Fig1b&d-Column2A	Fig1b&d-Column2B	Fig1b&d-Column2C
2	Time (min)	p-Erk/t-Erk, 30 pM PDGF, control vector, Expt. 1	p-Erk/t-Erk, 30 pM PDGF, control vector, Expt. 2	p-Erk/t-Erk, 30 pM PDGF, control vector, Expt. 3
3	0	0.194672394	0.201524091	0.339116171
4	5	0.395173883	0.389974466	0.555355249
5	15	0.690917146	1.236910363	1.632582883
6	30	0.394324884	0.72081196	1.488299981
7	60	0.38782972	0.38107614	0.428561181
8	120	0.384442827	0.216360469	0.458929493
9				
10				

Fig1b&d_raw.txt
Ready

EMBO excellence in life sciences

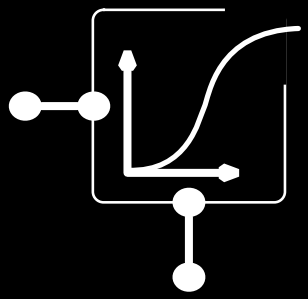
SourceData

- Issue 1: data availability for re-analysis

  ⇒ Source data behind the figures

- Issue 2: structure

  ⇒ Structured descriptive metadata



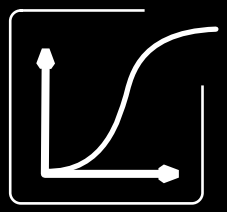
- Issue 3: discoverability

  ⇒ Data-oriented search

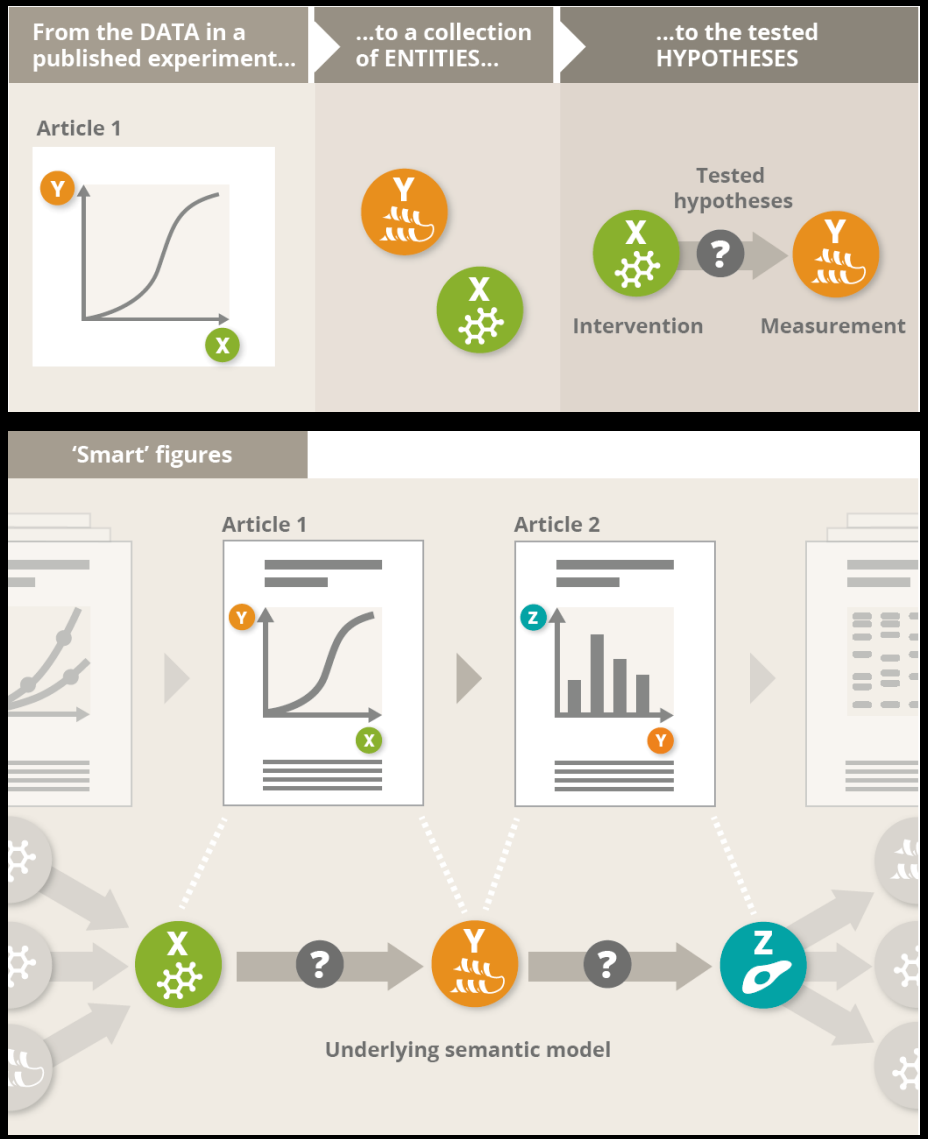
EMBO
excellence in life sciences

SourceData

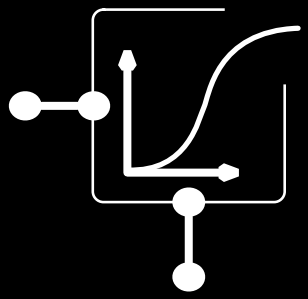
data



search

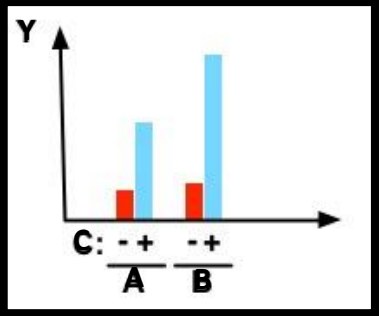


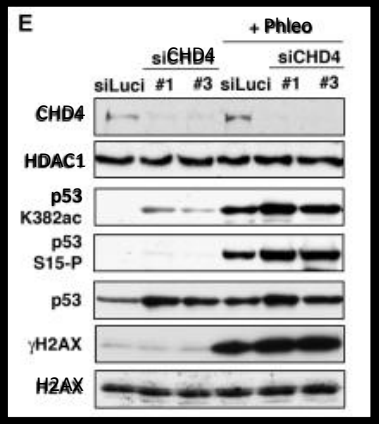
metadata

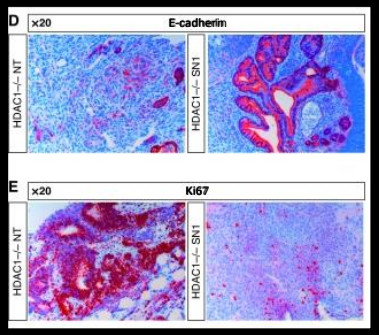


EMBO
excellence in life sciences

Descriptive metadata I



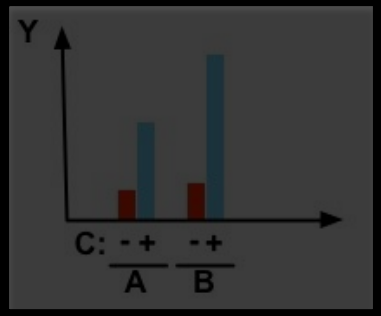


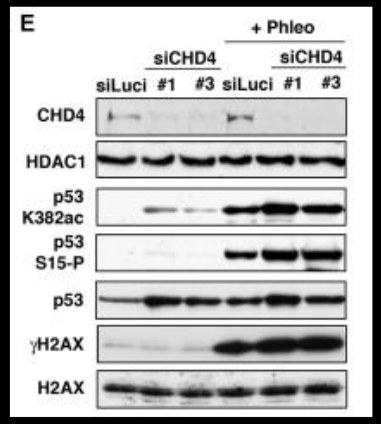


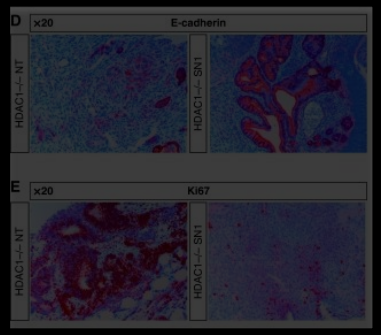
SOURCE DATA

EMBO excellence in life sciences

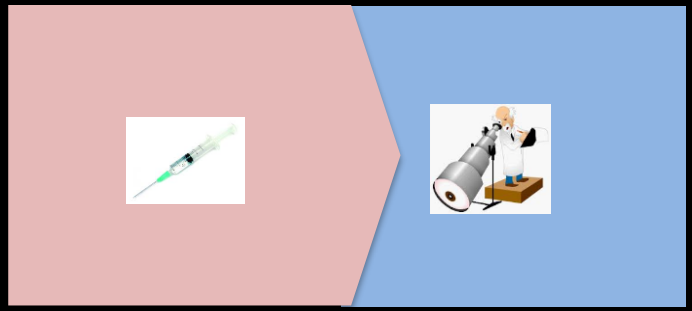
Descriptive metadata II

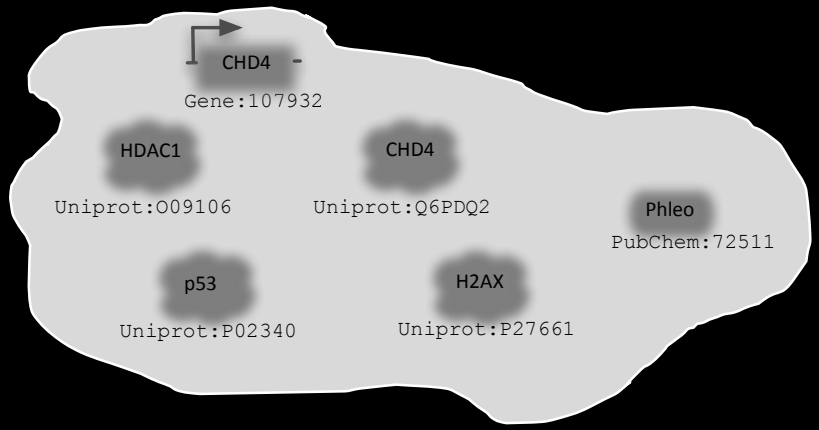






‘intervention’ ‘observation’





