BIOSEEK BETA

RESEARCH. DISCOVER. COLLABORATE

White Paper

www.bioseek.eu

BioSeek is **the most advanced search platform** for Life Sciences.

Using a **graph database** and **semantic search algorithms**, we solve two major problems scientific researchers face :

Keyword-based search gives poor results: Standard keyword-based search in relational databases simply doesn't deliver all relevant results: a number of non-formal matches are left out. Also, there are various ways each entity designation can be spelled, and varying IDs across different nomenclatures.

E.g.: the gene **NFKB1** has **225** different designations only on NCBI.
--> We developed a **semantic text analyzer** for **augmented search** (page 3), which **identifies all designations of given entity** and then groups the results regardless of the specific designation they contain, as long as it points to the searched entity.

Information is fragmented: Researchers and medical doctors need information which is spread across multiple, differently structured sources, like **NCBI**, **KEGG**, **UniProt**, etc. Cross-site navigation is a major timewasting task which steals precious energy from the research process.
--> We developed a **web aggregator** (page 4) which collects information from these sources and delivers it to our **unified portal**.
* We are continuously adding more sources

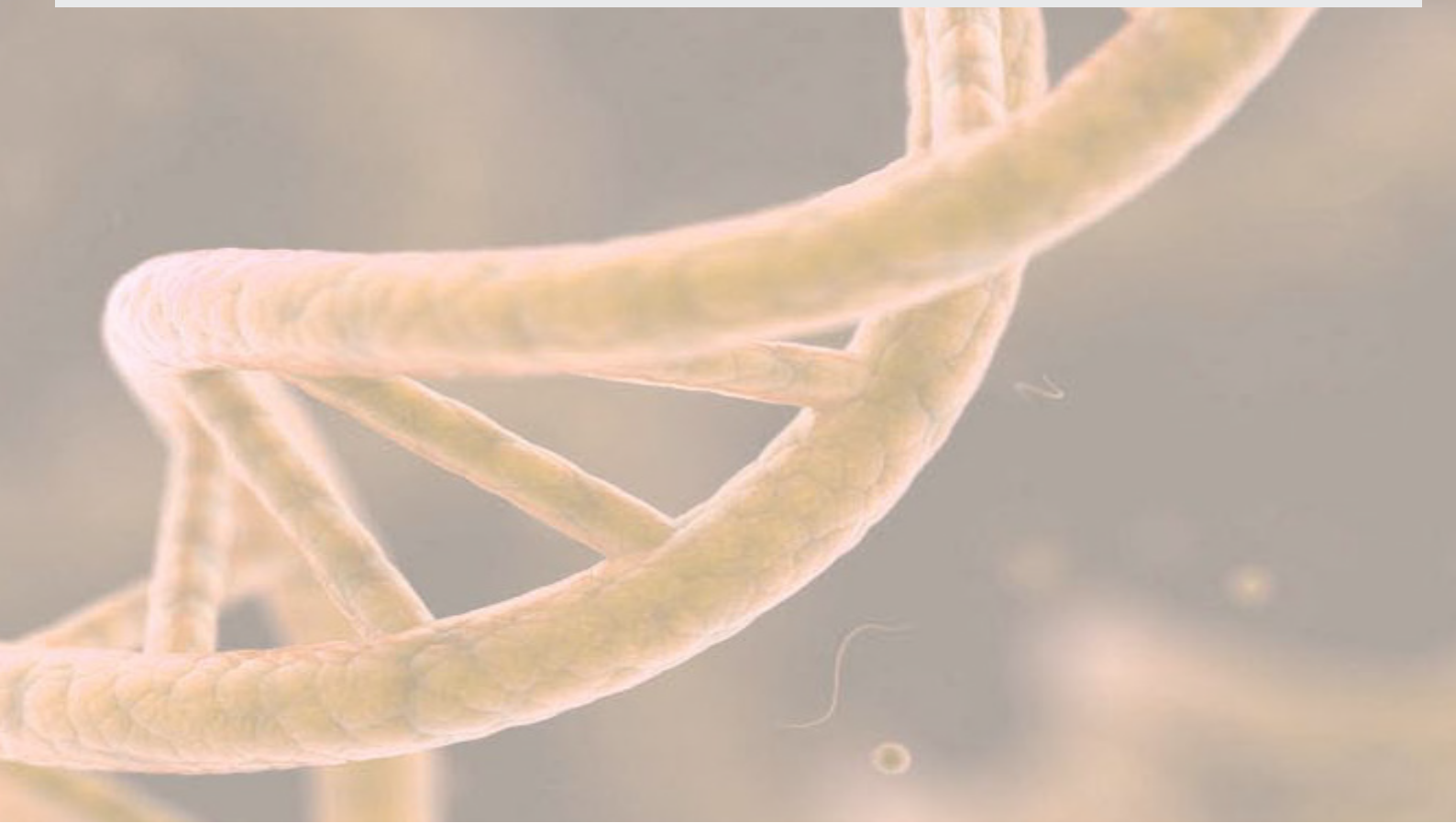
Augmented search

We created an algorithm which draws connections between the different designations of one entity. It's a text analysis engine which **indexes all bio terms within the text,** matches them against all the documents which contain **any of their spelling options**, and delivers all these documents as search results. **It converts each relevant term into an explorable object, spot on, by listing at the side of your screen all its relations** (via articles) to other entities such as genes, diseases, pathways, authors, etc. It recognizes more than **60 million terms so far**.

A **web browser extension** brings the analytical power of our semantic algorithm to a continuously growing list of web sources, among which are **Nature Journal**, **NCBI**, **Elsevier**, etc.

An independent **document reader** can analyze around 400 types of files, including **non-textual formats such as PDF.**

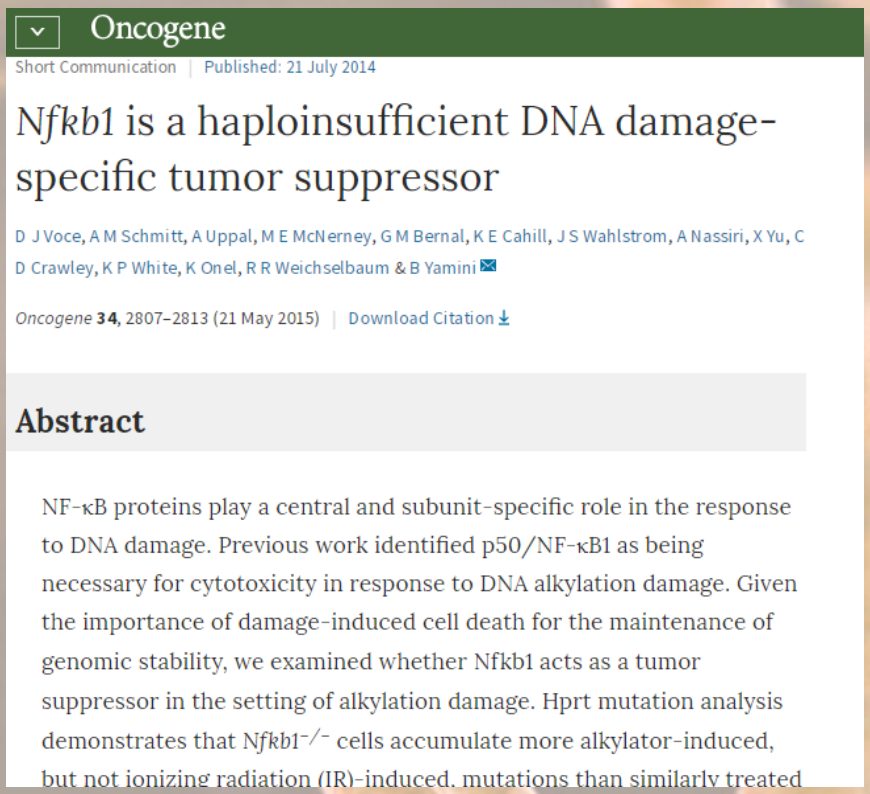
The BioSeek Browser Extension and the BioSeek Reader are **available to download on bioseek.eu**.



