Biomedical models Hub

Final results

DG JRC

29th National Contact Point meeting

12/11/2025

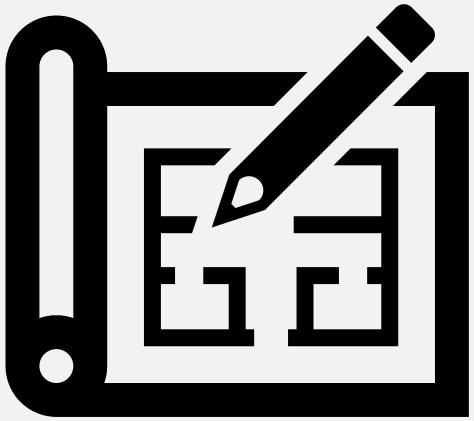
European Commission

BimmoH Project Results

Leveraging Artificial Intelligence technologies to publish the largest dataset of human biology-based models used in biomedical research keeping it updated over time

SECTIONS

1. **Introduction**
2. **Objectives**
3. **Method**
4. **Results**



Introduction

Promoting alternatives in biomedical research

- Half of the uses of animals takes place in biomedical research
- Projects needs to be authorised by member states based on justification that there is no equivalent alternative
- Researchers must ensure that the 3Rs are assessed

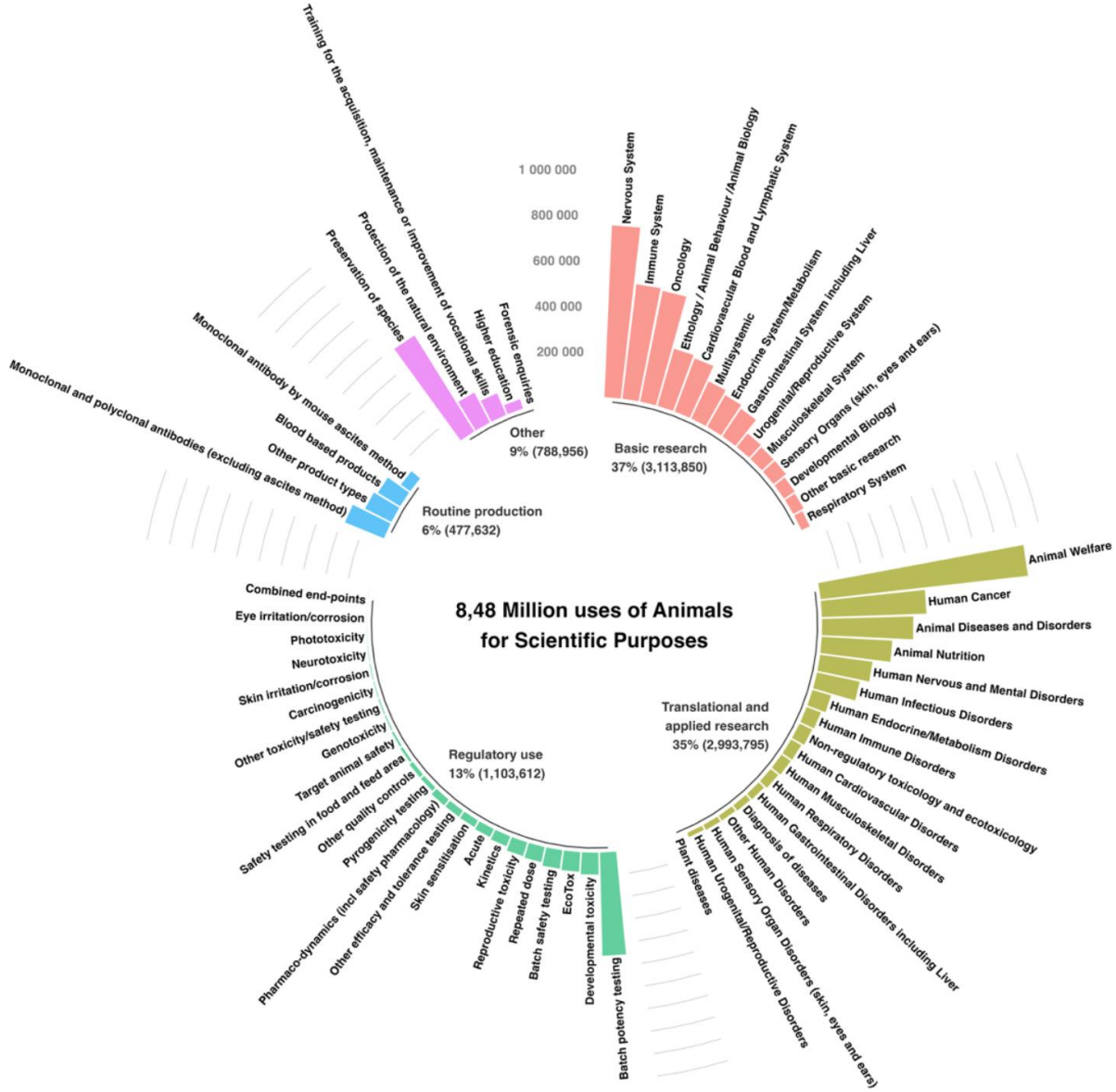


Figure 2: All uses of animals for research and testing in 2022

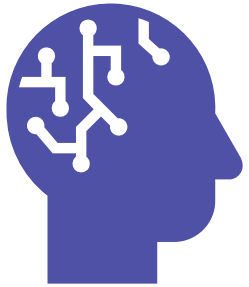
Biomedical reviews of non-animal models

Identify and describe specific research contexts where animal models have been put aside in favour of novel non-animal techniques (2013-2019)

