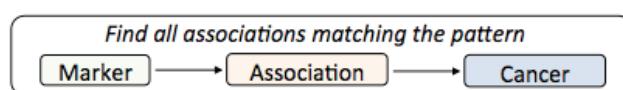




## Datasheet

### Identifying semantic patterns in text



[World J Surg Oncol.](#) 2007 Feb 7;5:17. Prognostic implications of immunohistochemically detected YKL-40 expression in breast cancer. Kim SH<sup>1</sup>, Das K, Noreen S, Coffman F, Hameed M.

#### BACKGROUND:

YKL-40 has been implicated as a mediator of collagen synthesis and extracellular matrix re-modeling as well as mitogenesis. Elevated serum levels of YKL-40 have been associated with worse survival in a variety of malignancies including breast cancer. We wished to determine if immunohistochemically detected expression had prognostic implications in breast cancer....

...YKL-40 expression was noted in 37 patients (34%). YKL-40 immunoreactivity significantly correlated with larger tumor size, poorer tumor differentiation, and a greater likelihood of being estrogen and/or progesterone receptor negative. .... In multivariate analysis, YKL-40 status was independent of T-stage and N-stage in predicting disease recurrence.... Immunoreactivity for YKL-40 was a significant predictor of breast cancer relapse in this subset of patients. This was independent of T or N-stage and suggests that tumor immunohistochemistry for this protein may be a valuable prognostic marker in breast cancer.

- **CUSTOMIZABLE:** identifies a huge variety of patterns from text
- **POWERFUL:** use “bundles” of complementary patterns to cover all possibilities
- **DYNAMIC:** optimise your patterns on the user interface before using the API

#### ‘Close, but not close enough’

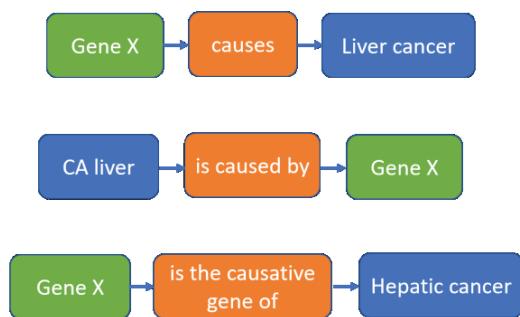
Imagine you are interested in gene X and disease D, and would like to look for articles describing the gene’s expression in disease cases. In a search box, you type in the gene’s name, the disease term and the word “expressed”, but disappointingly the results only list articles in which the terms merely co-occur in any order. Sounds familiar?

Biomedical text often describes entities (e.g. genes, diseases) in relation to one another (e.g. causation, repression, treatment). To aid knowledge discovery, such patterns of entities and relationships need to be captured semantically. Wouldn’t it be nice if a semantic search engine “understands” your query and returns hits, e.g. “gene X is highly expressed in disease D”?

#### Relationships matter

TExpress is an add-on module of TERMite, SciBite’s ultra-fast named entity recognition (NER) engine, to search for patterns of entities and verbs in text. Thanks to our flagship library of manually curated vocabularies behind TERMite, TExpress users can tap into the extensive coverage of entities or verbs and their synonyms to uncover less obvious matches.

#### Many ways to say the same thing





## TExpress Pattern Builder: create custom lexical patterns

Current Pattern `:(GENE) :(0,5) :(INDICATION)`

Add: **Type** | **Individual(s)** | **Spacer** | **Word Bag** | **Taxonomy** | **Word List** | **Regex** | **Catcher**

### Defining your own patterns

tExpress, which stands for “TERMite Expressions”, screens the entities that TERMite extracted by default with your pattern of interest. To define your first pattern, try the graphical pattern builder on the TExpress interface.

For data scientists who are familiar with regular expressions and JSON, once you are comfortable with the syntax, you can try creating patterns programmatically too.

### Multiple patterns? Use bundles!

To cover the variance in how relationships between entities can be described in text, a complementary collection of patterns is more effective than a standalone pattern. In TExpress, such collections can be saved neatly as “bundles”, which can be reused in future queries.

Ready-to-use bundles for biomarker identification, adverse event detection and phenotypic extraction are included to get you started.

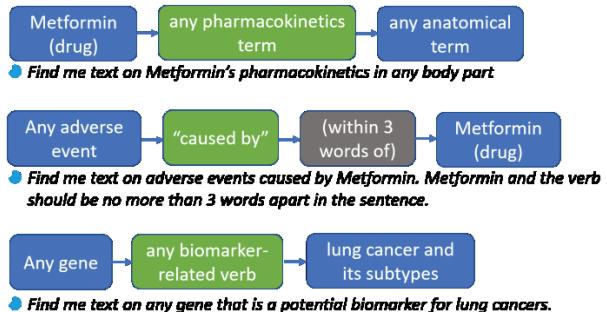
### Dynamic and fast

TExpress owes its speed and great deployment flexibility to the back-end TERMite engine. You can access TExpress via the user interface, the command line, or your own application by firing API calls, just like you would with TERMite.

Our customers particularly enjoy the dynamic user experience. To adapt to changing business requirements, they often experiment with patterns on the user interface before rolling them out to a large corpus, and run various custom post-processing scripts on the TExpress output as part of the TExpress API call.

Whatever your use case, from simple sentence co-occurrence search to detecting specific events, try TExpress today and new insights could be just a pattern or bundle away!

### Examples of Patterns TExpress Can Find



### About SciBite

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

#### 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