BioSeek| The Augmented Search

Here's an example of the augmented search in action:

Plain Site:



Oncogene

Short Communication | Published: 21 July 2014

Nfkb1 is a haploinsufficient DNA damage-specific tumor suppressor

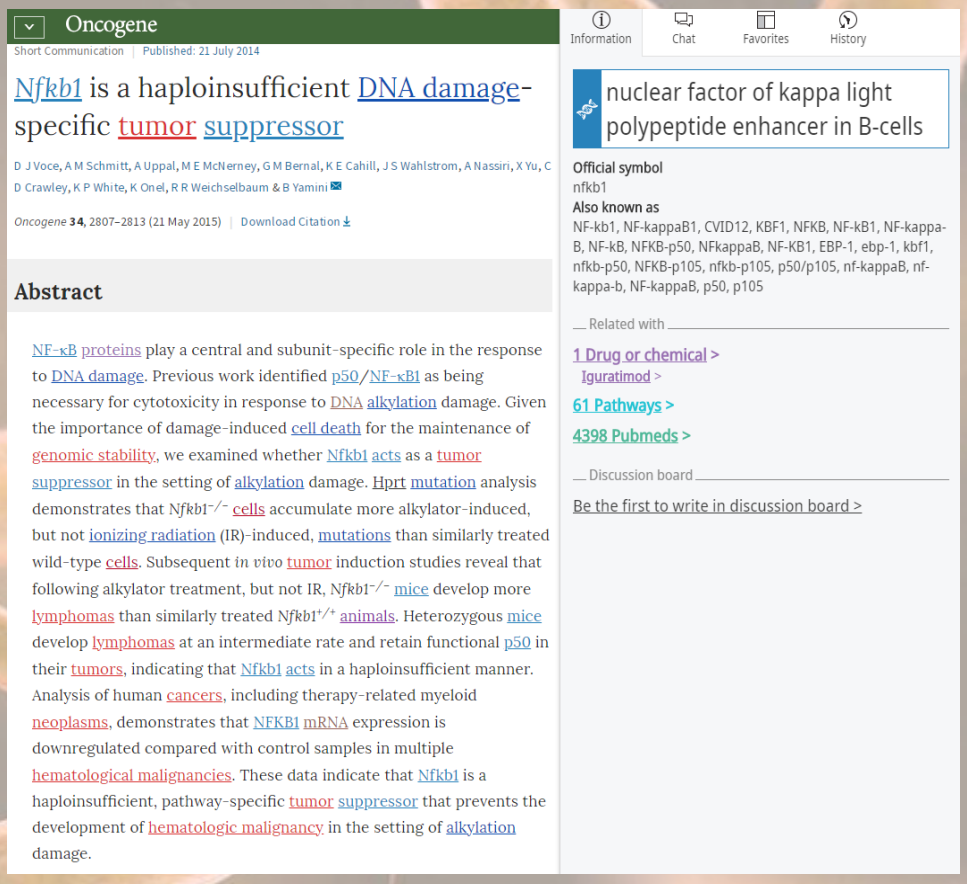
D J Voce, A M Schmitt, A Uppal, M E McNerney, G M Bernal, K E Cahill, J S Wahlstrom, A Nassiri, X Yu, C D Crawley, K P White, K Onel, R R Weichselbaum & B Yamini

Oncogene **34**, 2807–2813 (21 May 2015) | Download Citation

Abstract

NF-κB proteins play a central and subunit-specific role in the response to DNA damage. Previous work identified p50/NF-κB1 as being necessary for cytotoxicity in response to DNA alkylation damage. Given the importance of damage-induced cell death for the maintenance of genomic stability, we examined whether Nfkb1 acts as a tumor suppressor in the setting of alkylation damage. Hprt mutation analysis demonstrates that *Nfkb1*$^{-/-}$ cells accumulate more alkylator-induced, but not ionizing radiation (IR)-induced, mutations than similarly treated

Augmented Search ON:



Oncogene

Short Communication | Published: 21 July 2014

Nfkb1 is a haploinsufficient DNA damage-specific tumor suppressor

D J Voce, A M Schmitt, A Uppal, M E McNerney, G M Bernal, K E Cahill, J S Wahlstrom, A Nassiri, X Yu, C D Crawley, K P White, K Onel, R R Weichselbaum & B Yamini

Oncogene **34**, 2807–2813 (21 May 2015) | Download Citation

Abstract

NF-κB proteins play a central and subunit-specific role in the response to DNA damage. Previous work identified p50/NF-κB1 as being necessary for cytotoxicity in response to DNA alkylation damage. Given the importance of damage-induced cell death for the maintenance of genomic stability, we examined whether Nfkb1 acts as a tumor suppressor in the setting of alkylation damage. Hprt mutation analysis demonstrates that *Nfkb1*$^{-/-}$ cells accumulate more alkylator-induced, but not ionizing radiation (IR)-induced, mutations than similarly treated wild-type cells. Subsequent *in vivo* tumor induction studies reveal that following alkylator treatment, but not IR, *Nfkb1*$^{-/-}$ mice develop more lymphomas than similarly treated *Nfkb1*$^{+/+}$ animals. Heterozygous mice develop lymphomas at an intermediate rate and retain functional p50 in their tumors, indicating that Nfkb1 acts in a haploinsufficient manner. Analysis of human cancers, including therapy-related myeloid neoplasms, demonstrates that NFKB1 mRNA expression is downregulated compared with control samples in multiple hematological malignancies. These data indicate that Nfkb1 is a haploinsufficient, pathway-specific tumor suppressor that prevents the development of hematologic malignancy in the setting of alkylation damage.

Information | Chat | Favorites | History

nuclear factor of kappa light polypeptide enhancer in B-cells

Official symbol
nfkb1
Also known as
NF-kb1, NF-kappaB1, CVID12, KBF1, NFKB, NF-kB1, NF-kappa-B, NF-kB, NFKB-p50, NFkappaB, NF-KB1, EBP-1, ebp-1, kbf1, nfkb-p50, NFKB-p105, nfkb-p105, p50/p105, nf-kappaB, nf-kappa-b, NF-kappaB, p50, p105

Related with

1 Drug or chemical >
Iguratimod >

61 Pathways >

4398 Pubmeds >

Discussion board

Be the first to write in discussion board >

BioSeek| The web aggregator

Not simply data aggregation

We don't just move an item from point A to point B. We organize the data in a **graph structure** created especially for our advanced algorithms, which allows for **accurate, exhaustive and fast data retrievals**.

