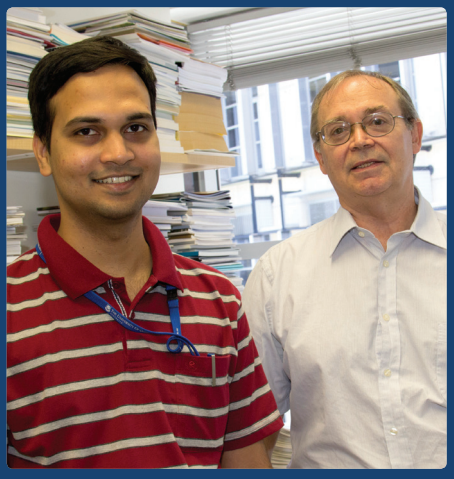
A computational approach to cancer

Professor Mark Ragan and **Dr Sriganesh Srihari** explain the intricacies of their work computationally assessing the gene regulatory networks involved in cancer initiation, as well as the importance of collaboration when combating the condition



A previous study has shown that BRCA genes provide a good example of synthetic lethality (SL), as their dysfunction can lead to the onset of cancer – notably some types of breast cancer. Could you briefly outline what is meant by SL and why this mechanism is an example of an SL relationship?

MR&SS: SL describes a genetic relationship between two genes in which the host cell remains viable if either of the two genes is inactivated individually, but dies if the two genes are inactivated in concert.

The relationship between BRCA1 and PARP1 is complex, and there are different views on why SL is observed between them. BRCA1 is a key player in a DNA damage response (DDR) pathway known as homologous recombination (HR). The main hypothesis is that in BRCA1-deficient (hence HR-deficient) cells, another DDR pathway known as NHEJ becomes active. Because NHEJ is error-prone, cells relying on NHEJ will accumulate damaged DNA and thereby be marked for apoptosis (programmed cell death). PARP1 curtails NHEJ by inhibiting Ku70/80. So by inhibiting PARP1, we release the brakes on NHEJ, leading to genomic instability and thence apoptosis.

How have you assessed the array of gene regulatory network (GRN) inference techniques that are available in order to settle upon the best-performing method for understanding cancer networks?

MR&SS: We selected nine state-of-the-art GRN-inference techniques spanning statistical, network-based and machine learning approaches. These include so-called unsupervised methods, which don't need to be trained (conditioned) on a reference dataset, and supervised methods, which learn parameter values through training. We assembled empirical datasets representing *Escherichia coli*, yeast and human, including some from an international network inference challenge, and computationally simulated other GRNs so we could assess the accuracy of the methods.

We applied the best-performing method, called SIRENE, to normal ovarian tissue and datasets of ovarian adenocarcinoma. This yielded a network of 144 interactions in normal and 108 in cancer, 47 of which overlapped. Manual investigation of these networks revealed novel regulatory interactions. In particular, we found a regulatory switch involving SP3, NFkB1 and E2F1 that controls angiogenesis-specific genes in ovarian cancer.

What issues arise when extending assessments from one type of tissue and cancer to another?

SS: Cellular networks are wired differently in different tissues – at least 20 per cent of human genes are susceptible to transcriptional variation that will rewire protein interaction networks. The genes and rewiring events differ from one tissue to another; what additional rewiring is caused by genomic instability in a particular tumour is largely unknown at this point.

Is it important for different working groups to have an in-depth understanding of each other's work within the project?

MR: In an ideal world, each of us would have exquisitely deep understanding of each other's perspective and work. In reality, our team members bring different skillsets and technologies, belong to different communities of practice and read different specialised literature. We've learned to present our ideas clearly, and ask lots of questions. This is our first collaboration, so a lot of this occurred as we wrote the grant application together, and continues as we co-supervise students.

You collaborate with several other research institutes. What benefit does such partnership bring to your work?

MR: There's no single canonical way to 'do' systems biology. Collaborations let us observe what approaches work in other systems, with other datasets. They give us access to ideas, expertise, biological resources, specialised technologies and infrastructure. System-level analyses are notorious for yielding too many predictions – even after prioritisation based on prediction quality or confidence, we can be left with far more candidate genes, proteins or pathways than we can feasibly take to wet-lab validation. Collaborations can be a source of ideas about what types of leads to follow or avoid, and why.

