Editorial
The continuing search for cancer-causing somatic mutations

Gillian L Dalgliesh and P Andrew Futreal

Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK

Corresponding author: P Andrew Futreal, paf@sanger.ac.uk

See related research article by Chanock *et al*., http://breast-cancer-research.com/content/9/1/R5

Abstract

It is known that cancer is caused by an accumulation of mutations in DNA. Many genes have been associated with tumour progression either through germline or somatic mutations, but mutations in these genes by no means account for all instances of the disease. The availability of the completed human genome sequence and reduced costs of sequencing have allowed large-scale screens to uncover genes that are somatically mutated in cancer. In this issue, Chanock and colleagues present a screen of 91 breast cancers for somatic variants in a set of 21 genes.

Over recent years, kinase genes have received much attention in this field because of their central role in many cellular processes, especially cell growth and proliferation [1-8]. Mutations in kinases have also been linked with cancer progression and have proved to be a successful target for therapeutic intervention. One such success was the treatment of *HER2/neu* overexpressing metastatic breast cancer with the anti-HER2/neu antibody trastuzumab (Herceptin). Groups are now focusing their efforts on other sets of genes to try to find new targets for cancer therapies. Chanock and colleagues [9] chose their gene set on the basis of either gene expression data, which they had previously published [10,11], or a known association with breast cancer. This is the first time that expression data has been used to guide systematic resequencing studies. A total of 87 somatic variants, spread across 16 genes, were uncovered by this study. About one quarter of the variants are in *TP53*, with the remainder spread through the other 15 mutated genes. The authors point out that there could be a proportion of rare SNPs in the reported set of 'somatic' variants because they did not have matched normal tissue for every tumour; overall conclusions relative to prevalence and pattern should therefore be drawn cautiously. The authors highlight that they uncovered non-synonymous somatic variants in genes that a recently published study by Sjöblom and colleagues [12] failed to find any variants in. They believe that the differing results may be due in part to the different oestrogen receptor status of the breast tumours used by

each group because they have previously observed differences in gene expression patterns between tumour classes. This observation serves to highlight the value of screening many different types of tumour.

Another unique point of this study is the sequencing of large areas of non-coding (intronic and 5′ and 3′ flanking) DNA, which has been largely avoided by other groups in favour of targeting resources to coding and splice site sequencing. Excluding non-coding regions could result in the overlooking of variants that alter gene expression; however, determining whether variants in non-coding regions provide a selective advantage to these tumours will be problematic. In fact, trying to discern which somatic variants have an active role in cancer progression (driver mutations) and which do not (passengers) is difficult even for coding variants. Some groups have taken a statistical approach. Greenman and colleagues [13] described methods to determine whether driver mutations are present in a mutation data set, to determine how many there are likely to be, and finally to give an indication of which genes they are likely to be in. These methods rely on screening for silent variants to obtain estimates of the background mutation prevalence, independent of selection. Comparing the observed to expected ratios of synonymous : non-synonymous variants enables selection pressures to be detected and estimated. Greenman and colleagues then used domain-specific and gene-specific methods to identify genes likely to be involved in tumour progression. Sjöblom and colleagues [12] screened for non-synonymous variants only, comparing the resulting gene-specific prevalences with background estimates from other studies to identify genes likely to be involved in tumour progression. Alternatively, various bioinformatics methods can be employed to give an indication of whether an amino acid substitution is likely to damage protein function on the basis of either conservation through species or whether or not the amino acid change is conservative. The study by Chanock and colleagues [9] favoured this type of analysis (Miyata score) to indicate which substitutions are more likely

to alter the protein structure and therefore be pathogenic. However, each group did concur that most variants uncovered through these large-scale screens are passengers and only a minority actually confer a selective advantage on the tumour. Ultimately, it is direct functional studies of genes implicated by large-scale screening projects that will yield conclusive proof of cancer gene status. This and other studies have yielded a wealth of such targets, and the ongoing work of these groups as well as the new Cancer Genome Atlas project funded by the National Institutes of Health are sure to provide many more.