- Chemical compounds & Reactions
- Molecular pathways
- Biological functions
- Genes and Gene products
- Gene Expressions
- Protein-to-Protein interactions
- Diseases & Cancer
- Protein similarities
- Drugs
- Tissue & Organs
- Publications
- Authors
- Citations & References
- Pathogens

KEGG

PubChem

National Center for Biotechnology Information
NCBI

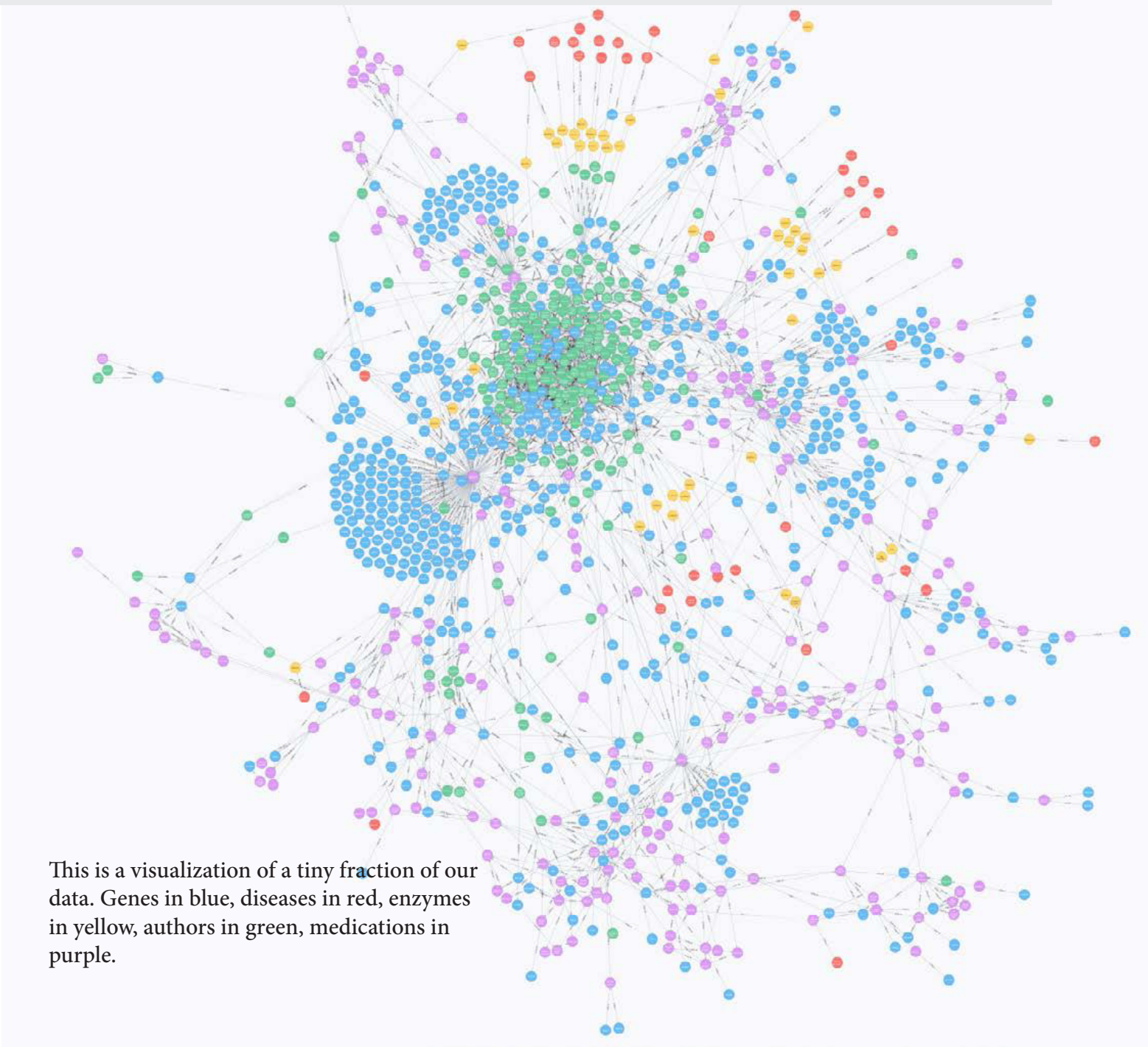
UniProt

PubMed.gov

REACTOME
A CURATED PATHWAY DATABASE

BioSeek| The Graph Database

The Graph Database is optimal for storing and processing such **massive volumes of data**, which contain highly **connected entities**. Its performance is considerably higher than the performance of the standard relational database model in terms of retrieving these connections between all sets of entities.
For example, we search a gene, and we want to see all diseases, related to this gene, all drugs, related to these diseases, and all clinical trials, related to these drugs. The Graph Database delivers all this information spot on.



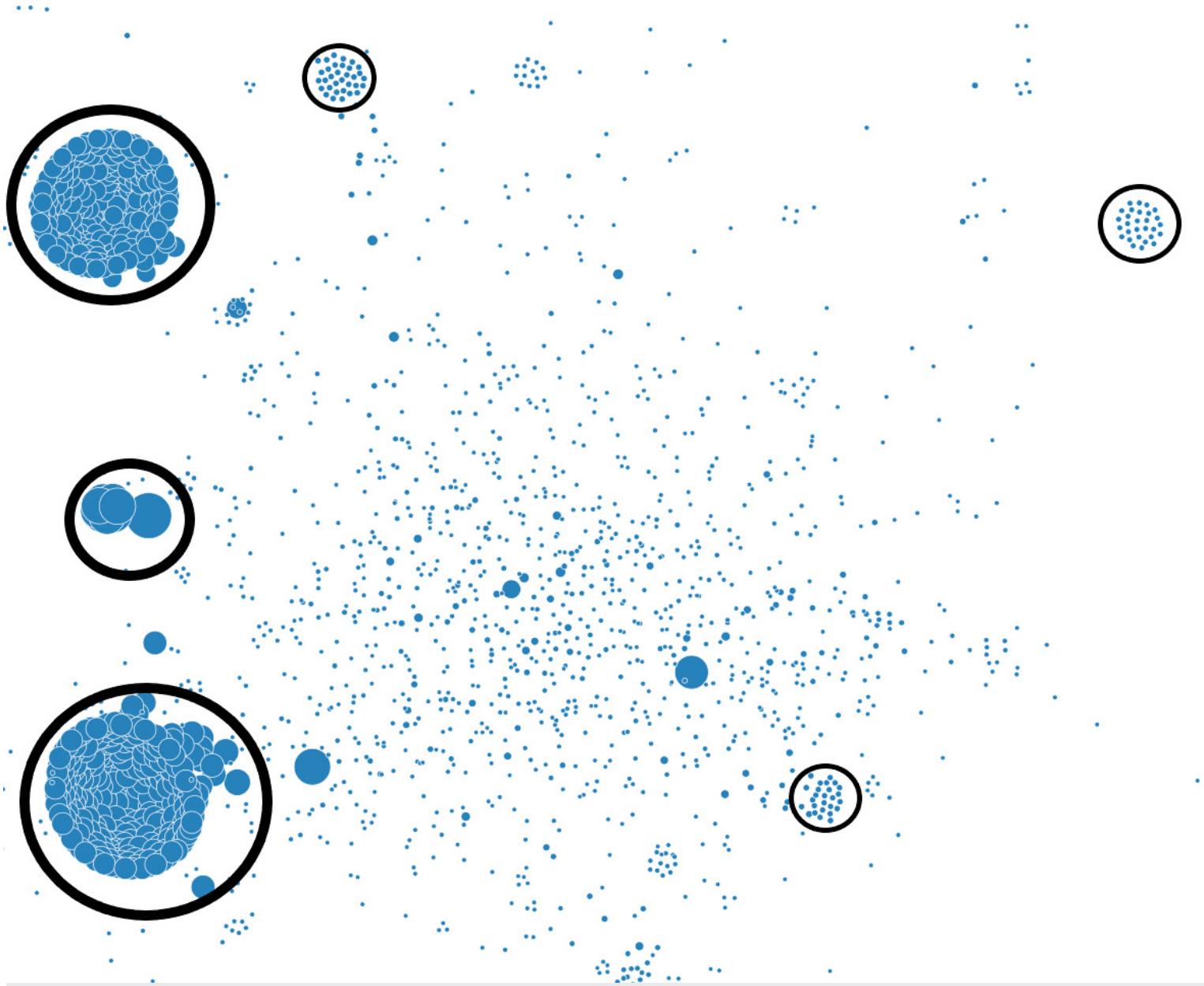
This is a visualization of a tiny fraction of our data. Genes in blue, diseases in red, enzymes in yellow, authors in green, medications in purple.

