

Exploring the Forgotten Genome of Cancer Research

A Garvan Institute sarcoma study identifies rare cancer-associated variants in the germline with the HiSeq® 2500 System.

Introduction

A decade ago, Professor David Thomas, FRACP, PhD noticed something interesting about the expanding field of precision medicine for cancer treatment. While there was an active and productive worldwide effort to sequence and analyze tumor DNA, there was scant research interest in the germline DNA of the subjects from whom the tumors were taken. The germline, he thought, had become the forgotten genome of cancer research.

Since then, Professor Thomas and his colleagues at the Garvan Institute's Kinghorn Cancer Center have been testing a hypothesis. Could analysis of germline DNA contribute to an improved understanding of the genetic determinants of this early onset cancer? Could it also enable earlier, more effective cancer detection and treatment?

They embarked on the creation of a new, international cohort of sarcoma subjects, called the International Sarcoma Kindred Study (ISKS), and developed a unique genetic panel of genes associated with cancer risk. Using the HiSeq 2500 System, they performed targeted exon sequencing of germline DNA from more than 1000 individuals.¹

iCommunity spoke with Professor Thomas about the unique aspects of the study and the importance of its findings. Specifically, how the "forgotten genome" might become an essential element of research cohort composition and a prominent feature of medical practice in the future.

Q: How did you become involved in cancer research?

David Thomas (DT): I trained to be a doctor and was interested in making a difference. Over the past 500 years, all science has been converging on a realistic, molecular view of the world, and that is translating into opportunities for cancer treatment. I see cancer as a worthy cause, and one in which I can make a difference.

Q: Why did you choose sarcoma as a research focus?

DT: I specialized in sarcomas early in my clinical practice as a consultant oncologist. My PhD studies focused on the cellular biology and biochemistry of osteosarcoma cell lines, which led me to sarcoma as an area of clinical interest. In addition to being a compelling area of unmet need, the fact that sarcomas are rare cancers provided me with a focus, which is essential for making a contribution in research.

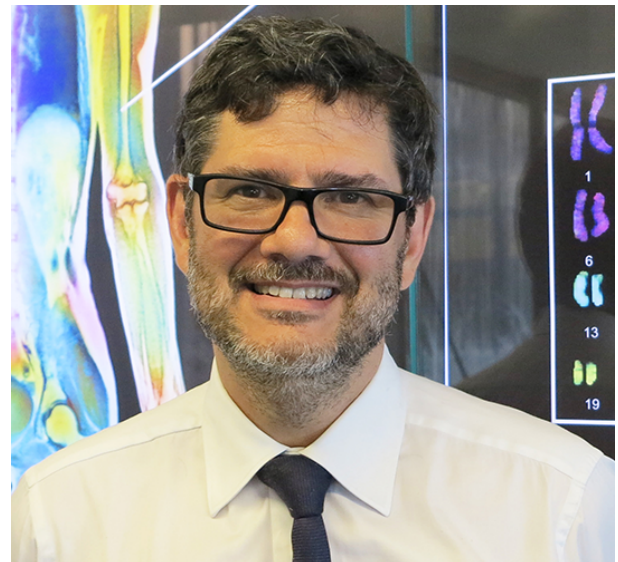
Q: What is sarcoma?

DT: Sarcomas are cancers of connective tissues, such as bone, muscle, and cartilage, which are diagnosed by imaging and biopsy. Only 1% of all cancers are sarcomas, but they're an intriguing, complex, heterogeneous group of diseases. There are approximately 50 different sarcoma subtypes within that 1%, all with different properties and molecular drivers.

Sarcomas particularly affect the young, comprising 20% of childhood cancer and 10% of cancer in young adults. Its victims are on average 20 years younger than those who are afflicted with most epithelial cancers.

Q: Why do you refer to the germline as the "forgotten genome" in cancer studies?

DT: In 2008, I realized that as precision medicine was evolving, it was becoming overly focused on the molecular analysis of tumors. There are 11,000 genomes or exomes in the International Cancer Genome Consortium and the Cancer Genome Atlas that have been sequenced and put into accessible databases. However, only 0.3% of those have attached germline genomes. The germline has become the forgotten genome in cancer research.



Professor David Thomas, FRACP, PhD is Director of the Kinghorn Cancer Center and Head of the Cancer Division of the Garvan Institute in New South Wales.

