

# Understanding drug toxicity through gene expression

Researchers based at the Science and Technology Facilities Council (STFC) Hartree® Centre worked with Syngenta - a leading agricultural and biotechnology company - to design bioinformatics tools for high throughput screening of candidate compounds.

## Challenge

Assessing the toxic potential of a molecule is an essential element in drug discovery. Toxicity, an undesirable effect caused by a drug at certain dose and time can be defined by a set of observable characteristics – phenotypes – caused by changes in underlying gene expression patterns. Assessing toxicity requires a deep understanding of the underlying gene expression. Toxicogenomics - an interdisciplinary, data-driven field uses various technologies to measure gene signatures. An emerging, cheap, rapid, high-throughput technology - L1000 - has the potential to become the standard method of generating gene expression profiles for toxicogenomic studies, allowing mass screening of molecules at a lower cost. Syngenta were interested in evaluating the potential of the L1000 platform to identify related data-driven methods that could influence future work in investigative and predictive toxicity.

## Approach

Researchers first developed a bioinformatics pipeline to automate the processing of L1000 data sets and went on to perform pathway enrichment, translating high dimensional gene expression profiles to relevant physiological processes to drug toxicity. This was followed by using high-dimensional network biology techniques enabling the translation of genotypes to phenotypes. Data from the L1000 platform was coupled with machine learning pipelines to generate in silico predictive models that could classify potential toxicity using gene expression profiles from cell lines treated for a range of candidate compounds.

## Benefits

Completed as part of the Innovation Return on Research (IROR) programme, a collaboration between STFC and IBM Research – the tools and techniques developed provided Syngenta with a fresh perspective regarding the adoption of data driven, machine learning-based techniques for further exploitation of L1000 datasets in their research programme, helping establish building blocks that can be continuously refined as Syngenta expands their library of L1000 datasets in future.

“ These tools provide the basis of the future development of our approach to interpret high-throughput transcriptomics data, helping us accelerate innovation of new technologies that help farmers to sustainably produce the food, feed, fuel, fibre, and related inputs society needs. ”

Richard Currie  
Syngenta

[www.hartree.stfc.ac.uk](http://www.hartree.stfc.ac.uk)

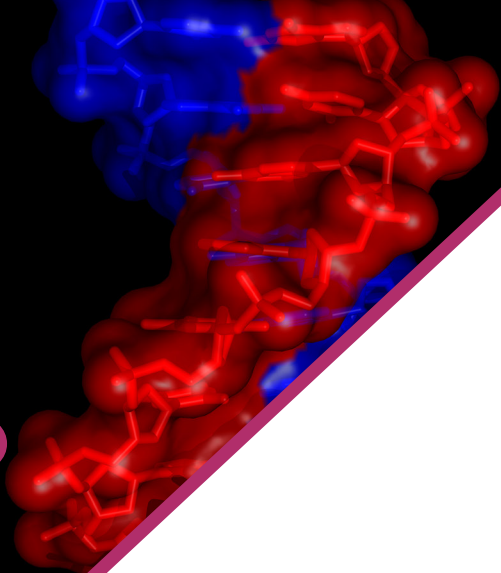
 @hartreecentre

 /company/stfc-hartree-centre

Partners







“ Our predictive toxicology exploratory programme looks at the use of high-throughput transcriptomics to help predict the toxicity of new crop protection chemicals.

There were no off-the-shelf solutions to analyse the high throughput data and provide insights we need. Researchers at STFC Hartree Centre demonstrated the ability of computational bioinformatic and machine learning approaches to combine transcriptomics, high dimensional network biology and chemical structure data to predict toxicity. ”

Richard Currie  
Syngenta

## At a glance

- Set up compute infrastructure to understand toxicity through gene expression datasets.
- Tools for bioinformatics analysis, pathway enrichment and translation of genotype to phenotype.
- Tools for exploiting natural language processing (NLP) techniques for knowledge discovery.
- Used machine learning for classifying drug toxicity using genomics and chemical structure.

## Who we are

- 60+ computational scientists and technologists
- World-leading supercomputing and AI infrastructure
- Bespoke small teams built around your project
- Tailored business development support
- Access to our network of industry, academic and technology partners

## What we do

- Boost productivity and enhance innovation for industry
- Big data analytics and artificial intelligence (AI)
- High performance computing and quantum simulation
- Training and skills development
- Insights into emerging technologies



## Our impact on UK industry and society

The Hartree Centre was created by UK Government to transform industry by accelerating the adoption of high performance computing (HPC), big data analytics and artificial intelligence (AI) technologies. We play a key role in realising UK Government's Industrial Strategy by stimulating applied digital research and innovation, creating value for the organisations we work with and generating economic and societal impact for the UK.

The Science and Technology Facilities Council (STFC) Hartree Centre is part of UK Research and Innovation.

[www.hartree.stfc.ac.uk](http://www.hartree.stfc.ac.uk)

 @hartreecentre

 /company/stfc-hartree-centre

 [hartree@stfc.ac.uk](mailto:hartree@stfc.ac.uk)