

## Cornelia de Lange Syndrome Gene Discovery

All of you reading this will have heard about the news of the discovery of a gene on chromosome 5 that is involved in Cornelia de Lange syndrome. Two articles describing this discovery were published in the June 2004 edition of the medical journal, *Nature Genetics*. Most people without a genetic science background would have difficulty understanding the articles, so this summary is to try to explain a little of what the research teams have shown and how they did this.

*Nature Genetics* Volume 36: Number 6: June 2004

- Krantz and others: pages 631-635  
Cornelia de Lange Syndrome is caused by mutations in *NIPBL*, the human homolog of *Drosophila melanogaster* *Nipped-B*

- Tonkin, Wang and others : pages 636-641  
*NIPBL*, encoding a homolog of *Scs-2* sister chromatid adhesion proteins and fly *Nipped-B*, is mutated in Cornelia de Lange syndrome

These articles are from the two main research teams that have been working to find the genetic cause of CdLS. One team was led by Dr Ian Krantz from The Children's Hospital of Philadelphia (CHOP), and the other was led by Dr Tom Strachan from the University of Newcastle, UK. Some of you would have met Dr Krantz and Dr Jackson when they visited Australia in **1999**, and collected research data and samples from a number of families. Dr Tom Strachan was an invited speaker at the International CdLS meeting held in Sydney in 2003.

The articles explain how each research group, using a combination of different methods, set about trying to identify the genetic changes causing CdLS. The human genome (our complete set of genetic information) is made up of DNA. It includes approximately 35,000 genes as well as long stretches of DNA between genes, all of which is packaged into 23 pairs of chromosomes. There are about 3 billion pieces of genetic code in the genome. The Human Genome Project is now at a stage where much of the DNA sequence making up the genetic code of the genome has been worked out. However, the function of most of these genes, and the correlation of particular

conditions with particular genetic changes (mutations) in genes, is far from complete. So searching for the genetic change(s) that could cause CDLS starts with trying to find which gene (or genes) were involved, then finding what sort of changes in that gene occurred in individuals with CdLS, and then looking into how those changes affect gene function.

We know that most of the time CdLS only occurs in one family member, which suggests it is due to a new genetic change in those families. In a few families, CdLS affects more than one person, and it can be passed from generation to generation. This information suggested that CdLS is an *autosomal dominant* genetic condition. That means it would be due to a genetic problem affecting a section of one chromosome of a particular pair, not both. This change was unknown, but it could have been a change within a gene (mutation), a disruption of a gene, or a change such as a deletion (missing section) or duplication (double copy) of a larger section of DNA, perhaps large enough to include several genes. Nearly all people with CdLS have normal appearing chromosomes (46 chromosomes, in 23 pairs). There are rare individuals with CdLS who have chromosome translocations, which can be clues to the location (locus) of disrupted genes (see later). There were no reported deletions or duplications that were associated with really definite CdLS (until very recently- see later). The 3q duplication patients shared many features with CdLS but some doctors felt they were not quite typical. Previous linkage studies by Dr Krantz's group, published in 2001, had shown that the 3q locus was not involved in some families with CdLS, but it was not ruled out for all.

One possibility the researchers wanted to explore was a deletion or duplication of a section of DNA that was big enough to include a gene or several genes, but too small to show up on a chromosome test. This is called a micro-deletion or micro-duplication. Dr Strachan's group reported how they looked into this possibility by using a relatively new technique, *high density BAC array*. This is a type of DNA "chip" technology, which targets thousands of small sections of DNA, chosen at points located at regular intervals across the whole set of chromosomes. This technique can be used to show if any of these randomly chosen areas is deleted or duplicated. If so, that would have been an immediate clue as to what causes CdLS. This

research did not show any hidden small deletions or duplications of the segments included on the "chip". It did not completely rule out the possibility of other very small imbalances, between the areas that were included on the "chip" but it meant this type of change was much less likely.

