Extracting research evidence from publications

Maria Levchenko

Francesco Talo

Aravind Venkatesan
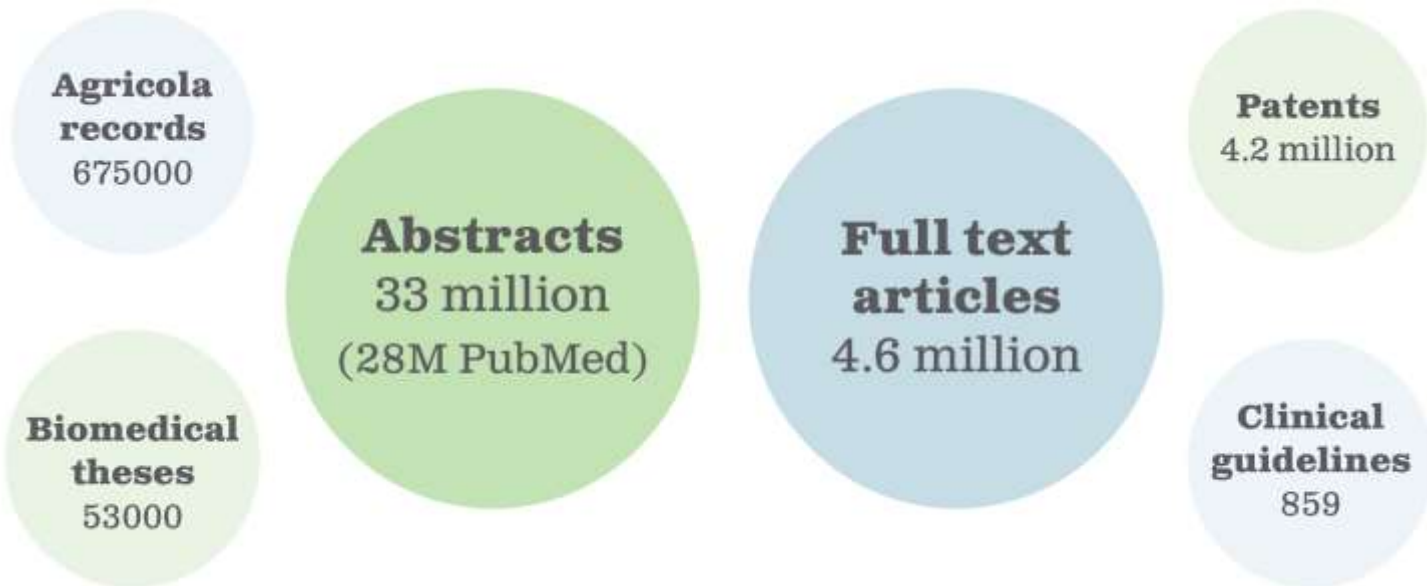
helpdesk@europepmc.org

EMBL-EBI

Contents

- Available data (annotation types and sources)

- API operations and parameters

- Web service outputs

- Use case examples

EMBL-EBI

What is Europe PMC?

- A public database of life sciences research literature

Agricola records
675000

Abstracts
33 million
(28M PubMed)

Full text articles
4.6 million

Patents
4.2 million

Biomedical theses
53000

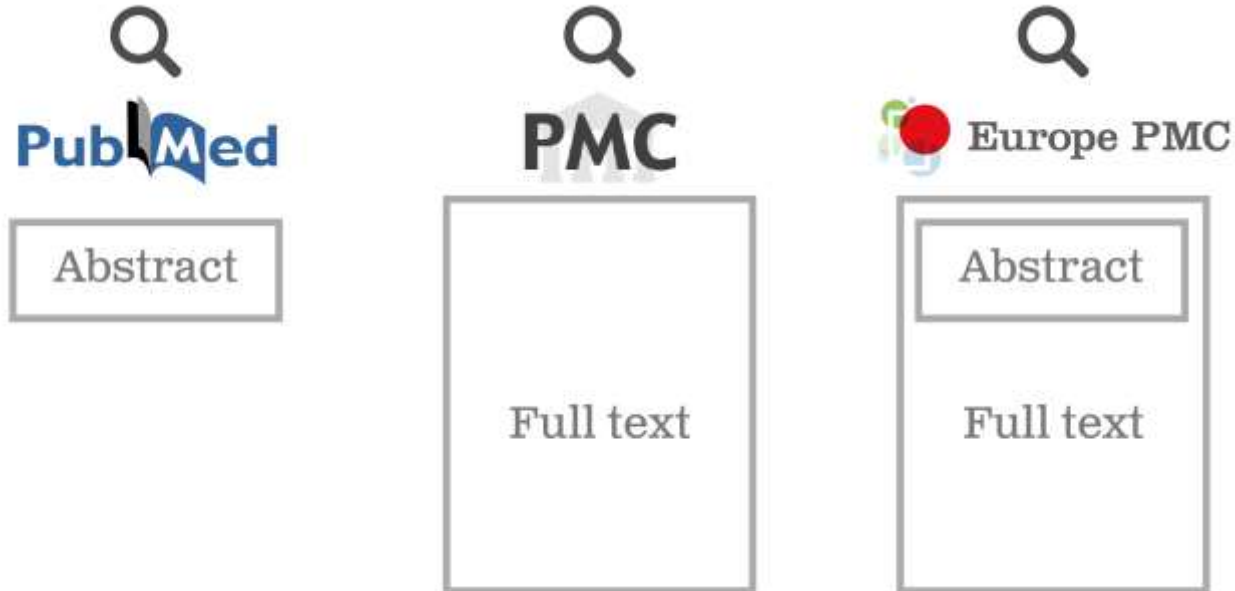Clinical guidelines
859

EMBL-EBI

What is Europe PMC?

