



Peter Nowell (left) and David Hungerford discovered the 'Philadelphia chromosome'.

## GENETICS

# Written in blood

*Technologies that rapidly sequence DNA reveal deep genetic diversity both within and among individuals with leukaemia.*

BY SARAH DEWEERDT

In 1959, David Hungerford of the Fox Chase Cancer Center and Peter Nowell from the University of Pennsylvania School of Medicine discovered that blood cells taken from two people with chronic myeloid leukaemia (CML) had a striking abnormality when viewed under the microscope: chromosome 22 in these cells had a big chunk missing.

That was the first glimpse of a genetic link to leukaemia — and, in fact, the first genetic abnormality associated with any form of cancer.

The stubby Philadelphia chromosome, named for the city of its discovery, is now known to be present in 95% of all people with CML.

The Philadelphia chromosome is formed when parts of chromosomes 9 and 22 swap places. This translocation brings together two genes, *BCR* and *ABL*, to create an abnormal entity, known as a fusion gene, which turns the cell malignant.

Despite the early identification of this fusion gene, knowledge of how single-gene changes contribute to leukaemia has been slow in coming. “We’ve known about these chromosome

changes for a long time, but not a lot else,” says Mel Greaves, who studies childhood leukaemia at the Institute for Cancer Research in London.

This has begun to change in the past decade or so, as high-speed DNA sequencing technology has produced a flood of information about genes mutated in various forms of leukaemia. The identification of these genes should stimulate the development of new treatments for leukaemia and other cancers.

Studies are revealing that CML, with its straightforward link to a single genetic abnormality, is an anomaly. Most forms of leukaemia are caused by a combination of genetic mutations, with considerable variability both within and among individuals. Overall, a few hundred genes — including several dozen fusion genes — have been implicated in different kinds of leukaemia. But each individual case involves only a handful of the possible mutations.

This contrasts with solid cancers, such as breast or colon cancer, in which there are often dozens of mutations in a single individual. Solid tumours also show widespread genomic instability — with duplications, losses and the exchange of large chunks of DNA — which is rarely seen in leukaemia.

The relatively small number of genetic changes per tumour may help to explain why some forms of leukaemia are so susceptible to treatment. For example, about 95% of patients diagnosed with CML are still alive after five years, thanks to drugs that target the *BCR-ABL* fusion protein (see ‘Target practice’, page S8).

## IN THE GENES

Leukaemia is classified as lymphoblastic or myeloid, according to the type of blood-cell precursors from which it arises, and as acute or chronic, depending on how aggressive it is (see ‘Living with leukaemia’, page S2). The list of contributing genes is distinct for each form of the disease, but there are substantial overlaps.

For example, the Philadelphia chromosome characteristic of CML is also found in about 5% of children with acute lymphoblastic leukaemia (ALL), the most common childhood cancer. About 25% of children with ALL carry a different chromosomal translocation, which results in the fusion of two genes called *ETV6* and *RUNX1*. *RUNX1* mutations are also common in acute myeloid leukaemia (AML), an aggressive form of the disease that primarily affects the elderly.

To find out which mutations cause disease, researchers look for ones that are ‘recurrent’, or present in multiple sufferers. In the largest leukaemia sequencing study undertaken so far, a team led by Timothy Ley, a cancer geneticist at Washington University in St Louis, Missouri, examined AML cells from 200 patients with the disease. Ley and colleagues decoded the cells’ exomes — the protein-coding portions of genomes — from 150 of the patients, and whole-genome sequences from the other 50.

They found a treasure trove of information.

Among their key findings was that 99.5% of the AML patients studied had mutations in at least one of nine categories of genes, including tumour-suppressor genes, signalling genes and those that regulate the development of myeloid cells, from which the cancer arises. “We’ve been able to define the gene sets, or pathways, that are mutated in almost every patient,” says Ley.

Notably, they found that more than three-quarters of people who develop AML have mutations in genes involved in epigenetics — chemical modifications to DNA that affect its function without altering the genetic code (see ‘Reversible tags’, page S10).

The importance of epigenetic modifications in leukaemia has been one of the most surprising insights to emerge from sequencing studies, says Ross Levine, an oncologist at Memorial Sloan-Kettering Cancer Center in New York. “These are not genes that were on anyone’s radar three or five years ago, but they’re highly recurrent,” he says.