What does the future hold for your research, particularly with regards to clinical applications?

MR: The goal of the computational side of our project is a computational model of DDR that can identify and prioritise druggable SL targets at $N = 1$ – that is, for single patients and tumours. Understanding how molecular subtypes of breast cancer differ in their reliance on DDR pathways would be a strong first step, but ultimately each patient is unique, and each tumour can evolve over time. Modelling the DDR will allow us to contribute in both directions – depth (individual patients and tumours) and breadth (common themes across cancers).

The bigger picture

Understanding the complexity of biological processes can give hope for treatment of the illnesses that occur when such processes break down. A group from **The University of Queensland**, Australia, aims to utilise one such mechanism – synthetic lethality – to attack cancer in new ways

THE POPULAR IDEA of genes passing on discrete, individual traits has long been accepted as a vast oversimplification. Instead, the genome should be thought of as a vast tapestry of interconnected threads, with the slightest movement of one affecting all of the rest. Illnesses that are caused when something in this network goes awry, such as cancer, are therefore incredibly difficult to characterise, let alone treat. However, this complexity has its benefits, as it can also yield numerous ways to disrupt the action of cancerous cells. One such exploitable effect is that of synthetic lethality (SL), first discovered by Dr Calvin Bridges in 1922, and given its name by Dr Theodosius Dobzhansky in 1946. SL occurs when a combination of two individually benign genetic events results in cellular or even organismal death. Along with this, combinations that only cause sickness are often grouped with SL, meaning it is a very broad effect.

There are thousands of known compounds capable of killing cancer cells. The key difficulty in developing tools to fight cancer is in targeting cancer cells while not harming healthy ones. This difficulty is manifest in the debilitating side-effects of conventional cancer treatments, such as chemotherapy. Fortunately, SL provides a natural targeting system. Cancer cells exhibit genetic mutations that differ from the host's cells, and if one of these mutations is susceptible to attack through an intervention that triggers a synthetically lethal response, then this will exclusively affect the cancer cells.

The technique has been shown to be experimentally viable, with studies revealing mutations in the BRCA1 and BRCA2 genes, commonly implicated in breast cancer, leave them vulnerable to inhibition of the enzyme coded by the PARP1 gene. This provides proof-of-concept for a new way of treating cancer. Further optimism comes from the fact that cancer cells undergo a multitude of mutations that differentiate them genetically from healthy cells, and any one of these could be vulnerable to targeting.

Hailing from The University of Queensland, Australia, Professor Mark Ragan and Drs Sriganesh Srihari and Peter Simpson aim to use these methods to find genes or genetic interactions which could be targeted using SL to kill cancerous cells, and their multidisciplinary team has been working on a variety of different projects within this area.

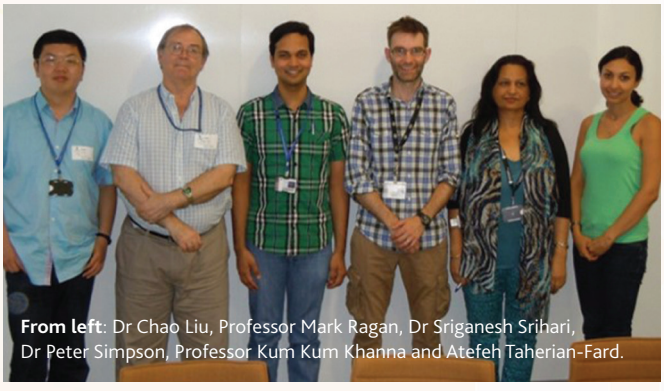
SYNTHETIC LETHALITY THROUGH COMPUTATION

Unfortunately, knowledge of molecular biochemistry in cancer cells (and even healthy cells) is still underdeveloped, so simply predicting synthetically lethal partners for particular mutations from first principles is currently impossible. Large-scale chemical and genetic screening processes have been considered, but would be incredibly resource intensive and are likely to yield many non-druggable targets. An alternative approach comes in the form of mathematical modelling.

