Professor Cisca Wijmenga: 'Our trio design will help us to determine how a certain variant fits within the paternal and maternal patterns.' (Photo: Rob Huibers)

Unique in-depth perspective on regional genetic variants

By Pieter van Megchelen

The Netherlands Genome Project ('het Genoom van Nederland') is truly unique. Never before have so many genomes from one country been sequenced using such a 'trio' design. By sequencing the genomes of 250 couples and their offspring (the trio's), this project will map genetic variation in the Netherlands. It will provide a solid foundation for 'imputing' rare variants (i.e. deriving them by computation) in the more than 100,000 samples already stored in Dutch biobanks and resulting from conventional genome-wide association (GWA) studies.

Project leader Prof. Cisca Wijmenga: 'In the next ten years, whole genome sequencing will become a standard tool in research and molecular diagnostics. We hope to kick start the nationwide implementation of such tools with this project and build up the expertise needed to enable our scientists and the public to benefit from these future developments'.

'Since the start of the Human Genome Project in 1990, we have gained many new insights into the genetic variation between individuals', says Wijmenga. 'This has been very beneficial for biomedical research. To find the biochemical pathways involved in disease, we want to study the differences between individuals with and without a specific disease. With genome-wide association studies, focussing on SNP's, we are getting closer to the genes and variants involved. But we have also found that there are major regional differences in these variants, and we now need to find the actual variants protecting groups from a disease or predisposing them to developing it - and get even closer to identifying the genes themselves.' To do this, a more in-depth way to study genomic variation is needed. The gold standard, of course, would be to sequence the complete genomes of many people. Nowadays, with faster and less expensive new sequencing techniques coming onto the market, this is becoming feasible. But brute force alone is not enough to do the job, it needs to be combined with a clever approach.

Trio design

One of the potential pitfalls in using whole genome sequences to assess genetic variation is the fact that large parts of our DNA make for very boring reading: larger and smaller sequences tend to be repeated over and over again. Because of this, it can be difficult to pinpoint the exact location of a specific nucleotide and to determine whether it is a true variation on that particular spot or simply a sequence from somewhere else. 'Our trio design, in which we analyse the genomes of two parents and one offspring, will help us to determine how a certain variant fits within the paternal and maternal patterns', says Wijmenga. Just as looking with two eyes adds depth and perspective, the trio design adds certainty about the overall sequential order.
The trio approach makes the Netherlands Genome Project unique, even in comparison with similar projects like the '1000 genomes project' (www.1000genomes.org). Its advantages in terms of reliability are obvious. So why aren't geneticists all over the world doing trios? 'Many biobanks do not have this kind of material. Dutch biobanks have collected many samples of parents and their offspring. In other countries, this has not often been done on such a large scale', ventures Wijmenga.

Large yield

In the next few months, the genomes in 750 samples from Dutch biobanks will be analysed by BGI, the Beijing Genomics Institute, based in Hong Kong, which today has the best experience in high throughput sequencing. From the autumn of 2010 onwards, Dutch scientists will be busy checking and analysing the enormous amount of data generated by their Chinese colleagues. The samples will be taken from population biobanks, representing healthy people living in all the different regions of the Netherlands (equal numbers from all the original 11 provinces and a few extra from Amsterdam and Rotterdam). The results will therefore show the amount of genetic variation among the different regions. Not just single DNA base variants (SNPs), but also insertions, deletions, copy number variations and other ways the genomes of individuals may differ. Wijmenga: 'These data will be interesting in themselves, for we truly don't know how much genomic difference there is between people from Groningen and those from the province of Zeeland. For biomedical purposes, however, there is much more to gain. If we know the specific local variants, we can go back to the data we have from over 100,000 samples already used in GWA studies. Using 'imputation' techniques, we can get a lot closer to identifying the genes involved in health and disease. It's like driving on a highway: those small hectometer signs on the side of the road show you the exact location – but only if you already know roughly where you are. In genomic data, your location may be far from trivial. That's why the sequential order and its validity are so important. Once we have more insight into the variants, we could design a cheap DNA chip for many specific regional variants, and test their association with health and disease. We are now building an enormous reference database that will be very useful in the future.'

The biobanks participating in the Netherlands Genome Project ('het Genoom van Nederland') are LifeLines, the ERGO Rotterdam Cohort Study, the Netherlands Twin Registry, and the Leiden Longevity Study. The project is supervised by a steering committee in which all the participating organizations are represented.

Link

Human genome Project - www.ornl.gov/sci/techresources/Human_Genome/home.shtml