Geoffrey Faulkner

Retrotransposons are global regulators of gene expresssion in mammals

Retrotransposons are genetic elements that have populated mammalian genomes via a copy-and-paste mechanism to such an extent that they comprise at least 50% of the human genome. Although erroneously dubbed "junk DNA" by some researchers in the past, retrotransposons in fact helped the human genome evolve and are now recognized to affect the function of thousands of genes when they occur in close proximity to those genes. It has also been noted that retrotransposons can insert themselves within genes and cause numerous diseases, such as breast cancer and haemophilia, and that the complement of retrotransposons differs substantially between individuals.

The work presented here confirmed that hundreds of thousands of retrotransposons in the human genome are transcribed to generate RNA and that this activity in a large number of cases served to regulate the expression of genes that were nearby on the genome. This discovery confirms the earliest theories of retrotransposon function *in vivo*, where they act as "controlling elements," and counters recent assertions that they are neutrally evolving and non-functional elements. The work was also of major biomedical relevance because many regulatory retrotransposons were found to occur next to genes associated with disease susceptibility in humans.

Mathias Francois

SOX18 induces development of lymphatic vasculature in mice

We have discovered the gene that is the master regulator of the development of lymphatic vessels. This work (François et al, *Nature* 2008) demonstrated that the transcription factor SOX18 plays a key role in initiating lymphangiogenesis in the embryo. We have also established that SOX18 is not required for maintenance of lymphatic vessels in the adult. Embryonic mice with loss of function of Sox18 do not activate the lymphatic hallmark gene *Prox1*, do not generate lymphatic endothelial progenitor cells and completely lack a lymphatic vasculature. Our subsequent data show that SOX18 directly activates *Prox1* transcription in a VEGF-C dependent manner. This work has repositioned the field of lymphatic research and provided a potential new avenue for therapeutic modulation of neo-lymphangiogenesis in the adult, without compromising existing lymphatic vessels.

Outcomes and Significance

The human syndrome Hypotrichosis-Lymphedema-Telangiectasia (HLT) is caused by mutations in the *SOX18* gene. Our work in elucidating the pivotal role of *Sox18* in initiating lymphatic endothelial differentiation from vascular precursors explains the aetiology of lymphatic defects arising in HLT.

Our findings are also directly relevant to secondary lymphedema, which is an acquired disorder caused by damage to the lymphatic system either by surgery, injuries or parasitic infections. Few advances have been made in the comprehension of the signalling pathways involved in lymphatic endothelial development, and our studies have illuminated how these processes are directed by specific endothelial transcription factors and growth factors. Our discovery will lead to pioneering approaches to improve lymphedema recovery.

Further, the lymphatic system is thought to be the major route of tumour metastasis during cancer progression. Understanding the molecular basis of lymphangiogenesis is crucial because blocking the formation of new lymphatic vessels is vital to prevent secondary tumour formation. Outcomes of this work will show the way to innovative strategies for modulating neo-lymphangiogenesis to prevent tumour metastasis during cancer progression.

Future Directions

Our discovery will translate into biomedical research that centres around the molecular and cellular events that initiate lymphatic development *in vivo*, studying both lymphangiogenesis in normal embryonic development and in animal models of cancer metastasis and secondary lymphedema. We are also translating our knowledge of the SOX biology into a clinical context by analysing the prevalence of SOX18 single nucleotide polymorphism in women developing lymphedema after breast cancer surgery. Further, our discovery is likely to be applicable to therapeutic strategies involving the prognosis of risk in developing secondary lymphedema. The long-term view of our project is to develop innovative therapeutic strategies that will constitute a solid adjunct to pre-existing cancer treatment.

Stuart Macgregor

Common sequence variants on 20q11.22 confer melanoma susceptibility

Melanoma is the deadliest skin cancer and is a major public health concern in Australia. Early detection and treatment is critical to outcomes in affected individuals. An understanding of the genetic factors influencing risk of developing melanoma is crucial to the identification of at-risk individuals. By identifying a set of genetic variants underlying the genetic risk, early detection (and hence treatment) of the disease can be improved.

Rare melanoma risk variants have been identified but these only account for a very small proportion of cases. In contrast, common variants with moderate effect on disease susceptibility are of substantial greater importance in terms of overall public health. Until recently, only one gene (MC1R) had been validated as harbouring common susceptibility risk variants.