Another class of genes commonly mutated in leukaemia are those involved in the development and differentiation of various types of blood cells. In ALL, these genes vary depending on whether the cancer arises from precursors of B or T cells, two of the main classes of lymphocytes. It takes the B-cell route in 85–90% of cases in children and 75% in adults<sup>1</sup>. It usually involves mutations in genes, such as *PAX5* or *IKZF1*, which control B-cell development. Cases triggered by T-cell mutations, on the other hand, usually involve genes in a signalling pathway called Notch, which is important in T-cell development.

### GROWING BACK

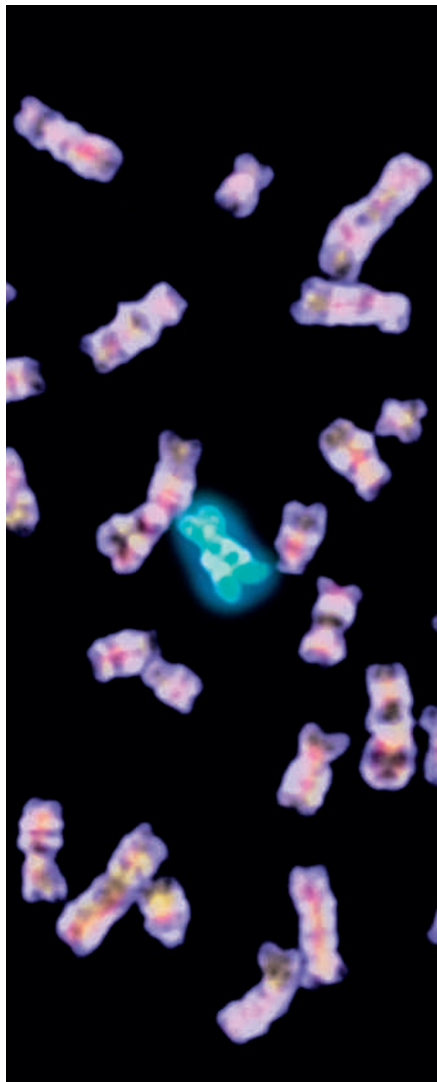
Of the four major leukaemia types, the one with the most diffuse genetics is chronic lymphocytic leukaemia (CLL). Sequencing studies suggest that a large number of genes can be involved, each mutated in a small proportion of cases. For example, mutations in a Notch pathway gene called *NOTCH1* are found in 12% of CLL cases<sup>2</sup>, and mutations in the *POT1* gene appear in another 3.5% of cases<sup>3</sup> (and 9% of those with the most aggressive form of this leukaemia).

The *POT1* gene binds to DNA at the telomere, the region at the tip of a chromosome, and protects it from damage. Telomeres have long been known to be involved in cancer, but this finding, reported in *Nature Genetics* in March 2013, is the first example of a telomere-protecting protein being implicated in cancer.

Assembling lists of genes that may be mutated in leukaemia is only the beginning of understanding its genetics. Even within a tumour there is enormous variability, with subsets of cells carrying different sets of mutations.

“Within each patient you’ve got this evolutionary tree of cancer,” says Greaves. “It’s not sufficient to list the mutations, because they segregate in different branches of the tree.”

One line of evidence for this evolutionary tree comes from studying leukaemia that



Parts of chromosomes 9 (blue) and 22 swap places to create the Philadelphia chromosome.

recurs after treatment, which is often genetically distinct from the original cancer. Greaves and his colleagues investigated this recurrent leukaemia, using fluorescent probes to analyse mutations in five people with ALL. They found that some of the tumour mutations in the initial sample had been replaced with others when they analysed the recurrent disease<sup>4</sup>. Ley’s group observed a similar pattern in AML, using whole-genome sequencing to compare DNA from leukaemia cells in eight people at diagnosis and at relapse<sup>5</sup>.

The implications of this slippery genetics for treatment are profound. Drugs targeting one set of mutations eradicate a portion of the leukaemia cells, but other subsets of cells then expand to take their place. “I think this explains a lot of treatment failures with targeted therapy,” Ley says.

Greaves compares the process to pruning a rosebush: chopping off one large branch will simply stimulate the growth of the other branches. “The challenge is to get rid of all the unwanted growth in a way that it doesn’t ever

come back,” he says. “So we need to chop it off at the base.”

In clinical terms, this means that therapies need to target not the tumour cells that are most prolific at the time of diagnosis, but rather the cells that carry the original mutations — because those mutations will be present in the entire tumour. But to do this, scientists must work out how to identify these original mutations.