- A designated repository of 28 life science funders

EMBL-EBI

Europe PMC and PubMed Central

- Europe PMC is a partner of PubMed Central USA
- Both databases share full-text content
- Europe PMC differs in the range of content & additional services

EMBL-EBI

Literature as part of big data in biology

33 million
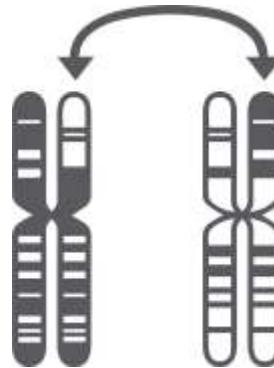biomedical abstracts

3 million
genomic variants

138 000
protein structures

image by
Jason D. Rowley
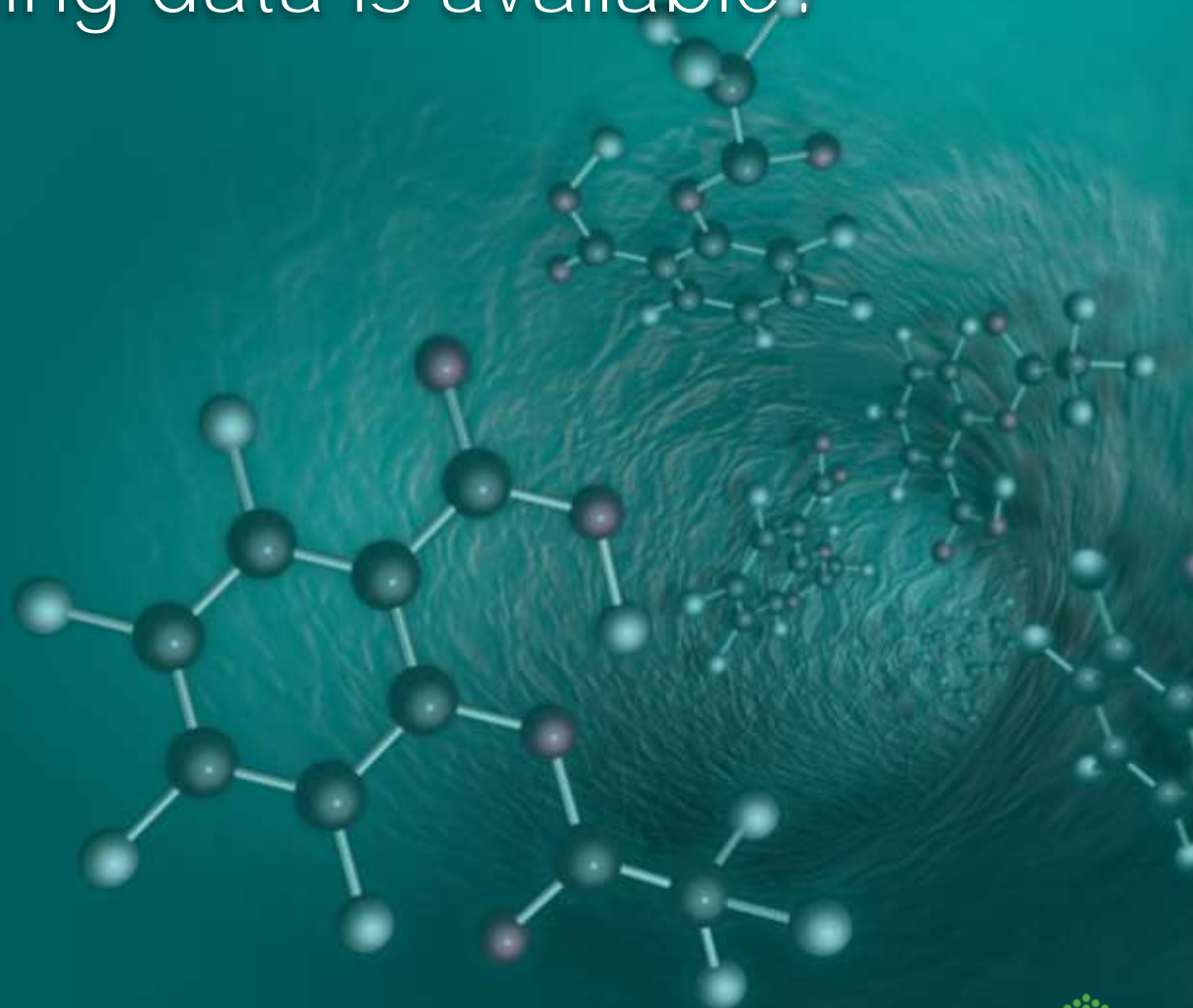
in Europe PMC