Q: What is the potential value of the germline in cancer studies?

DT: In the Genomics Cancer Medicine Program at the Kinghorn Cancer Center, our hypothesis is that by understanding the genetic determinants of early onset cancer, we could potentially identify cancers earlier and at a curable stage. Because we cure 85% of individuals who get cancer under the age of 40 today, we could improve that percentage even more by introducing the secondary prevention and management strategies that are already used in other areas of medicine.

For example, after someone has a myocardial infarction, there is a standard treatment program to manage their subsequent risk of having a second heart attack. We don't do that in cancer. Genomic analysis of the germline could help us understand why an individual got cancer, enabling us to initiate a follow-up medical program in a risk-stratified way.

“We found a statistically significant correlation between the number of pathogenic variants that people carry and an earlier age of cancer onset.”

Q: What sparked the creation of ISKS?

DT: The genesis of ISKS was the need for a patient-centered cohort that was annotated with information relevant to understanding the genetic risk for cancer. The patterns of cancer in a family tell us how genes and alleles are transmitted. That information is rarely collected so we had to make a special effort to create a cohort that fit the goal of genetic risk analysis. The ISKS started in Australia and is now open at 21 centers across all continents. The study reaches 1900 families and is growing.

Q: How did you choose which genes to include in the sequencing panel used in the study?

DT: We took a superset of commercial and academic heritable gene panels that were available at the time and added a few genes that we knew were sarcoma-related, such as exostoses genes *EXT1* and *EXT2*. There were some genes, like *ERCC2*, which were included because they are involved in DNA repair. It was intended to be a panel of genes known to be associated with other diseases that would enable us to ask what happens when those genes are mutated in a sarcoma population.

Q: What were the unique features of the study?

DT: Instead of looking directly at tumors, we analyzed the germline DNA of a 1162 proband cohort of subjects with sarcoma. We performed targeted exon sequencing on a panel of 72 genes associated with cancer risk. Most of the cohort was composed of individuals who were not selected for family history. For the ISKS, we chose consecutive cases coming into specialist sarcoma units around the world. It was the same with the other cohorts, except for the 10 subjects from kConFab (Kathleen

Cunningham Foundation Consortium) that were familial breast cancer cases who had sarcoma.

The study also included a case-controlled rare variant analysis of 6545 cancer-free individuals. Using a case-controlled design across populations to identify individual genes associated with sarcoma risk was a novel component of our study.

Q: What were the major findings of this 5-year pathogenic variation study?

DT: As the ISKS cohort grew, it became clear that one important quantitative aspect of the phenotype was the age of cancer onset. In our study, we found a statistically significant correlation between the number of pathogenic variants that people carry and an earlier age of onset. That told us that age of onset, and not necessarily sarcoma type, was a robust predictor of the burden of genetic heritability.

We performed a rare variant burden analysis to identify which genes among the 72 on our panel were particularly enriched. The obvious candidate for a pan-sarcoma panel was *TP53*, which is known to be associated with sarcoma risk, and it was powerfully enriched across the entire cohort. There were other genes that we didn't expect to be associated with sarcoma, including *ATM*, *ATC*, and *ERCC2*, which were strongly enriched.

Among people with sarcoma, there were 61 individuals who had pathogenic variation in genes like *APC*, *MLH1*, *MSH2*, and *MSH6*, which are associated with bowel cancer. We also identified 28 individuals in the sarcoma cohort who have *BRCA1* or *BRCA2* mutations, which is well known in breast cancer medicine. The importance of these genes is that we already have established programs for managing risk. I think that's a very important outcome.

“We used the HiSeq 2500 System, which was perfectly suited for the study and made sequencing incredibly affordable.”

Q: Was it surprising to identify polygenic as well as monogenic determinants of sarcoma risk?

DT: Our finding of polygenic risk was completely novel. It arose out of an observation about the progressive earlier age of onset for individuals with multiple pathogenic variants. What was striking to me was that if we looked at people who carried two or more of the strongest alleles that are prominently associated with cancer risk, for example in genes like *BRCA1*, the age of onset for those individuals was even younger than for the *TP53* mutation carriers. That means the combined effect of this previously unrecognized polygenic rare variation is at least as great as the strongest known monogenic determinant of sarcoma risk.

What's also notable is that we have about 20 people in our cohort with *TP53* mutations, and 36 who have this polygenic pattern. It implies that not only is the polygenic effect size greater, but its contribution as a cohort is almost twice as large. Twice as many people are carrying these multiple variants.

