Lists of cancer mutations awash with false positives

But researchers show artefacts can be reduced by taking differing rates of mutation into account.

Heidi Ledford

17 June 2013

Some call them the 'fishy genes': errors in DNA that seem to be associated with tumours, but which researchers trawling through cancer genome data cannot explain. For instance, why would mutations in genes involved in the sense of smell be linked to lung cancer?

Tumour cells, such as these from a lung cancer, are riddled with genetic mutations, but many of those are caused by the cancer rather than being involved in causing the disease.

A reanalysis of cancer genome data, published on *Nature*'s website [1], finally does away with the fishy genes — and a host of others thought linked to cancer — by accounting for how mutation rates vary between different locations in the genome. The effects are drastic: in one analysis, using the researchers' improved models whittled down the list of genes potentially associated with lung cancer from 450 to 11. The study could help cancer researchers to focus their efforts on the genes that matter, thus preventing them from wasting time by investigating blind alleys.

"This will affect future cancer genome projects," says Gad Getz, a computational biologist at the Broad Institute in Cambridge, Massachusetts, who is a lead author on the study. The major sequencing projects for cancer genomes have incorporated the new models into their pipeline, he says, noting that previous analyses — many of which have been reported in high-profile publications — are now being redone.

Suspicious patterns

All cells accumulate mutations, but cancer genomes tend to be particularly riddled with errors — in part because cancer involves the deactivation of the cell's own repair mechanisms, and because tumour cells reproduce at a faster rate than those in healthy tissue. Most mutations do not affect the cell's life cycle, however, and thus are probably not involved in cancer.

To determine which mutations are important drivers of the disease, researchers compare the genomes of many tumours and look for mutations that occur more frequently in cancerous tissue than one would expect due to random chance.

Many researchers had expected that as analyses began to include more tumour genomes, this search would become more refined and the lists of candidate cancer genes would get shorter. Instead, the lists grew longer. To find out why, Getz and his colleagues looked for other factors that might boost mutation rates.

In some cases, that higher rate of mutation was not linked to cancer, but to confounding factors, the team demonstrated. To identify the false positives, they took into account the fact that genes that are less often transcribed into RNA are more susceptible to mutating. This is because the transcription process is coupled with a DNA repair process that can undo some mutations. Genes that produce very low levels of RNA — as would be the case for an olfactory gene in lung cells, say — are therefore more likely to show as false positives.

The team also took into account the different rates at which mutations can occur during DNA replication. Genes that are copied later in the process are more likely be copied incorrectly because the enzymes that construct new DNA become error-prone when supplies of DNA's chemical building blocks run low.

Once these and other confounding factors were incorporated into their analyses, Getz's team found far fewer mutations that seemed to be linked to cancer. Getz thinks that the new models could be used to improve trawls through genome data from patients with other diseases as well.

The analysis upgrade is vital, says Tom Hudson, president of the Ontario Institute for Cancer Research in Toronto, Canada. It is especially useful for researchers who must pick through the lists of putative cancer-associated genes in search of the next drug target. "It takes so much work to follow up on these hits," says Hudson. "Being able to weed out false positives is obviously very important —

certainly for the life of many postdocs."

References

1. Lawrence, M. S. *et al. Nature* http://dx.doi.org/10.1038/nature12213 (2013).