SOURCE DATA

EMBO excellence in life sciences

SourceData

- Issue 1: data availability for re-analysis

  ⇒ Source data behind the figures

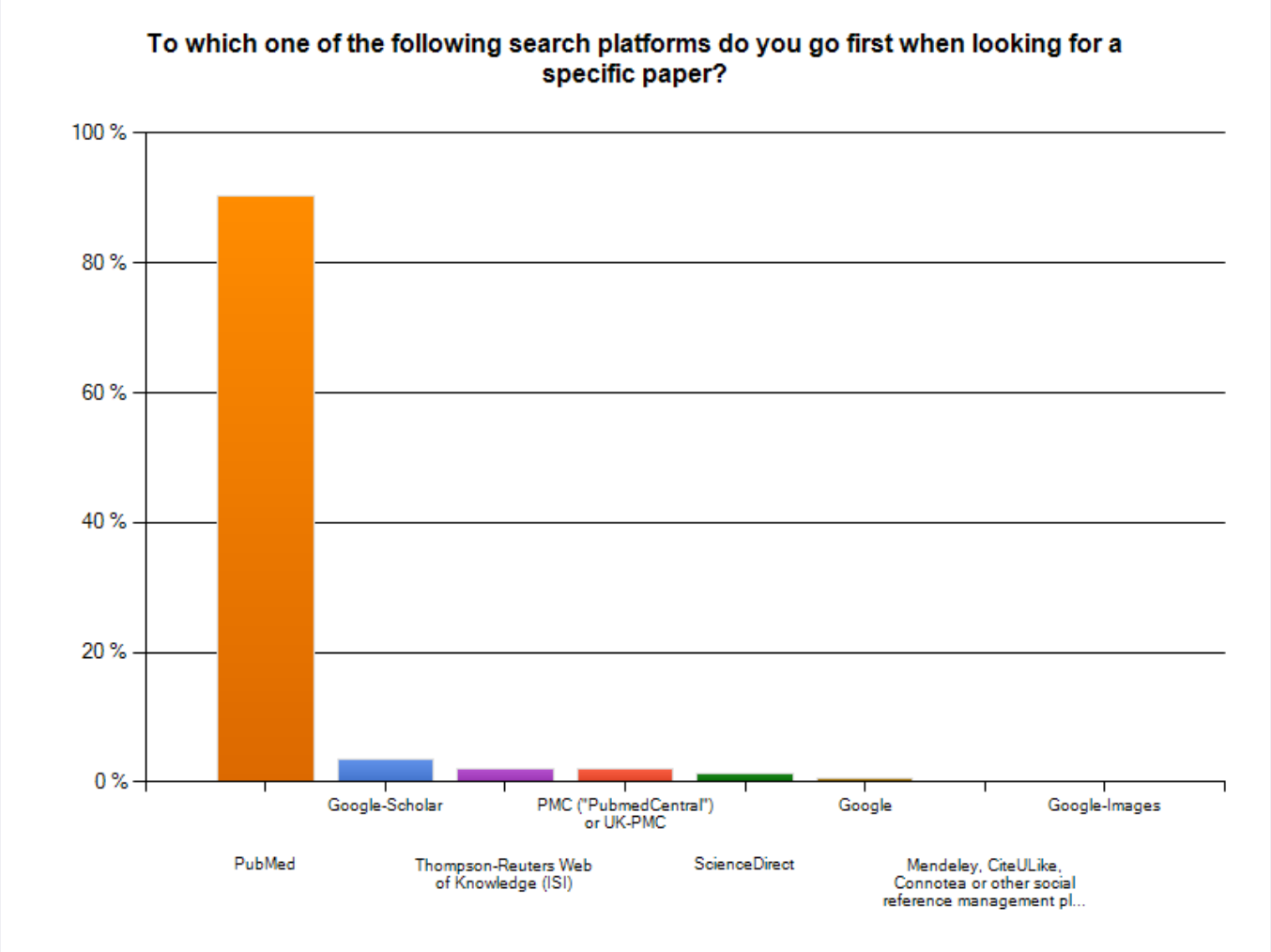
- Issue 2: structure

  ⇒ Structured descriptive metadata

- Issue 3: discoverability

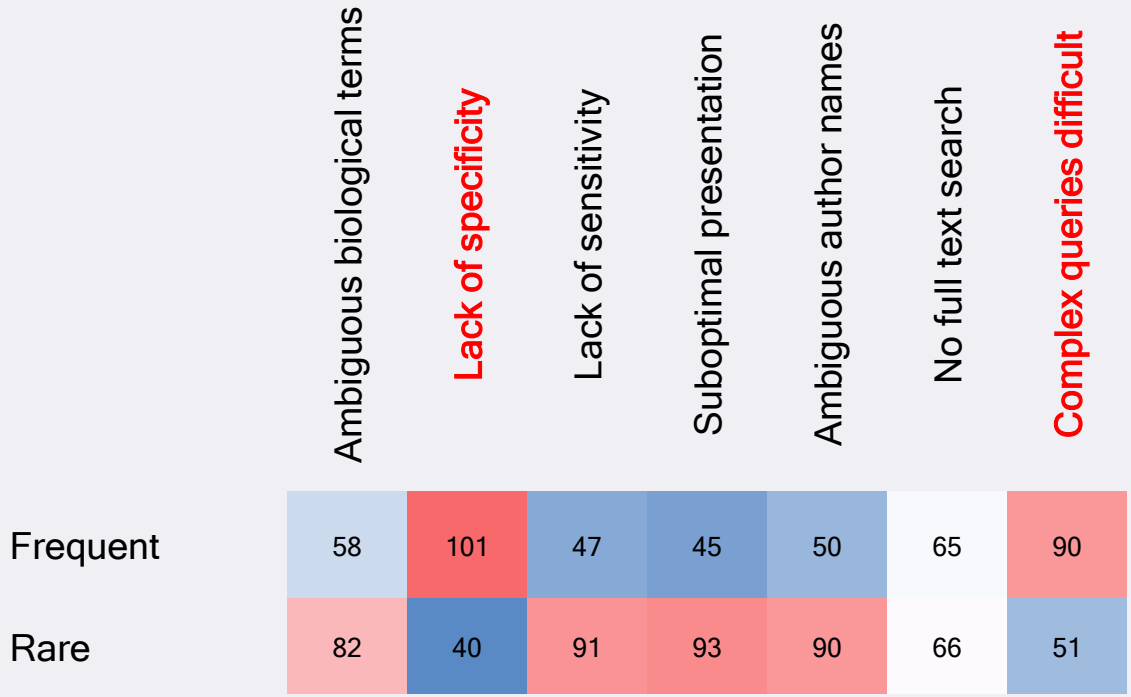
  ⇒ Data-oriented search

EMBO
excellence in life sciences



EMBO
excellence in life sciences

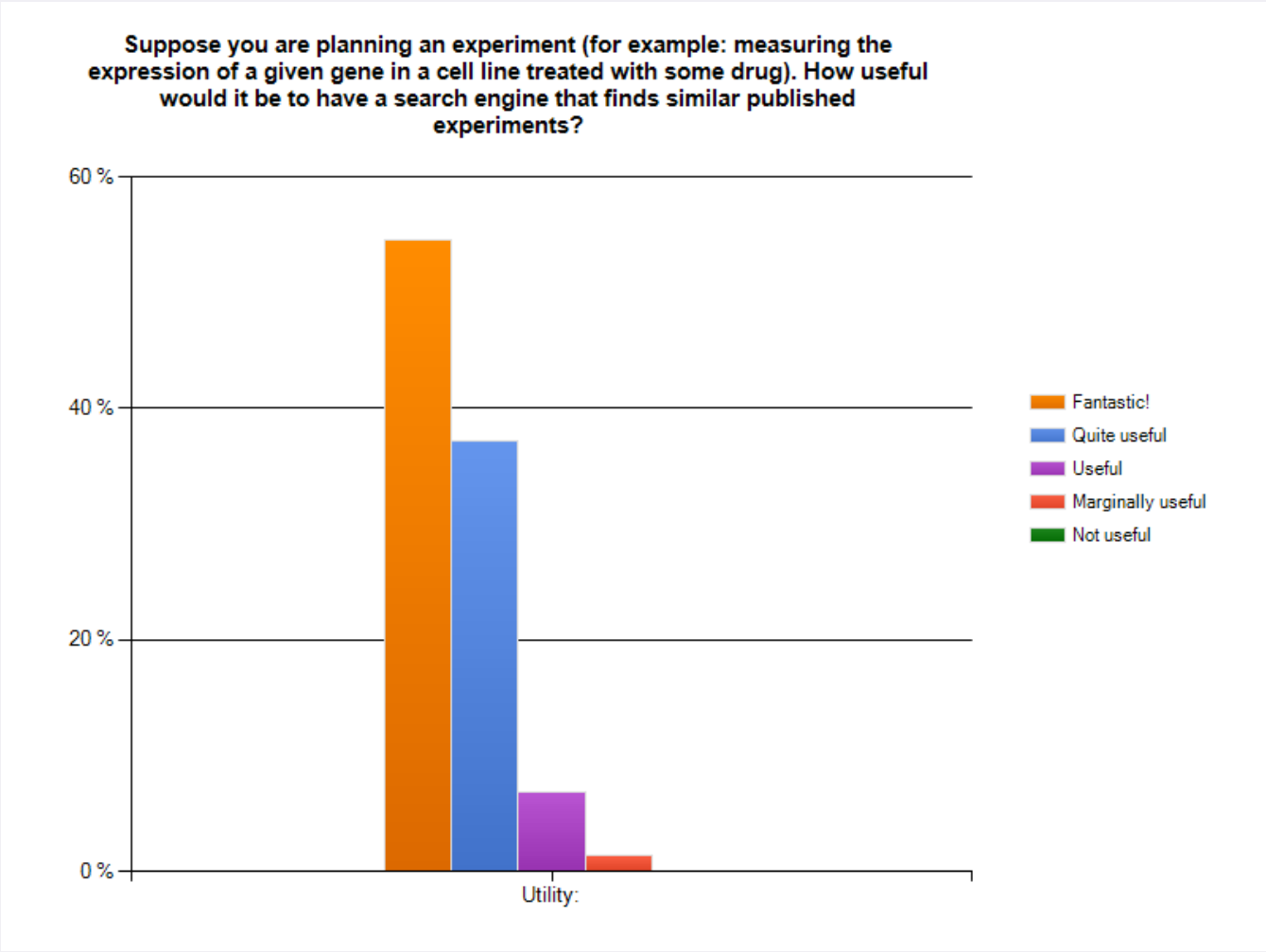
How frequently did the following issues negatively affect your ability to find relevant scientific literature?



EMBO
excellence in life sciences

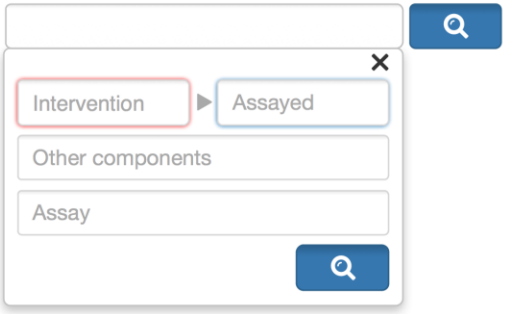
Using PubMed only, how much time (in minutes, real time or estimated) would you need to solve the following search 'exercises':

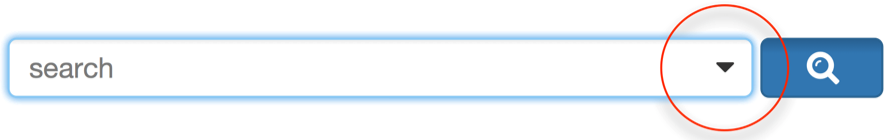
“Did anybody publish an experiment involving the siRNA-mediated knock down of gene X [X=some gene you do NOT know well]?”	25 min
“Find five substrates of the cyclin-dependent kinase CDK1 in mammalian cells?”	30 min
“Find more than three papers that include a time-course of p53 activation upon DNA damage”	45 min

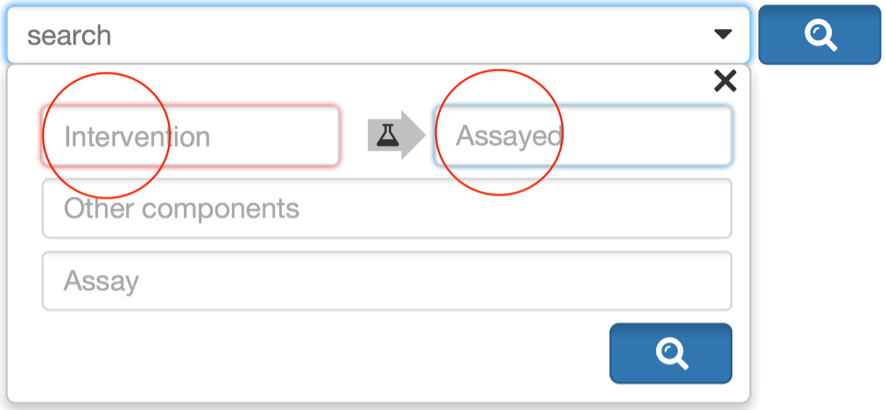


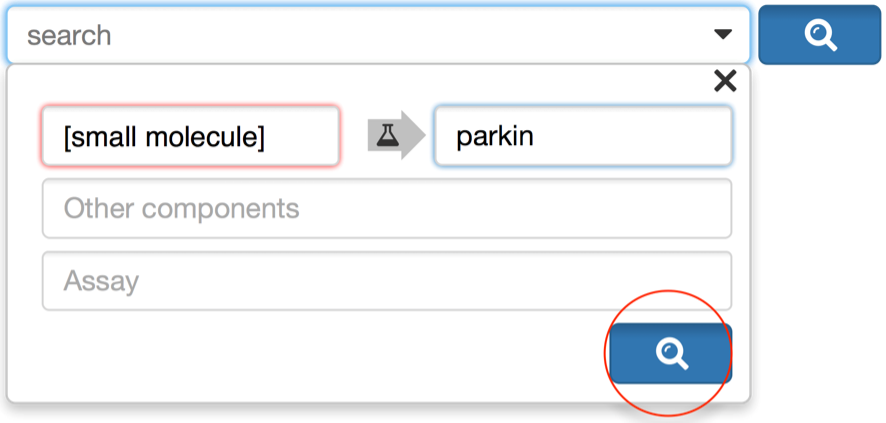
EMBO
excellence in life sciences

Data-oriented search




search


search
Intervention
Assayed
Other components
Assay

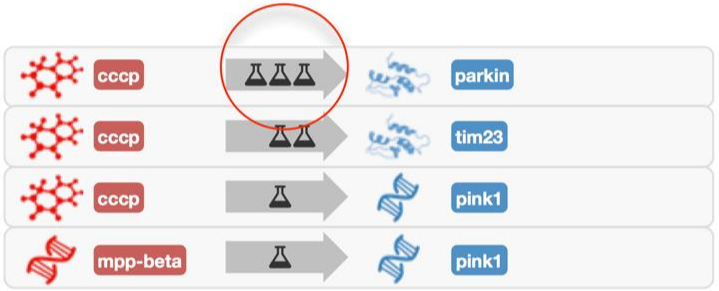

search
[small molecule]
parkin
Other components
Assay

intervention:[small molecule] assayed:parkin

Mitochondrial processing peptidase regulates PINK1 processing, import and Parkin recruitment

Greene AW1, Grenier K, Aguileta MA, Muise S, Farazifard R, Haque ME, McBride HM, Park DS, Fon EA

