

## Use case

# Comprehensive competitive intelligence monitoring in real time



## Comprehensive competitive intelligence monitoring in real time

To remain competitive, pharmaceutical companies must keep up to date with scientific research relevant to their business. This typically involves individuals or review teams routinely scanning a range of sources such as the biomedical literature, patents, publicly available grant information and a range of biotech-focussed news websites in an attempt to identify articles of interest.

Manually searching a range of sources with multiple keywords or phrases covering all the therapeutic areas or targets of interest is resource intensive, which constrains the number of different sources that can be scanned, the frequency at which the scanning can be performed and the depth of review possible for potentially interesting articles.

The exponentially growing amount of literature and increasingly diverse range of sources make it almost impossible to maintain a comprehensive and up-to-date understanding. The result is that the legacy approach to literature scanning is no longer practical.

SciBite enables Pharmaceutical companies to automate this process and cast a wide net to continuously monitor a comprehensive set of data sources and be alerted when interesting scientific and medical advances are published.

### **SciBite: automating competitor intelligence scanning**

#### **Semantic enrichment of unstructured data**

Automating the process of scanning multiple sources of data is challenging because terms or phrases of interest can be spread out in an article and different authors use different terminology when describing the same thing.

At the core of the SciBite platform are the established controlled vocabularies which apply an explicit, unique meaning and description to scientific terms. Comprising tens of millions of synonyms, SciBite's manually curated vocabularies, or VOCabs, have unrivalled depth and breadth, ensuring comprehensive coverage of relevant terminology and providing the robust foundation necessary for an effective and impactful literature monitoring strategy.

SciBite's ontology management platform, CENTree, enables organisations to maintain up to date ontologies representing

evolving scientific language. Subject matter experts can easily contribute to keeping things current and augment our manually curated standard reference vocabularies (such as Cell, Gene and Indication), with bespoke terms, including those relating to novelty and specific technologies such as biomarkers and diagnostics. This enables SciBite to cope with the evolving language in new scientific fields, such as "CRISPR", "10x Genomics" and "Drop-Seq". CENTree also leverages machine learning techniques to suggest potential new terms that are similar to other related words because they are used in a similar context.

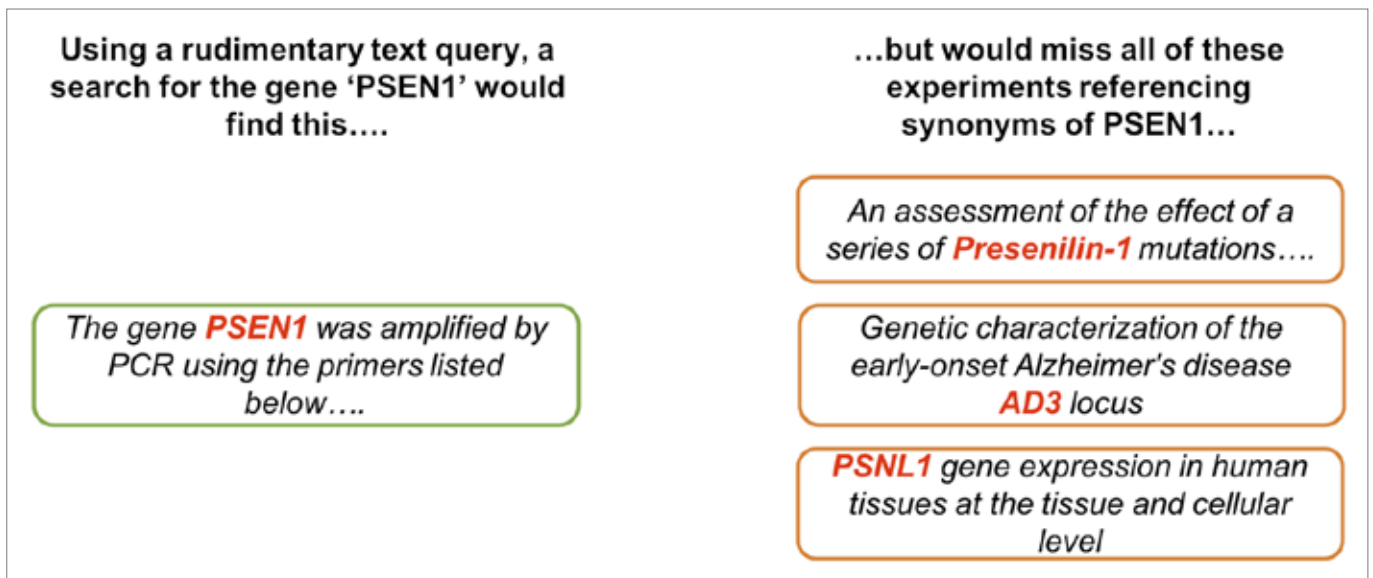
When used in combination with CENTree, SciBite's Named Entity Recognition (NER) engine, TERMite, applies standard, well established ontologies and controlled vocabularies to scientific documents, generating a semantic index and endowing text entries with an explicit, specific meaning. This transforms unstructured document text into a structure that can be queried in a simple fashion to answer questions that would otherwise require time-consuming, error-prone manual work.