Features

*presented in detail on pages to follow

- **Analytics** - A number of innovative analytical tools, created with focus on cutting-edge **information visualization**.
- **Visual Search** - a tool for multicomponent queries, with possibility for specifying connections between them. It performs a very large number of operations fast.
- **Mind Map -** a tool for visualization and sharing of an idea. It is unique because it relies upon the aggregated data from our database and allows you to follow natural links between entities in Life Sciences.
- **Messaging platform** - fully integrated with all BioSeek's components, it enables the users to exchange industry-specific information much more efficiently- it recognizes bio terms and converts them into explorable objects directly in the chat window.
- **Publishing Platform -** a next generation document-editing and sharing platform, which envisions open science.

BioSeek Analytics| Clusterization



Clusterization

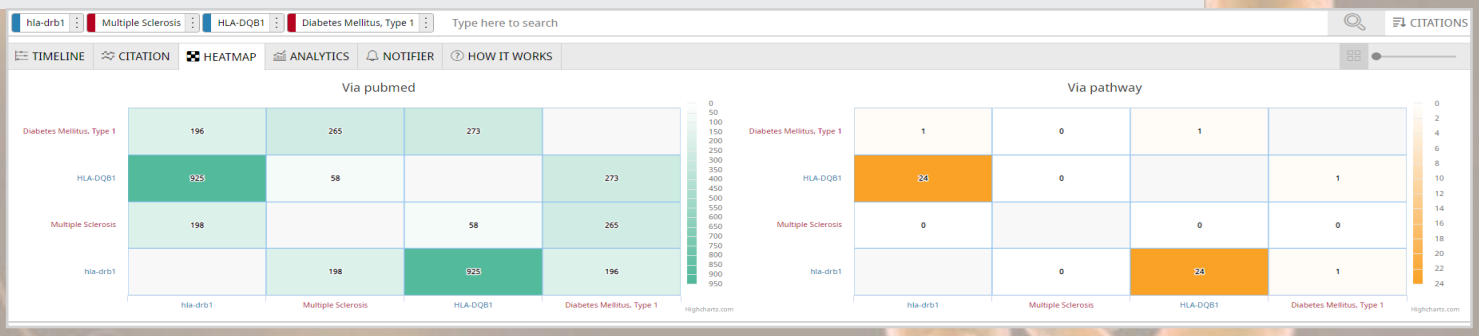
This image shows all genes related to colorectal cancer. Each individual bubble represents one single gene. The larger the bubble, the more mentions this gene has. The article mentions act as a gravitational force between the bubbles, therefore clusters are formed by genes which are often co-mentioned in the same article. The clusters are signs of established or emerging theories .

This algorithm is based on **T-distributed Stochastic Neighbor Embedding (t-SNE) methodology**. More about it in this video.

BioSeek Analytics| Heatmap & VENN Diagram

Heatmap

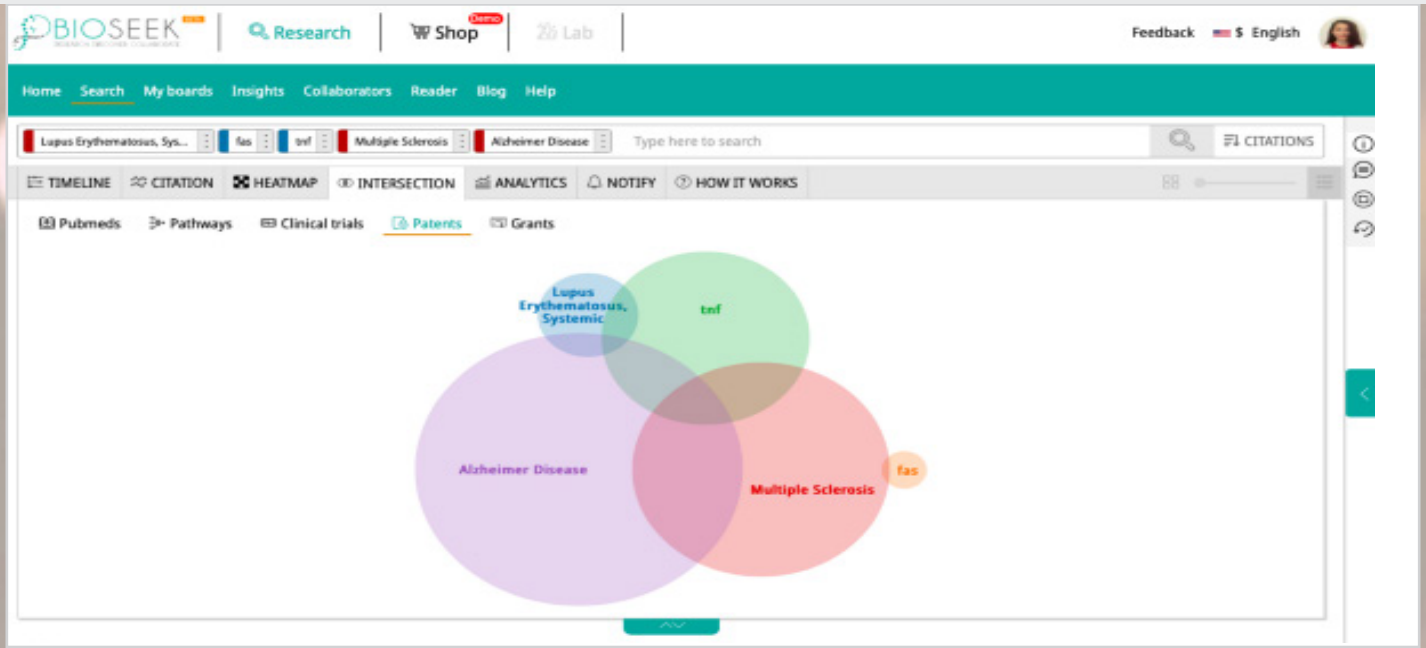
Automatically performs searches for articles which mention each possible pair of entities forming your original search. Click a box to view results.



VENN Diagram- The Intersection Analysis

Delivers a clear visualization of the number of articles which mention given entity AND the inetersections between them. All these sets of combinations are easily accessible- just click the zone of interest to view results.

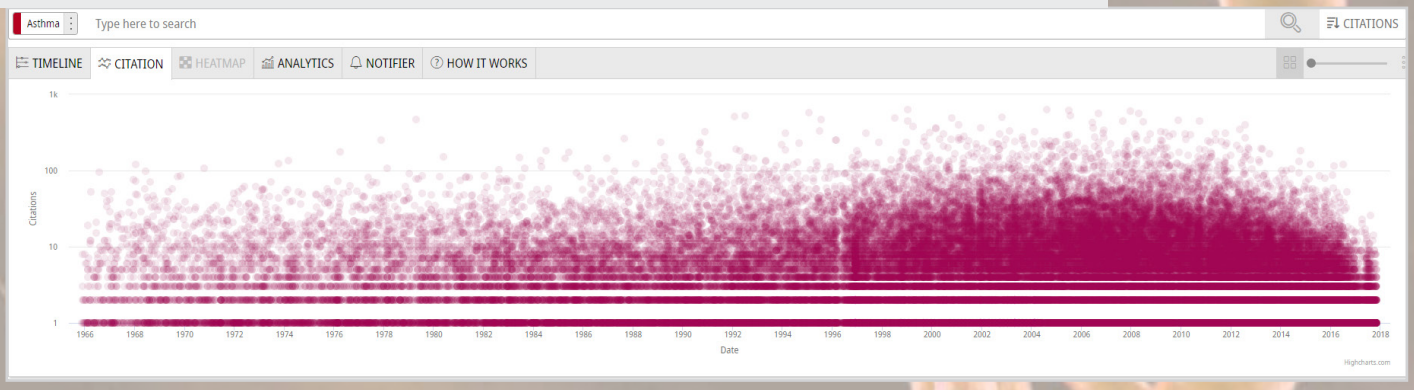
Read more



BioSeek Analytics| Citations and distribution

Citations

Number of mentions of given entity in time. Dots are clickable and retrieve corresponding results.