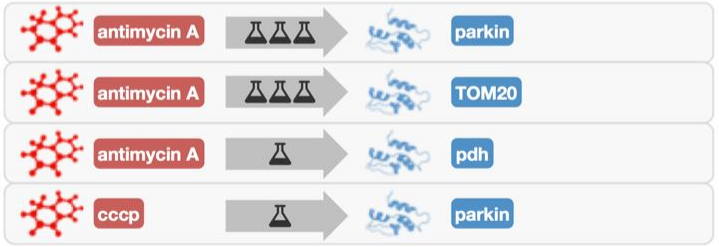
EMBO Rep. 2012 Apr;13(4):378-85. doi: 10.1038/embor.2012.14



Parkin and PINK1 function in a vesicular trafficking pathway regulating mitochondrial quality control

Greene AW1, Grenier K, Aguileta MA, Muise S, Farazifard R, Haque ME, McBride HM, Park DS, Fon EA

EMBO Rep. 2012 Apr;13(4):378-85. doi: 10.1038/embor.2012.14

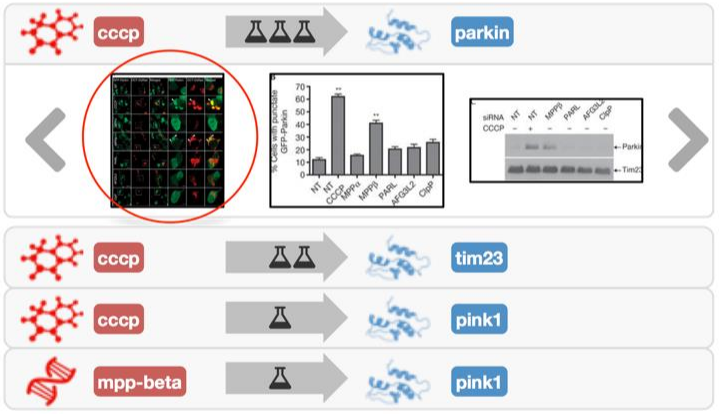


intervention:[small molecule] assayed:parkin

Mitochondrial processing peptidase regulates PINK1 processing, import and Parkin recruitment

Greene AW1, Grenier K, Aguileta MA, Muise S, Farazifard R, Haque ME, McBride HM, Park DS, Fon EA

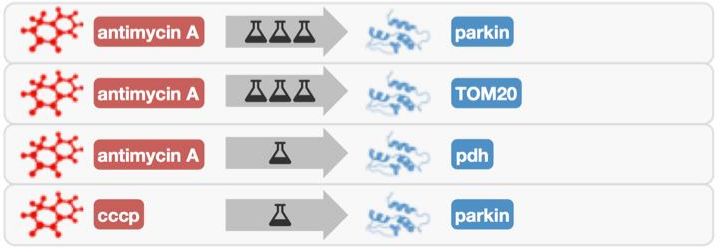
EMBO Rep. 2012 Apr;13(4):378-85. doi: 10.1038/embor.2012.14



Parkin and PINK1 function in a vesicular trafficking pathway regulating mitochondrial quality control

Greene AW1, Grenier K, Aguileta MA, Muise S, Farazifard R, Haque ME, McBride HM, Park DS, Fon EA

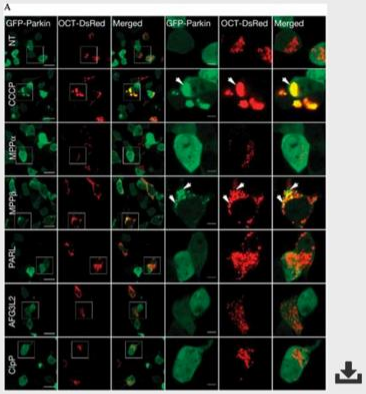
EMBO Rep. 2012 Apr;13(4):378-85. doi: 10.1038/embor.2012.14



From: **Figure 5. MPPβ knockdown induces Parkin recruitment and mitophagy**

Interventions

- CCCP
- AFG3L2
- ClpP
- MPP alpha
- MPP beta
- PARL



Assayed

- OCT
- Parkin

« A B C D E »

Live confocal micrographs of siRNA-, GFP-Parkin- and OCT-DsRed-transfected HEK293T cells ±CCCP; Scale bars, 20 μm (low magnification) and 5 μm (high magnification).

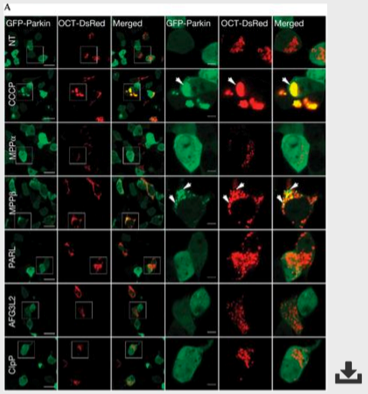
Related data ○ This paper only ◉ Global search See all (12)...

FIP200 complex. (**A**) Immunoblot analysis of Atg proteins in the indicated MEFs. (**B**) m5-7 cells were cultured in regular medium containing 10 ng/ml Dox for 8 days, and then cultured in the absence of Dox for a further 2 days. (**C**,**D**) m5-7 cells were cultured in regular medium with or without 10 ng/ml Dox for 4 days, and then cultured in the presence of 50 μg/ml CHX for the indicated time periods (**C**). mRNA levels of the indicated genes were measured by quantitative PCR following the 4-day Dox treatment. Data represent mean±s.e. (**P*<0.05) (**D**). CHX, cycloheximide; Dox, doxycycline; MEFs, mouse embryonic fibroblasts; mRNA, messenger RNA; WT, wild-type.

From: **Figure 5. MPPβ knockdown induces Parkin recruitment and mitophagy**

Interventions

- CCCP
- AFG3L2
- ClpP
- MPP alpha
- MPP beta
- PARL



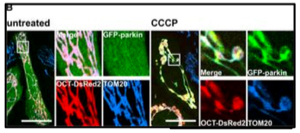
Assayed

- OCT
- Parkin

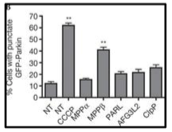
« A B C D E »

Live confocal micrographs of siRNA-, GFP-Parkin- and OCT-DsRed-transfected HEK293T cells ±CCCP; Scale bars, 20 μm (low magnification) and 5 μm (high magnification).

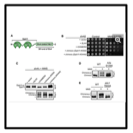
Related data ○ This paper only ◉ Global search See all (12)...



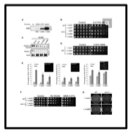
CCCP | Parkin OCT TOM20



CCCP AFG3L2 ClpP | Parkin mitochondria



Parkin PARL Baf A



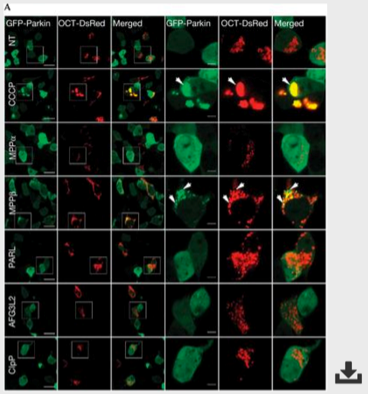
CCCP OCT Drp1

From: **Figure 5. MPPβ knockdown induces Parkin recruitment and mitophagy**

> PARL

Interventions

- CCCP
- AFG3L2
- ClpP
- MPP alpha
- MPP beta
- PARL



Assayed

- ~~OCT~~
- Parkin

« A B C D E »

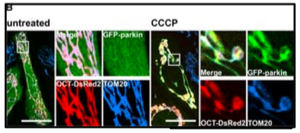
Live confocal micrographs of siRNA-, GFP-Parkin- and OCT-DsRed-transfected HEK293T cells ±CCCP; Scale bars, 20 µm (low magnification) and 5 µm (high magnification).

Upstream of PARL [Uniprot:XS002]

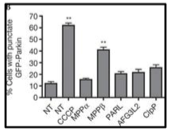
PINK1 PARL

Mitochondrial processing peptidase regulates PINK1 processing, import and Parkin recruitment. Greene et al, *EMBO rep.* 2012

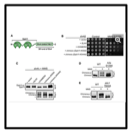
Related data ○ This paper only ◉ Global search See all (12)...



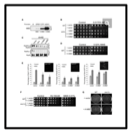
CCCP | Parkin OCT TOM20



CCCP AFG3L2 ClpP | Parkin mitochondria



Parkin PARL Baf A



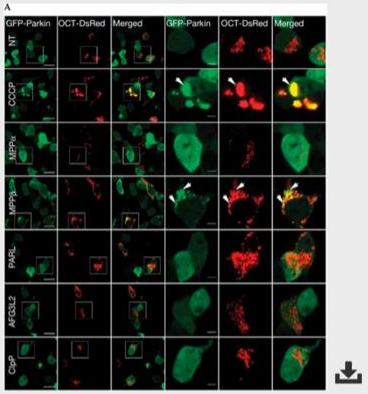
CCCP OCT Drp1

From: **Figure 5. MPPβ knockdown induces Parkin recruitment and mitophagy**

> parkin

Interventions

- CCCP
- AFG3L2
- ClpP
- MPP alpha
- MPP beta
- PARL



Assayed

- OCT
- Parkin

« A B C D E »

Live confocal micrographs of siRNA-, GFP-Parkin- and OCT-DsRed-transfected HEK293T cells ±CCCP; Scale bars, 20 μm (low magnification) and 5 μm (high magnification).

Downstream of **Parkin** [Uniprot:O60260]

Parkin **Alpha-synuclein**

Ubiquitination of a new form of alpha-synuclein by parkin from human brain: implications for Parkinson's disease. Shimmer et al, *Science 2001*

Parkin **Pael-R**

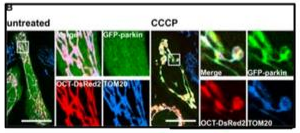
An unfolded putative transmembrane polypeptide, which can lead to endoplasmic reticulum stress, is a substrate of Parkin. Imani et al, *Cell* 2001

Parkin **Human**

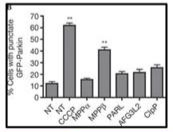
Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. Kitada et al, *Nature J.* 1998

See all...

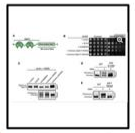
Related data ○ This paper only ◉ Global search See all (12)...



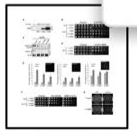
CCCP **Parkin** **OCT** TOM20



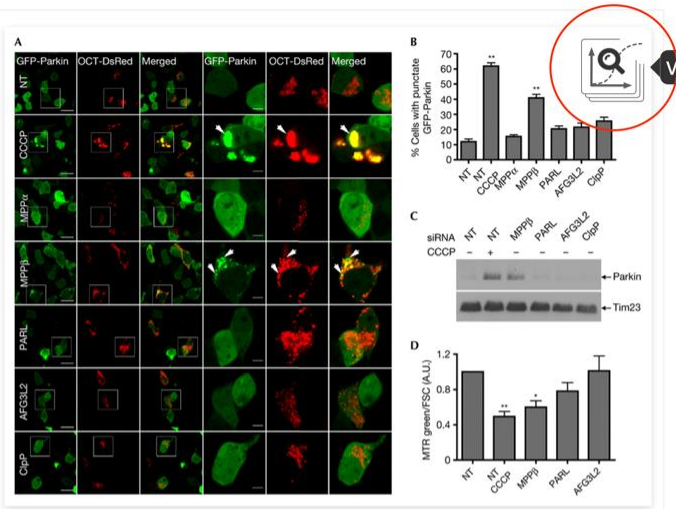
CCCP **AFG3L2** **ClpP** **Parkin** mitochondria



Parkin **PARL** Baf A



CCCP **OCT** Drp1



View as SmartFigure

Download figure | Open in new tab | Download powerpoint

Figure 5.

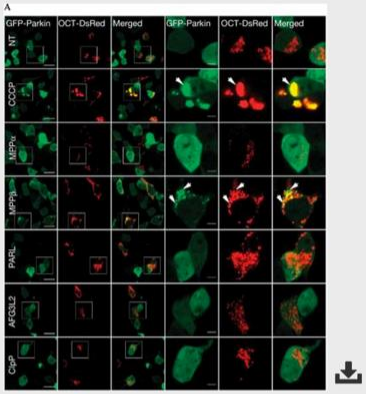
MPPβ knockdown induces Parkin recruitment and mitophagy. (**A**) Live confocal micrographs of siRNA-, GFP-Parkin- and OCT-DsRed-transfected HEK293T cells ±CCCP; Scale bars, 20 μm (low magnification) and 5 μm (high magnification). (**B**) Percentage of transfected cells displaying a punctate mitochondrial pattern of GFP-Parkin distribution, n=3; $^{*}P$<0.05; $^{**}P$<0.01 compared with NT siRNA. (**C**) Immunoblots of endogenous Parkin in mitochondrial fractions from siRNA-transfected cells ±CCCP. (**D**) Mitochondrial mass measured by flow cytometry of MTR Green fluorescence in siRNA-transfected cells treated ±CCCP for 24 h, n=3. A.U., arbitrary unit; CCCP, carbonyl cyanide m-chlorophenyl hydrazone; GFP, green fluorescent protein; MPP, mitochondrial processing peptidase; MTR, MitoTracker; NT, non-targeting; PARL, presenilin-associated rhomboid-like protease; PINK1, phosphatase and tensin homologue-induced kinase 1; siRNA, short interfering RNA.