in dbVar

in PDBe

EMBL-EBI

Europe PMC community text-mining platform

EMBL-EBI
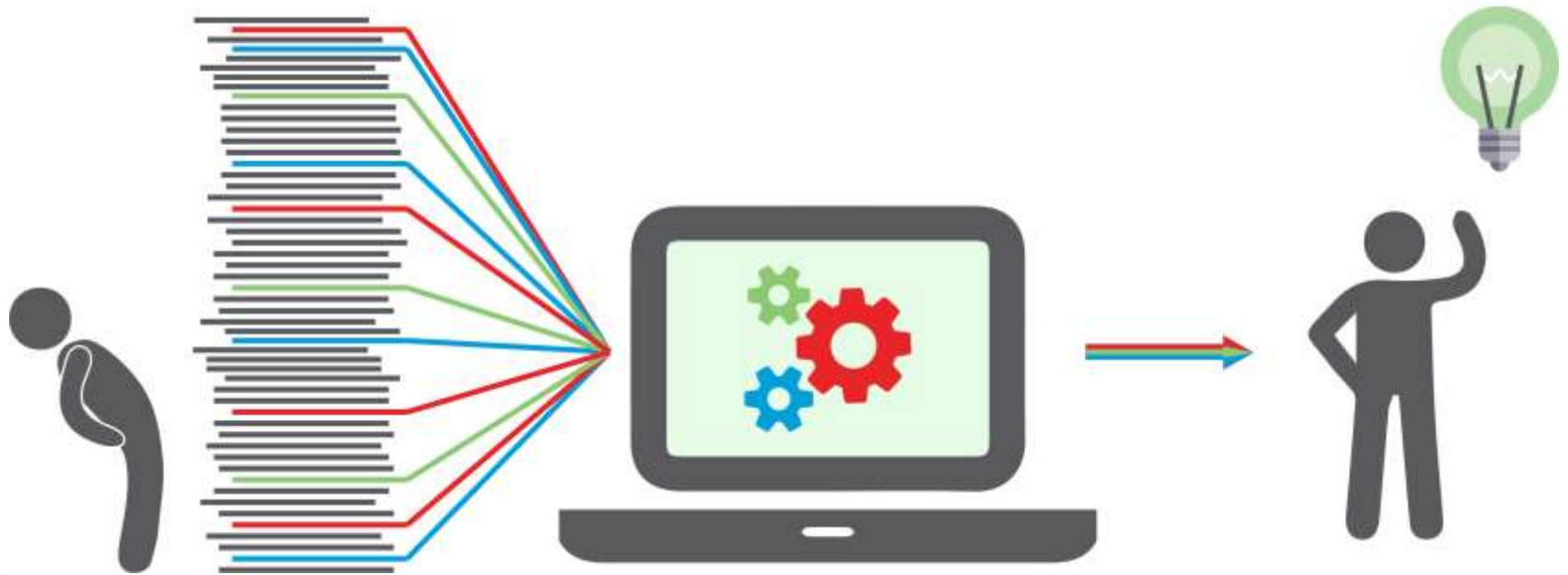
What text-mining data is available?

EMBL-EBI

What literature is covered?

Content in Europe PMC

Content for text-mining

Abstracts 33 million

Full text articles 4.6 million

Open Access 1.8 million

Abstracts 33 million

Full text articles 4.6 million

Open Access 1.8 million

EMBL-EBI
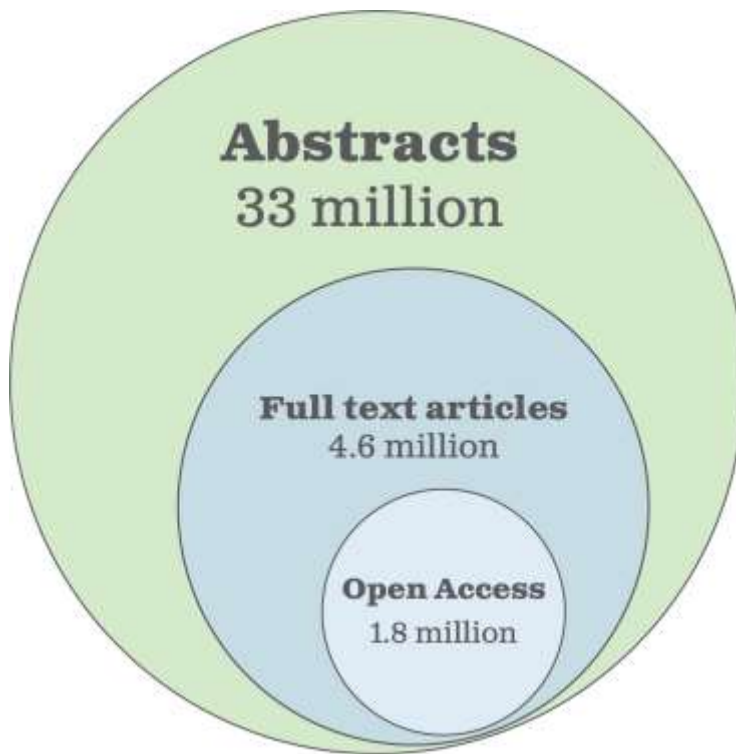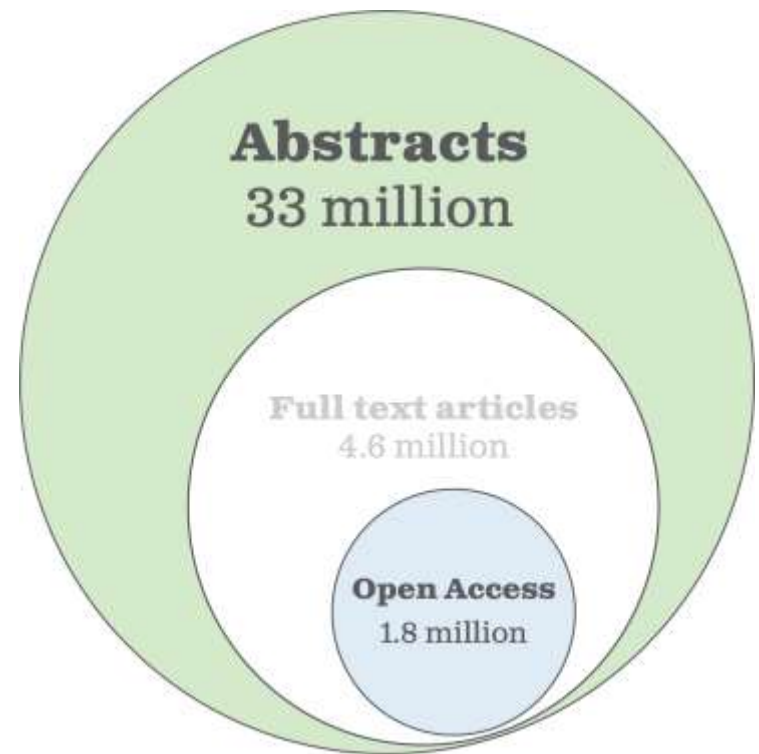
Text-mining outputs