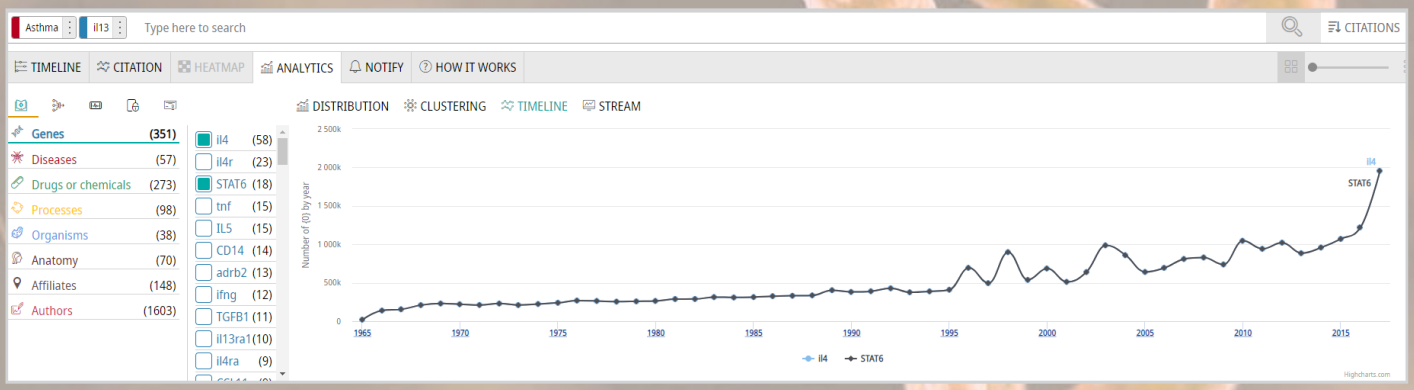
Distribution

Faceted representation of search results- by type of entity. “Distribution” shows the top 20 most mentioned entities in each category - genes, diseases, processes, etc.

BioSeek Analytics| Timeline and Stream

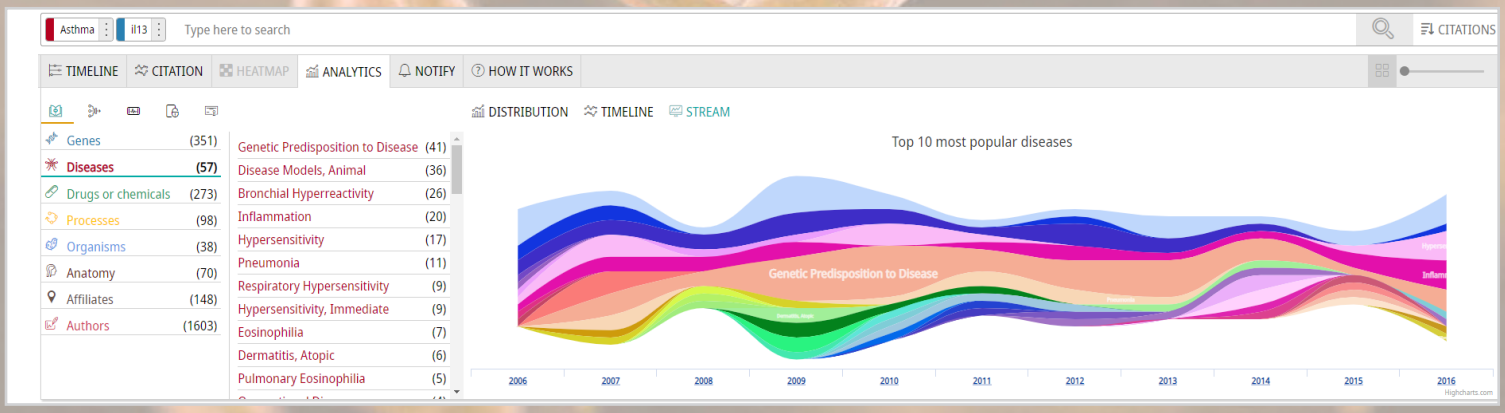
Timeline Analytics

Shows number of mentions of given entity in articles for each year. Dots are clickable and retrieve corresponding results.



Stream Analytics

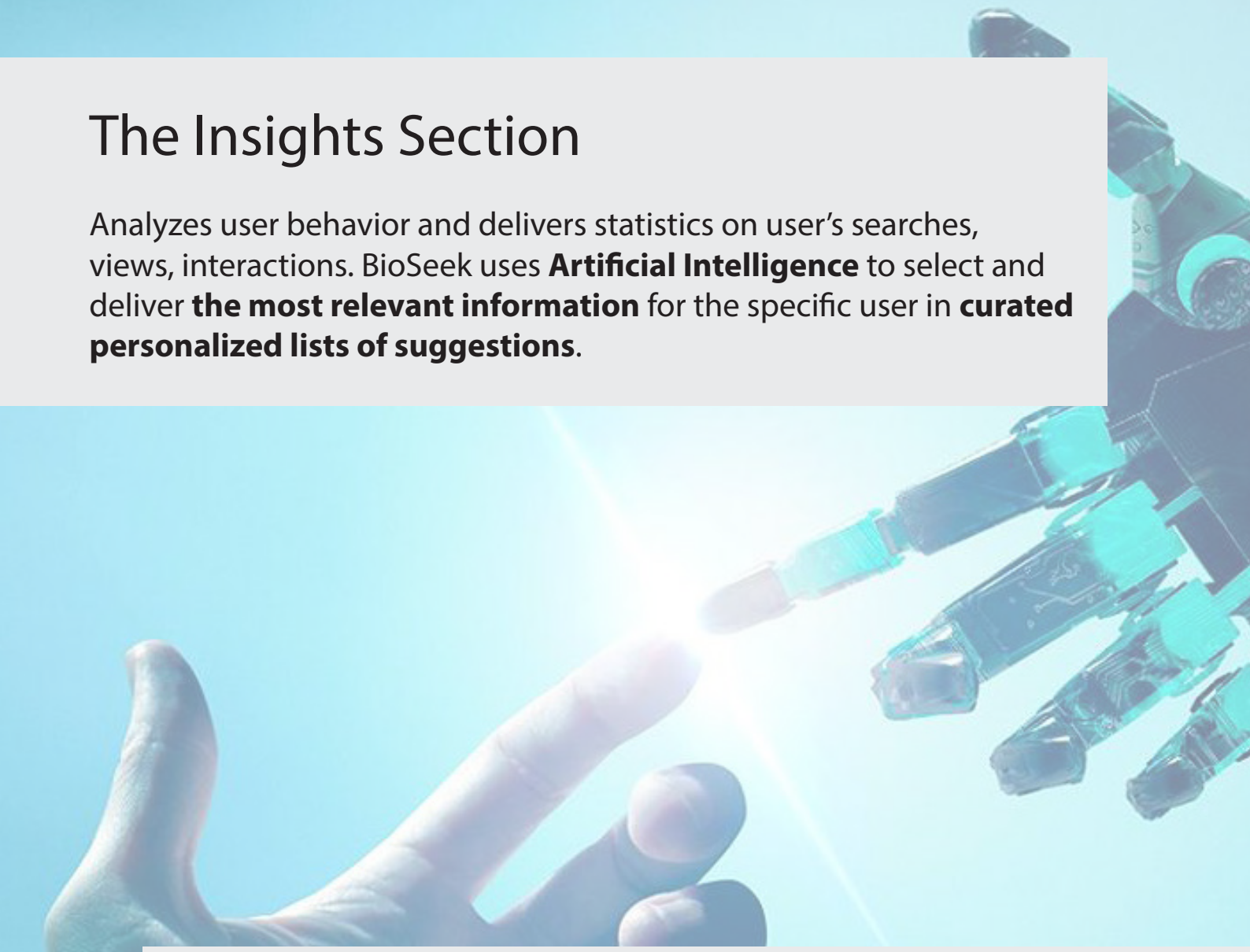
Shows popularity tendencies of entities within given group in time. Click on a color to see details of entity in the Right Panel.