Finally, we examined whether Parkin recruitment induced by MPPβ knockdown could

From: **Figure 5. MPPβ knockdown induces Parkin recruitment and mitophagy**

Interventions

- CCCP
- AFG3L2
- ClpP
- MPP alpha
- MPP beta
- PARL



Assayed

- OCT
- Parkin

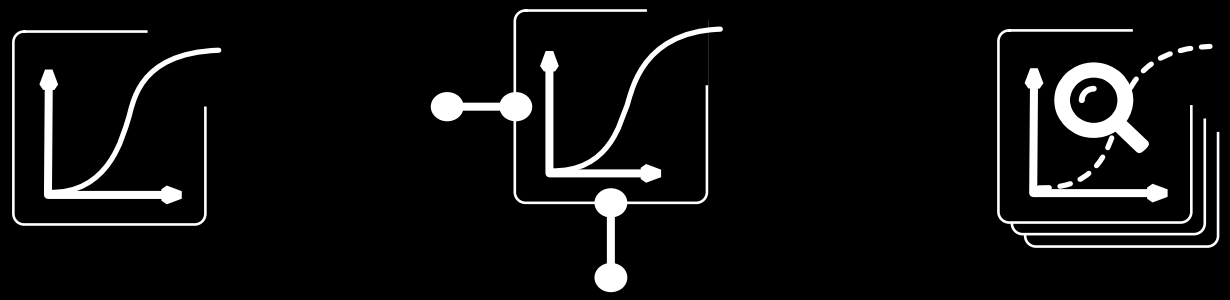
« A B C D E »

Live confocal micrographs of siRNA-, GFP-Parkin- and OCT-DsRed-transfected HEK293T cells ±CCCP; Scale bars, 20 μm (low magnification) and 5 μm (high magnification).

Related data ○ This paper only ◉ Global search See all (12)...

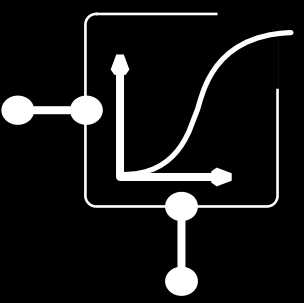
FIP200 complex. (**A**) Immunoblot analysis of Atg proteins in the indicated MEFs. (**B**) m5-7 cells were cultured in regular medium containing 10 ng/ml Dox for 8 days, and then cultured in the absence of Dox for a further 2 days. (**C**,**D**) m5-7 cells were cultured in regular medium with or without 10 ng/ml Dox for 4 days, and then cultured in the presence of 50 μg/ml CHX for the indicated time periods (**C**). mRNA levels of the indicated genes were measured by quantitative PCR following the 4-day Dox treatment. Data represent mean±s.e. (**P*<0.05) (**D**). CHX, cycloheximide; Dox, doxycycline; MEFs, mouse embryonic fibroblasts; mRNA, messenger RNA; WT, wild-type.

SourceData



- Provide tools adapted to the workflow of researchers and of publishers
- Couple data availability to increased visibility of the papers
- Current scope: hypothesis-driven research in cell & molecular biology

EMBO
excellence in life sciences



Towards integrating biocuration into the publishing workflow

Pre-acceptance stage → Tools for Data editor → Authors' approval → Publish

EMBOpress

Team

EMBO

Sara El-Gebali

Nancy George

Thomas Lemberger

Robert Bosch Stiftung

OUR PARTNERS

EMBOpress

WILEY

HighWire

SIB Swiss Institute of Bioinformatics

nature publishing group npg

SIB

Robin Liechti

Lou Götz

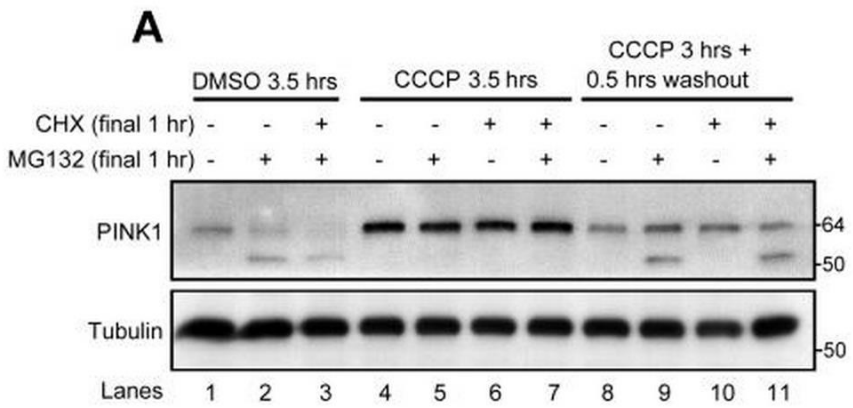
Isaac Crespo

Ioannis Xenarios

Vital-IT High Performance Computing Center

Tagging & 2D annotation

Figure 2-A



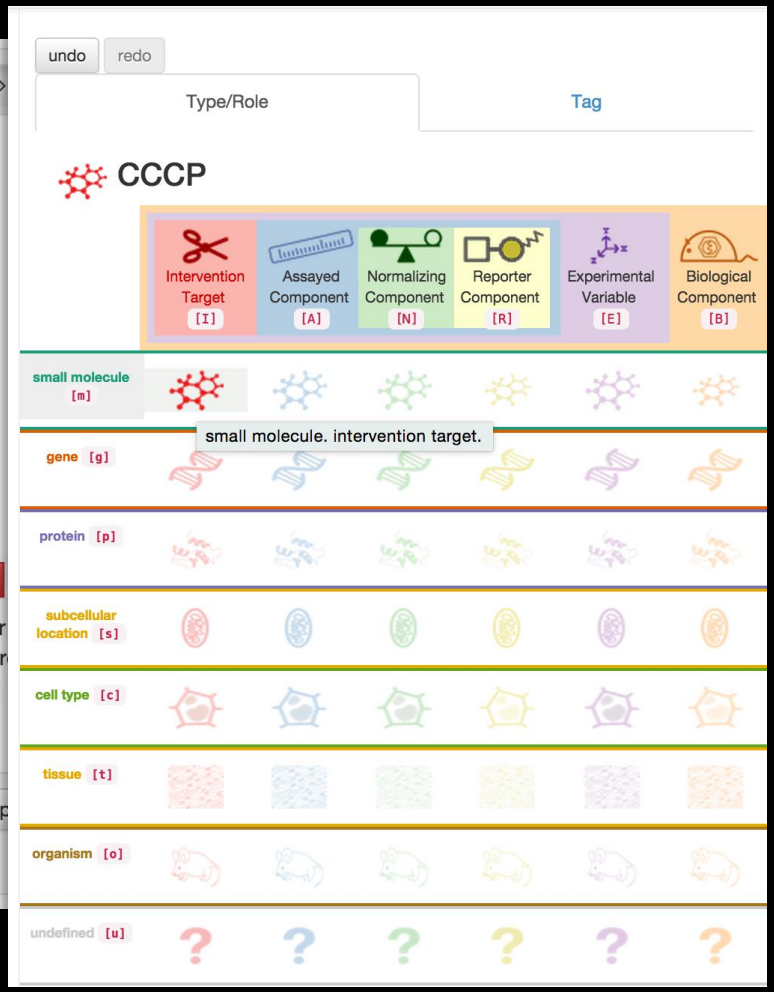
confirm delete

A) HeLa cells stably expressing YFP-Parkin were treated with DMSO for 3.5 h, 2 μM CCCP for 3.5 h, or CCCP for followed by washout of CCCP for 0.5 h in the absence of serum. 50 μM MG132 and/or 100 μM cyclohexamide wer added for the last 1 h of treatment. Whole-cell lysates (WCL) run on SDS gels and immunoblotted for endogenous PINK1 and tubulin.

Click on a tag and hold the ALT key to confirm / delete all tags or suggestions of the same text and type.

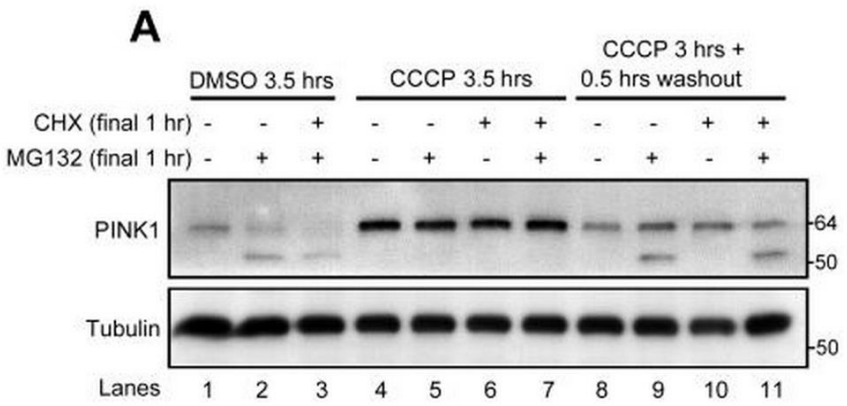
perform on all tag suggestions of this panel: confirm delete

delete p



Tagging & 2D annotation

Figure 2-A



A) HeLa cells stably expressing YFP-Parkin were treated with DMSO for 3.5 h, 2 μM
3 h followed by washout of CCCP for 0.5 h in the absence of serum. 50 μM MG132 and/
were added for the last 1 h of treatment. Whole-cell lysates (WCL) run on SDS gels and im
PINK1 and tubulin.

Click on a tag and hold the ALT key to confirm / delete all tags or suggestions of the same text and type.

perform on all tag suggestions of this
confirm delete

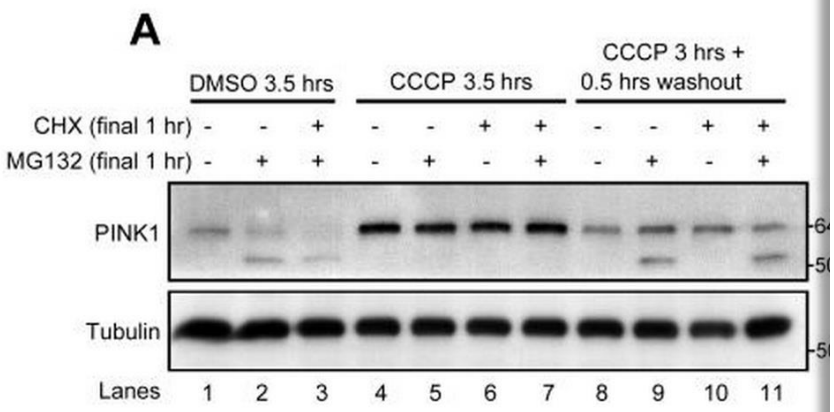
List | Condensed | Summary | new tag

Tag	External ID	Taxon
Figure 2-A		
CCCP		
CCCP		
CCCP		
cyclohexamide		
MG132		
PINK1		Homo sapiens
tubulin		Homo sapiens
DMSO		
Parkin		Homo sapiens
YFP		Homo sapiens
HeLa cells		Homo sapiens

EMBO excellence in life sciences

Visual feedback

Figure 2-A



A) HeLa cells stably expressing YFP-Parkin were treated with DMSO for 3.5 h, 2 μ
3 h followed by washout of CCCP for 0.5 h in the absence of serum. 50 μM MG132 a
were added for the last 1 h of treatment. Whole-cell lysates (WCL) run on SDS gels and
PINK1 and tubulin.