Dr Strachan's group also used another approach. They concentrated on 3 children with CdLS who had visible chromosomal rearrangements called reciprocal translocations. Translocations are where two pieces from two different chromosomes have broken away, then swapped places and rejoined in the new location. The "breakpoints" are regions where a gene could have been disrupted. Breakpoints do not always disrupt a gene, as they can be located in the long sections of DNA between genes. We know that most people with reciprocal translocations are healthy, so we assume that their translocations have not disrupted an important gene. We know that approximately 1 in 500 apparently healthy people have balanced translocations. By chance, someone with CdLS could also have a harmless balanced translocation that had nothing to do with their CdLS. Alternatively, a translocation in a person with CdLS could be a real clue as to where the CdLS gene could be located. Many genes have been found by following up these "coincidences" of translocations and genetic conditions.

Of the three CdLS individuals with translocations, one had a translocation involving chromosomes 3q and 17q, the next 14q and 21 q, and the last between 5p and 13q. Any of those six regions could be "candidate" regions for CdLS. Exploring a translocation breakpoint is very difficult work, sometimes taking several years, and can take a lot of time and resources. What we see as a breakpoint under a microscope leads to a rather large region at the DNA level, containing many genes. Dr Strachan's groups started with the 3q breakpoint region, as this had previously been suggested to be a good candidate region for CdLS because of the similarity to patients with 3q duplications. Much hard work later, as he announced at the CdLS meeting in Australia, they found that the 3q area in the first translocation patient did contain a big disrupted gene, but they could not show any problems in this gene in other people with CdLS.

Meanwhile, Dr Krantz's group was using another approach, which was to concentrate on families where there was more than one person with CdLS. Some families have affected siblings and in other rare families, CdLS has been passed from parent to child. Linkage analysis is a type of DNA fingerprinting, looking at patterns that identify or "mark" particular chromosome locations. These patterns often have variations between the two copies of a chromosome pair, so they can be used to track which of the two chromosomes a parent has passed on to a child. It is possible to compare these markers and work out which seem to track with the CdLS, and which do not. For example if a parent with CdLS passed on one copy of chromosome 3 to one affected child, but passed on their other copy of chromosome 3 to the other affected child, we would exclude chromosome 3 as tracking with the CdLS in that family. If each affected child got the same copy of chromosome 3 from the parent, it is still a possibility that the gene is on that chromosome 3 (but not proven). This technique can only be used when there are multiple affected family members, and the more affected people in one family, the more accurate the technique at localising a possible area. In CdLS there are few families with more than one affected person, but combining the data from all these families can also help.

Dr Krantz's team used this method to analyse the DNA from 12 families with more than one affected person. From this they were able to narrow down the possible CdLS-linked region to 4 different locations, including regions on chromosome 2q, 5p, 10p and 14q. The 5p region seemed to be the most likely, but that still did not prove it was the right area to focus on - as linkage data like this can, unfortunately, just be by chance! However, one of the translocation breakpoints in the translocation patients (which everyone knew about) happened to be in the region 5p13. Dr Krantz's team's linkage data was strongest at 5p13. This was unlikely to be just by coincidence, so they knew at that time that they were probably on the right track, and that 5p13 was the best region to work on. This work was reported in "poster" at the American Society of Human Genetics meeting in Los Angeles in November 2003.

At the same meeting, there was another important poster. Doctors from Salt Lake City, Utah reported the birth of a baby very severely affected with CdLS, but this baby had already been shown before

birth to have a visible deletion of a tiny chromosomal section including the 5p13 region. A chromosomal deletion of just that region had never been reported before. Both research groups knew what this meant: this was strong evidence that the 5p13 region must contain a CdLS gene.

The papers then describe, in VERY technical language, how each research group then intensely studied the translocation breakpoint in 5p13, to find which gene in that region (which contains many genes) was disrupted by the translocation. Eventually both research groups independently found the disrupted gene, and then showed a number of other CdLS patients had mutations in that gene, proving that it was a gene involved in CdLS.

The researchers named the gene *NIPBL*, which is short for "Nipped-B-like". This is because the gene is very similar to a gene called *Nipped-B*, previously identified in the fruit fly (scientific name *Drosophila melanogaster*). Scientists name things this way (sometimes!), to help keep track of these similarities. Research of gene variations in the fruit fly has been immensely useful to studying human genes, as there are strong basic similarities in many of the genes essential for growth and development of many living things. These genes tend to be conserved through evolution. This means they will be present and similar in a range of organisms, for example, from yeast and plants to animals and humans, because they are so important. A gene "homolog" means a similar version of a gene, in a different species. *NIPBL* (humans) is a homolog of *Nipped-B* (fruit fly) and they are homologs of *Scs-2* (the name of the yeast version of the gene).

