Linguamatics

Background

Syntaxin Ltd is a biotechnology company that has developed a proprietary technology to create novel recombinant proteins termed Targeted Secretion Inhibitors (TSI). These engineered molecules are selectively able to deliver an endopeptidase into defined target cells and specifically cleave SNARE proteins driving secretion from that cell. This technology is based upon the endopeptidase activity found within clostridial neurotoxins, which cleaves SNARE proteins and inhibits vesicular cell secretion.
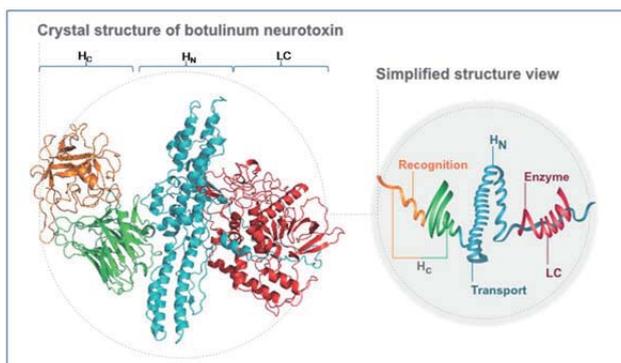
For a biotechnology company with focused resources, using target and disease information from the existing scientific literature is an important aspect of identifying new therapeutic opportunities. Syntaxin's TSIs have a unique mode of action, therefore a combined knowledge of disease, tissue, receptor, SNARE protein and secreted mediator is required to identify new therapeutic opportunities - either for existing TSIs or for new TSI molecules. Stand alone text queries would not pro-vide the breadth or depth of analysis required within the specified time frames. Having access to advanced text mining capabilities to make target selection decisions was identified as a key informatics technology to integrate into Syntaxin's selection process.

This case study outlines how combined text mining queries via I2E allowed an informed disease selection process to be implemented.

Challenge

A key challenge for Syntaxin's informatics team was to be able to interrogate the breadth of scientific literature linking key aspects related to the TSI molecule function. Particularly when focusing on cell secretion mechanisms and disease information, the scientific literature provides a richer knowledgebase than many genomic-based resources. Having traceable literature statements to validate therapeutic hypotheses was also an important aspect for Syntaxin. Additionally as the analysis moved from disease area through to secretion, SNARE protein and cell type, a key requirement was to be able to link queries to provide a view of the available disease space.

I2E showed it could provide organised, structured and comprehensive analysis of the large corpus of scientific literature, linking overlapping queries together. I2E offered the level of usability that Syntaxin required, from basic queries through to advanced text mining approaches supported by a responsive team at Linguamatics.

"Rapid identification of new therapeutic opportunities is a key activity in deploying a new platform technology such as Syntaxin's TSIs. Applying an advanced text mining approach via I2E allows SME sized companies to access the literature at scale and rapidly define targeting and therapeutic approaches for their technology. Having I2E and now I2E OnDemand really puts text mining approaches within reach of the biotech sector allowing the types of data analysis to be carried out that have largely been the domain of pharmaceutical companies" - Keith Foster, Founder & Chief Technology Officer, Syntaxin

Importantly for Syntaxin's needs, traceable statements could be easily derived and shared with co-workers via common formats such as Microsoft Excel spreadsheets.

The first query challenge was to look at repositioning of a current TSI molecule based upon the receptor it uses for cellular entry and secretions targeted by the endopeptidase domain. The goal was to identify other disease areas and hence therapeutic opportunity for that class of TSI.

Solution

An advanced text mining approach via Linguamatics I2E was identified as a core activity to rapidly, and with focused resource, mine the scientific literature for new therapeutic opportunities focused on inhibiting cell secretion. Combining the breadth of MEDLINE with focused ontologies derived from EntrezGene, Panther, MeSH and SNOMED meant large many-to-many term queries could be constructed with ease and results returned for analysis within minutes. Having structured output linking to key statements meant the results could be validated quickly and searches refined and combined with the next aspect of TSI biology that Syntaxin wanted to explore.

The initial query used five key genes in a signalling pathway of interest. I2E utilizes multiple gene synonyms via EntrezGene and these could easily be combined with MeSH and SNOMED disease terms to provide an in depth focused dataset of over 150 disease terms. Further filtering allowed them to exclude disease where the gene terms played a positive role focusing down to around 60 diseases of interest. Further selection was made based on expression of a key receptor establishing 11 disease types for further exploration.

The disease terms could then be combined with lists of query terms focused on secreted mediators, both protein and small molecule based. Moving from query to query allowed Syntaxin to build up a picture of pathway, disease, cell surface receptors and secreted mediators. Importantly, the researchers could validate the associa-

tions through direct links into the underlying text. Having the output in a familiar format such as Microsoft Excel meant results could be easily shared in the organisation.

Based on the results obtained through I2E, a list of disease target areas were identified linked to associated secretions. This formed the basis of a laboratory based programme.

Summary

I2E has provided Syntaxin with the capability to span the required knowledge domains at scale and effectively with a focused informatics team – a key requirement in a resource constrained biotech environment. I2E searches are now routinely used to explore new areas of biology and opportunities for the company.

For more information or a demonstration, please call us today
☏ +44 1223 421360
email enquiries@linguamatics.com