- Respiratory tract diseases
- Breast cancer
- Neurodegenerative diseases
- Immuno-oncology
- Immunogenicity testing for advanced therapy medicinal products
- Cardiovascular diseases
- Autoimmune diseases

Automated approaches for systematic reviews

Recent progress in machine learning, artificial intelligence and natural processing becoming mature enough to be implemented



- Automated Literature Search and Screening (NLP)
- Meta-analysis and Data Interpretation (ML)
- Semi-automation and Human-in-the-loop Approaches (ML/AI)
- Active Learning and Reinforcement Learning (AI)

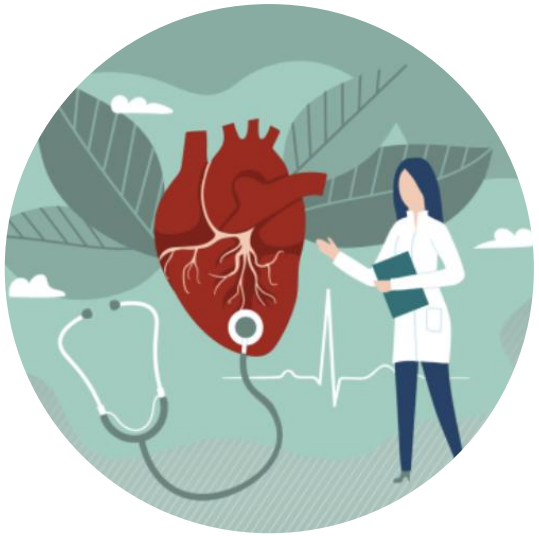
Pilot Project

Initiative or scheme designed to test and explore the feasibility, effectiveness, or usefulness of a particular idea, concept, policy, or program within a limited scope or timeframe.

- Use of automated approaches to mine the vast body of published literature
- Enable the creation and maintenance of an up-to-date, state of the art knowledge source collating Non-Animal Models applied to biomedical research
- Allow the extension of models already collected by EURL ECVAM both in time and in scope

BimmoH’s Objectives

BimmoH’s primary goals



Extending biomedical reviews

Focusing on human-based biology models

Supporting stakeholders

Definitions

BimmoH Dataset

Database of scientific articles references making use of human-based biology models selected by BimmoH's Machine Learning Classifier and indexed with BimmoH vocabularies.

Human biology-based model

Systems or methods used to replicate or simulate human biological processes, diseases, or drug responses for research purposes, based on human material or data.

Machine Learning classifier

Subset of Artificial Intelligence algorithms which build a model based on training data, to make predictions or decisions without being explicitly programmed to do so.

Non-Animal model

Research method that studies biological processes without making use of animals

BimmoH Vocabularies

List of key terms along with their definitions, used to index articles present in the BimmoH dataset.

Human biology-based model identification

In BimmoH, a model is defined as any system capable of receiving an input and that allows recording an output. This includes human and/or animal models based on data or biological or biochemical materials, and may be:

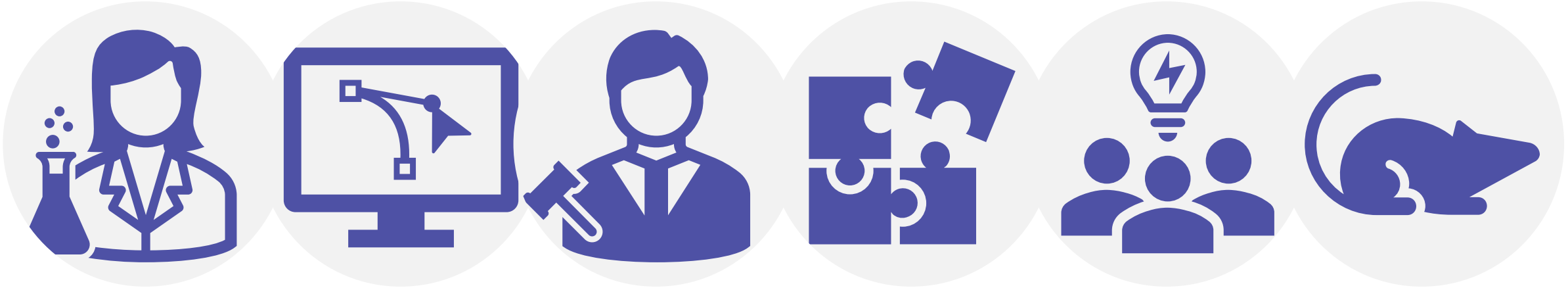
- ***In vitro***: Living biological components outside a living organism.
- ***In silico***: Computational-based models.
- ***In chemico***: Chemical systems used for testing and analysis.

Studies using biological materials or data solely for analysis (i.e., biomarker detection in histological studies, omics studies, etc.) are not considered as models. Humans are not considered models for ethical reasons. They are test subjects under the governance of clinical trial regulations.