**Figure 1:** Examples of different phrases used to reflect novelty

### Smart, comprehensive searches

Most literature searches are typically limited to finding the specific terms used by the author. For example, a literature search for the Alzheimer's related gene, PSEN1, would miss references to synonyms such as Presenilin-1, AD3 and PSNL1.



**Figure 2:** Literature search tools miss synonyms of search terms of interest

Through semantic enrichment, SciBite ensures that all relevant data is found, regardless of which synonym is used as the search term. Users can create specific searches containing multiple relevant terms and entities forming so called search patterns.



## Summary

**Diagnosis** of lethal or prenatal-onset autosomal recessive disorders by parental exome sequencing. Summary **Rare** genetic disorders resulting in prenatal or neonatal death are genetically heterogeneous but **testing** is often limited by the availability of fetal DNA, leaving couples without a potential prenatal **test** for **future** pregnancies. **We describe** our novel **strategy** of exome sequencing parental DNA samples to diagnose recessive monogenic disorders in an audit of the first 50 couples referred. Exome sequencing was carried out in a consecutive series of 50 couples who had one or more pregnancies affected with a lethal or prenatal-onset disorder. In all cases there was insufficient DNA for exome sequencing of the affected fetus. Heterozygous **rare** variants (MAF < 0.001) in the same gene in both parents were selected for analysis. Likely disease-causing variants were **tested** in fetal DNA to confirm co-segregation. Parental exome analysis identified heterozygous pathogenic (or likely pathogenic) variants in 24 different genes in 26/50 couples (52%). Where two or more fetuses were affected a genetic **diagnosis** was obtained in 18/29 cases (62%). In most cases the clinical features were typical of the disorder, but in others they **result** from a hypomorphic variant or represent the most severe form of a variable phenotypic spectrum. We **conclude** that exome sequencing of parental samples is a powerful **strategy** with high clinical **utility** for the genetic **diagnosis** of lethal or prenatal-onset recessive disorders.

## Summary

Diagnosis of lethal or prenatal-onset autosomal recessive disorders by parental exome sequencing. Summary Rare genetic disorders resulting in prenatal or neonatal **death are genetically heterogeneous but testing is often limited by the availability of fetal DNA, leaving couples without a potential prenatal test for future** pregnancies. **We describe our novel strategy** of exome sequencing parental DNA samples to diagnose recessive monogenic disorders in an audit of the first 50 couples referred. Exome sequencing was carried out in a consecutive series of 50 couples who had one or more pregnancies affected with a lethal or prenatal-onset disorder. In all cases there was insufficient DNA for exome sequencing of the affected fetus. Heterozygous rare variants (MAF < 0.001) in the same gene in both parents were selected for analysis. Likely disease-causing variants were tested in fetal DNA to confirm co-segregation. Parental exome analysis identified heterozygous pathogenic (or likely pathogenic) variants in 24 different genes in 26/50 couples (52%). Where two or more fetuses were affected a genetic diagnosis was obtained in 18/29 cases (62%). In most cases the clinical features were typical of the disorder, but in others they result from a hypomorphic variant or represent the most severe form of a variable phenotypic spectrum. We conclude that exome **sequencing of parental samples is a powerful strategy with high clinical utility for the genetic diagnosis** of lethal or prenatal-onset recessive disorders.