Annotation: **chemical**

Methotrexate is an antimetabolite, a structural analogue of folic acid, which acts as an inhibitor of the enzyme dihydrofolate reductase (DHFR), leading to the depletion of

Annotation: **protein**

EMBL-EBI

Annotation providers

	Named Entities (Accessions, Genes/Proteins, Chemicals, Organisms, Diseases, Gene Ontology)	Gene Mutations	Gene-Disease relationships	Gene Functional annotations	Protein-protein interaction	Transcription factor - Target gene relationships	Biological event (Phosphosrylation events)
Europe PMC	✓						
HES-SO/SIB				✓			
DisGeNET			✓				
Open Targets Platform			✓				
IntAct					✓		
NaCTEM							✓
PubTator (NCBI)		✓					
NTNU/CNIO/BSC						✓	

EMBL-EBI

Annotation types

Genes/Proteins	Genetic mutations
Chemicals	Gene-Disease OpenTargets
Organisms	Gene-Disease DisGeNET
Diseases	Gene Function
Gene Ontology	Protein-protein Interactions
Accession Numbers	Transcription factor - Target gene
	Phosphorylation event

EMBL-EBI

Annotation types

PDBe › 2bk9

X-ray diffraction
1.2Å resolution

Drosophila Melanogaster globin

Source organism: *Drosophila melanogaster*

Primary publication:
Bishistidyl heme hexacoordination, a key structural property in Drosophila melanogaster hemoglobin.

de Sanctis D, Dewilde S, Vonrhein C, Pesce A, Moens L, Ascenzi P, Hankeln T, Burmester T, Ponassi M, Nardini M, Bolognesi M

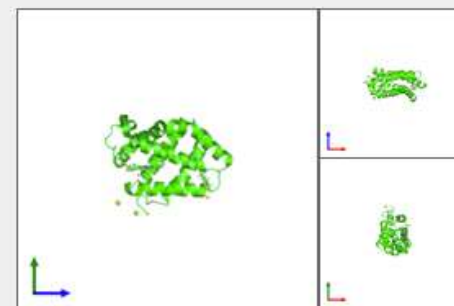*J. Biol. Chem.* **280** 27222-9 (2005)
PMID: 15917230

Released: 20 May 2005

Model geometry
Fit model/data

EMBL-EBI

Annotations API

Annotations API

API operations

annotations-api-controller : Europe PMC Annotations API

GET	**/annotationsByArticleIds**

Get the annotations contained in the list of articles specified

GET	**/annotationsByEntity**

Get the annotations of the articles which have at least one annotation tagging the specified entity

GET	**/annotationsByProvider**

Get the annotations of the articles which have at least one annotation provided by the specified provider

GET	**/annotationsByRelationship**

Get the annotations of the articles which have at least one annotation tagging both the specified entities (i.e. Gene-Disease relationship)

GET	**/annotationsBySectionAndOrType**

Get the annotations of the articles which have at least one annotation of a type (if specified) inside an article section (if specified). At least one value between section and type must be specified.

EMBL-EBI

Navigating results: pageSize and cursorMark

Parameter	Value
pageSize	4

Search results:

Page 1

Article 1:	Annotation 1
	Annotation 2
	Annotation 3

| Article 2: | Annotation 1 |

| Article 3: | Annotation 1 |
| | Annotation 2 |

| Article 4: | Annotation 1 |

Page 2

Article 1:	Annotation 1
	Annotation 2
	Annotation 3

| Article 2: | Annotation 1 |

| Article 3: | Annotation 1 |
| | Annotation 2 |

| Article 4: | Annotation 1 |

EMBL-EBI

Navigating results: pageSize and cursorMark

Parameter	Value
cursorMark	

Response Body

```
{
    "nextCursorMark": 43.72716719,
    "cursorMark": 0,
```

EMBL-EBI

Filter

Parameter	Value
filter	1 (default)