Genetic variability also stems from the fact that leukaemia, like any cancer, involves two genomes: the host genome, representing the inherited set of genes, and the tumour genome, comprising the mutations acquired on the path to cancer. Most studies of genes involved in leukaemia have focused on the tumour genome.

In paediatric leukaemia, tumour genome mutations are acquired early — perhaps before birth. As lymphocyte precursors expand rapidly in the liver during fetal development, one of them may acquire a translocation or other mutation that sets it on the path to malignancy. The *ETV6-RUNX1* translocation linked to ALL, for example, has been found in newborns’ blood.

After birth, blood stem cells divide more slowly, about once a month on average, so it may take many decades for a cell to acquire the combination of mutations necessary to turn malignant. This explains why other forms of leukaemia, such as AML, primarily affect the elderly. “It really is just a disease of randomness and ageing,” Ley says.

### INHERITED RISK

Despite the importance of these acquired mutations in triggering leukaemia, some researchers prefer to focus on inherited risk. “We’ve spent a lot of time looking at the mutations in the tumour genome, but we really haven’t spent much time looking at the host genome,” says Jun Yang, a pharmaceutical scientist at St Jude Children’s Research Hospital in Memphis, Tennessee.

Yang is among those who are trying to redress the balance. In the largest study of inherited risk genes so far, published in March 2013, his team scanned the genomes of 2,450 children with ALL and 10,977 controls<sup>6</sup>.

Three of the four risk genes they identified — *IKZF1*, *CEBPE* and *CDKN2A/2B* — were already known to have a role in leukaemia or blood-cell development. The fourth, *ARID5B*, had never been identified in studies of leukaemia tumour genomes, but preliminary evidence suggests that it may be involved in lymphocyte differentiation.

The gene variations identified in the study as being associated with ALL are common: some are carried by more than 20% of the population, most of whom do not develop the disease. Still, having even one high-risk variant almost doubles the risk of ALL; carrying six or more adds up to a ninefold higher risk.

Another approach to studying the role of the host genome involves families in which several members develop the same form of leukaemia.



These families are rare: familial AML accounts for only about 1% of all cases of this cancer, for example. But working out the genetics of familial leukaemia provides powerful evidence that a given gene is truly important in the development of cancer, says Marshall Horwitz, a geneticist at the University of Washington in Seattle. Horwitz was part of a team that linked mutations in a gene called *GATA2* to familial AML. Other researchers have identified AML families with mutations in *RUNX1* and *CEBPA*.

Together, those three genes — *GATA2*, *RUNX1* and *CEBPA* — account for about half of all familial AML. Other forms of familial leukaemia are also starting to give up their genetic secrets: Horwitz says that he and a group of collaborators have unpublished data pinpointing the first gene to be implicated in familial ALL.

### MULTIPLE MUTATIONS

To understand the genetics of leukaemia, it is not enough to identify which individual mutations are possible, however. We also need to understand how multiple mutations collude to trigger a cancer (see 'Related disorders'). In mice, this process can be traced step by step.

George Vassiliou, a haematologist and cancer geneticist at the Wellcome Trust Sanger Institute in Hinxton, UK, and his colleagues generated mice with a defect in a gene called *NPML1*, which is mutated in 35% of patients with AML. About one-third of these mice develop AML late in life<sup>7</sup>. But when Vassiliou and his team induced additional mutations in blood-cell precursors, he says, "you get dramatic acceleration of leukaemia".

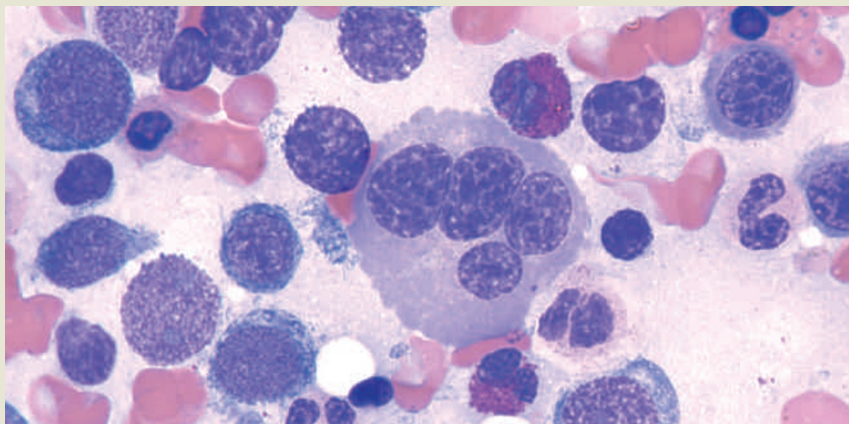
The same thing happens when the mutant mice are crossbred with mice bearing duplications of *FLT3*. These two genetic alterations are the ones most frequently seen together in AML. "Two mutations in isolation do a little bit but nothing dramatic," Vassiliou says. "Suddenly you put them together and the whole thing goes on fire, the mice get leukaemia very fast."