This definition was used by biomedical research experts to create select relevant scientific articles to train the Artificial Intelligence classifier and to validate its performance.

User communities support

BimmoH's final goal is to promote further human-based biology models with the aim of increasing the translatability of basic and applied research while reducing the use of animal models.



Academic Researchers

Biomedical Data scientists

Regulators

Policy Makers

Industry Innovators

Investment Analysis

Three Rs Organisations

Requirements

1. Include articles that utilise human biology-based models, even if they also incorporate animal models
2. The database must be comprehensive, covering a broad range of diseases and conditions.
3. Support advanced search capabilities, including filters and keywords.
4. Provide detailed metadata for each model, by including study methodology.
5. Provide direct link access to original publications.
6. Ensure data consistency and accuracy through regular updates and expert validation.
7. Allow for the export of data and search results in various formats.
8. Offer training materials and resources for new users.
9. Ensure high performance and responsiveness of the database search functionality.
10. Develop a responsive and intuitive web interface for easy navigation.
11. Include visual aids and infographics to enhance data presentation.

Methodological approach

SWOT analysis and Stakeholder workshop

Included requirements

- Include both reviews and articles
- Go beyond identified disease areas to cover the whole biomedical research
- Include past and future article references (regular updates) compared to biomedical reviews
- Include articles even if animal models are used alongside human-biology based models
- Limit the analysis the title and abstract selection allowing wide screening

Excluded or postponed

- Provision of full papers (but DOI links are included)
- Assessing Material and Methods section
- Provision of bibliographic and citation data
- Provision of information on human-based model (performance, validation, guideline adherence, etc.)
- Perform manual curation (inclusion/exclusion) of the AI collected data
- Identification of animal derived material

Data process pipeline

Step 1 PubMed Data selection: retrieval of relevant PMIDs from baseline query

Step 2 Retrieval of PubMed titles and abstracts

Step 3 Application of Machine learning classifier excluding irrelevant / animal only

Step 4 Apply indexing scripts to enrich articles descriptions

Step 5 Upload the BimmoH dataset to the web application user interface

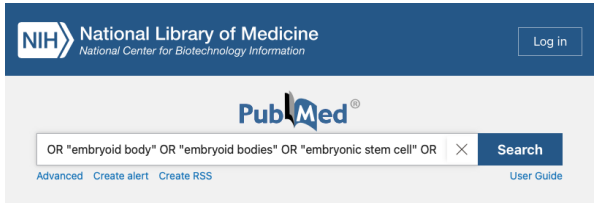
Data preparation

Biomedical reviews data consolidation

- Single Excel file containing of seven biomedical reviews models
- Harmonisation of data (Disease, Organ, Tissue, Model type, Cell type, Omics use, Title, Abstract)

PubMed baseline dataset creation

- PubMed query identifying Human-biology based models and excluding non relevant articles (clinical trial, case studies, etc…)
- Retrieving 4.8 million candidate articles to classify
- Using PubMed API to retrieve articles PMID



