

Missing breast cancer genes may soon be discovered

We are closer to finding the missing 80% of breast cancer genes than ever before thanks to the success of the COSMIC database (Catalogue Of Somatic Mutations In Cancer) the 5th European Breast Cancer Conference (EBCC-5) was told today.

The genes BRCA1 and BRCA2 account for approximately 20% of the familial risk of breast cancer, leaving 80% to be explained. The COSMIC database developed at the Sanger Institute, Cambridge, UK was created in 2004 to provide free up to the minute genetic data to scientific communities and to prevent the duplication of research. The data in COSMIC has expanded to include data on 538 genes, 124,367 tumours with 23,157 mutations.

So far in genetic research 350 cancer genes have been identified and of these, 311 have a mutation in the somatic cell. Scientists believe that many somatic mutations are caused by a number of things – the way that DNA repairs and maintains itself, or past exposure to a harmful substance such as a virus. Alternatively it may be a combination of both.

This for example could explain why some smokers get lung cancer and some don't – there must be a combination of faulty gene and harmful exposure to tobacco smoke to activate the cancer cells.

The challenge for scientists is telling the difference between, say the 28 genes that are frequently mutated, beyond what would be expected by chance and are almost certainly cancer causing genes, and the 180 'bystander' genes which may be coincidentally involved. It's highly likely that some 'bystander' genes may turn out to be damaging as more researchers exploit this extraordinary free collection of data.

The data has been drawn together from the scientific literature on cancer cell lines (including those master lines held by the National Cancer Institute in the US) and tumour biopsy analysis. The information is stored in a standardised fashion to facilitate searching and compilation by anyone not expert in the area. Tissue samples are extracted and histology for each sample noted and the definitions are plotted in the database. A single DNA sequence is held for each transcript, which in turn is translated to give the protein sequence used by COSMIC. Mutations are mapped to these standard sequences.

Recently, information about patients lifestyle, ethnicity and tumour characteristics have been incorporated into tables. Grouping of information, for example, ethnicity can be done and even more complex features such as cigarette smoking history. The values that have been stored so far for this feature are expressed as pack years along with less specific comments, such as smoker, nonsmoker, ex-smoker and never-smoker. This system allows COSMIC to capture the wide range of information reported in the literature. It also accepts different data content for different genes, for example, drug response information for tumours with and without EGFR mutations. This can be crucial in explaining why a drug works against a particular cancer, or, more commonly fails to work!

Mike Stratton, creator of COSMIC comments, "About 5 -10% breast cancer cases are due to a genetic predisposition to the disease. A number of breast cancer genes, such as BRCA1 and BRCA2, have been identified. However, it is clear that the genes underlying most genetic susceptibility to breast cancer is yet to be uncovered. Major efforts are now being made to find these additional genes and progress on this front will be reviewed."

Source: Federation of European Cancer Societies

This document is subject to copyright. Apart from any fair dealing for the purpose of private study, research, no part may be reproduced without the written permission. The content is provided for information purposes only.