BioSeek Analytics| Artificial Intelligence

The Insights Section

Analyzes user behavior and delivers statistics on user's searches, views, interactions. BioSeek uses **Artificial Intelligence** to select and deliver **the most relevant information** for the specific user in **curated personalized lists of suggestions**.



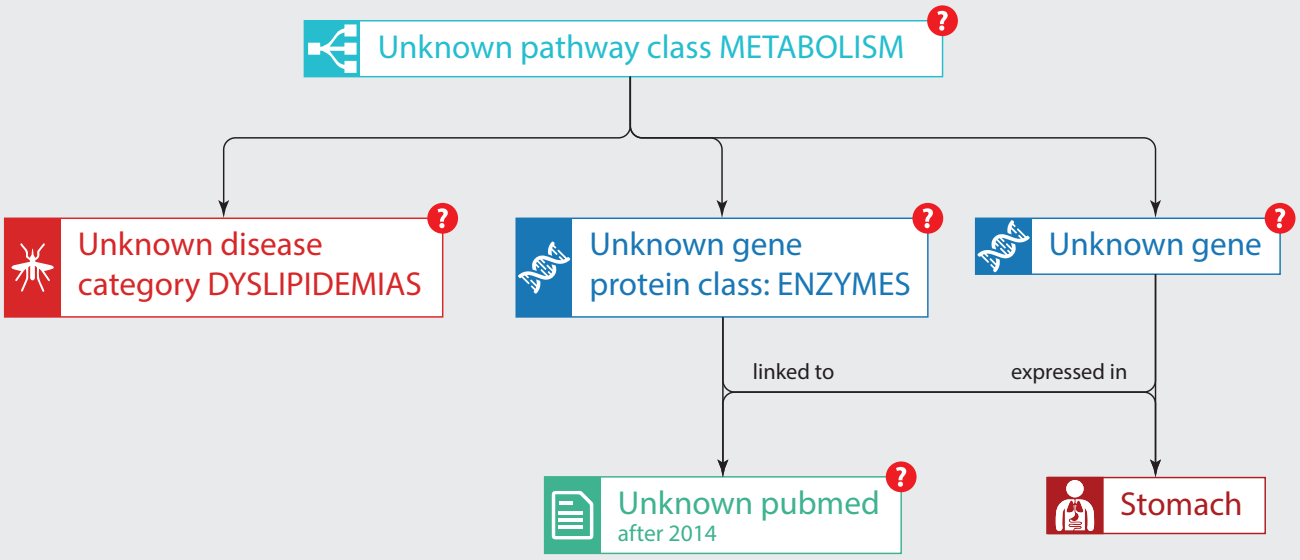
Alerts

The **notification system** allows you to set alerts on search criteria, including **multicomponent searches**. For example, you enter "Atopic Dermatitis", "FLG", and "il13", click "Notify" and get emails whenever we find matching results- new articles which mention these three entities.

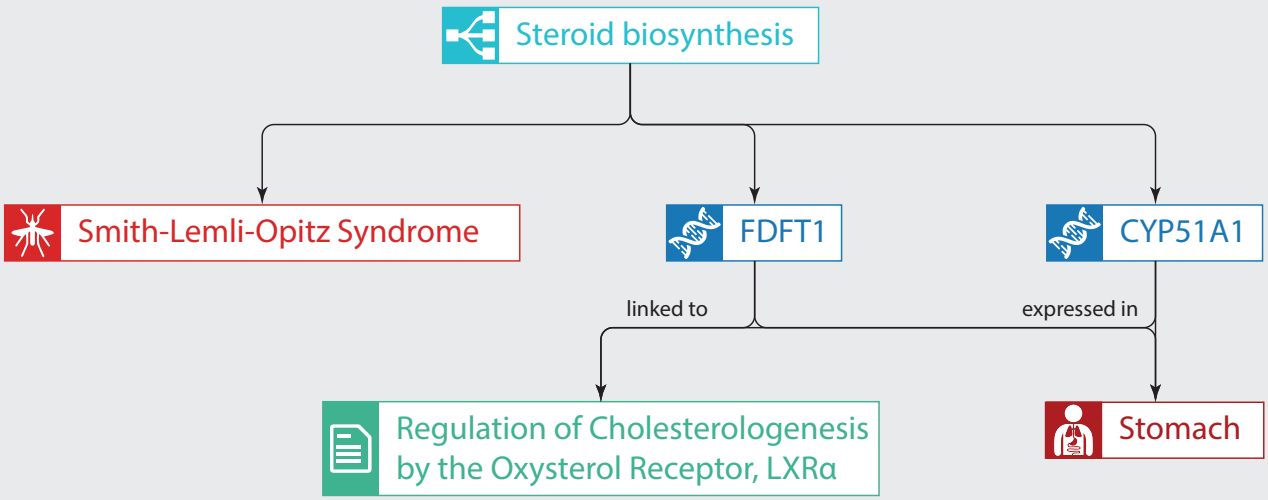
BioSeek Tools| Visual Search

Since you have the data structured as a graph, it is best to structure your search as a graph too.