Click on a tag and hold the ALT key to confirm / delete all tags or suggestions of the same text and type.

perform on all tag suggestions of
confirm delete

List Condensed Summary new tag

Biological Component

HeLa cells YFP Parkin DMSO

Experiment Variable

Intervention Target

CCCP MG132

cyclohexamide

Assayed Component

PINK1

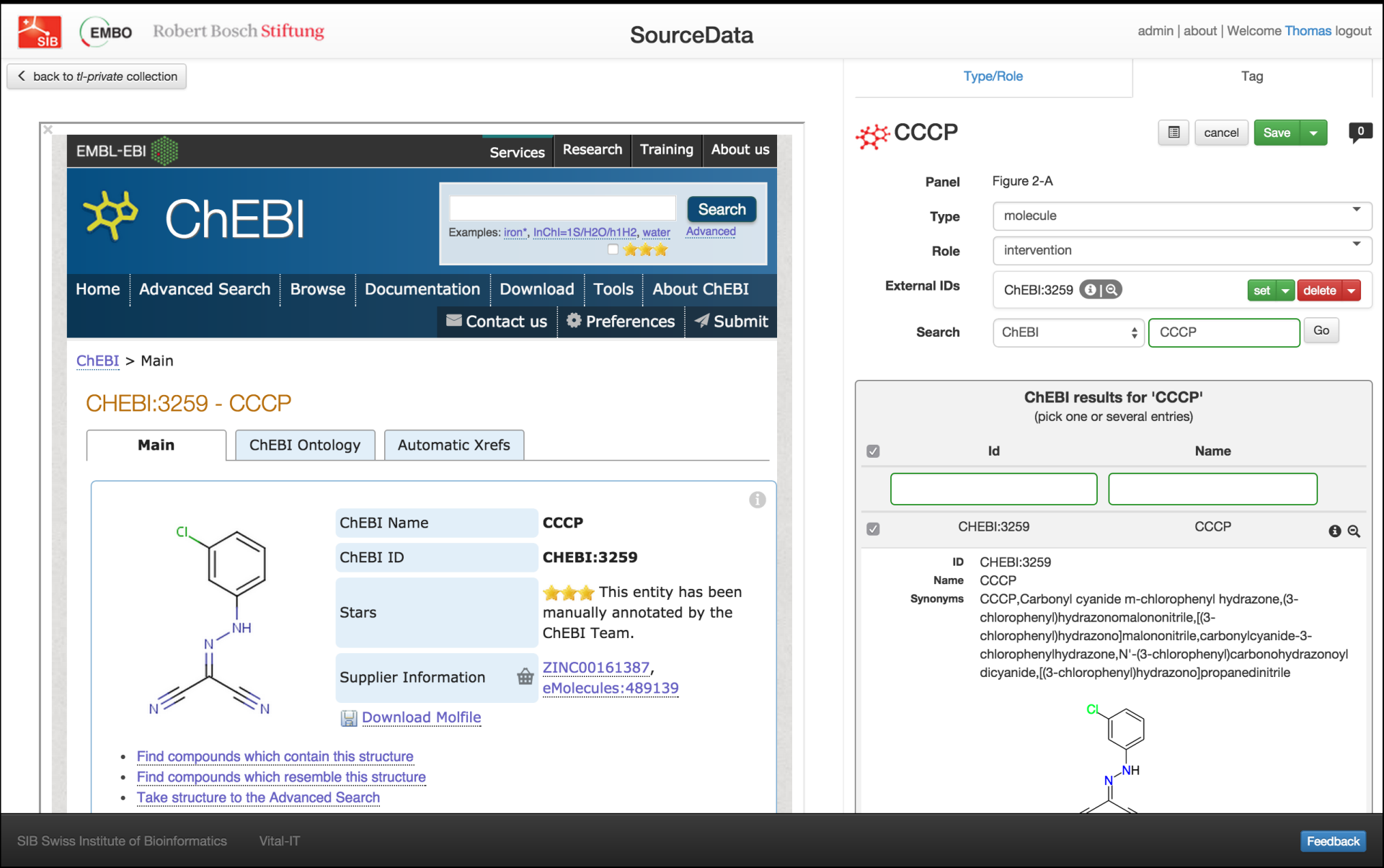
Normalizing Component

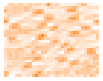
tubulin

Reporter Component

Category Variable

Links to external databases

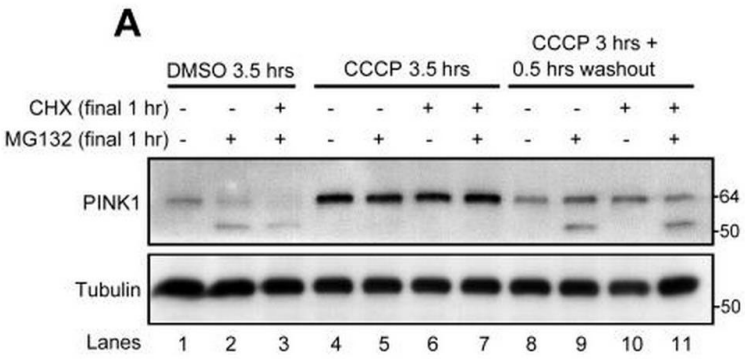


	ENTITY TYPES	EXTERNAL RESOURCES
	Simple molecules	CHEBI
	Genes	NCBI Gene
	Proteins	UniprotKB/Swiss-Prot
	Subcellular structures	Gene Ontology – cellular components
	Cell lines & types	Cellosaurus & Uberon
	Tissues	Uberon
	Organisms	NCBI Taxonomy
	Undefined	None

Normalized entities

Figures and panels

Figure 2-A



A) HeLa cells stably expressing YFP-Parkin were treated with DMSO for 3.5 h, 2 μM CCCP for 3.5 h, or CCCP for 3 h followed by washout of CCCP for 0.5 h in the absence of serum. 50 μM MG132 and/or 100 μM cyclohexamide were added for the last 1 h of treatment. Whole-cell lysates (WCL) run on SDS gels and immunoblotted for endogenous PINK1 and tubulin.

Click on a tag and hold the ALT key to confirm / delete all tags or suggestions of the same text and type.

perform on all tag suggestions of this panel: confirm delete

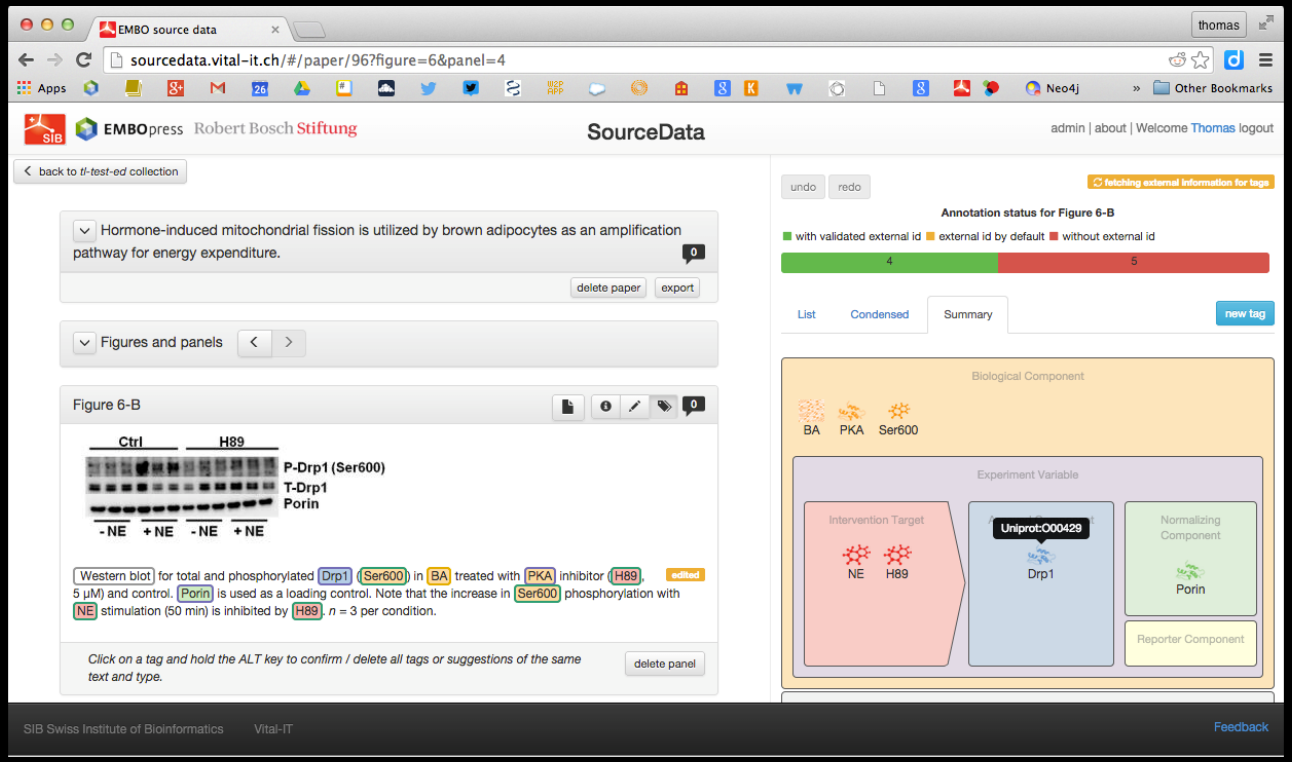
delete panel

Annotation status for Figure 2-A

with validated external id | external id by default | without external id

List | Condensed | Summary | new tag

Tag	External ID	Taxon	(x)
CCCP	ChEBI:3259		(x3)
MG132	ChEBI:75142		(x2)
cyclohexamide	ChEBI:27641		
PINK1	Uniprot:Q9BXM7	Homo sapiens	
tubulin		Homo sapiens	
DMSO			
Parkin	Uniprot:O60260	Homo sapiens	
YFP			
HeLa cells	CL/CVCL:CVCL_0030	Homo sapiens	



We **DO NOT** (yet) annotate:

Schemes, drawings, pathways, sequences

processes, attributes, activities, properties such as cell cycle progression, apoptosis rate, metabolic activity, temperature, fever, phosphorylation levels
...

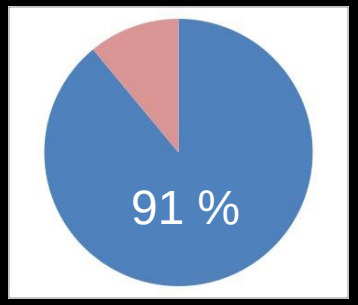
time, space, physical forces
the value of parameters
concentrations, amounts
specific genetic variants
disease state

Pervasive structures

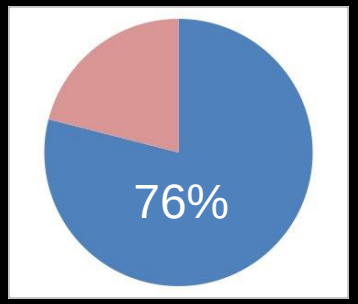
- Curation performed at the level of panels
- Focus on ‘objects’ (entities)
- When possible, represent experimentally tested (causal) hypotheses

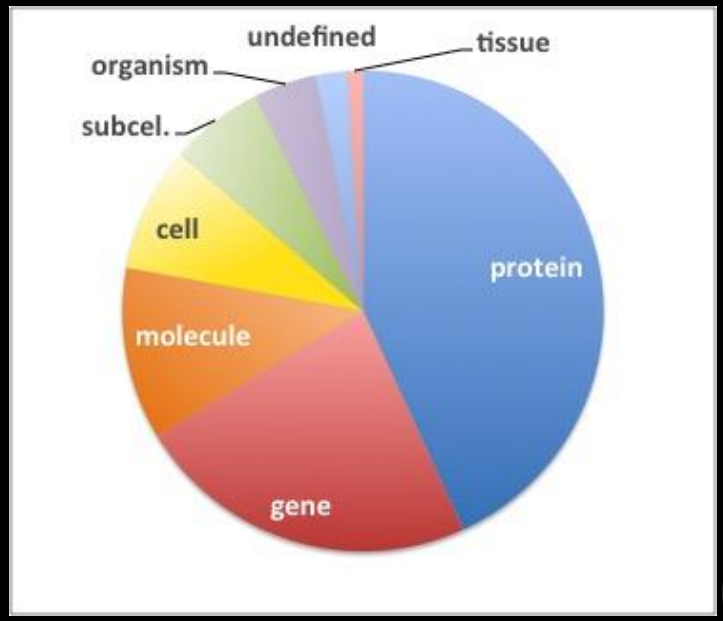
Preliminary analysis of 561 figures (23855 tags, cell & mol biol ‘autophagy’-related papers):

Fraction of figure with multiple panels



Fraction of panels with a ‘causal hypothesis’

