



The Norwegian Genomics Consortium

A National Platform for High-Throughput Genomics

NGC NEWSLETTER

Nr.01/2014

The Norwegian Genomics Consortium (NGC) is a collaboration between groups at the Oslo University Hospital/University of Oslo, the Norwegian University of Science and Technology and the University of Bergen. Since its foundation in 2000, initially as the Norwegian Microarray Consortium, it has been the leading provider of high-throughput genomic analyses in Norway, running one of the most successful FUGE program technology platforms for the Research Council of Norway.

Today, NGC provides the Norwegian and international scientific community with state-of-the-art high-throughput genomic analysis services. NGC is built on three regional Genomics Core Facilities in Oslo, Bergen and Trondheim, and collaborates with the microarray unit at the University of Tromsø. Together we offer a wide range of analysis services using high-throughput sequencing (illumina and SOLiD) and commercial microarrays (Affymetrix, Agilent and illumina) to study genome structure, dynamics and function. Our services include experimental design, laboratory and bioinformatics analysis.



WHOLE EXOME SEQUENCING - NEW BOOST FOR DIAGNOSTICS AND RESEARCH ON MENDELIAN DISORDERS

By Rita Holdhus and Torunn Fiskerstrand

Genomics Core Facility, Department of Clinical Science, University of Bergen, and Center for Medical Genetics and Molecular Medicine, Haukeland University Hospital.

Next generation sequencing technology represents an important leap forward for diagnostics and research on genetic disorders. The protein coding sequences of the genome - the exome - make up only 1.5 % of the total 3.1 billion basepairs in our DNA. Because 85 % of all pathogenic variants are located in, or in close proximity to, the coding sequences, whole-exome sequencing (WES) is an efficient strategy to track down mutations causing Mendelian disorders. Despite some technical shortcomings, WES is now used in diagnostic routine in several large genetic centers around the world^{1,2}. It is particularly useful in very heterogeneous disorders, and/or when the diagnosis is difficult to establish, which is often the case with rare disorders. Notably, the molecular cause is known only for about half of the 7000 described Mendelian disorders³, and the discovery of new disease genes for these may frequently provide important knowledge also about the pathogenic mechanisms of more common diseases⁴.

In our department at Haukeland University Hospital, several teams of scientists and doctors have been using WES for nearly four years in the search for new disease genes for Mendelian disorders. In December 2012 we

started producing data in-house, using a SOLiD 5500xl system. Technical, bioinformatic, interpretive and validation pipelines for WES using this system were established, in collaboration with scientists from the Computational Biology Unit, UiB and Radboud University Nijmegen Medical Centre. Since then, WES has been performed (on this system) in DNA from 22 cases with suspected Mendelian disease, and a mutant allele highly likely to be causative have been found in six of these, with only one in a previously described disease gene (27 % molecular diagnostic rate).

Cases were thoroughly characterized clinically and genetically, and a hypothesis of inheritance pattern was established. The latter is crucial - depending on whether it is a case with suspected autosomal dominant, X-linked or autosomal recessive inheritance, the filtering strategies of the WES data will be different⁵, as well as the number of family members included in the analysis. Genetic counseling is offered to all individuals who have WES performed, and approval by the regional ethics committee is obtained as necessary. We give them the choice (written consent) of getting back information on incidental findings of rare risk variants

www.genomics.no

Parameters	#variants left
Unfiltered	57,615
Intronic and UTR variants	23,890
Synonymous variants	13,148
Known SNP's (dbSNP)	1,261
Less than 5 reads contain variant	1,090
Less than 80% of reads contain variant	127
In-house db SNP's	34
In-house db Indels	11
Manual check of variant in genome browser	3

Table 1. Filtering of WES data - recessive inheritance model

In a boy with juvenile osteoporosis, a total of 57615 variants were detected in the exome. With the hypothesis that the condition could be due to homozygosity for a variant (autosomal recessive inheritance) or due to a variant on the X chromosome, the data were filtered according to the parameters in the table. The reduction in number of candidate variants after each filtering step is shown. One of the three resulting candidate variants was located in a gene (*PLS3*) on the X chromosome, known to cause juvenile osteoporosis. Thus, a molecular diagnosis could be established for the patient.

causing severe Mendelian disease, for which there are preventive measures.

