



Keith tuned into genefinding. Image: George Stawicki, My Photos

## FINDING THE GENE IN THE GENOME

The human genome has been sequenced, but finding where the actual genes are in the long sequence of DNA bases is a laborious task. Research by Dr Keith Knapp in the School of Information Technology, Deakin University, will speed up the search for the particular sequences of DNA that make up individual genes.

"DNA stores information like words in a text file," says Keith. "The software I'm working on analyses the DNA sequence to find patterns which can be used to predict where a gene starts and finishes, and what its components are."

Keith first evaluated leading, open-source, gene-finding software, by comparing programmatic predictions with actual lab results. "Some geneticists were surprised at my results," he says, "because I showed that there were some significant weaknesses in gene-finding programs that they thought were reliable".

His next step was to look for ways to improve the software. Concentrating on exons, DNA sequences which contain the instructions for protein production, Keith found that some exons were more information-rich than other sections.

"The first exon in a gene usually has more information in it than subsequent exons in that gene," he explains. "So, the software can search for these high-information exons which are likely to be the start of a gene".

Keith is now re-engineering the software to include his refinements. Once complete, this new version of the software will in turn be made available for others to use and improve further. It could even solve a long-standing question.

"Scientists don't currently have a generally accepted definition of a gene," Keith says. "If we could improve the software to be 100% accurate as to where a gene starts and stops, then that could be used as the definition of a gene."

Keith graduated with his PhD in April 2009 and is taking up a position at Whitworth University in Spokane, Washington, USA, in July.

### FURTHER INFORMATION:

School of Information Technology  
Principal supervisor: Associate Professor Phoebe Chen  
E: [phoebe.chen@deakin.edu.au](mailto:phoebe.chen@deakin.edu.au)  
[www.deakin.edu.au/scitech/it](http://www.deakin.edu.au/scitech/it)

**'The software can search for these high-information exons which are likely to be the start of a gene.'**

