

novo|seek

Accelerating searches of research grants and scientific literature with novo|seekSM

Bioalma's novo|seek is a powerful alternative to traditional literature searching on the web. Designed to help scientists with their research projects and proposals within the life science community, novo|seek is a dynamic information extraction system for quickly identifying key concepts found in published papers. Here we demonstrate how a single concept can initiate complete exploration of scientific literature, current research grants and external web-based resources.

With thousands of scientific journals, thousands of current research grants and millions of abstracts in Medline, it is difficult for any grant-writing scientist to data-mine all of the resources that will provide support for his research projects. There is clearly a need for an improved alternative to performing time-consuming literature mining and visiting multiple, unrelated websites. We developed novo|seek specifically to address these bottlenecks and help scientists derive important concepts from all of the available abstracts while making the results easy to explore.

The novo|seek search box interface allows the user to enter a term (or a collection of terms), which the system automatically recognizes as a disease, gene, chemical substance, tissue type, scientific technique and others. The system returns an interactive web page that enables the user to navigate through the many related abstracts (Fig. 1).

Scientific abstract searching

Using novo|seek, we explored the previous scientific literature and current research grants for the arachidonate 5-lipoxygenase-activating protein (ALOX5AP or FLAP). The *ALOX5AP* gene product has been associated with regulating the production of substances that cause inflammation, such as leukotrienes¹.

When we entered ALOX5AP as a search term, novo|seek returned a results page with a list of nearly 300 abstract titles, the oldest from 1982. Additionally, novo|seek provided interactive tabs, links, menus and columns to explore (Fig. 2).

Four small icons enabled us to choose how we preferred to view the list of abstract titles. The choices included a simple view (title, journal and author), a snippet view (a simple view and the search term with flanking words), a sentence view (a simple view and the search term relative to a sentence) and a traditional abstract view.

Ramón Alonso Allende

Bioalma, Centro Empresarial Euronova, Ronda de Poniente, 4, 28760 Tres Cantos, Madrid, Spain. Correspondence should be addressed to R.A.A. (allende@bioalma.com).

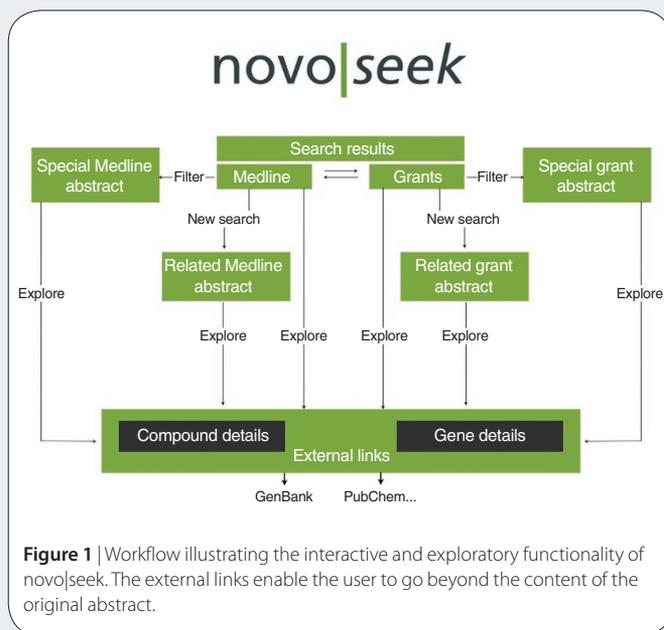


Figure 1 | Workflow illustrating the interactive and exploratory functionality of novo|seek. The external links enable the user to go beyond the content of the original abstract.

In all cases, ALOX5AP and its synonyms were highlighted automatically in each view. To highlight disease terms in these views, we expanded the highlight panel and selected Diseases and Syndromes. The highlighting options enabled us to quickly identify the disease terms in the published abstracts.

The highlighted terms in novo|seek abstracts not only are designated with different colors but are also interactive. When selecting the highlighted ALOX5AP term, a new web page appeared that permitted us to explore more about the gene. There were useful external web links to genomic data, sequences, pathways, microarray identifiers, structural domains and ontologies, all on a single page.