PubMed data content and access

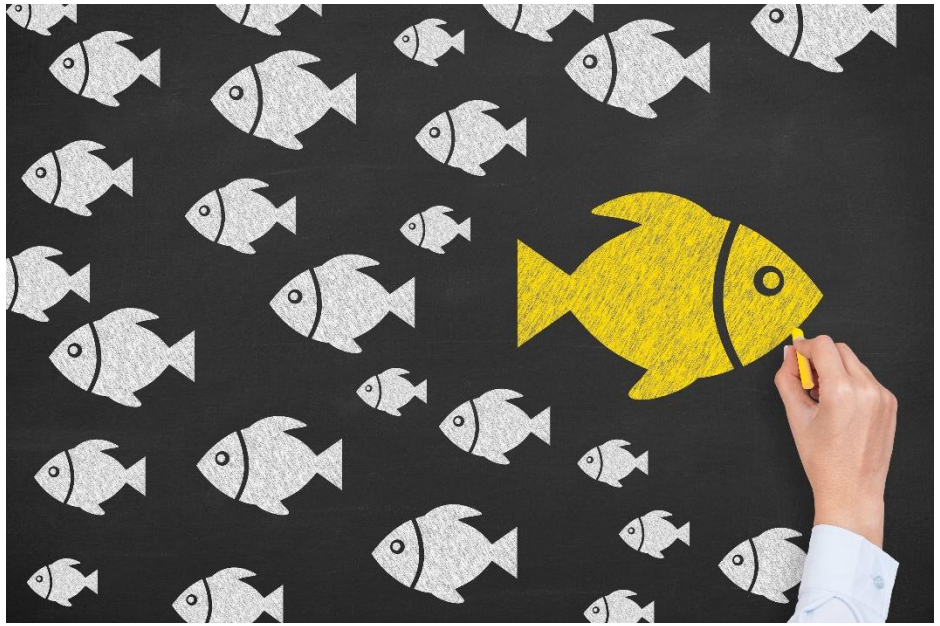
- PubMed APIs use for single article information retrieval
- Standard information (Title, Abstract, Author, DOI, PMID, etc.)

```
PMID- 40926989
OWN - NLM
STAT- In-Process
LR  - 20250912
IS  - 1663-9812 (Print)
IS  - 1663-9812 (Electronic)
IS  - 1663-9812 (Linking)
VI  - 16
DP  - 2025
TI  - Organoids: their emerging essential role in pathological mechanisms and drug
      discovery of diabetes and its complications.
PG  - 1650200
LID - 10.3389/fphar.2025.1650200 [doi]
LID - 1650200
AB  - Diabetes mellitus is a metabolic disease with a high global prevalence, which
      affects blood vessels throughout the entire body. As the disease progresses, it
      often leads to complications, including diabetic retinopathy and nephropathy.
      Currently, in addition to traditional cellular and animal models, more and more
      organoid models have been used in the study of diabetes and have broad
      application prospects in the field of pharmacological research. We summarized the
      organoid models that have been developed for the study of diabetes mellitus and
      its complications, and describe their sources, establishment and maturation
      measures with a focus on pancreatic organoids. For the first time, we summarized
      the contribution of organoids in diabetes and its complications in terms of
      mechanism studies, drug screening, and cellular replacement therapies, in the
      hope of providing a feasible direction for personalized medicine and precision
      treatment of diabetes and its complications. In addition, we discuss the
      strengths and limitations of organoids in the field of diabetes and its
      complications. Nowadays, people strongly advocate personalized medicine and
      precision medicine, and in this regard, organoid technology has advantages that
      are unmatched by any conventional experimental models. By combing organoid
      technology with high-throughput technologies, "patient-specific" drug screening
      can be achieved faster and more accurately. Organoids are also becoming a
      potential source of transplantable tissues and functional cell types for cellular
      replacement therapies in regenerative medicine. With further development of
      assembly and vascularization techniques, organoids will gradually mature and
      improve. In conclusion, the 3D organoid system greatly complements the existing
      modeling system and may play a significant role in future basic and clinical
      research.
CI  - Copyright © 2025 Xu, Zhang, Geng, Luo and Sun.
FAU - Xu, Xiaoyu
AU  - Xu X
AD  - State Key Laboratory for Quality Ensurance and Sustainable Use of Dao-di Herbs,
      Beijing, China.
AD  - Institute of Medicinal Plant Development, Peking Union Medical College and
      Chinese Academy of Medical Sciences, Beijing, China.
AD  - Key Laboratory of Bioactive Substances and Resource Utilization of Chinese Herbal
      Medicine, Ministry of Education, Beijing, China.
FAU - Zhang, Yunqi
AU  - Zhang Y
```

Classification (identifying relevant articles)

Developing a classifier based on supervised Machine Learning

- Low data requirement for training
- Allowing task specific optimisation
- Production of predictable and stable results
- Allowing transparency and interpretability
- Low computational cost
- Ease of update over time



Indexing of selected articles

Anatomy, Histology, and Cells

Anatomical classification
lung, liver, skin, etc.

Histological characteristics
epithelial, connective tissue, etc.

Cell types
fibroblasts, neurons, hepatocytes, etc.

Clinical Conditions, Disease, and Pathophysiology

Specific diseases
diabetes, cancer

Clinical conditions
inflammation, fibrosis

Pathophysiological processes
immune dysregulation

Models

in vitro
living biological components outside a living organism

in silico
computer-based simulations or predictions

in chemico
purely chemical or biochemical test systems

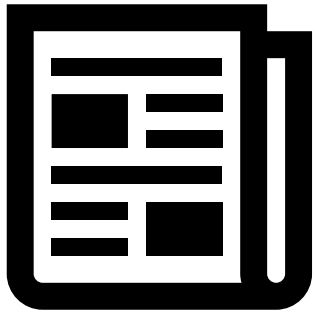
BimmoH Dataset – JRC Data Catalogue

Data coming from PubMed:

- PMID
- DoI
- Authors information
- Bibliographical details
- Title
- Mesh terms

Data coming from BimmoH:

- Model
- Disease
- Human anatomy
- Omics
- Cell line type
- Animal mentions

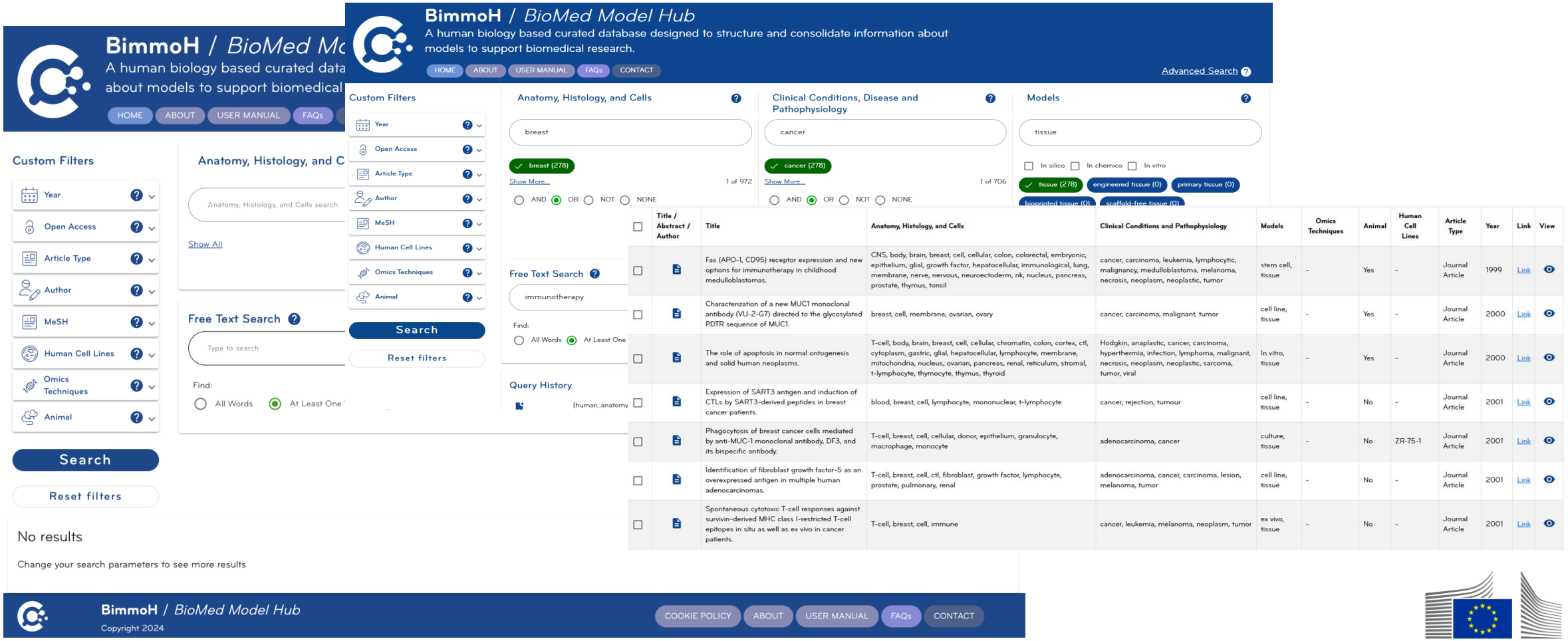


Data coming from OpenAlex:

- Open Access availability
- Abstract (Key terms)