Filter = 1

Article 1:	p53 (1)
	p53 (2)
	p53 (3)

| Article 2: | p53 |

| Article 3: | p53 (1) |
| | p53 (2) |

| Article 4: | p53 |

EMBL-EBI

Filter

Parameter	Value
filter	1 (default) ⇕

Filter = 1   Filter = 0

Article 1:	p53 (1)
	p53 (2)
	p53 (3)

| Article 2: | p53 |

| Article 3: | p53 (1) |
| | p53 (2) |

| Article 4: | p53 |

Article 1:	p53 (1)	cancer
	p53 (2)	human
	p53 (3)	MTX

| Article 2: | p53 | mouse |

| Article 3: | p53 (1) | tumour |
| | p53 (2) | GPCR |

| Article 4: | p53 | 2gx9K |

EMBL-EBI

Output

Parameter	Value
format	JSON (default) ⇕

- JSON will produce a JSON representation of the articles and relative annotations
- XML will produce a XML representation of the articles and relative annotations
- JSON-LD will produce a JSON linked Data representation of the annotations. To see details about JSON-LD go to http://europepmc.org/AnnotationsApi#jsonLD
- ID_LIST will produce a list of articles identifiers including pmcid if available

EMBL-EBI

API operations

annotations-api-controller : Europe PMC Annotations API

GET	**/annotationsByArticleIds**

Get the annotations contained in the list of articles specified

GET	**/annotationsByEntity**

Get the annotations of the articles which have at least one annotation tagging the specified entity

GET	**/annotationsByProvider**

Get the annotations of the articles which have at least one annotation provided by the specified provider

GET	**/annotationsByRelationship**

Get the annotations of the articles which have at least one annotation tagging both the specified entities (i.e. Gene-Disease relationship)

GET	**/annotationsBySectionAndOrType**

Get the annotations of the articles which have at least one annotation of a type (if specified) inside an article section (if specified). At least one value between section and type must be specified.

EMBL-EBI

Get annotations by entity

Get annotations by provider

Get annotations by relationship

Get annotations by section and/or type

Get annotations by article iD

GET	/annotationsByArticleIds

Get the annotations contained in the list of articles specified

Parameters

Parameter	Value
articleIds	Provide multiple values in new lines (at least one required).
type	
section	
provider	
format	JSON (default)

← MED:14670548, PMC:19992...

(max 8)

EMBL-EBI

How to find the article iD

Article iD = SOURCE:EXTERNAL_ID

☐ Mechanisms and Physiological Roles of Mitophagy in Yeast.
(PMID: 29370687 PMCID:PMC5792711)

| Abstract 🔴 | Citations ❓ | BioEntities ❓ | Related Articles ❓ | External Links ❓ |

Fukuda T[1], Kanki T[1]

Affiliations ▸

Molecules and Cells [23 Jan 2018, 41(1):35-44]

Type: review-article, Review, Journal Article
DOI: 10.14348/molcells.2018.2214 ❓

Source   External_iD

Article iD = MED:29370687

Source:
MED: PubMed/MEDLINE
PMC: PubMedCentral
PAT: Patents
AGR: Agricola (USDA/NAL)
CBA: Chinese biological abstracts
HIR: NHS Evidence
CTX: CiteXplore submission
ETH: EThOS theses
CIT: CiteSeer
PPR: Preprints
NBK: NLM Books

EMBL-EBI

Use cases

EMBL-EBI
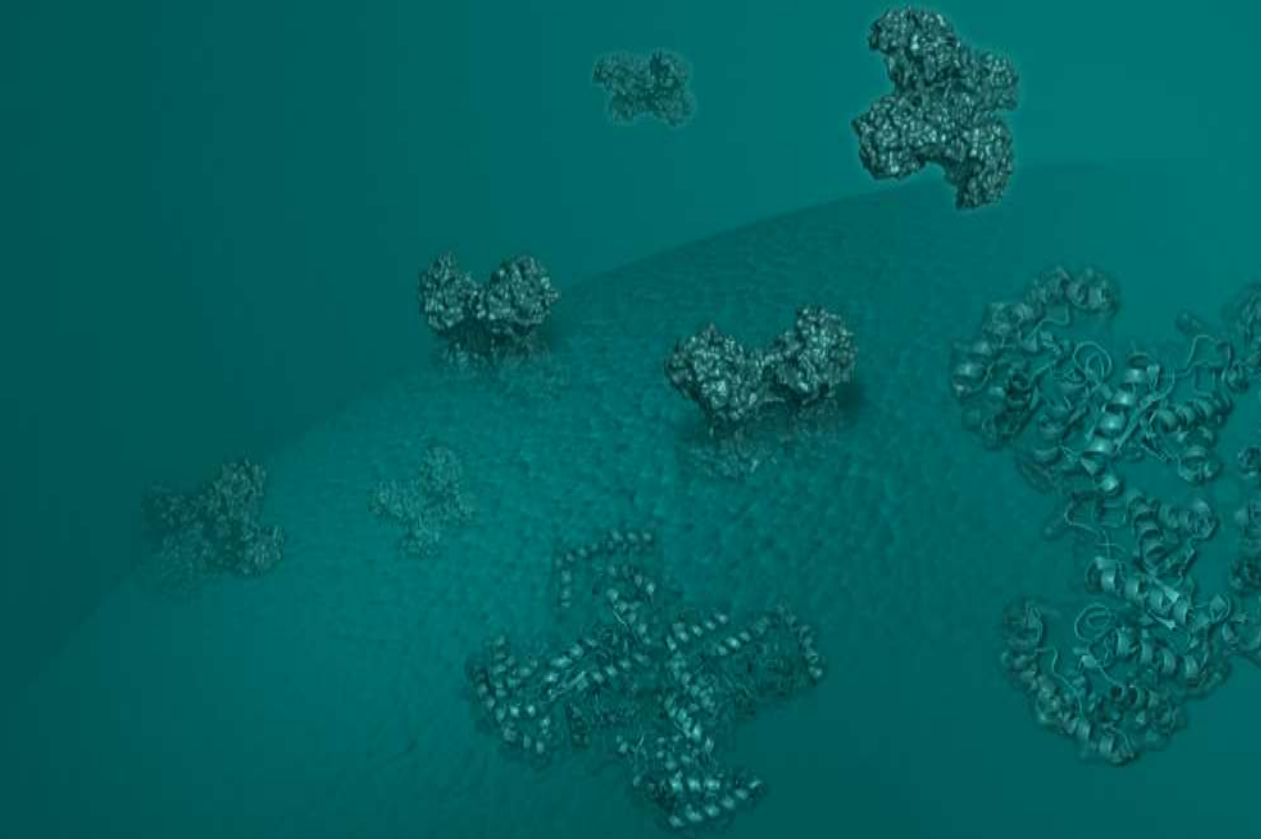
Get chemicals mentioned in methods