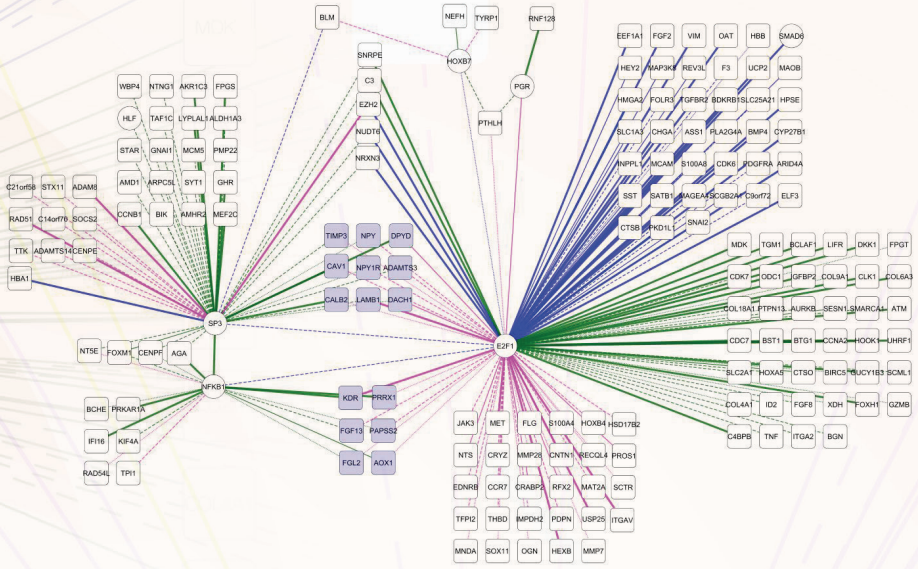
Graph theory has a long history of efficiently finding optimal paths through incredibly complex networks. By adapting already well-developed tools for studying similar networks from the fields of computer science, mathematics and statistics, it is possible to predict potential targets within cancer cells which can then be subjected to empirical study. This approach has the advantage of allowing cancer researchers to focus on a relatively small number of pathways over a wide variety of different cancer types, significantly simplifying a highly diverse set of illnesses.

TESTING OLD METHODS

In cancer, the important element is not individual genes but rather the complex interactions between them. Therefore, a genome-wide approach must be taken. Gene regulatory networks (GRNs) can be thought of as the units which 'execute the cellular code'. By modelling genes as nodes and the interactions between them as edges, it should be possible to see the emergent patterns that cause genes to become cancerous. However, creating mappings of genetic processes onto graphs is far from a rigid science; every map is also an abstraction,



From left: Dr Chao Liu, Professor Mark Ragan, Dr Sriganesh Srihari, Dr Peter Simpson, Professor Kum Kum Khanna and Atefeh Taherian-Fard.



Gene regulatory network showing oncogenic transcription factors and gene modules inferred for large-scale datasets in ovarian cancer. The 15 genes represented in violet-shaded boxes (centre-left) are involved in angiogenesis.

In cancer, the important element is not individual genes but rather the complex interactions between them. Therefore, a genome-wide approach must be taken

and choosing which parts of the system to include and how to accurately represent their interactions is as much of an art as a science. As such, there is a wide variety of techniques which have been developed, and Ragan has contributed to studies which evaluate these methods in a clinical context.

In a recent paper, nine state-of-the-art methods for making inferences about GRNs were tested over 38 different datasets, and the quality of their (highly variable) predictions was assessed. The most successful method – called Supervised Inference of Regulatory NEtworks (SIRENE) – was then applied to a variety of ovarian cancer datasets, implicating numerous genetic mechanisms, with many already having existing support in the literature. This shows the power of systems biology, identifying relevant targets without having to set foot in a wet lab, but also its drawbacks as, with a huge variety of methods to choose from, there is a distinct possibility of falling afoul of 'look-elsewhere effects'.