**Figure 3:** Comparison of a search using keywords or dictionary terms and a search using the same keywords and terms in specified patterns

To accommodate the reality that there are often multiple ways of describing an outcome of interest, multiple patterns can be aggregated into 'bundles' which can be run across the same data simultaneously. SciBite calculates

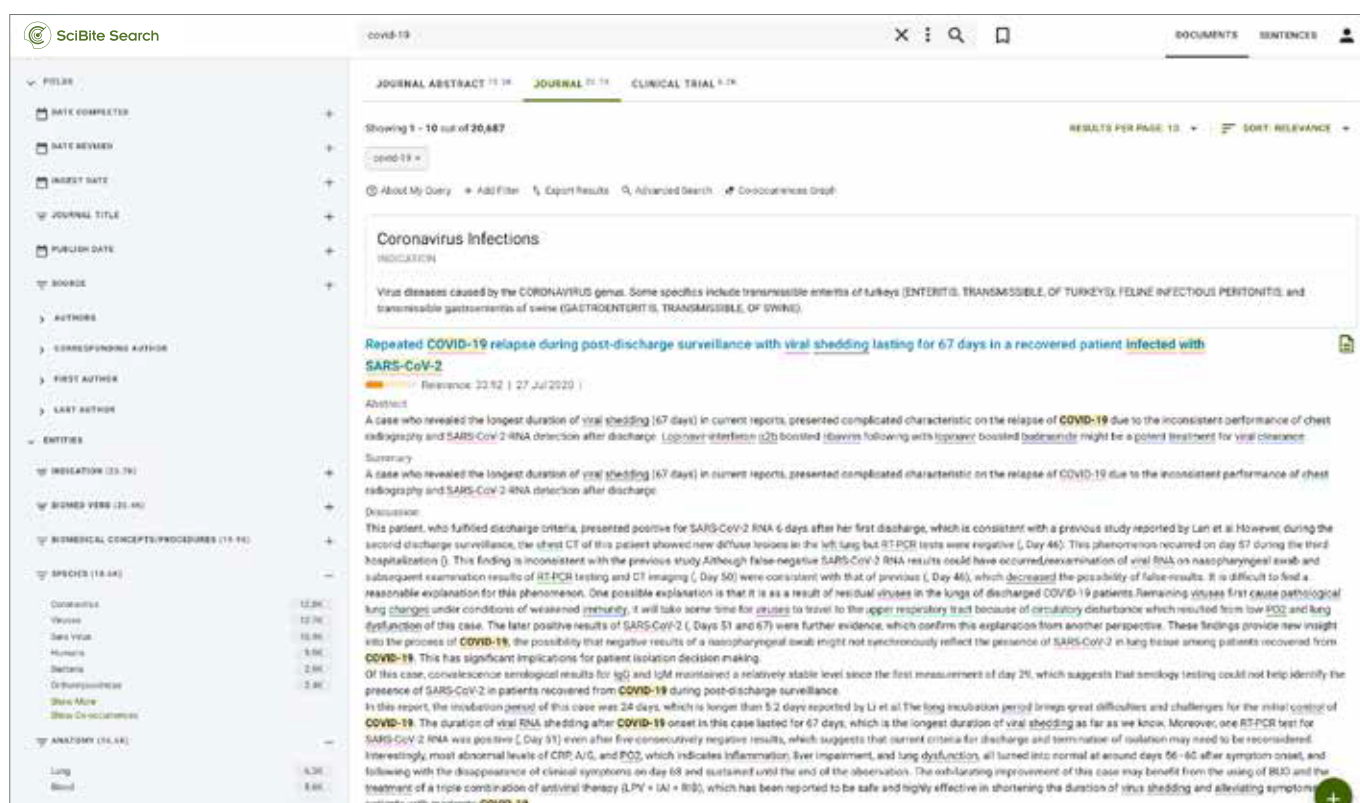
a score for a piece of text based on how many patterns match and whether those matches are complementary or competing, providing an incredibly powerful yet clear method to identify relevant data.