Get chemicals mentioned in methods

GET **/annotationsBySectionAndOrType**

Get the annotations of the articles which have at least one annotation of a type (if specified) inside an article section (if specified). At least one value between section and type must be specified.

Parameter	Value
type	Chemicals
section	Methods
filter	1 (default)
format	JSON (default)
cursorMark	
pageSize	4

EMBL-EBI

Get chemicals mentioned in methods

GET **/annotationsBySectionAndOrType**

Get the annotations of the articles which have at least one annotation of a type (if specified) inside an article section (if specified). At least one value between section and type must be specified.

Request URL

```
https://www.ebi.ac.uk/europepmc/annotations_api/annotationsBySectionAndOrType?type=Chemicals&section=Methods&fil
```

Response Body

```
{
   "nextCursorMark": 910.725210858,
   "cursorMark": 0,
   "articles": [
      {
         "source": "PMC",
         "extId": "PMC3926537",
         "pmcid": 3926537,
         "annotations": [
```

EMBL-EBI

Get chemicals mentioned in methods

nextCursorMark: 910.725210858
cursorMark: 0
articles:
    0:            {…}
    1:            {…}
    2:            {…}
    3:            {…}

EMBL-EBI

Get chemicals mentioned in methods

nextCursorMark: 910.725210858
cursorMark: 0
articles:
  0:
      source: "PMC"
      extId: "PMC3926537"
      pmcid: 3926537
    annotations: [...]

EMBL-EBI

Get chemicals mentioned in methods

```
nextCursorMark:     910.725210858
cursorMark:         0
articles:
  0:
    source:         "PMC"
    extId:          "PMC3926537"
    pmcid:          3926537
    annotations:
      0:            {…}
      1:            {…}
      2:            {…}
      3:            {…}

              …

      1254:         {…}
      1255:         {…}
```

EMBL-EBI

Get chemicals mentioned in methods

```
nextCursorMark:    910.725210858
cursorMark:        0
articles:
  0:
    source:        "PMC"
    extId:         "PMC3926537"
    pmcid:         3926537
    annotations:
      0:
        prefix:    ""
        exact:     "Calixarene"
        postfix:   " Based Ion Selective Sensors "
        tags:      [...]
        type:      "Chemicals"
        section:   "Methods (http://purl.org/orb/Methods)"
        provider:  "Europe PMC"
```

EMBL-EBI

Get chemicals mentioned in methods

```
nextCursorMark:     910.725210858
cursorMark:         0
articles:
  0:
    source:         "PMC"
    extId:          "PMC3926537"
    pmcid:          3926537
    annotations:
      0:
        prefix:     ""
        exact:      "Calixarene"
        postfix:    " Based Ion Selective Sensors "
        tags:       [...]
        type:       "Chemicals"
        section:    "Methods (http://purl.org/orb/Methods)"
        provider:   "Europe PMC"
```

EMBL-EBI

Get chemicals mentioned in methods

```
nextCursorMark:      910.725210858
cursorMark:          0
▼ articles:
  ▼ 0:
      source:        "PMC"
      extId:         "PMC3926537"
      pmcid:         3926537
    ▼ annotations:
      ▼ 0:
          prefix:    ""
          exact:     "Calixarene"
          postfix:   " Based Ion Selective Sensors "
        ▶ tags:      [...]
          type:      "Chemicals"
          section:   "Methods (http://purl.org/orb/Methods)"
          provider:  "Europe PMC"
```

4.1. Calixarene Based Ion Selective Sensors

The structure of calixarene ligands to interact with cations, anions, neutral species and their various combinations, is determined by the character of other functional groups bonded to the basic skeleton of

EMBL-EBI

Get chemicals mentioned in methods

```
nextCursorMark:     910.725210858
cursorMark:         0
articles:
  0:
    source:         "PMC"
    extId:          "PMC3926537"
    pmcid:          3926537
    annotations:
      0:
        prefix:     ""
        exact:      "Calixarene"
        postfix:    " Based Ion Selective Sensors "
        tags:
          0:
            name:   "Calixarene"
            uri:    "http://purl.obolibrary.org/obo/CHEBI_51198"
        type:       "Chemicals"
        section:    "Methods (http://purl.org/orb/Methods)"
        provider:   "Europe PMC"
```