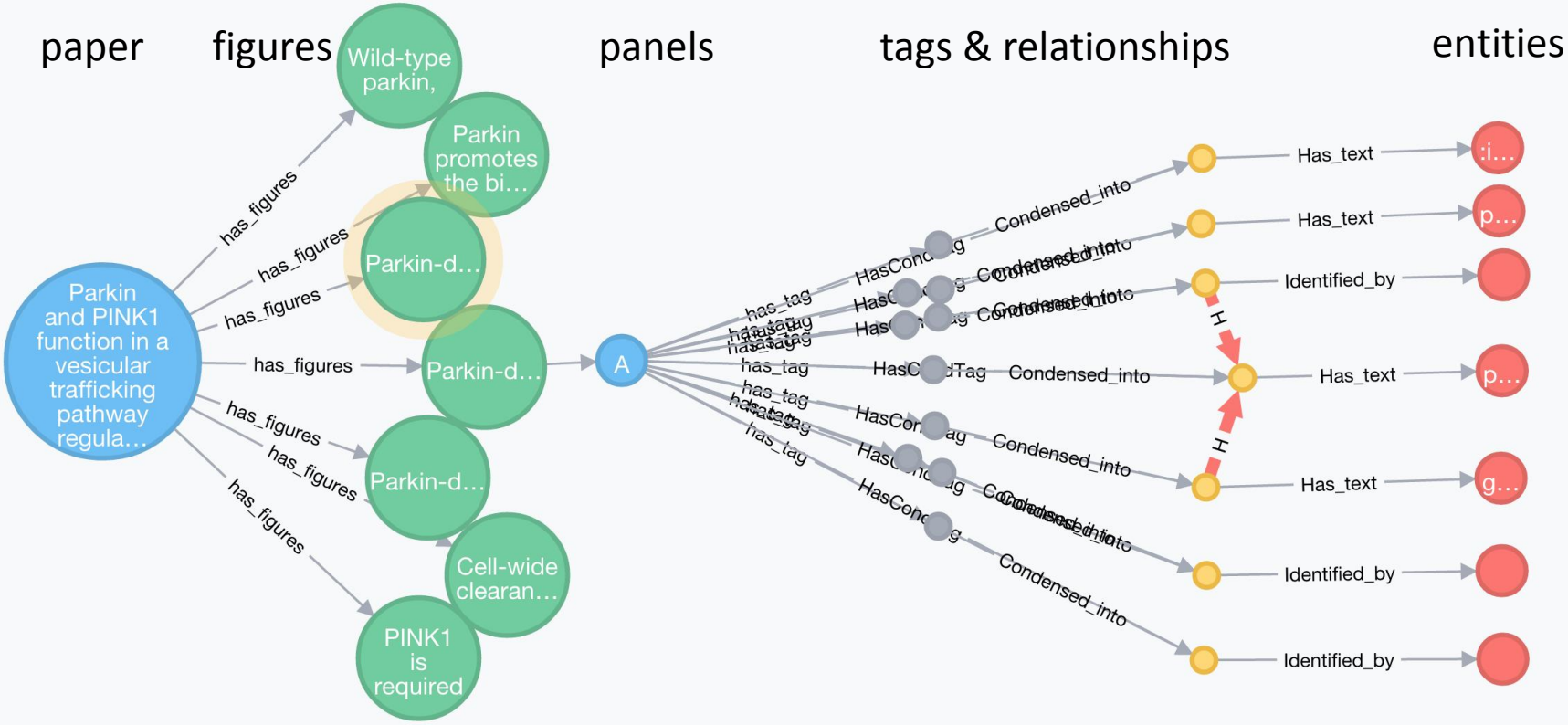




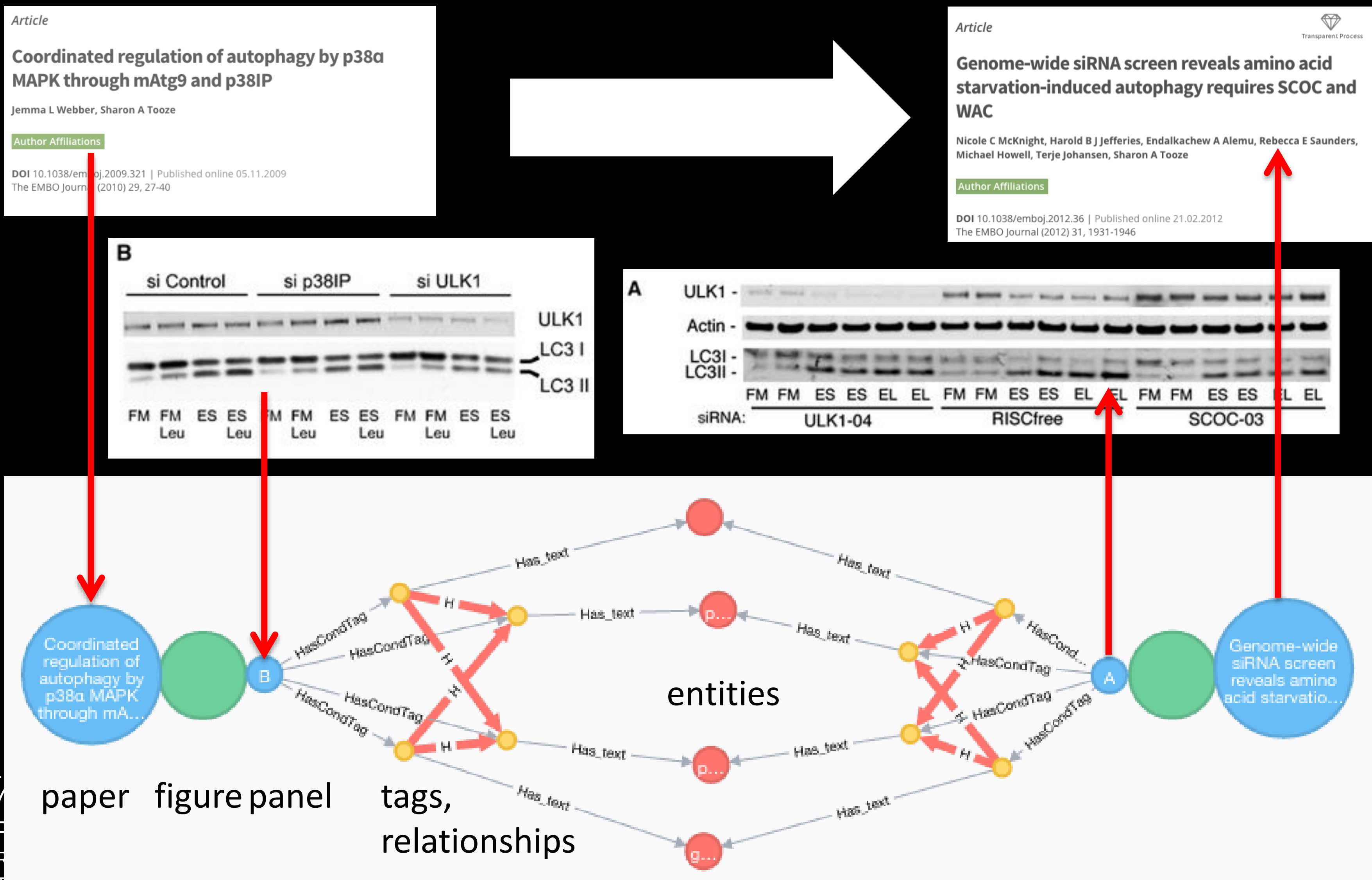
Distribution of entity types

EMBO
excellence in life sciences

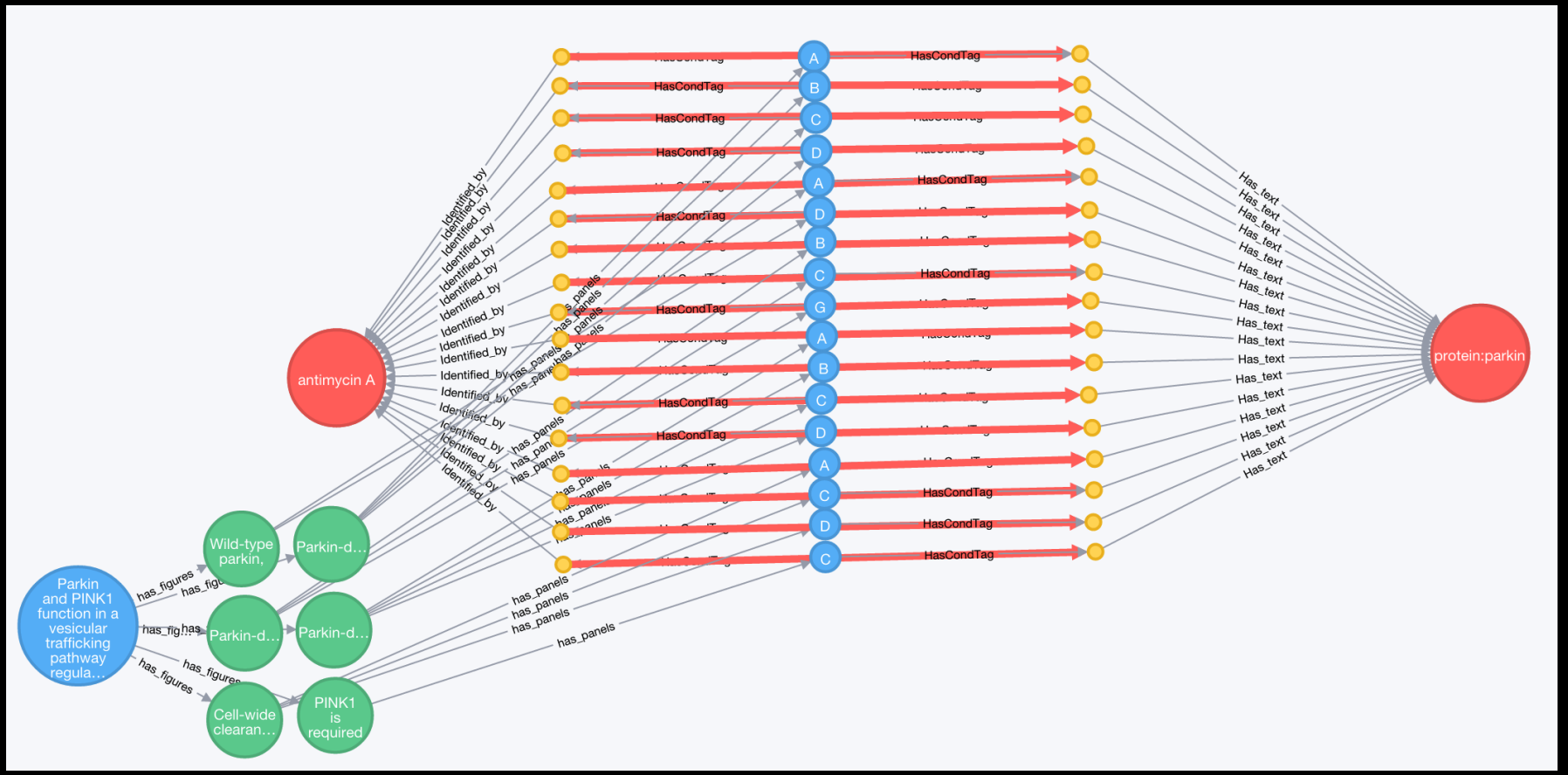
Connect papers to defined entities



Connect figures via defined entities

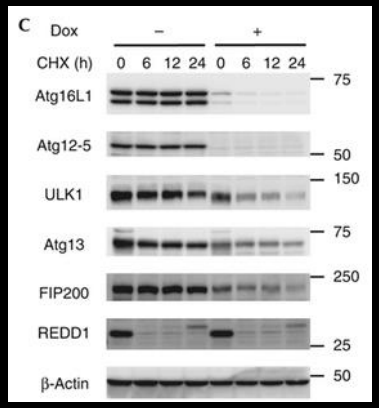


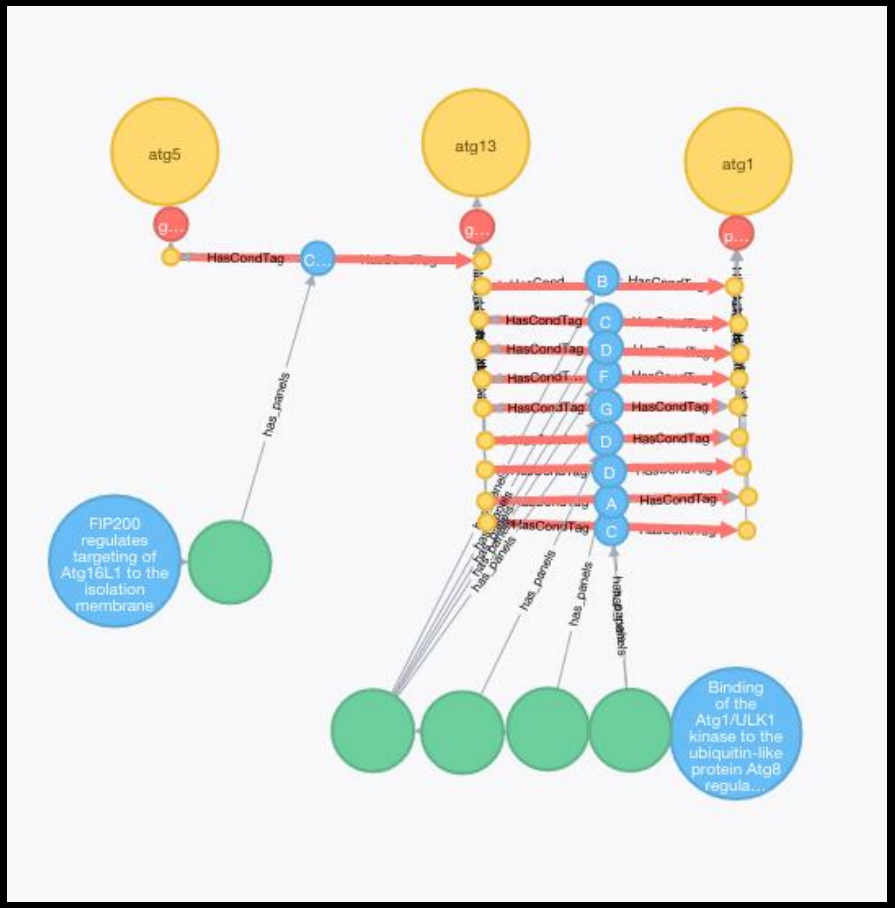
Intervention-Assay



⇒ Summarization with ‘hypothesis profiles’

Data-oriented search






Article

Binding of the Atg1/ULK1 kinase to the ubiquitin-like protein Atg8 regulates autophagy