Search query

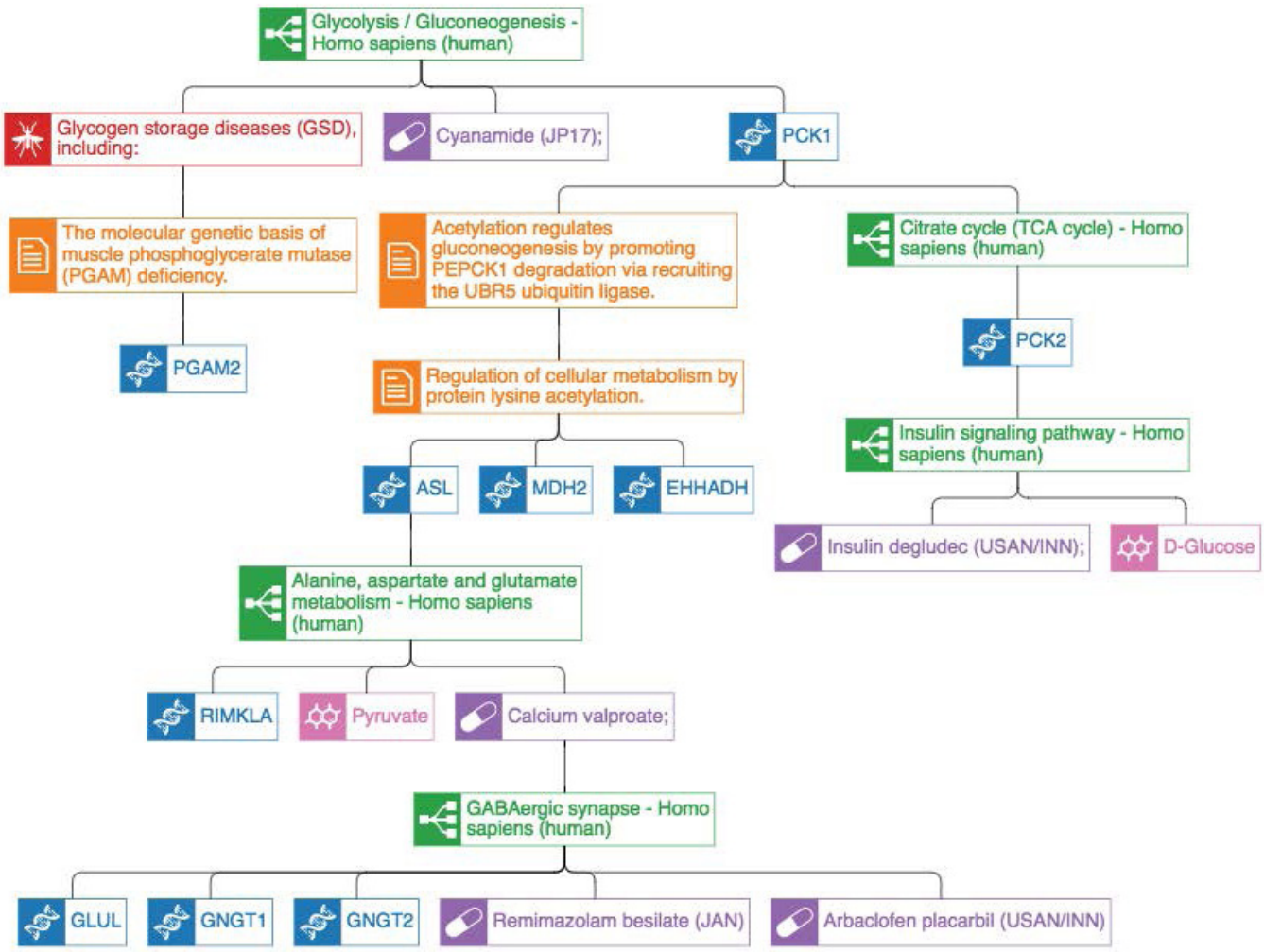


Search result



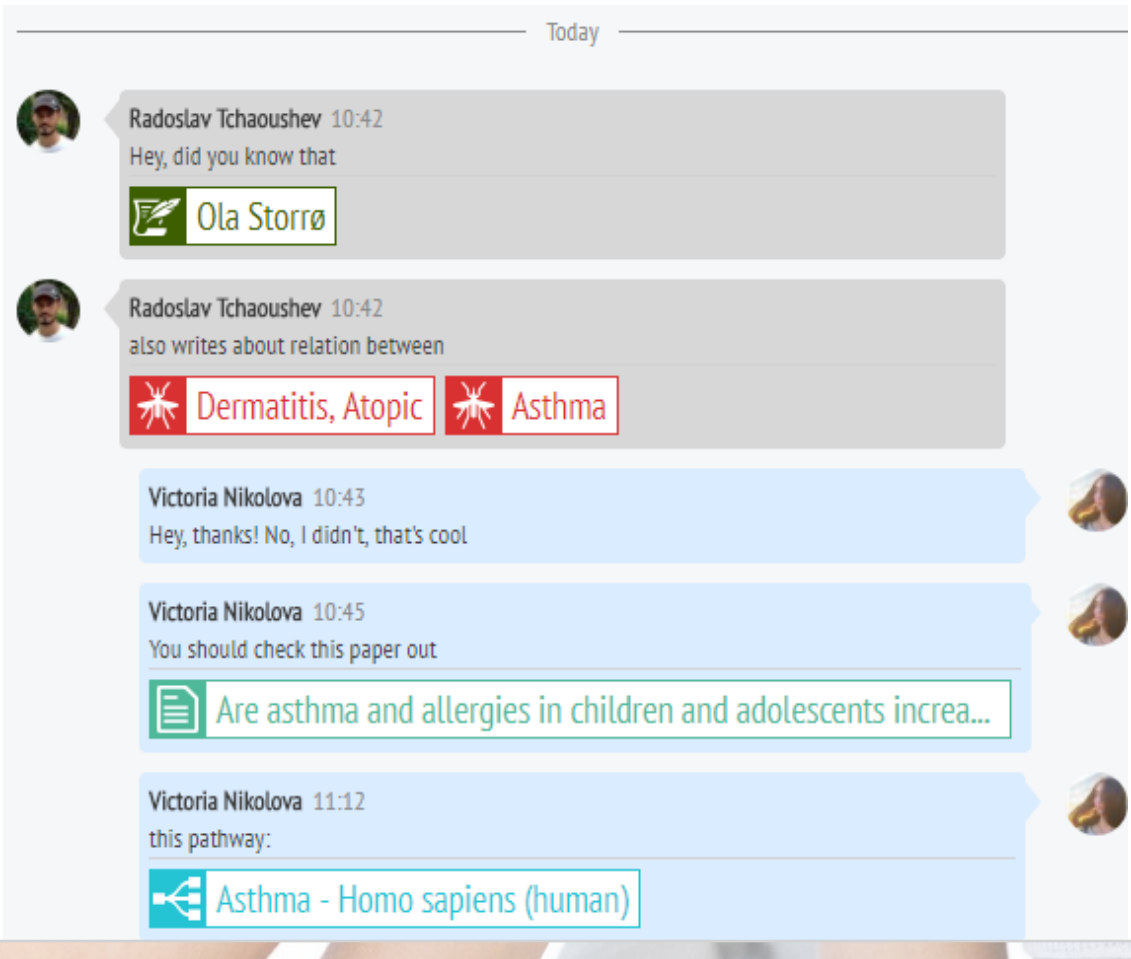
BioSeek Tools| Mind Map

The Mind Map is a visualization tool which helps you **structure and develop your idea**. It can include **multiple entities** and **identify connections** between them. It is an efficient means of **production and transmission of knowledge**.