Another interesting point raised by this group is the difficulty of detecting heterozygous variants in DNA samples from primary tissues because of their heterogeneous nature. They point out that pre-screening, such as the temporal temperature electrophoresis that was used on a limited set of their samples, is impractical for high-throughput screens. This problem could be tackled by developing software to improve the detection of variants by using current sequencing technologies. Alternatively, the new sequencing technologies, which allow single molecules of DNA to be analysed, could be a more attractive proposition. These new technologies should also reduce the time and cost of large-scale resequencing projects such that the sequencing of entire cancer genomes should soon be feasible.

Large-scale sequencing strategies to uncover genes that are somatically mutated in cancer are producing a wealth of such data and promise to continue doing so. All of these studies have given us greater insights into the complexity of cancer at the molecular level. It is to be hoped that the combined effort of all of the groups concerned should uncover some as yet unknown genes or groups of genes amenable to therapeutic intervention and build on previous successes to discover new treatments for cancer.

Competing interests
The authors declare that they have no competing interests.

References
1. Bignell G, Smith R, Hunter C, Stephens P, Davies H, Greenman C, Teague J, Butler A, Edkins S, Stevens C, *et al.*: **Sequence analysis of the protein kinase gene family in human testicular germ-cell tumors of adolescents and adults.** *Genes Chromosomes Cancer* 2006, **45:**42-46.
2. Davies H, Hunter C, Smith R, Stephens P, Greenman C, Bignell G, Teague J, Butler A, Edkins S, Stevens C, *et al.*: **Somatic mutations of the protein kinase gene family in human lung cancer.** *Cancer Res* 2005, **65:**7591-7595.
3. Greenman C, Stephens P, Smith R, Dalgliesh GL, Hunter C, Bignell G, Davies H, Teague J, Butler A, Stevens C, *et al.*: **Patterns of somatic mutation in human cancer genomes.** *Nature* 2007, in press.
4. Hunter C, Smith R, Cahill DP, Stephens P, Stevens C, Teague J, Greenman C, Edkins S, Bignell G, Davies H, *et al.*: **A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy.** *Cancer Res* 2006, **66:**3987-3991.
5. Stephens P, Edkins S, Davies H, Greenman C, Cox C, Hunter C, Bignell G, Teague J, Smith R, Stevens C, *et al.*: **A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer.** *Nat Genet* 2005, **37:**590-592.
6. Bardelli A, Parsons DW, Silliman N, Ptak J, Szabo S, Saha S, Markowitz S, Willson JK, Parmigiani G, Kinzler KW, *et al.*: **Mutational analysis of the tyrosine kinome in colorectal cancers.** *Science* 2003, **300:**949.
7. Parsons DW, Wang TL, Samuels Y, Bardelli A, Cummins JM, DeLong L, Silliman N, Ptak J, Szabo S, Willson JK, *et al.*: **Colorectal cancer: mutations in a signalling pathway.** *Nature* 2005, **436:**792.
8. Davies H, Bignell GR, Cox C, Stephens P, Edkins S, Clegg S, Teague J, Woffendin H, Garnett MJ, Bottomley W, *et al.*: **Mutations of the *BRAF* gene in human cancer.** *Nature* 2002, **417:**949-954.
9. Chanock SJM, Burdett LP, Yeager MP, Llaca VP, Langerod AP, Presswala S, Kaaresen RM, Strausberg RLP, Gerhard DSP, Kristensen VP, *et al.*: **Somatic sequence alterations in twenty-one genes selected by expression profile analysis of breast carcinomas.** *Breast Cancer Res* 2007, **9:**R5.
10. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, *et al.*: **Molecular portraits of human breast tumours.** *Nature* 2000, **406:**747-752.
11. Sorlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, *et al.*: **Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.** *Proc Natl Acad Sci USA* 2001, **98:**10869-10874.
12. Sjöblom T, Jones S, Wood LD, Parsons DW, Lin J, Barber TD, Mandelker D, Leary RJ, Ptak J, Silliman N, *et al.*: **The consensus coding sequences of human breast and colorectal cancers.** *Science* 2006, **314:**268-274.
13. Greenman C, Wooster R, Futreal PA, Stratton MR, Easton DF: **Statistical analysis of pathogenicity of somatic mutations in cancer.** *Genetics* 2006, **173:**2187-2198.