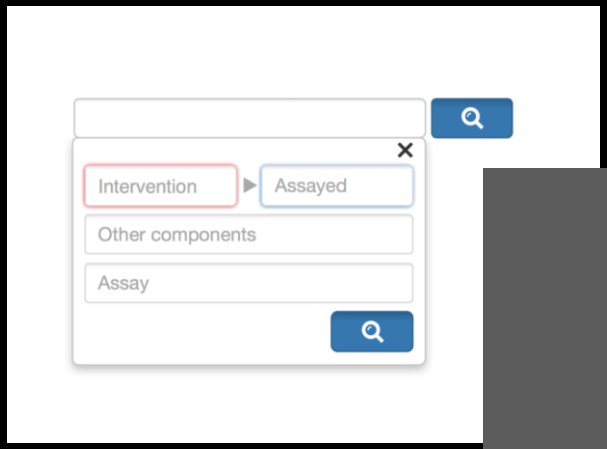
Claudine Kraft, Monika Kijanska, Eyal Kalie, Edyta Siergiejuk, Sung Sik Lee, Giuseppe Semplicio, Ingrid Stoffel, Andrea Brezovich, Mayanka Verma, Isabella Hansmann, Gustav Ammerer, Kay Hofmann, Sharon Tooze, Matthias Peter

Author Affiliations

DOI 10.1038/emboj.2012.225 | Published online 10.08.2012
The EMBO Journal (2012) 31, 3691-3703

⇒ Use of directed relationships

Future App: ‘smart figures’



From: **Figure 2. The Atg12—Atg5–Atg16L1 complex regulates the stability of the ULK1–Atg13–FIP200 complex.**

Interventions: DOX, CHX, Atg5

Assayed: Atg13, Atg16L1, Atg12-5, ULK1, FIP200, REDD1

Fig. 2C

m5-7 cells were cultured in regular medium with or without 10 ng/ml Dox for 4 days, and then cultured in the presence of 50 μg/ml CHX for the indicated time periods (**C**). mRNA levels of the indicated genes were measured by quantitative PCR following the 4-dayDox treatment. Data represent mean±s.e. (*P0.05)

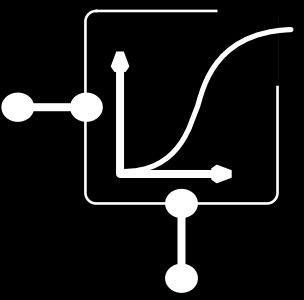
Downstream of Atg13 [Uniprot:UX098]

- **Atg1** — **Binding of the Atg1/ULK1 kinase to the ubiquitin-like protein Atg8 regulates autophagy.** Kraft et al, *EMBO J.* 2012
- **Dhhf** — **Major histocompatibility complex class II-restricted presentation of a cytosolic antigen by autophagy.** John et al, *Science* 2012
- **Mitochonria** — **The role of Xrf in mitochondrial respiration.** Jones et al, *Nature J.* 2024

See all...

Related data — This paper only / Global search — See all (12)...

Atg5 CHX ULK1 | ULK1 Atg13 Xzz | Crt Atg5 UPP | ULK1 Atg13 Atg16 | BafA Atg5 | LC3 Atg13 | Atg5 Atg13 Xrt | Atg5 Atg13 Nix

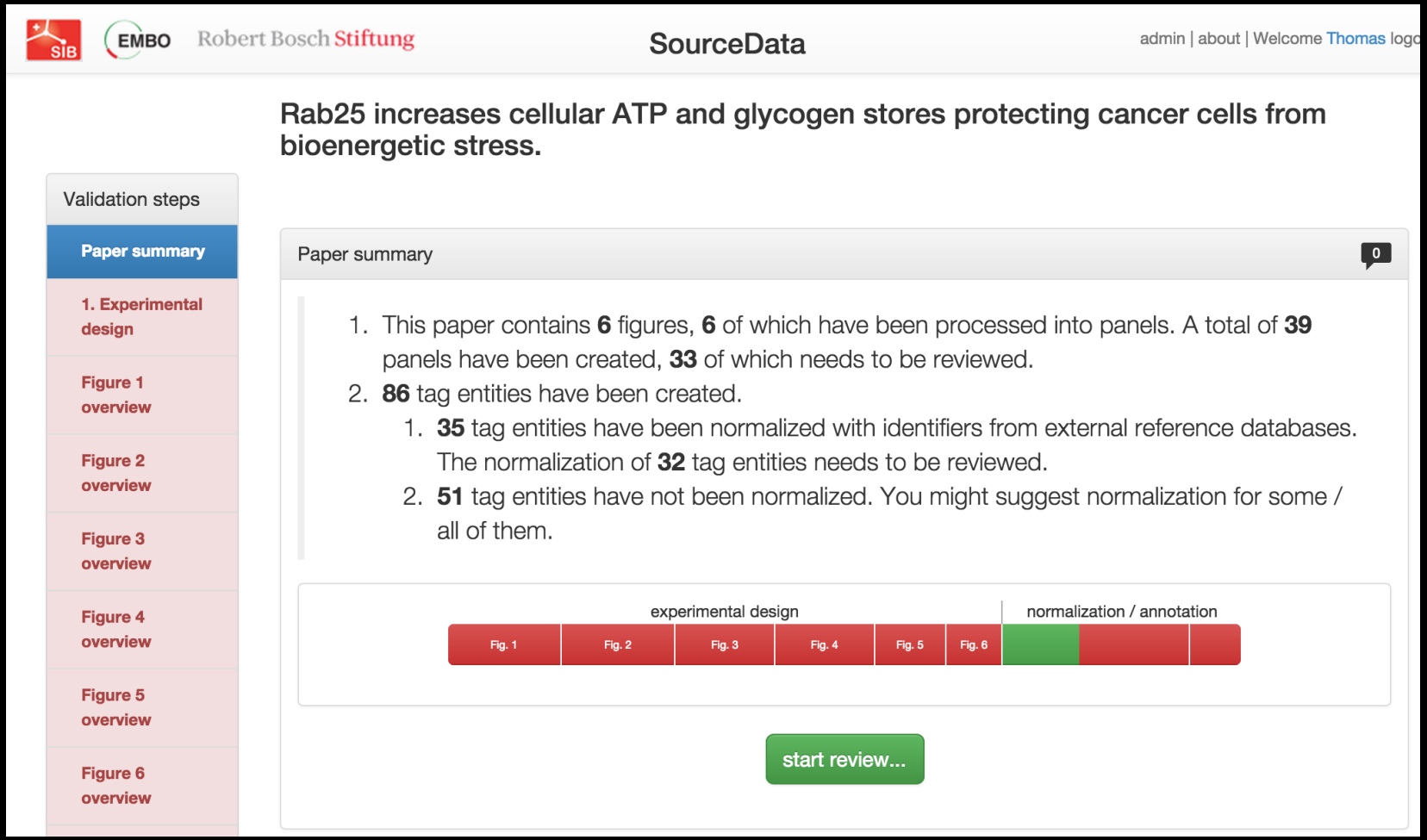


Towards integrating biocuration into the publishing workflow

Pre-acceptance stage → Data editor → **Authors' approval** → Publish

EMBOpress

Author-side validation interface



Author-side validation interface

SIB EMBO Robert Bosch Stiftung SourceData admin | about | Welcome Thomas logo

Rab25 increases cellular ATP and glycogen stores protecting cancer cells from bioenergetic stress.

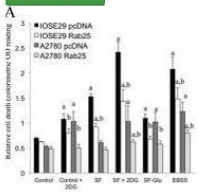
Validation steps

- Paper summary
- 1. Experimental design
- Figure 1 overview
- Figure 1-A
- Figure 1-B
- Figure 1-C
- Figure 1-D
- Figure 1-E
- Figure 1-F
- Figure 1-G
- Figure 1-H
- Figure 2 overview
- Figure 3 overview
- Figure 4 overview
- Figure 5 overview
- Figure 6 overview
- 2. Normalization
- normalization

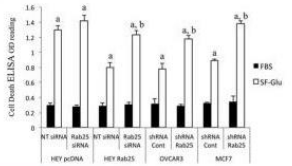
Figure 1

This figure has been divided into **8** panels, **6** of which need to be reviewed.