EMBL-EBI

Get chemicals mentioned in methods

Get gene mutations in papers on diabetes

EMBL-EBI

Get gene mutations in papers on diabetes

Parameter	Value
entity	diabetes
format	ID_LIST

EMBL-EBI

Get gene mutations in papers on diabetes

Get gene mutations in papers on diabetes

```python
#!/usr/bin/env python
'''
Created on 05 March 2018

@summary: This script retrives all gene mutations annotations along with "diabetes" annotations
@param: Entity name, Annotation type and Base URL

'''

import re
import pprint

''' For REST request '''
import urllib2
import json

pp = pprint.PrettyPrinter(indent=4)

def main():
    baseQueryURL = "https://www.ebi.ac.uk/europepmc/annotations_api/" # Base URL
    entity = "diabetes" # Entity name
    ann_type = "&type=Gene%20Mutations" # Annotation type

    # Get PMIDs for a given entity name
    ids = getArticleIDbyEntity(baseQueryURL, entity)

    # Get Annotations for a given type
    annotations = getAnnotationsbyType(baseQueryURL, ids, ann_type)

    # Print annotations
    pp.pprint(annotations)
```

EMBL-EBI

Get gene mutations in papers on diabetes

```python
#!/usr/bin/env python
'''
Created on 05 March 2018

@summary: This script retrives all gene mutations annotations along with "diabetes" annotations
@param: Entity name, Annotation type and Base URL

'''

import re
import pprint

''' For REST request '''
import urllib2
import json

pp = pprint.PrettyPrinter(indent=4)

def main():
    baseQueryURL = "https://www.ebi.ac.uk/europepmc/annotations_api/" # Base URL
    entity = "diabetes" # Entity name
    ann_type = "&type=Gene%20Mutations" # Annotation type

    # Get PMIDs for a given entity name
    ids = getArticleIDbyEntity(baseQueryURL, entity)

    # Get Annotations for a given type
    annotations = getAnnotationsbyType(baseQueryURL, ids, ann_type)

    # Print annotations
    pp.pprint(annotations)
```

EMBL-EBI

Get gene mutations in papers on diabetes

```python
'''Get list of PMIDs that contain "diabetes"  mentions'''
def getArticleIDbyEntity(basequeryURL, entityName):

    get_entity = "annotationsByEntity?entity="
    format_type = "&format=ID_LIST"

    # API query
    query = basequeryURL + get_entity + entityName + format_type

    # List to store PMIDs
    id_list = list()

    # Call the API
    api_response = urllib2.urlopen(query, timeout=10)

    # Process the response
    response_data = api_response.read()
    json_data = json.loads(response_data)

    for key in json_data:
        if key == 'articles':
            for article in json_data[key]:
                source = article['source']
                pmid = article['extId']
                formated_id = source + ":" + pmid
                id_list.append(formated_id)

    # Returns PMID list
    return id_list
```

EMBL-EBI

Get gene mutations in papers on diabetes

```python
'''Get gene mutation annotation for list of PMIDs'''
def getAnnotationsbyType(basequeryURL, pmid_list, annotation_type):

    get_annotions = "annotationsByArticleIds?articleIds="
    format_type = "&format=JSON"
    pmids = ",".join(pmid_list)

    # API query
    query = basequeryURL + get_annotions + pmids + annotation_type + format_type


    # Call the API
    get_response = urllib2.urlopen(query, timeout=10)
    response_data = get_response.read()

    # Returns annotations
    return response_data

if __name__ == '__main__':
    main()
```

EMBL-EBI

Get gene mutations in papers on diabetes

```
[
  {
    "source": "MED",
    "extId": "28729637",
    "pmcid": 5519666,
    "annotations": [
      {
        "prefix": "association between ",
        "exact": "rs499974",
        "postfix": " (MOGAT2) and a diac",
        "tags": [
          {
            "name": "rs499974",
            "uri": "http://identifiers.org/dbsnp:rs499974"
          }
        ],
        "type": "Gene Mutations",
        "section": "Abstract (http://purl.org/dc/terms/abstract)",
        "provider": "PubTator (NCBI)"
      },
```

EMBL-EBI

… Other use cases?