In humans, researchers have observed patterns in which genetic mutations are seen together in leukaemia. For example, the *ETV6-RUNX1* translocation and deletion of *PAX5* commonly occur together in B-cell ALL. These patterns affect a patient's prognosis and response to specific treatments, so researchers have begun to classify leukaemia on the basis of its genetic profile, rather than just the outward appearance of cells. "In the past ten years, we have gone from having 5 or 6 subtypes of childhood leukaemia to having 11 or 12, each of which has a different constellation of genetic changes," says Charles Mullighan, who studies the genomics of childhood leukaemia at St Jude Children's Research Hospital in Memphis.

By contrast, other mutations seem to be mutually exclusive. For example, *RUNX1* is commonly mutated in AML — but never in those AML cases involving a translocation called inversion 16. This is because inversion 16 results in a fusion gene called *CBFB-MYH11*,

## RELATED DISORDERS

### Looking for clues to leukaemia



Cells from a patient with myelodysplastic syndrome may help researchers understand leukaemia.

The genes involved in leukaemia overlap with those responsible for a group of other disorders in which the bone marrow is either unable to make healthy blood cells or makes too many of the wrong kind. Among the most important of these conditions is a non-aggressive form of cancer called myelodysplastic syndrome (MDS).

About one-third of people who have MDS go on to develop AML, and these patients have a particularly poor prognosis. A better understanding of MDS genetics might not only improve the diagnosis of MDS, but also help predict the progression from MDS to AML. More broadly, studying MDS genetics may shed light on how different combinations of mutations can make the difference between a cancer that is relatively quiescent and one that is aggressive.

Many mutations associated with AML have also been seen in people with MDS. Researchers have found that patients who have MDS and have mutations in AML-linked genes, such as *RUNX1* or *FLT3*, are more likely to progress to AML. Beyond that, however, little was known

about MDS genetics in its own right.

"Until recently, most of what we knew about the genetics of MDS was learned first in AML and other myeloid leukaemias, and then retrospectively applied to MDS," says Timothy Graubert, a haematologist at Washington University in St Louis, Missouri.

A handful of studies in the past few years has begun to identify mutations that are hallmarks of MDS. For example, more than half of those who have MDS carry a mutation in the spliceosome, a set of proteins involved in processing RNA. These mutations have also been found in other forms of leukaemia and in some solid tumours, but their link to MDS seems to be more robust. Researchers next plan to analyse how the spliceosome and other newly discovered mutations influence the prognosis of those who have MDS.

"This pathway will be important in cancer biology in general," Graubert says, "but it's the first example that we have of a genetic pathway that's disproportionately involved in MDS." — S.D.

which requires *RUNX1* to be functional to turn a cell malignant, says Pu Paul Liu, a molecular biologist at the National Human Genome Research Institute in Bethesda, Maryland, who first described this fusion gene in 1993. Liu is attempting to capitalize on this observation to develop a new treatment for this subtype of AML. He has identified a compound that disrupts the interaction between the fusion protein and *RUNX1*, and delays the development of leukaemia in a mouse model. He says he hopes to begin human trials of the drug soon.

Levine, whose lab is investigating clusters of mutations associated with poor prognosis, agrees that working out how leukaemia mutations combine is a promising approach. "Our

hope is that we're not studying a gene, we're studying a genotype," he says. "And that will lead to better models of leukaemia, and better insights into how leukaemia develops." ■

**Sarah DeWeerd** is a science writer based in Seattle, Washington.

1. Van Vlierberghe, P. & Ferrando, A. J. *Clin. Invest.* **122**, 3398–3406 (2012).
2. Puente, X. S. *et al. Nature* **475**, 101–105 (2011).
3. Ramsay, A. J. *et al. Nature Genet.* **45**, 526–530 (2013).
4. Anderson, K. *et al. Nature* **469**, 356–361 (2011).
5. Ding, L. *et al. Nature* **481**, 506–510 (2012).
6. Xu, H. *et al. J. Natl Cancer Inst.* **105**, 733–742 (2013).
7. Vassiliou, G. S. *et al. Nature Genet.* **43**, 470–475 (2011).