Q: How did you perform the targeted gene panel sequencing?

DT: We used the HiSeq 2500 System, which was perfectly suited for the study and made sequencing incredibly affordable. The cost of sequencing the panel was about \$200 AUD per sequence, which meant that we had the capacity to do this on a scale of more than 1000 cases. The study was the first of its kind in a rare disease, and getting as much as we could from every dollar we spent was an important part of our strategy.

We performed 10 batches of sequencing over a three-year period. For the later batches, the sequencing was completed in three days using the high-output mode of the HiSeq 2500 System. We were 96-plexing in a single lane and ending up with ample read depth. All the sequencing was completed about three years ago. We've been analyzing it since that time.

"With WGS, we'll be able to fill in the heritability data that is missing from what we've detected with targeted exon sequencing."

Q: Will you be using whole-genome sequencing (WGS) in your studies?

DT: We're just about to embark on performing WGS of this cohort. That is incredibly exciting because we can start to ask broader and deeper questions than we could when we were looking at a panel of known genes. For example, there are many genes, in addition to the ones on our panel, involved in response to DNA damage. With WGS, we can look for pathogenic variation across the totality of that pathway and ask whether we see a signal from every gene, or just the ones we picked for the panel. Also, looking at the whole genome allows us to discover new genes that have not previously been associated with cancer.

WGS might solve the mystery of missing heritability. For example, we have about 130 individuals who have the clinical characteristics of Li Fraumeni-like syndrome, but we have only about 20 people in whom we've discovered *TP53* mutations by looking at coding sequences. We wonder how many of those other individuals, who apparently do not have a *TP53* mutation, will turn out to have a mutation that lies 20 bases upstream of the coding sequence of the gene. With WGS, we'll be able to fill in the heritability data that is missing from what we've detected with targeted exon sequencing.

It will be interesting to see how many more biomarkers we can discover that are associated with early onset of sarcoma.

Q: Is sarcoma risk determined at birth?

DT: Based on a published study of environmental and heritable factors,² and the analysis of 72 genes in our sarcoma study, I think 1 in 4 people who get sarcoma will be found to have a genetic mutation or mutations that are responsible for it. That's not to say that there aren't environmental influences. The strongest of those is radiation exposure. If a breast cancer patient receives radiation therapy, their risk of secondary sarcoma is much higher. Some of the genes we identified, like *ATM* and *ATR*, are directly involved in repairing DNA damage from radiation. So, someone's sarcoma risk from radiation exposure might be affected by the genes they inherit. If they have the *ATM* or *ATR* variants we identified, their chance of radiation-induced sarcoma might be much higher than someone who doesn't have those variants.



The HiSeq 2500 System in use at the Garvan Institute's Kinghorn Cancer Center.

Q: Could some of the variants you identified ultimately be used to predict therapeutic response?

DT: There are drugs being developed whose usefulness in treating cancer patients is predicated upon the presence of a germline mutation. An example is PARP (poly ADP ribose polymerase) inhibitors, which have been shown to trigger responses in ovarian and breast cancer patients with *BRCA1* and *BRCA2* mutations.^{3,4} There's a good reason to believe that individuals in whom we found the same mutations might benefit from those treatments.

If I extrapolate from the PARP story, Dr. Susan Domcheck at the University of Pennsylvania published a study several years ago looking at prostate and pancreatic cancer subjects who had a *BRCA2* mutation.⁵ They were treated with olaparib, a PARP inhibitor drug that is an accepted treatment for breast and ovarian cancer patients with the same mutations. The data supported the hypothesis that therapy directed against a genetically defined target has activity regardless of anatomic organ of origin.

I believe the same rules about response-to-therapy predictions will hold true for sarcoma as well. I would love to see the information we've discovered with ISKS used for PARP inhibitor studies on sarcoma subjects.

Q: How might your findings change clinical practice?

DT: We're at a time in history where the best way to treat patients is by combining bioresearch with clinical care. I believe in evidence-based, research-led clinical care. The Kinghorn Cancer Center supports a translation focus on human-centered cancer research. Now we have genomics, an enormously powerful tool for understanding human disease that is providing us with opportunities to improve health outcomes. We've never been able to see interactions between multiple genes before. With gene sequencing panels, we can identify individual genes and look at interactions between genes. In the past, we might have said that the significance of a variant was uncertain. Yet, when someone has two variants occurring, what does that mean? Our data suggest that the effect of two variants of uncertain significance amounts to more than the threshold that we regard as clinically important if it were a single gene. It occurs frequently enough to be considered part of what will shape clinical genetics practice.

