infected with the virus, which was unknown at the time. Five of the SV40-positive non-Hodgkin's lymphoma patients in the study of Vilchez *et al.* were born after the contaminated poliovirus vaccine was administered, indicating other mechanisms of viral transmission.

SV40 has previously been associated with solid cancers in humans, including brain tumours, osteosarcomas and malignant mesotheliomas. As it is unlikely that the presence of SV40 large T-antigen alone is sufficient to induce cancer, further work is required to determine exactly how this protein fits into the cascade of transformation events.

*Kristine Novak*

🌐 **References and links**
**ORIGINAL RESEARCH PAPER** Vilchez, R. A. *et al.* Association between simian virus 40 and non-Hodgkin lymphoma. *Lancet* **359**, 817–823 (2002) | Shivapurkar, N. *et al.* Presence of simian virus 40 DNA sequences in human lymphomas. *Lancet* **359**, 851–852 (2002)
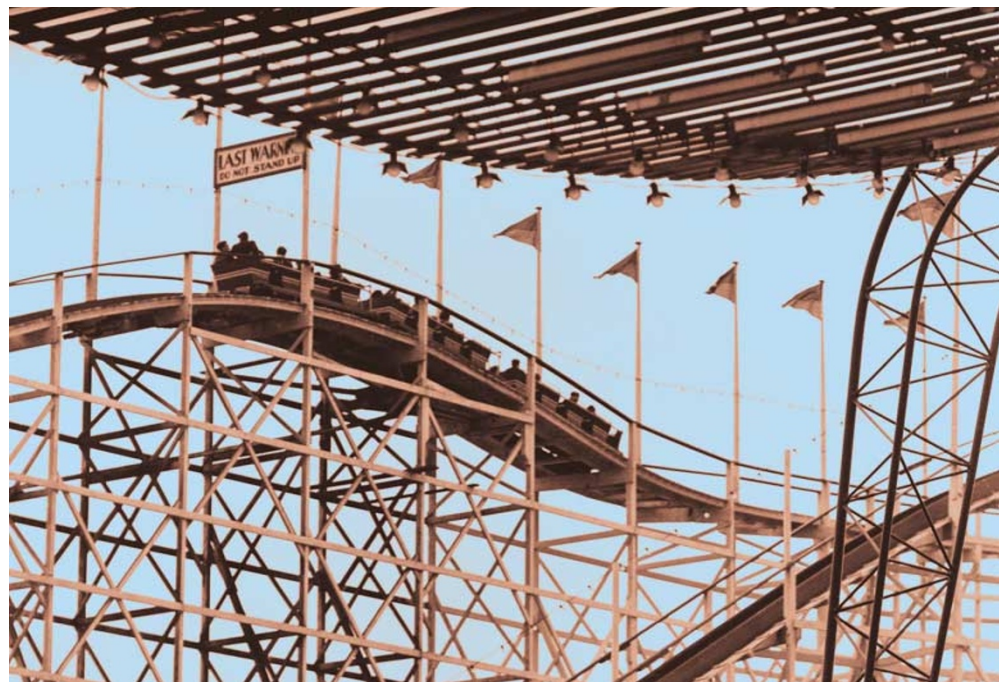**FURTHER READING**
Malkin, D. Simian virus 40 and non-Hodgkin lymphoma. *Lancet* **359**, 812–813 (2002)
**WEB SITES**
Adi Gazdar's lab:
http://www.swmed.edu/cancer/Labs/Gazdar.htm
**Regis Vilchez's department:** http://public.bcm.tmc.edu/departments/medicine.html

# Highs and lows of prediction

Breast cancer — like many forms of cancer — runs in families, and twin-studies indicate that genetics, rather than shared environmental factors, accounts for much of this familial clustering. We've identified a few genes that, when mutated, cause a vastly increased risk of breast cancer, but these genes account for only about 25% of the excess risk seen in families with a high predisposition to breast cancer. In theory, the remaining 75% of this excess risk could be accounted for by a few genes, each contributing a relatively large excess risk, or many genes, each contributing a small excess risk. If the latter case is true, will genetic testing be able to identify women with the highest risk of developing breast cancer? In the March issue of *Nature Genetics*, Paul Pharoah and colleagues argue that it will.

The authors analysed breast cancer occurrence in the relatives of nearly 1,500 individuals with breast cancer, and developed genetic models to fit breast cancer incidence in this population. Two models fit the data well. In the first, a large number of co-dominant alleles accounts for breast cancer susceptibility. Each allele is associated with a small increase in risk, but the effect of more than one allele is multiplicative. In the second model, a single, common recessive allele accounts for breast cancer susceptibility. The authors prefer the polygenic model because it better fits the data in multiple-case families: mothers and siblings have a similar excess risk in these families, which would not be the case if a recessive gene (or genes) accounted for much of the risk.

In the polygenic model, the log of the risk in the population follows a normal distribution. The higher the standard deviation about the mean, the easier it is to discriminate between individuals at high risk and those at low risk. Pharoah and colleagues estimate the standard deviation to be 1.2. If this is the case, the 20% of the population at highest risk is 40 times more likely to develop breast cancer than the 20% at lowest risk.

How does this theoretical ability to identify high-risk individuals compare with the discriminatory power of known risk factors that do not require genotyping? The authors used established risk factors — including age at menarche, number of full-term pregnancies, age at first full-term pregnancy, contraceptive use and family history — to estimate the risk distribution in the population used for the genetic modelling. Again, the distribution was log normal, but this time the standard deviation was only 0.3. This means that the 20% of the population at highest risk is only 3.5-fold more likely to develop breast cancer than the 20% at lowest risk, making it much harder to identify those at the highest risk.

But what if we cannot identify all the genetic factors responsible for the broad risk distribution seen in the polygenic model? Even if only 50% of the factors are known, the model predicts that they are still better at discriminating high- from low-risk individuals than are established non-genetic factors.

So polygenic screening could be an effective way of indentifying those individuals who would benefit most from regular screening and preventive strategies. The next challenge will be to identify these genes. This will be tough, as each gene probably contributes only a tiny proportion of each person's risk, but cancer geneticists can at least console themselves with the thought that their efforts will be of true clinical value.

*Cath Brooksbank*

🌐 **References and links**
**ORIGINAL RESEARCH PAPER** Pharoah, P. D. P. *et al. Nature Genet.* 2002 Mar 4; [epub ahead of print].
**WEB SITE**
**Bruce Ponder's lab:** http://www.hutchison mrc.cam.ac.uk/Ponder.htm