

Executive Viewpoint | GenomeQuest, Inc.

You Really Should See Your Lawyer's Software

Unexpected value for research biologists awaits within sequence search solutions developed to meet the needs of patent attorneys and other intellectual property experts. Many researchers regard patents and intellectual property issues from a distance, as these concerns fall outside the scope of hypothesis formation, data production, laboratory problem solving, and other typical elements of scientific investigation. That's why they're often surprised to discover that the intellectual property experts working at their own institutions are using a unified sequence data, search, and analysis solution with an enviable selection of analysis options, a powerful interface, a fast, automated workflow, and an automatically updating database containing more sequences than Genbank, more annotations than RefSeq, and sequence data that the researchers themselves have produced. This year, GenomeQuest, Inc. is making a deliberate effort to tell biologists about these resources and how to access them at reasonable cost. At the same time, we're updating our GenomeQuest™ sequence search solution with features of specific interest to biologists, such as multiple sequence alignment and a new, web-enabled platform that supports research collaboration.

Ron Ranauro
President & CEO
GenomeQuest, Inc.

Tap into Premium Sequence Search Tools and Content

Patent sequence search is a critical element of pharmaceutical and biotech intellectual property (IP) portfolio management. GenomeQuest, Inc. has captured the leadership position in the market for patent sequence search solutions by providing a powerful and popular sequence content, search, and analysis platform for the specialized needs of patent attorneys and other IP professionals. Those specialized needs include the ability to search many more sequences and annotations than those available in GenBank and other public databases, the ability to search by sequence similarity rather than by homology, the ability to filter sequence similarity results according to detailed annotation and sequence composition criteria, and a fast, straightforward workflow construction. Each of GenomeQuest's answers to these needs is also of substantial utility to research biologists, and in many cases ought to be used as a supplement to, or in place of, their existing research databases and software solutions.

Most research biologists are already painfully aware that GenBank and other publicly accessible sequence databases are incomplete, out-of-date, and inconsistently annotated. But unless they've made a specific effort, those biologists might not know just how much more sequence and related information has already been recorded, re-annotated, indexed, and made available to patent law professionals. GenomeQuest's GQ-PAT database, developed to enable faster, more complete sequence searches, contains more patent number-sequence pairs than GenBank and more annotations than RefSeq. Intellectual property issues aside, the GQ-PAT

database is a vast resource of biological sequence research information, much of which is absent from public sequence databases and too expensive to collect or access from other sources.

Popular search algorithms, such as Smith-Waterman and BLAST, are of great value to research biologists, who use them to uncover potential ancestral relationships between sequences. However, these heuristic algorithms often do not do a very good job of answering questions unrelated to homology. The frustration of using heuristic algorithms for short sequence searching is well

ROI

Savings directly attributable to GenomeQuest are easy to calculate. The data component alone eliminates costs of acquiring new sequence data, database rebuilds, annotation, indexing, and multiple database searching. Payback for GenomeQuest usually occurs within a few weeks.

known to biologists who have tried tweaking and re-tweaking BLAST parameters in an effort to answer a seemingly straightforward sequence similarity question. Because patent claims are based on sequence similarity, rather than homology, legal experts rely on a different algorithm, GenePAST. The GenePAST algorithm searches according to percent identity between sequences, rather than the likelihood of finding homologous relationships. This makes the GenePAST algorithm perfect for short sequence searching projects,

such as probe and motif analysis. While BLAST returns varied and confusing short sequence search results, GenePAST delivers correct, unambiguous answers in minutes.

Analyzing search results is a significant aspect of any sequence search project. Even a carefully thought out sequence search query can produce a report that goes on for one thousand pages or more. Manual sorting and sifting through these reports for sequences of actual research interest brings laboratory workflow and research productivity to an agonizing crawl. From an intellectual property portfolio management perspective, a delay of this kind could have a serious negative impact on portfolio value. And in law, the consequences of human error in sequence search results analysis are unacceptable. To address this need, GenomeQuest developed rapid, powerful search results filtering. Complex Boolean filtering criteria are easily constructed by a GenomeQuest software interface featuring drop-down

menus and text-entry boxes. GenomeQuest reduces thousand-page search result reports down to the few sequences of interest in minutes.

Life science researchers and biological IP experts often ask the same kinds of sequence search questions, although for different reasons. Researchers are understandably more interested in scientific problems than

legal ones, but they have much to gain by expanding their repertoire of skills to exploit resources developed for biological sequence patent law. Applied to research, sequence search solutions developed for intellectual property portfolio management can yield faster, more complete answers to research questions and increase life science productivity.

There is a wealth of information about genes in patents.

1) Gene Expression data are typically first published in patents.

- a. Patent application number US20070141587 (Gene expression profiling in biopsied tumor tissues). Apart from BRCA2, expression levels of more than 20 genes including hENT1 and GATA3 are indicative of long term survival after breast cancer surgery.
- b. Patent application number US20070072209 (Methods of treatment and diagnosis of Kaposi's sarcoma (KS) and KS related diseases). Through expression profiling, the orphan G-protein coupled receptor CXCR7 is implicated in Kaposi's Sarcoma.

2) Chemical inhibitors of genes are usually first published in patents.

- a. Patent application US20070142388 (Inhibitors of Akt activity). 2,3-diphenylpyrazine containing chemicals can inhibit isoforms of the Akt gene.
- b. Patent application US20070032519 (Compounds modulating c-kit and c-fms activity and uses therefor). New families of chemicals targeting c-kit and c-fms.

3) Patents are the best indicators of competitive landscape. Companies actively researching a given gene are readily accessible as assignees of patents.