**Figure 4:** Example of a bundle of search patterns

Search patterns and bundles can be created, tested and refined quickly, easily and transparently. This flexibility enables users to remain responsive to the changing terminology and language used in a particular area of interest and support a wide range of different use cases and search strategies.

SciBite instantly marks up the text of an article so that reviewers can easily interpret it, rapidly identify the topics covered and instantly get a feel for what it is about. This enables them to derive more value from the reading experience and focus on understanding interesting findings rather than searching for them.



**Figure 5:** SciBite Search marks up all relevant terms within an article to facilitate interpretation

By identifying the key themes referenced in a document, SciBite can automate the process of categorizing and distributing documents proactively notifying users when there is new material of interest.

### Gain a holistic view and identify trends sooner

SciBite is not limited to publicly available data, users can also apply semantic enrichment to internal and third party data sources such as internal document repositories<sup>1</sup>, ELNs<sup>2</sup>, patents and commercially available databases<sup>3</sup>. Since the resulting data will be as well structured and interoperable as public data, it becomes facile to integrate multiple

disparate sources and gain a holistic view of everything that is known, both internally and externally, about any compound, target or disease. SciBite provides users with a single, consistent and simple user interface, that enables them to ask questions across data sources that would have otherwise been time consuming or impossible to answer.

The results of such analyses deliver valuable business insight and enable companies to understand the research landscape, such as identifying trending topics, and reveal how the volume of information associated with the current 'hot topic' has changed over time.

1 See SciBite's publication 'Unlock the Full Potential of Departmental Scientific Documents', for details

2 See SciBite's publication 'Unlock the Full Potential of ELN Data', for details

3 See SciBite's library for examples of semantic enrichment of the biomedical literature



It becomes straight forward for a company to define its internal strengths and to assess the competitive landscape relating to specific targets or diseases. Similarly, it is possible to identify which companies or institutions are working in which disease or technology area of interest to explore options for collaboration or acquisition. In each case, alerts can be setup to ensure information is highlighted to the right people in a timely manner.

### **Summary**

Most Pharmaceutical companies struggle to maintain an up-to-date awareness of the latest biomedical research relevant to their own therapeutic programmes. To ensure competitiveness, they must move on from the prevailing manual approach involving the time consuming, piecemeal review of a small range of data sources.

SciBite reduces both the time and uncertainty involved in evaluating the vast body of research and news to track trends, gain early insight into potentially groundbreaking scientific advances. SciBite provides a resource-effective solution to enable Pharmaceutical companies to unlock unique findings hidden in unstructured text and efficiently and comprehensively monitor a wide range of data sources and derive valuable scientific and business insights.

### **About SciBite**

SciBite is an award-winning semantic software company offering an ontology-led approach to transforming unstructured content into machine-readable clean data. Supporting the top 20 pharma with use cases across life sciences, SciBite empowers customers with a suite of fast, flexible, deployable API technologies, making it a critical component in scientific data-led strategies. Contact us to find out how we can help you get more from your data.

To learn how SciBite can unlock the value of your data, speak to one of our experts today or email us at [contact@scibite.com](mailto:contact@scibite.com)

SciBite's data-first, semantic analytics software is for those who want to innovate and get more from their data. At SciBite we believe data fuels discovery and we are leading the way with our pioneering infrastructure that combines the latest in machine learning with an ontology-led approach to unlock the value of scientific content. Supporting the world's leading scientific organisations with use-cases from discovery through to development, SciBite's suite of fast, flexible, deployable API technologies empower our customers, making it a critical component in scientific, data-led strategies. Contact us to find out how we can help you get more from your data.

To learn how SciBite can unlock the value of your data, speak to one of our experts today or email us at [contact@scibite.com](mailto:contact@scibite.com)

**Head Office:**

SciBite Limited  
BioData Innovation Centre  
Wellcome Genome Campus  
Hinxton, Cambridge CB10 1DR  
United Kingdom

 [www.scibite.com](http://www.scibite.com)  
 [contact@scibite.com](mailto:contact@scibite.com)  
 LinkedIn: SciBite  
 Twitter: @SciBite  
 +44 (0)1223 786 129