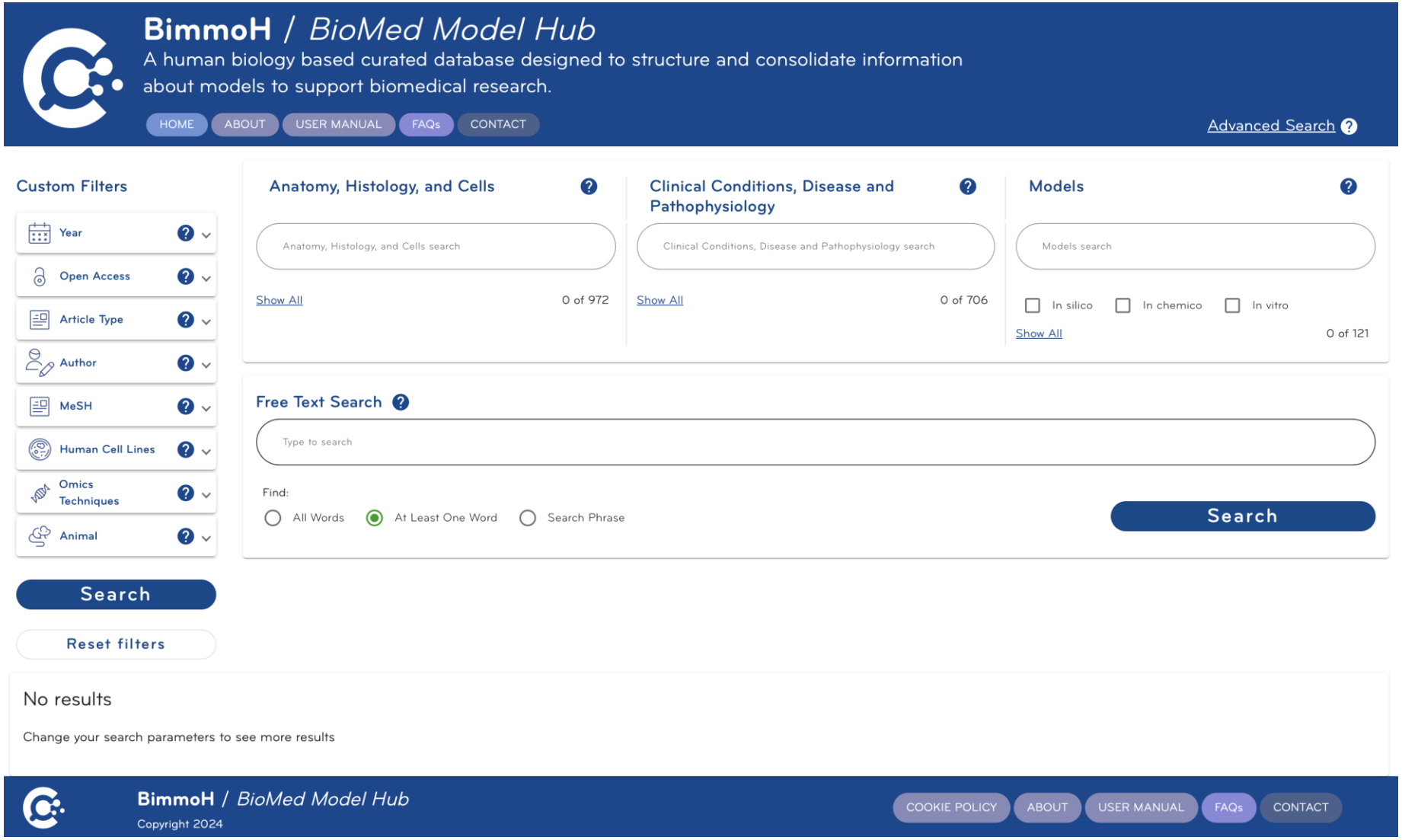
BimmoH – Web User Interface

Available soon at https://bimmoh.eu

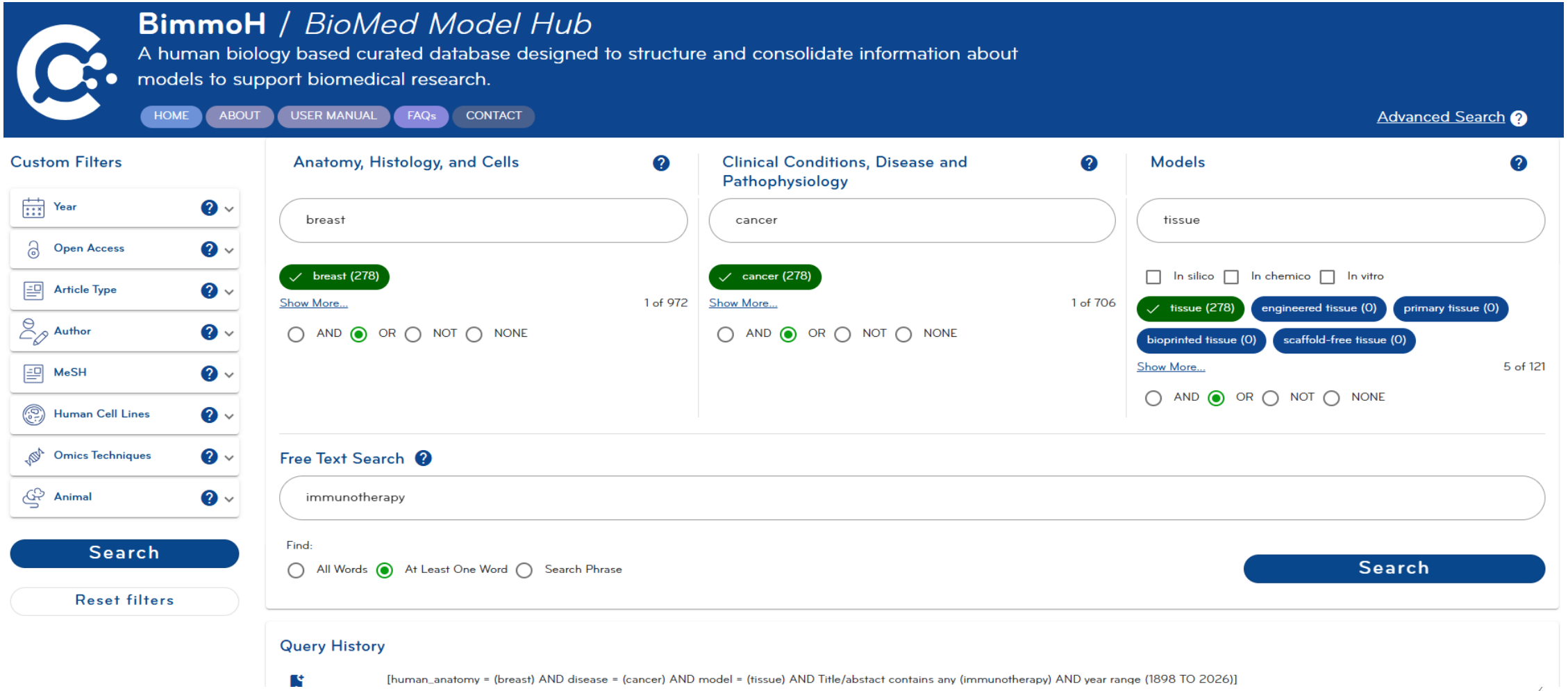


Results

BimmoH – Web User Interface



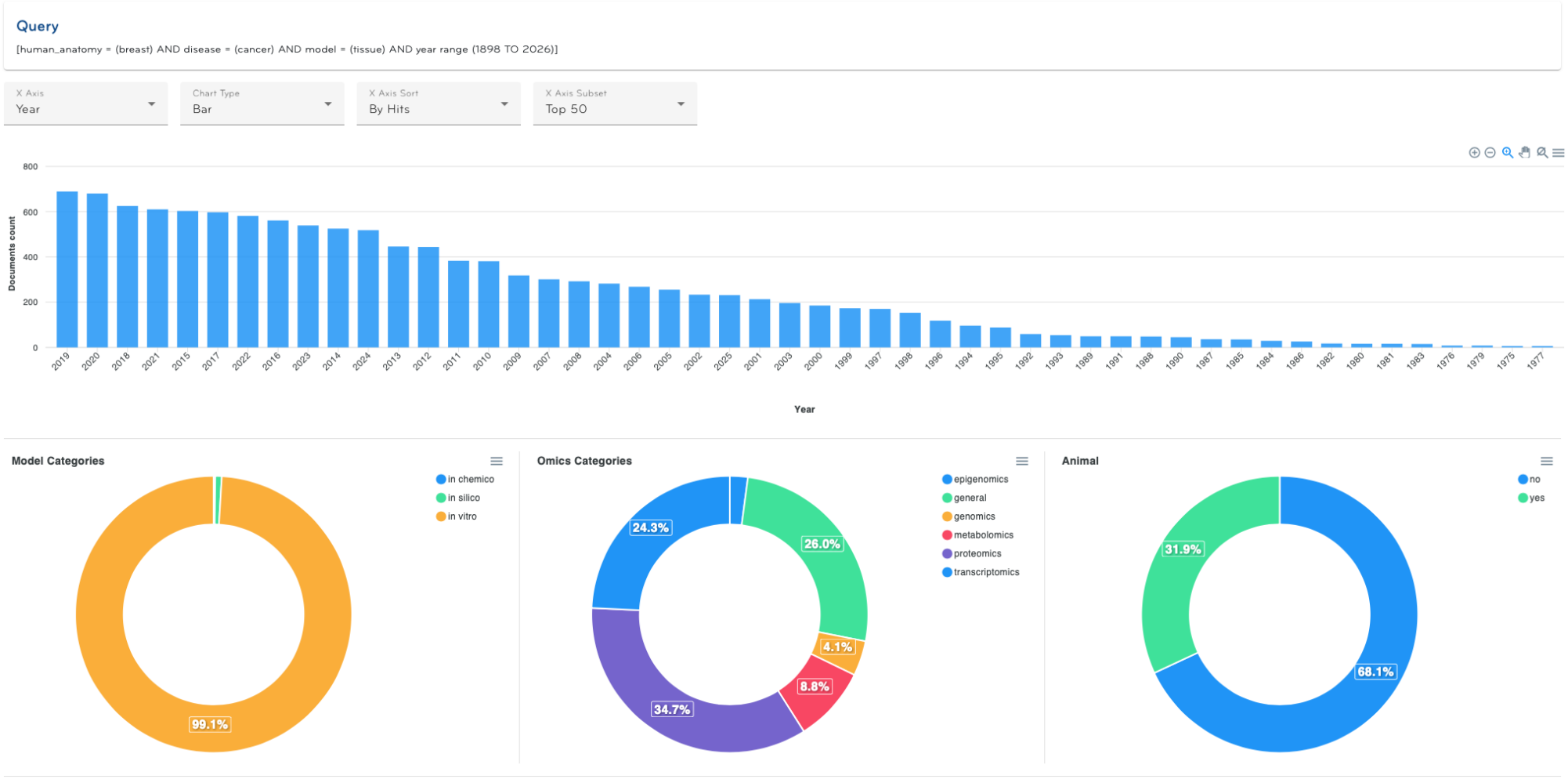
BimmoH – Web User Interface



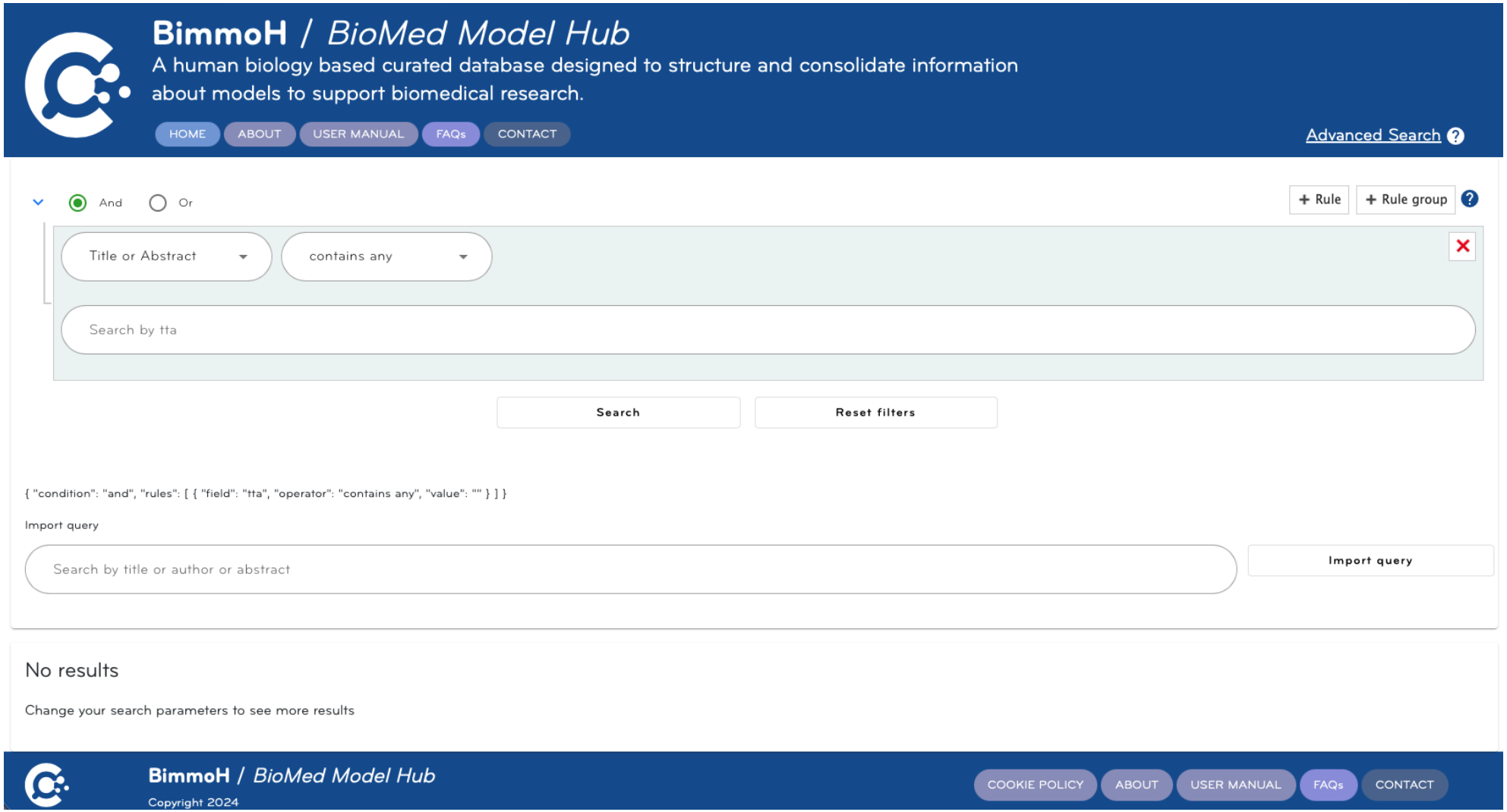
BimmoH – Web User Interface

Title / Abstract / Author	Title	Anatomy, Histology, and Cells	Clinical Conditions and Pathophysiology	Models	Omics Techniques	Animal	Human Cell Lines	Article Type	Year	Link	View
	Fas (APO-1, CD95) receptor expression and new options for immunotherapy in childhood medulloblastomas.	CNS, body, brain, breast, cell, cellular, colon, colorectal, embryonic, epithelium, glial, growth factor, hepatocellular, immunological, lung, membrane, nerve, nervous, neuroectoderm, nk, nucleus, pancreas, prostate, thymus, tonsil	cancer, carcinoma, leukemia, lymphocytic, malignancy, medulloblastoma, melanoma, necrosis, neoplasm, neoplastic, tumor	stem cell, tissue	-	Yes	-	Journal Article	1999	Link	
