

GENOMICS

The great pretenders

“ the effects of pseudogene expression need to be further examined ”

Pseudogenes — ancestral copies of protein-coding genes that have become mutated over time — are common throughout the genome. Several lines of evidence indicate that pseudogene transcripts can influence the function of their protein-coding counterpart by behaving as regulatory RNAs. They also complicate massively parallel sequencing and gene expression studies, as they are highly similar in sequence to the fully functioning gene. Arul Chinnaiyan and colleagues have attempted to map the position of pseudogenes that are expressed throughout the genome.

The authors analysed RNA sequencing transcriptome data compiled from 248 cancer and 45 benign samples using a bioinformatics approach that focused on detecting pseudogene transcription. They identified 2,156 unique pseudogene transcript clusters, and their start and end points were mapped based on the coordinates of pseudogenes that have been identified in two databases: encyclopedia of DNA elements (ENCODE) and Yale pseudogene resources. These and additional analyses indicated that they had identified 2,082 distinct transcripts that corresponded to 1,437 wild-type genes, indicating that pseudogenes are expressed much more frequently than was previously thought.

In addition to analysing ubiquitous and tissue-specific expression, because most of the samples used in this study were tumour samples, the authors also analysed tumour-specific expression. They found that 218 pseudogenes were expressed only in the tumour samples, the majority of which were expressed across cancer types. Forty were expressed in specific cancer types, and the authors analysed some of these, specifically those from breast and prostate cancer, in greater detail. Expression of the

coxsackievirus and adenovirus receptor pseudogene (*CXADR-Ψ*) is increased in approximately 25% of prostate cancers, unlike the wild-type gene, which has been implicated as a tumour suppressor gene. Interestingly, almost all of the prostate cancer samples with increased *CXADR-Ψ* expression did not have gene fusions involving the ETS oncogenes, whereas expression levels of *CXADR* were fairly consistent in ETS-positive and ETS-negative patients. The authors also identified a chimeric RNA transcript that consisted of the first two exons of a protein-coding gene, *KLK4*, and the last two exons of a pseudogene, *KLKPI*. This transcript was also expressed as a protein and warrants further study given its high levels of expression in 30–50% of prostate cancer samples used in this study.

In the breast cancer samples, the authors focused on *ATP8A2-Ψ*, a pseudogene on chromosome 10 of a LIM domain-containing protein. The expression of the wild-type gene on chromosome 13 varied across the breast cancer samples, whereas *ATP8A2-Ψ* was highly expressed in samples with a luminal histology. Knockdown of this pseudogene in two breast cancer cell lines that overexpressed it reduced proliferation, migration and invasion, and overexpression of *ATP8A2-Ψ* in benign breast cancer cells increased their proliferation and invasive growth. Interestingly, knockdown of this pseudogene had no effect on expression levels of the wild-type gene, indicating that

there is no regulation of the wild type gene by the pseudogene, unlike that seen for *BRAF* and *Oct4* transcripts that have an inverse relationship with their pseudogenes.

These findings indicate the need to consider pseudogene expression in all high-throughput gene expression studies, and that the effects of pseudogene expression need to be further examined as their regulation and their effects on cell biology seem to be complex.

Nicola McCarthy



ORIGINAL RESEARCH PAPER
Kalyana-Sundaram, S. et al. Expressed pseudogenes in the transcriptional landscape of human cancers. *Cell* **149**, 1622–1634 (2012)