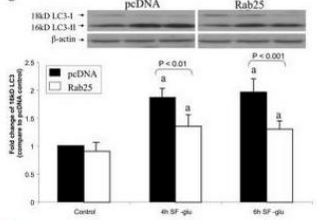
Validated



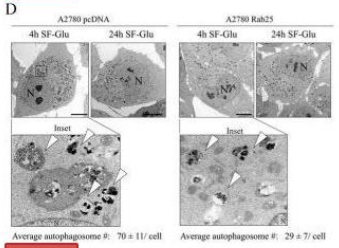
Validated



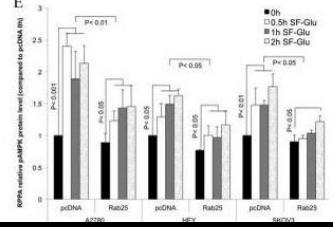
Check



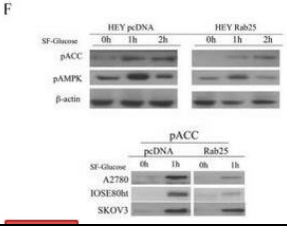
Check



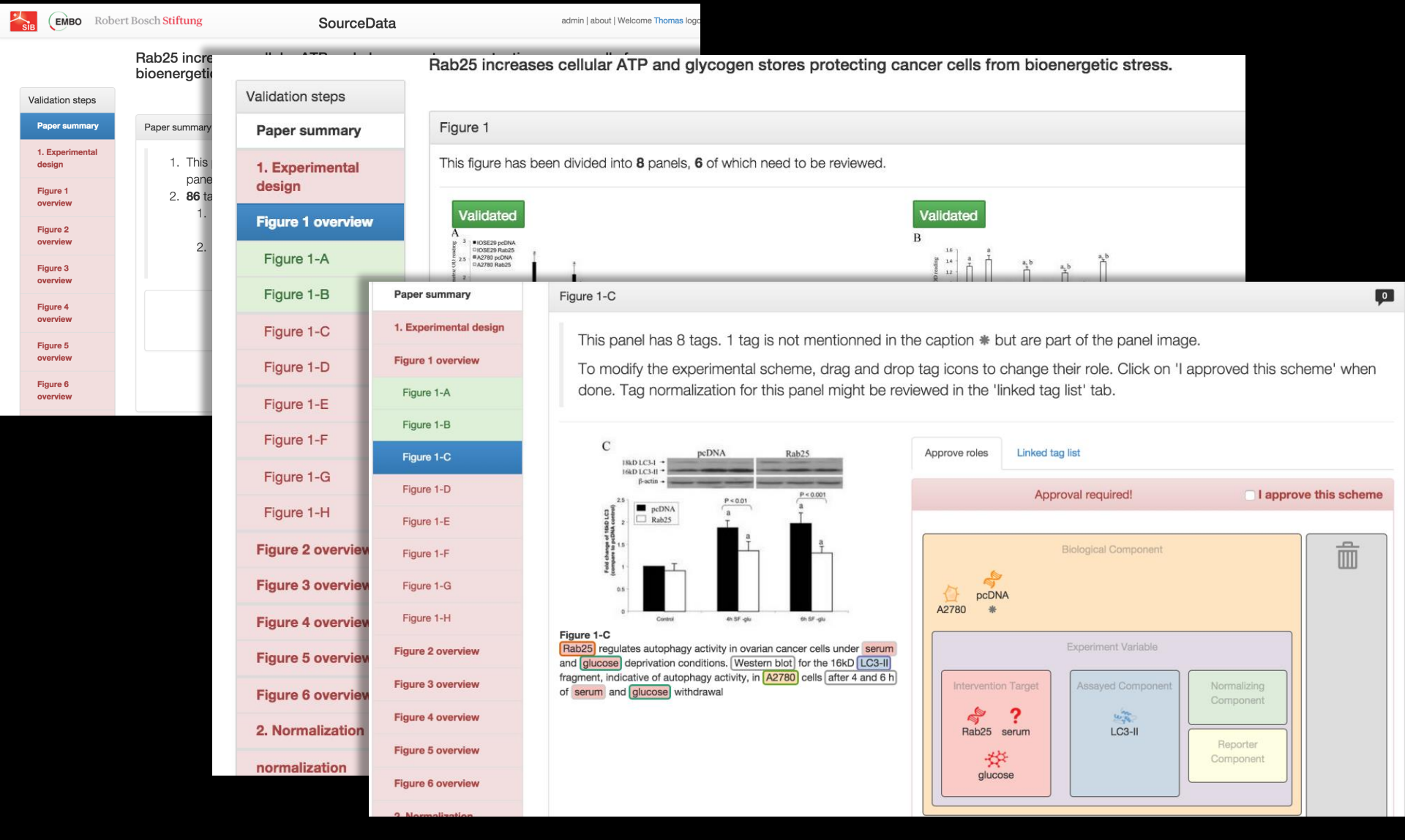
Check



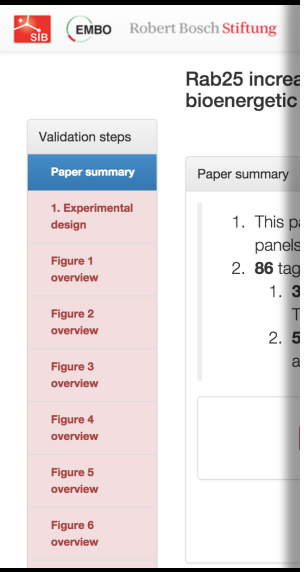
Check



Author-side validation interface



Author-side validation interface

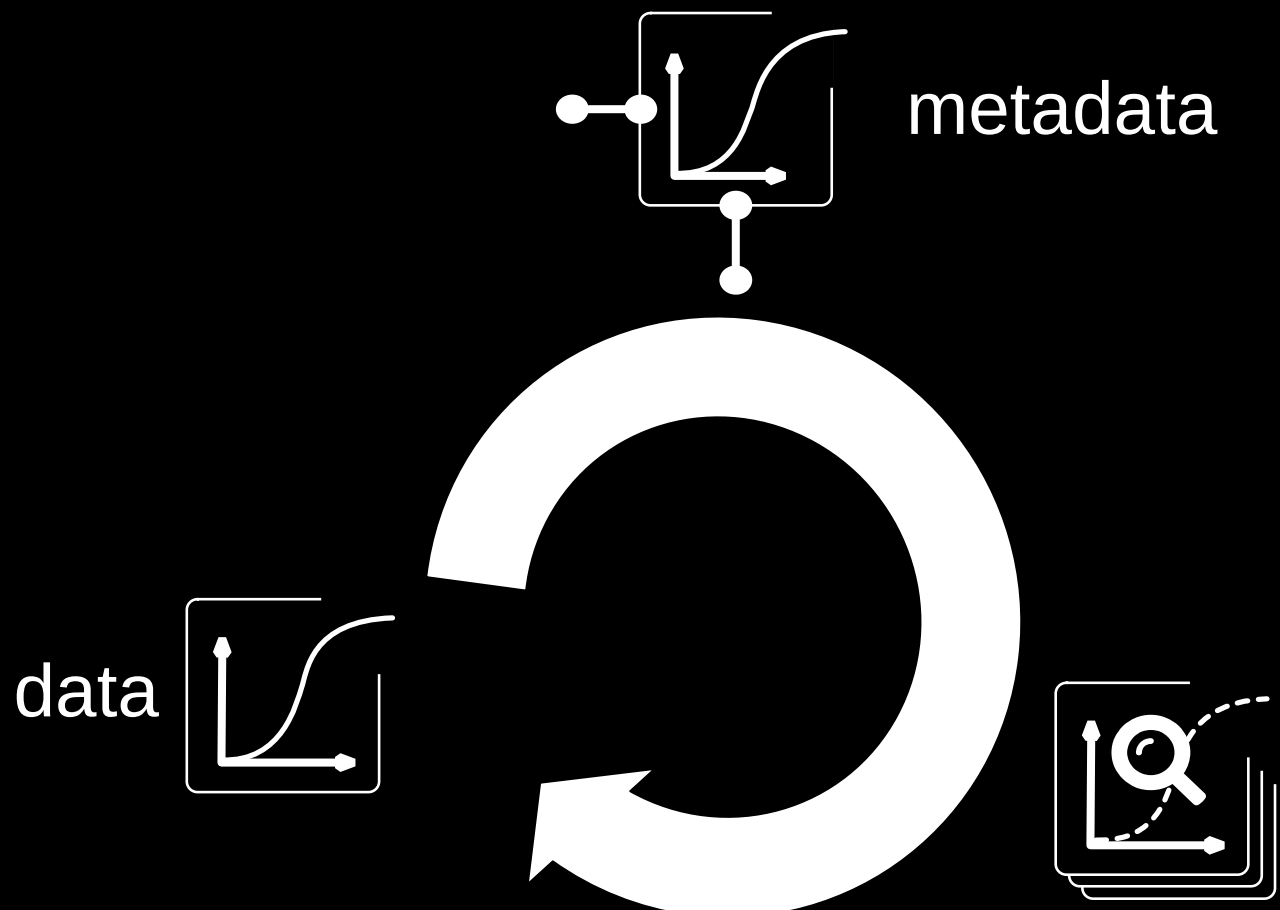


normalization

Use this form to validate or reject the association of one or several identifiers from reference resources (databases) with each tag. Click on an identifier to get a summary of the database entry.

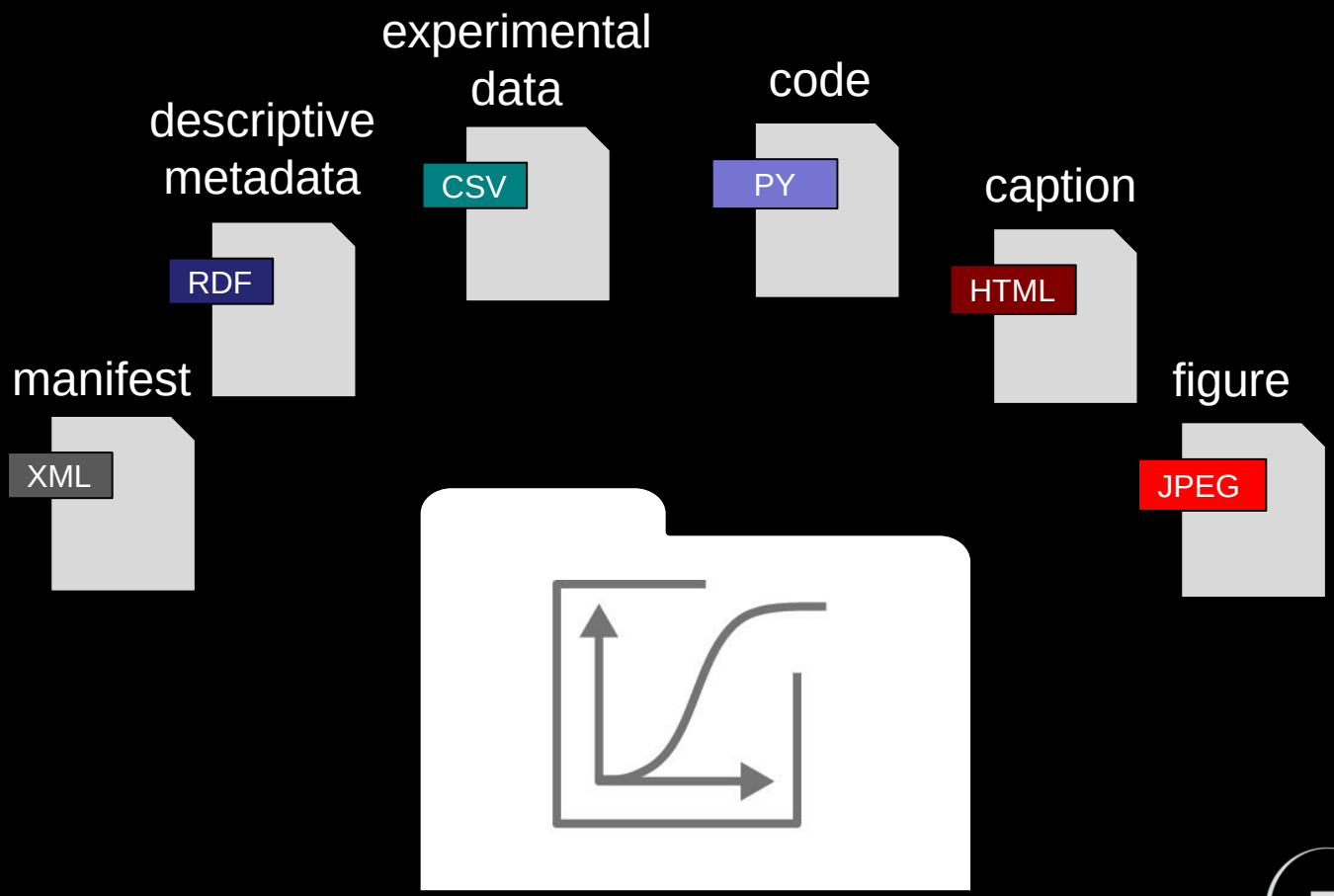
Tag	Species	Database link	Actions	Panels	
2DG, DG		ChEBI:15866	approved	8 panels	0
amino acid		ChEBI:33709	approved	Figure 1-A	2
ATP		ChEBI:15422	approved	8 panels	0
glucose		ChEBI:17234	approved	19 panels	0
glycogen		ChEBI:28087	approved	11 panels	0
MK2206		ChEBI:67271	approved	7 panels	0
oligomycin		ChEBI:25675	approve reject	2 panels	0
Rab25	Homo sapiens	NCBI gene:57111	approved	2 panels	0
Rab25, rab25	Homo sapiens	NCBI gene:57111	approve reject	32 panels	0
acetyl-CoA carboxylase, ACC	Homo sapiens	Uniprot:Q13085	approve reject	4 panels	0
actin		Uniprot:Q7ZVI7	approve reject	Figure 1-F	0
AKT	Homo sapiens	Uniprot:G3V2I6	approve reject	3 panels	0
AKT	Homo sapiens	Uniprot:P31749	approve reject	10 panels	0
GLUT1	Homo sapiens	Uniprot:P11166	approve reject	2 panels	0
GS	Homo sapiens	Uniprot:P13807	approve reject	Figure 3-A	0
GSK3	Homo sapiens	Uniprot:P49840	approve reject	3 panels	0
GSK3		Uniprot:J9Y056	approve reject	Figure 3-C	0
LC3-II		Uniprot:A6NCE7	approve reject	Figure 1-C	0

SourceData



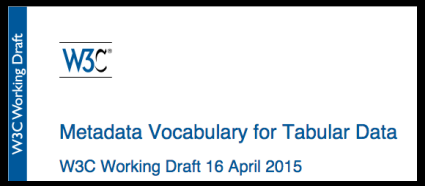
EMBO
excellence in life sciences

Figures as data packages



EMBO
excellence in life sciences

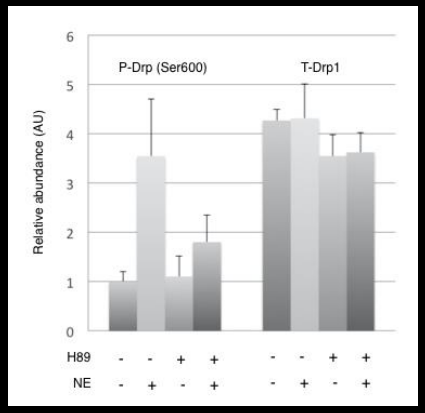
Figures as data packages



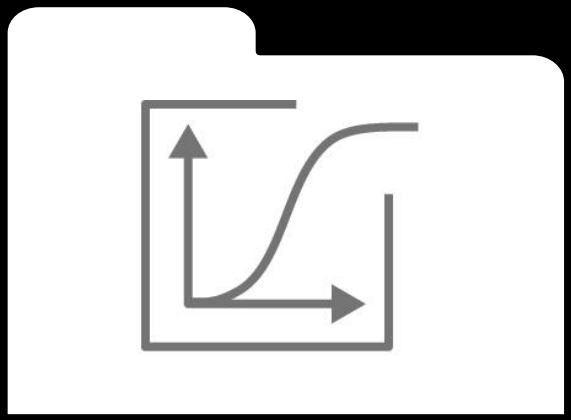
metadata

replicate	H89	NE	P-Drp1 (Ser600)	T-Drp1	Porin
1	0	0	2	11	5
2	0	0	3	12	6
3	0	0	3	11	5
1	0	1	10	13	6
2	0	1	5	10	4
3	0	1	12	9	5
1	5	0	3	7	4
2	5	0	2	12	6
3	5	0	4	11	7
1	5	1	5	11	6
2	5	1	4	12	6
3	5	1	6	8	5

data



illustration



Total and phosphorylated Drp1 (Ser600) in BA treated with PKA inhibitor (H89, 5 µM) and control. Porin is used as a loading control. Note that the increase in Ser600 phosphorylation with NE stimulation (50 min) is inhibited by H89. *n* = 3 per condition.

legend

Home

thomas

sourcedata.embo.org

SOURCEDATA

About Partners News Contact Beta test Newsletter

Figure source data

Publishing source data alongside figures helps authors to archive research data and readers to analyse published results.

Read more

Descriptive metadata

The description of published experiments using recognized standardised terminologies reduces ambiguities.

Read more

Data-oriented search

The web of connected data will open new possibilities for scientists to browse, search and mine the scientific literature in a more efficient and intuitive manner.

Read more

What is SourceData?

Our goal is to provide new ways for scientists to efficiently find research data in the published literature. We develop tools to make the data underlying published papers easier to find, to promote their reuse and to connect data across the scientific literature.

Read more

An initiative by EMBO

EMBO excellence in life sciences

EMBO excellence in life sciences