	Characterization of a new MUC1 monoclonal antibody (VU-2-G7) directed to the glycosylated PDTR sequence of MUC1.	breast, cell, membrane, ovarian, ovary	cancer, carcinoma, malignant, tumor	cell line, tissue	-	Yes	-	Journal Article	2000	Link	
	The role of apoptosis in normal ontogenesis and solid human neoplasms.	T-cell, body, brain, breast, cell, cellular, chromatin, colon, cortex, ctl, cytoplasm, gastric, glial, hepatocellular, lymphocyte, membrane, mitochondria, nucleus, ovarian, pancreas, renal, reticulum, stromal, t-lymphocyte, thymocyte, thymus, thyroid	Hodgkin, anaplastic, cancer, carcinoma, hyperthermia, infection, lymphoma, malignant, necrosis, neoplasm, neoplastic, sarcoma, tumor, viral	In vitro, tissue	-	Yes	-	Journal Article	2000	Link	
	Expression of SART3 antigen and induction of CTLs by SART3-derived peptides in breast cancer patients.	blood, breast, cell, lymphocyte, mononuclear, t-lymphocyte	cancer, rejection, tumour	cell line, tissue	-	No	-	Journal Article	2001	Link	
	Phagocytosis of breast cancer cells mediated by anti-MUC-1 monoclonal antibody, DF3, and its bispecific antibody.	T-cell, breast, cell, cellular, donor, epithelium, granulocyte, macrophage, monocyte	adenocarcinoma, cancer	culture, tissue	-	No	ZR-75-1	Journal Article	2001	Link	
	Identification of fibroblast growth factor-5 as an overexpressed antigen in multiple human adenocarcinomas.	T-cell, breast, cell, ctl, fibroblast, growth factor, lymphocyte, prostate, pulmonary, renal	adenocarcinoma, cancer, carcinoma, lesion, melanoma, tumor	cell line, tissue	-	No	-	Journal Article	2001	Link	
	Spontaneous cytotoxic T-cell responses against survivin-derived MHC class I-restricted T-cell epitopes in situ as well as ex vivo in cancer patients.	T-cell, breast, cell, immune	cancer, leukemia, melanoma, neoplasm, tumor	ex vivo, tissue	-	No	-	Journal Article	2001	Link	

BimmoH – Web User Interface



BimmoH – Web User Interface



Conclusion

BimmoH dataset - Statistics

More than 750,000 articles with a 90% precision

This is the largest dataset of scientific articles making use of human biology-based models kept updated over time

Thank you