To identify genes conferring disease susceptibility we examined, in the first published study of its type in melanoma, over a half a million genetic variants simultaneously in a large sample of melanoma cases and controls. This approach, known as a genome wide association study or GWAS is expensive to implement. A relatively unusual approach, known as DNA pooling, was successfully used to reduce the cost of the first stage of the study. The results from this first stage were replicated in two subsequent samples.

The major finding was that common genetic variants (on chromosome 20), carried by 1 in 6 in the caucasian population, almost double melanoma risk. This region on chromosome 20 appears to be the single most important contributor to melanoma risk and will be invaluable in helping to predict disease risk in the near future. As well as better assessment of risk, characterization of the identified genetic variants involved in disease initiation and progression may also lead to the development of new treatments.

Common sequence variants on 20q11.22 confer melanoma susceptibility

Stuart Macgregor, Kevin M. Brown, Grant W. Montgomery, Nicholas G. Martin, The International Melanoma Genetics Consortium (GenoMEL), The Australian Melanoma Family Study, Nicholas K. Hayward.

Cutaneous melanoma is an important health problem in fair-skinned populations worldwide. An understanding of the genetic factors influencing melanoma risk and the identification of susceptible individuals may aid in increasing sun protection and early detection of the disease in populations at risk. Rare, high penetrance melanoma risk variants have been identified but these only account for a very small proportion of cases. In contrast, common variants with moderate effect on disease susceptibility are of substantial greater importance in terms of overall public health. Until recently, only one gene (MC1R) had been validated as harbouring common low-penetrance susceptibility risk alleles.

To identify disease susceptibility loci, we conducted a genome-wide association pooling study for cutaneous melanoma. A three stage study design was used, with samples totaling 2,019 melanoma cases and 2,105 controls. Using DNA pooling in the first stage (N=864 cases, N=864 controls), we examined 550,000 single nucleotide polymorphisms for association with melanoma. We identified a new melanoma risk locus on chromosome 20q11.22. Excellent agreement was found between the pooling results and subsequent confirmatory individual genotyping. The association was replicated in two further samples (combined P < 1e-15). The per allele odds ratio was 1.75 (1.53, 2.01), with 1 in 6 in the Australian caucasian population carrying risk alleles. The effect size was larger in early-onset cases.

The locus at chromosome 20q11.22 appears to be one of the single most important contributors to melanoma risk and will be invaluable in helping to predict disease risk in the near future. These findings are a striking demonstration of the power of genome-wide association studies to identify new loci underlying common complex disease.

This work was published in Nature Genetics, 40(7):838-40, July 2008 (corresponding author S. Macgregor). Since publication we have run a more detailed analysis increasing the number of single nucleotide polymorphisms to over 1 million. Detailed fine mapping analysis is underway for both these samples and other independent samples

Retrotransposons are global regulators of gene expresssion in mammals

A fundamental goal of the post-genome era has been to characterise the transcriptome, or full complement of RNA transcripts, generated by each eukaryotic genome. To date, massive scale sequencing projects have identified most human protein-coding genes, as well as more than 200,000 unique messenger RNAs and non-protein coding RNAs (ncRNAs). If we consider the transcriptome as a machine, its gross parts have largely been identified; our focus has now shifted towards defining the interactions (regulation) between each part. *Elucidating these interactions is a crucial step towards understanding the fundamental genetic basis of virtually all human diseases.*

Retrotransposons comprise 30-50% of the human genome, are one of the largest sources of intergenic transcription and were first characterised as regulators of nearby genes1. They spread via a "copy-and-paste" mechanism, are considered a prime factor in the expansion of the mammalian genome and have been implicated in numerous diseases. In the past year, several prominent works have provided a snapshot of the regulatory potential of retrotransposons, including via control of chromatin structure[2] and the generation of endogenous RNA interference (RNAi)[3]. These works created great interest in retrotransposon transcription amongst the genomics community.

Here, as part of the international Genome Network Project, we present a genome-wide study of transcription from within human and mouse retrotransposons. This activity was found to pervade each genome and generated 6-30% of the total RNA transcripts detected, depending on cell type4. An analysis of 250,000 retrotransposon-derived promoters revealed that these frequently functioned as alternative promoters for nearby genes and/or expressed ncRNAs. Finally, a genome-wide screen identified 23,000 candidate regulatory regions derived from retrotransposons. Overall, the extent of retrotransposon transcription far exceeded prior estimates, with a common theme of strong tissue specificity and association with nearby protein-coding genes. Therefore, the global transcription of retrotransposons serves to regulate numerous human genes and should be considered in modelling the underlying genetics of human diseases.