In the future, clinical geneticists might review the results of our panel for both monogenic and polygenic factors to identify individuals at risk for sarcoma. Those individuals might be screened by MRI for presymptomatic, curable cancer. For people who never had cancer, but carry a genetic predictor of risk, we can use that information to prevent or cure cancer. We want to be able to detect the cancer when a surgeon can remove or treat it easily.

"Applying WGS broadly in fighting cancer will enable us to determine what fraction of cancer cases are modifiable."

Q: Is there clinical value in the germline?

DT: We've shown there is clinical value in the germline or "forgotten genome." As cancers grow, there are mutations present in one metastasis, but not in another. Pathology labs simplify it by subtracting the germline, leaving only the unique variation that has been generated during tumor progression. Those sorts of tests have produced few actionable variants. The irony is, we might be removing a rich source of information that we could use to increase the opportunities to treat people. By definition, a germline mutation is truncal. It is shared by all the subsequent metastases and subclones. I think the germline biomarkers for treatments might actually be more useful clinically.

Q: What is your vision of the future in sarcoma treatment?

DT: My vision for the future is that for anyone under the age of 40 who gets cancer, part of their diagnostic workup should include a genetic test to find out why. That information could inform treatment and the follow-up plans, and the preventive cancer management of the individual's family. This approach could change oncology fundamentally. It would shift the focus away from treating the disease in a narrow window towards

considering the total opportunity to improve health for the significant fraction of the community that is affected by cancer.

Q: What are the next steps in sarcoma research?

DT: First, WGS of sarcoma subjects and their families is essential. We need to build a vast library of human diseases and their genetic underpinnings. We need to do so in a way that is not limited to targeted exomes or even whole exomes. We need to do this properly and regard it as an investment.

When we sequence a sarcoma subject's genome, we create an enduring resource. In comparison, our panel will become redundant within two years because there will be new genes identified. More than 40% of our community will develop cancer, with 30% of those ultimately dying from the disease.⁶ Applying WGS broadly in fighting cancer will enable us to determine what fraction of cancer cases are modifiable. That information will drive government investment in what should be a public-health approach to this disease.

Second, research programs need to perform functional validation of the genes associated with cancer risk in a way that provides a library of independent information. If I were investing as a research organization, I would increase the rate at which we could derive all possible information to validate the variations we identify in screening.

We can generate thousands of variants, but only a fraction of those are interpretable around epidemiology currently. Case-controlled designs will provide evidence that something is pathogenic. If we tell an individual that a variant is significant, it would be helpful to provide information about the effect that variant produces, based on a companion library of mutations. We'll also need to track cancer subjects routinely. Does everything we've been taught to expect about a variant hold true in practice? That will require long-term follow-up of individuals to see whether our predictions hold true.

Learn more about the Illumina system mentioned in this article:

HiSeq 2500 System, www.illumina.com/systems/hiseq_2500_1500.html

References

1. Ballinger ML, Goode DL, Ray-Coquard I, et al. Monogenic and polygenic determinants of sarcoma risk: an international genetic study. *Lancet Oncol.* 2016; 17:1261-1271.
2. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer—analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med.* 2000; 343: 78-85.
3. Audeh MW, Carmichael J, Penson RT, et al. Oral poly(ADP-ribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and recurrent ovarian cancer: a proof-of-concept trial. *Lancet.* 2010; 376: 245-251.

4. Tutt A, Robson M, Garber JE, et al. Oral poly (ADP-ribose) polymerase inhibitor olaparib in patients with *BRCA1* or *BRCA2* mutations and advanced breast cancer: a proof-of-concept trial. *Lancet*. 2010; 376: 235-244.
5. Kaufman B, Shapira-Frommer R, Schmutzler RK, et al. Olaparib monotherapy in patients with advanced cancer and a germline *BRCA1/2* mutation. *J Clin Oncol*. 2015; 33: 344-250.
6. Australian Government, Australian Institute of Health and Welfare. Cancer Risk. www.aihw.gov.au/cancer/. Accessed February 8, 2017.