DEVELOPING NEW METHODS

As well as testing existing models, the group is also developing their own. Building on ideas from control systems theory, Srihari is helping to design models that simulate the progression of cancer. This work uses Boolean modelling, where states are expressed through collections of points which can either be on or off. The system is novel because Boolean systems are usually unable to properly model gene expression data which are time-dependent or require shades of grey between the two states, but the developed system, known as BoolSpace, can deliver results in both of these situations. When applied to pancreatic and breast cancer systems, BoolSpace sheds light on why cancer cells are so robust. The model reveals significant 'baton passing' between active genes, with different genes driving growth at separate stages. This means that drugs targeted at individual genes will fail to have therapeutic effect if they arrive at the wrong point in the process. BoolSpace can provide information on which ones to target and offer a 'cover set' of genes, one of which is always active. Targeting this entire cover set could therefore be sufficient to disrupt the cancer's progression.

An absolute necessity for the modelling of genetic networks, and for systems biology in general, is an accurate set of empirical data around which to apply computer models. To this end, Ragan's team uses 'highly annotated' maps of DNA damage response (DDR) channels produced by colleagues Professor Kum Kum Khanna and Dr Chao Liu. When DDR channels are compromised, cancer is more prevalent. These channels are relevant to the BRCA genes discussed earlier, as BRCA1 is known to be an important part of the homologous recombination DDR system. DDR channels are also being modelled, with work being undertaken to find druggable SL targets. Currently, several potential targets are in cell-line testing.

The mathematics and complex modelling involved in the Australian scientists' research may seem incredibly abstract at first glance, and somewhat removed from the realities of cancer care. However, the interdisciplinary nature of this work and the extent to which it is inspiring researchers from far-flung fields to move into new areas show just how far cancer research has come. Improvement in cancer treatments will also have great human benefit, not only because it will offer new ways to treat patients, but also because it will reduce the debilitation currently associated with cancer treatment.

INTELLIGENCE

COMPUTATIONAL METHODS TO UNDERSTAND SYNTHETIC LETHALITY RELATIONSHIPS IN CANCER

OBJECTIVES

- To assess existing and develop novel methods for making inferences about the involvement of gene regulatory networks in the initiation and progression of cancer
- To use graph theory to discover new synthetic lethality relationships that can be exploited to specifically target cancer cells

KEY COLLABORATORS

Dr Fares Al-Ejeh; **Professor Georgia Chenevix-Trench**; **Professor Kum Kum Khanna**, QIMR-Berghofer Medical Research Institute • **Professor Sunil Lakhani**, Queensland Health • **Dr Chao Liu**, Institute for Molecular Bioscience, The University of Queensland

FUNDING

National Breast Cancer Foundation, Australia
• National Health & Medical Research Council (NHMRC), Australia

CONTACT

Professor Mark Ragan
Head, Genomics and Computational Biology
Institute for Molecular Bioscience

The University of Queensland
St Lucia, QLD 4072, Australia

T +61 7 3346 2616
E m.ragan@uq.edu.au

www.imb.uq.edu.au/mark-ragan

MARK RAGAN is founding Head of Genomics and Computational Biology at the Institute for Molecular Bioscience, and adjunct Professor in the School of Information Technology and Electrical Engineering, The University of Queensland. A graduate of the University of Chicago and Dalhousie University, he was with the National Research Council Canada for 28 years before relocating to Australia in 2000.

SRIGANESH SRIHARI is a postdoctoral fellow at the Institute for Molecular Bioscience, The University of Queensland, in an NHMRC-funded project on breast cancer. His research interests are computational biology and bioinformatics, systems biology, data mining and database design. He holds a PhD in Computer Science from the National University of Singapore, and spent three years as a software engineer with Oracle in Bangalore, Warsaw and London.

PETER SIMPSON has more than 10 years' experience investigating genomic dysregulation in breast cancer. He earned his PhD from the University of Liverpool, UK, and spent five years as a Postdoctoral Research Fellow at the Breakthrough Breast Cancer Research Centre within the Institute of Cancer Research, London. He is now based in the translational research environment of the UQ Centre for Clinical Research, supported by a fellowship from National Breast Cancer Foundation, Australia.