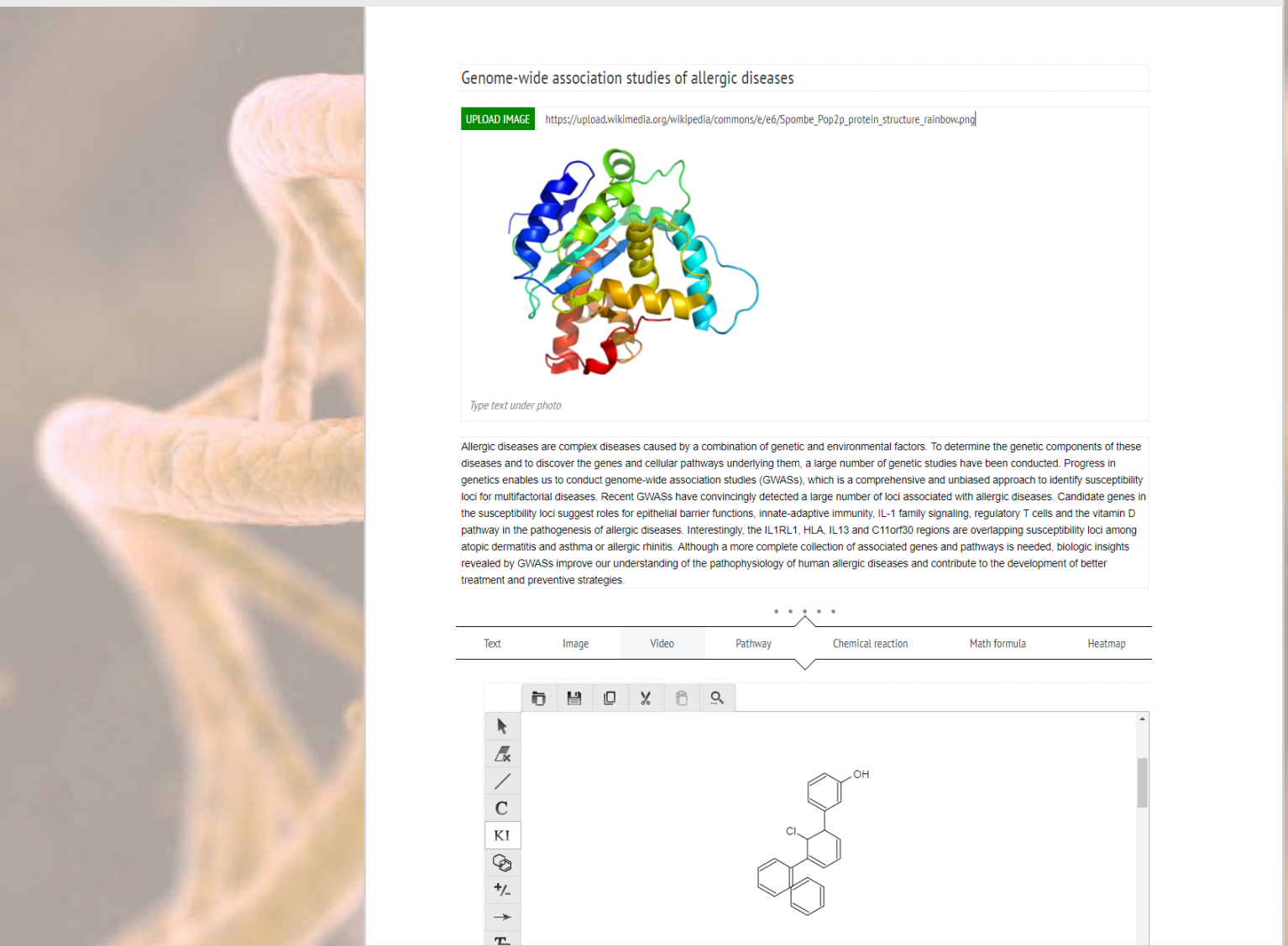
The BioSeek Messenger

- The Messenger is **deeply integrated** with all components of the platform.
- The **semantic algorithm** which converts terms into interactive entities (allowing for further navigation) operates here too.
- Recognizes links from other websites and converts them into **interactive entities** – shareable, clickable, drag-droppable.
- **Mathematical formulas** and **chemical reactions** are also drag-droppable and interactive.



BioSeek Modules| Publishing

- A document-editing and publishing platform designed for **collaboration.**
- The **pluggable architecture** of our Publishing Platform allows for the **embedding** of **mathematical formulas, chemical reactions, biological pathways, 3D images** and **any other relevant elements** without the need of a third party software.
- Any **embedded** object is **interactive** and can be included in another document even by drag-and-drop.
- **Export** to industry standard formats like **PDF**, **LaTeX**, etc.
- **Citations**' sources are automatically listed in the destination document.



Enterprise Solution

The **Enterprise package** is paid and offered **on premise** for corporations:

- Access to all information available online in **complete privacy**
- The **algorithmic foundation** of BioSeek can be applied to all **internal documentation, teams**, etc.
- We provide an **API** which enables the integration of our technology into existing software solutions.

BioSeek| The Team

Rossen Genchev
CEO
Years of experience: 22

Boyan Dimitrov
COO
Years of experience: 19

Victoria Nikolova
Project Manager
Years of experience: 9

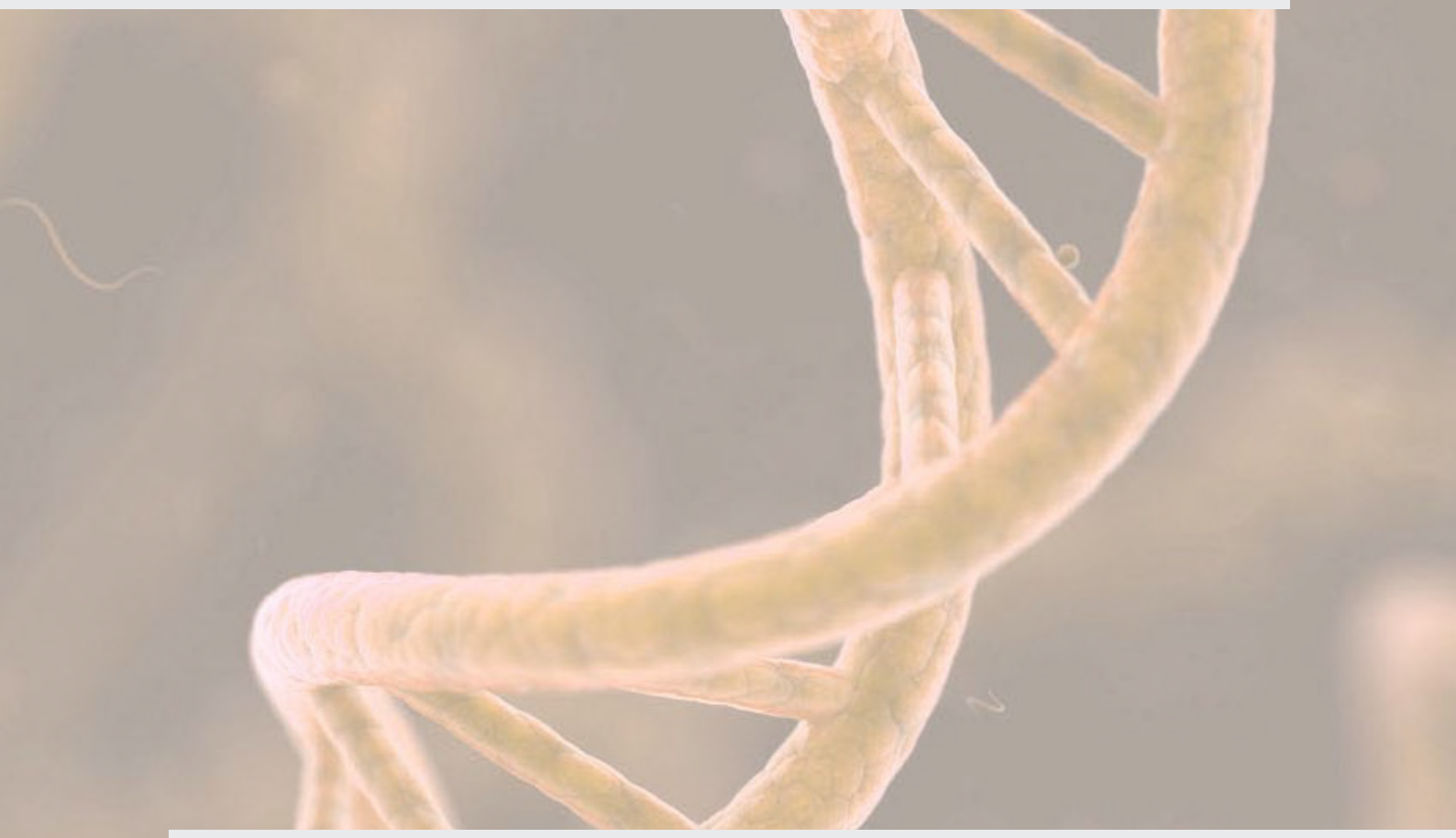
Tsvetan Panagonov
Dev Lead
Years of experience: 8

Prof. Roumen Pankov, Ph. D.
Molecular Biology,
Cell Biology
Sofia University "St. Kliment Ohridski"
Bulgaria

Sider Penkov, Ph. D.
Molecular Biology and Genetics of Model Organisms
Kurzchalia laboratory
Max Planck Institute of Molecular Cell Biology and Genetics
Germany

We are still in **BETA**.

The modules and functionalities not yet available will be released according to BioSeek's roadmap in Q1 2019.



BIOSEEK BETA
RESEARCH. DISCOVER. COLLABORATE

rossen@bioseek.eu
+359 88 428 8336
victoria@bioseek.eu
+359 98 884 7228

LinkedIn
Facebook

www.bioseek.eu