Help and contacts
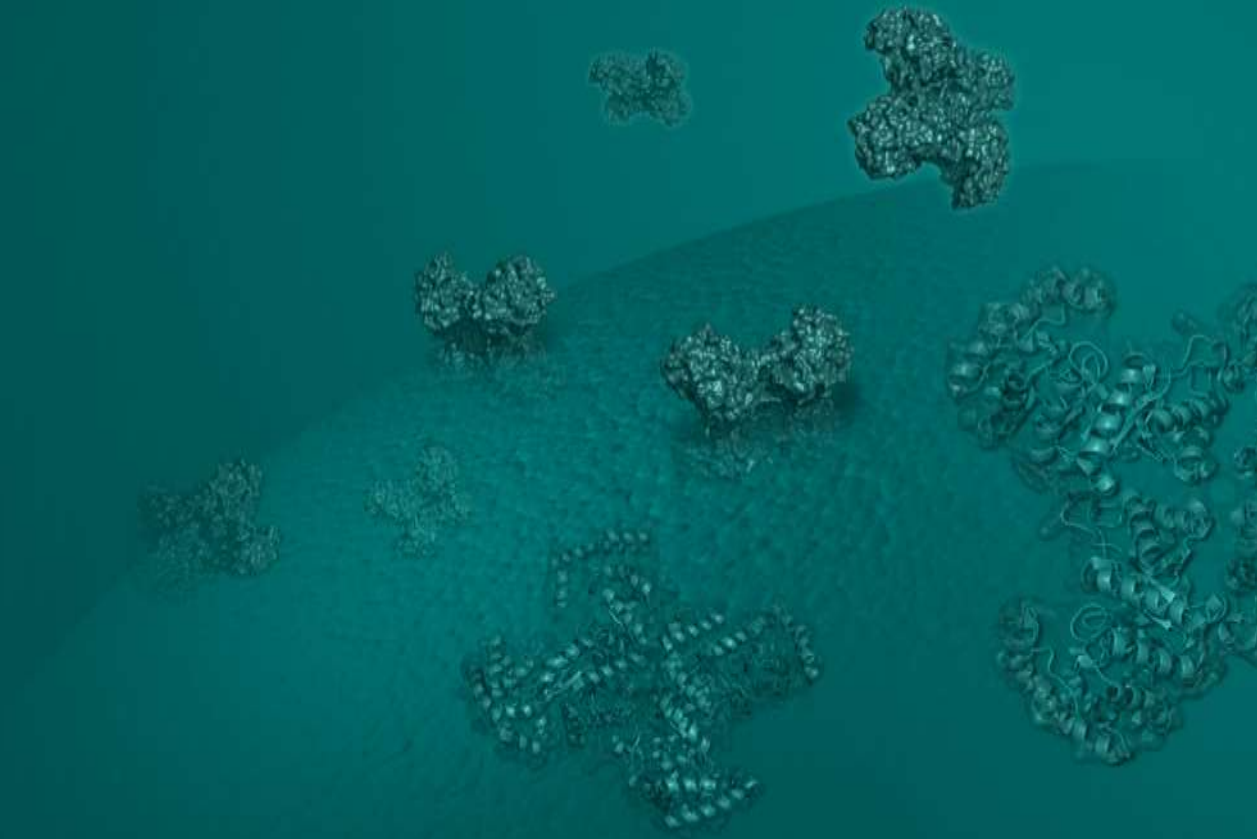
EMBL-EBI

Contact us

Europe PMC Webservice Users Group

https://groups.google.com/a/ebi.ac.uk/forum/#!forum/epmc-webservices

Questions about Europe PMC
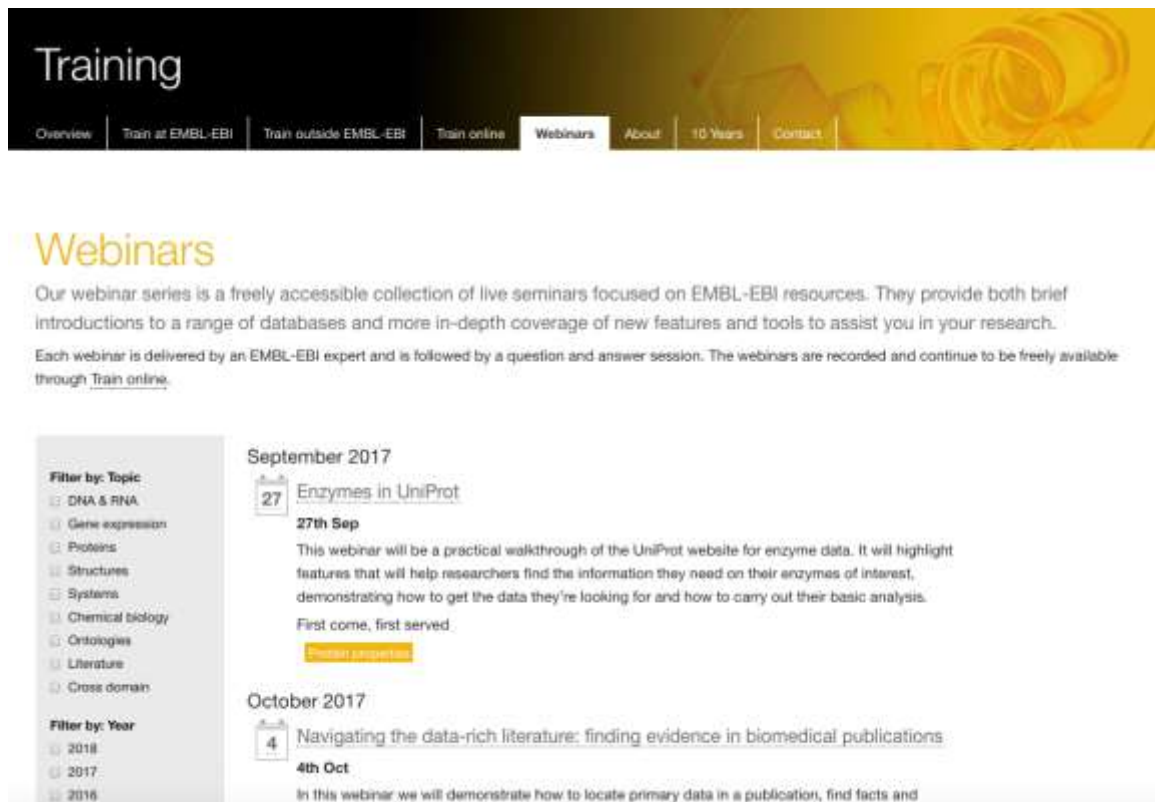
helpdesk@europepmc.org

@EuropePMC_News

EMBL-EBI

Upcoming webinars

See the full list of upcoming webinars at
http://www.ebi.ac.uk/training/webinars

Feedback

Tell us what you think

EMBL-EBI