The main result page had a side column of related concepts that enabled us to visualize which terms frequently appear in conjunction with ALOX5AP. The concepts are grouped as follows: diseases or

APPLICATION NOTES

Figure 2 | Screenshot of ALOX5AP results page generated by novo|seek search. The expanded highlight panel shows the available bioentity types. Genes and proteins are highlighted in red, and diseases and syndromes are highlighted in green. The highlighted terms are interactive and give additional information when selected.

syndromes, pharmacological substances, genes and proteins, signs and symptoms, chemical substances, organisms, organs and body parts, tissues, cell components, biological functions, and procedures and techniques. Some of the top concepts associated with ALOX5AP identified by novo|seek were asthma, arachidonic acid, MK-886, aspirin intolerance, leukotriene, lung tissue, leukotriene synthesis and zymosan. For many years, scientists knew that ALOX5AP was associated with inflammation and asthma, but only in the last decade has there been correlation with pulmonary conditions. After expanding the diseases and syndromes list, we chose to explore ischemic stroke. At the time of writing, the number of abstracts that included both ALOX5AP and ischemic stroke was only a dozen, and the oldest paper was published in 2005.

In addition to exploring related diseases, we also used novo|seek to explore the various pharmacological substances listed in the concepts column. Some of the identified substances included MK-886, MK-591, BW A4C, WY 14643, SKF104353, tepoxalin and AA861. Specifically, clicking on the MK-886 term reduced the number of Medline abstracts to less than 20, and it revealed only one grant abstract. Additional filtering of the term to MK-591 resulted in two abstracts. These abstracts contained the terms FLAP (an alias of ALOX5AP), MK-886 and MK591. Selecting MK-886 and MK-591 in novo|seek allowed us to navigate to NCBI's PubChem and visualize the compounds (**Fig. 3**).

Next we searched for published reports describing sequence variants associated with ALOX5AP and disease conditions. We used the following query statement to find 31 research abstracts: "ALOX5AP AND (SNP or polymorphism)". The publication dates for the articles ranged from 2003 to 2009. When we selected the term *mus sp.* from the filter by concepts side column, a single Medline abstract

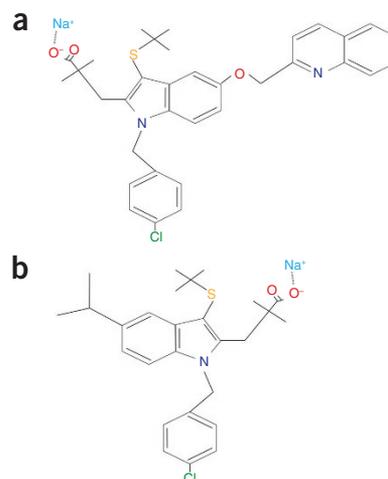


Figure 3 | Chemical structures downloaded from the PubChem website. (a) MK-591. (b) MK-886.

was presented. The identified abstract reports how ALOX5AP gene polymorphisms may be associated with Alzheimer's disease².

Grant abstract searching

After exploring the numerous published abstracts related to ALOX5AP, we continued our search for additional information by selecting the GRANTS link in the results page. Similar to the Medline results, the GRANTS web page included highlighted terms and additional interactive views, menus and links to other related grants. Specifically, the grants focused on studying how FLAP is related to lung inflammation and pulmonary fibrosis. When we selected the Diseases and Syndromes bioentity, the phrase "idiopathic pulmonary fibrosis (IPF)" was highlighted, and a new search revealed over 300 grants that mention this disease. When we selected a grant title, a new web page revealed more details about the grant. Moreover, the web page had a link to SciSearch.

Summary

novo|seek is a powerful tool that researchers studying specific diseases can use to expedite the search for related drugs and genes in published abstracts. It can help scientists discover and explore previously unrecognized relationships in scientific literature that are not easily revealed with typical information retrieval approaches.

1. Dixon, R.A. *et al.* Requirement of a 5-lipoxygenase activating protein for leukotriene synthesis. *Nature* **343**, 282–284 (1990).
2. Manev, H. & Manev, R. 5-Lipoxygenase (ALOX5) and FLAP (ALOX5AP) gene polymorphisms as factors in vascular pathology and Alzheimer's disease. *Med. Hypotheses* **66**, 501–503 (2006).

This article was submitted to *Nature Methods* by a commercial organization and has not been peer reviewed. *Nature Methods* takes no responsibility for the accuracy or otherwise of the information provided.