The research groups then looked at some of the functions of the *NIPBL* gene. Genes are recipes which our cells use to make proteins, each with a particular function. Depending on the protein, it might function as a hormone (such as insulin), a "building block" in certain tissues (such as the many different collagen proteins), or a regulator of other genes (increasing or decreasing activity). Dr Strachan's group suggested that the protein product of the *NIPBL* gene be named "delangin". Newly discovered gene products are often named this way - after the condition they can be associated, with the ending "-in", which signifies it is the name of a gene product.

Some genes are always active (housekeeping genes), while others are only active at certain times in development (developmental regulators). Some genes have more than one type of function, or different effects at different times in development. The research teams showed that *NIPBL* was strongly expressed during embryo development, in the areas of the body that we know are significantly affected in CdLS, including the developing limb buds, the face and heart, so one function of *NIPBL* is as a developmental regulator. Another function of *NIPBL* was suggested by a known function of the *NIPBL* yeast homolog, *Scp2*, which was known to have some role in the way chromosomes function in the cell. .

The other aspect studied by both groups was the effect of the mutations on the gene product (delangin). Although it was all in the same gene, *NIPBL*, there were different mutations found within this gene in all the individuals or families studied. Could different mutations have different effects on the gene product, and thereby be responsible for the variation we see between different people with CdLS? Mutations in genes can stop a protein being produced at all, decrease the amount of protein produced, allow production of a faulty protein that does not work (all "loss of function"), or produce new abnormal forms of the protein that have new, unwanted functions ("gain of function"). So far the evidence about CdLS mutations suggests that they cause loss of function. When loss of function of only one copy of the gene is enough to lead to a clinical condition, this is called haploinsufficiency. The papers went on to start to explore why this haploinsufficiency could cause problems, by looking at some of the interactions delangin might have on other development regulator genes. This is where a lot of the future research will be focussed. Perhaps some of the variability in features is due to the type of *NIPBL* mutation, while other variability might be due to the way the delangin interacts in the setting of the unique general genetic background possessed by each individual.

Most of the individuals with CdLS reported in these papers were the only affected person in their family. The research showed what was expected: that the mutation in the affected person was new (*de novo*), not present in the DNA from the blood samples taken from either parent. One more important point was established by Dr

Krantz's group. Their research included several rare families where unaffected parents had more than one child with CdLS. They reported two families where each affected child had the same mutation, but examination of the parents' DNA (from blood samples) did not show the mutation. In one family, the three children had the same mother but different fathers. This suggests that the mutation was carried in a parent, but only in a limited population of cells, which in these cases must have been in the reproductive organs. "Mosaicism" is the term used to describe the genetic situation of a mixture of normal and abnormal cells in the one individual. "Germinal mosaicism" refers to mosaicism being present in the reproductive organs (ovaries or testes). We know that germinal mosaicism must be rare in CdLS, as most families do not have more than one affected child, but it does exist. This means we cannot be 100% sure that an unaffected parent, with a normal gene test from a blood sample, could not have another affected child. If a mutation has been identified in the person with CdLS, this does mean that there is a genetic test for that family, which could be used for extra reassurance in a subsequent pregnancy.

This is just the start of a new phase of research for CdLS. This gene seems to be involved in at least 50% of people so far tested in the research studies. This might not be the only gene that can be associated with CdLS - there are a few more translocations to check out!

As Dr Krantz has written elsewhere, the discovery of this gene is important for many reasons:

- To confirm the diagnosis, especially where uncertain clinically
- To offer reassurance, through genetic testing (including prenatal testing) that other family members are unaffected
- To generate interest in CdLS in the general medical and scientific community, fostering more awareness and future research
- To understand the role of this new important regulatory gene in normal and abnormal development
- To understand more about what causes CdLS, with the aim of leading to improved therapies

**Dr Meredith Wilson**  
**Head, Department of Clinical Genetics**  
**The Children's Hospital at Westmead**  
**Australian CdLS Association Medical Advisor**

**June 2004**