Whole exome sequencing was performed by extracting the exons in all human protein coding genes (exome), using SureSelect Human All Exon kit. The exome was then exposed to massively parallel sequencing using the SOLiD 5500xl system. On average, 40-50,000 variants are called by the algorithms per exome, but by removing all variants not very likely to be disease causing, we are usually left with 1-10 variants, depending on the type of analysis (Table 1). All common variants described in databases such as dbSNP and our in-house database were filtered out, as well as synonymous variants (not causing amino acid change) and variants in introns and UTR regions except for those affecting canonical splice sites. The possible causative pathogenic variants on our final list were subject to thorough evaluation using Alamut, a gene browser-like software with annotation gathered from public databases and different prediction tools for both missense mutations and splicing events. Also, data on gene and protein function and expression patterns, previously described pathogenic mutations in the candidate genes and any animal models are collected. The variants presumed to be highly likely detrimental were then validated by Sanger sequencing.

Most of the 22 probands investigated by WES suffered from Mendelian diseases with unknown molecular causes, i.e. the probability of identifying new disease genes was high. In several cases we used more than one test strategy, because it was difficult to determine the inheritance, e.g., autosomal dominant de novo or autosomal recessive. In one case (juvenile osteoporosis), a pathogenic allele was found in a known gene, *PLS3*, for this type of condition (Table 1). In five other cases we found a highly likely pathogenic allele in genes not previously linked to Mendelian diseases. One of these, a case with arthrogyrosis (joint stiffness) due to mutation in the *PIEZO2* gene was recently published in PNAS⁶. In the last four cases where we identified a likely pathogenic variant, additional experiments are ongoing, and/or manuscripts are in preparation.

WES technology is continuously improving to overcome challenges such as the difficulty of

sequencing regions with a higher GC-content (typically in exon 1 of the genes) or repeated sequences, and confidently detecting indels. It has been estimated that up to 80 % of mutations may be detected by exome sequencing⁵. Importantly, the technology also implies ethical challenges, particularly related to incidental findings of risk variants for other genetic conditions than the one investigated⁷. Despite much international concern, the risk of finding another rare variant causing a Mendelian condition is between 1-3 %⁸, and for parents with a severely disabled child looking for a diagnosis, the prospect of incidental findings is a minor issue in our experience.

In conclusion, WES is offering powerful new opportunities for identifying the molecular cause of Mendelian disorders, and has been introduced both in diagnostics and research on these conditions. In 22 probands with rare Mendelian disorders, we identified a highly likely molecular cause in 6 of them using WES, corresponding to a molecular detection rate of 27 %. 5 of these presumed pathogenic variants were located in genes not previously associated with a Mendelian disorder, and are subject to further investigations and publication.

References

1. Yang, Y. et al. Clinical whole-exome sequencing for the diagnosis of mendelian disorders. *N Engl J Med* 369, 1502-11 (2013).
2. de Ligt, J. et al. Diagnostic exome sequencing in persons with severe intellectual disability. *N Engl J Med* 367, 1921-9 (2012).
3. VA., M. Online Mendelian Inheritance in Man, OMIM: <http://www.ncbi.nlm.nih.gov/omim>. (2014).
4. Fiskerstrand, T. et al. Familial diarrhea syndrome caused by an activating GUCY2C mutation. *N Engl J Med* 366, 1586-95 (2012).
5. Gilissen, C., Hoischen, A., Brunner, H.G. & Veltman, J.A. Disease gene identification strategies for exome sequencing. *Eur J Hum Genet* 20, 490-7 (2012).
6. Coste, B. et al. Gain-of-function mutations in the mechanically activated ion channel PIEZO2 cause a subtype of Distal Arthrogyrosis. *Proc Natl Acad Sci U S A* 110, 4667-72 (2013).
7. Burke, W. et al. Recommendations for returning genomic incidental findings? We need to talk! *Genet Med* 15, 854-9 (2013).
8. Dorschner, M.O. et al. Actionable, pathogenic incidental findings in 1,000 participants' exomes. *Am J Hum Genet* 93, 631-40 (2013).

NGC Services

NGC services include experimental design, laboratory and bioinformatics analysis related to microarray and high-throughput sequencing.

We currently offer

- "Whole genome" gene expression, SNP genotyping and DNA methylation analysis using microarrays.
- DNA/RNA sequencing analysis, including:
 - RNA sequencing
 - Small RNA sequencing
 - Exome resequencing
 - Amplicon resequencing
 - ChIP sequencing
 - Metagenom sequencing

THE NORWEGIAN CANCER GENOMICS CONSORTIUM

By Ola Myklebost

Consortium Leader, Oslo University Hospital

www.genomics.no

The Norwegian Cancer Genomics Consortium (NCGC) was founded by a group of dedicated oncologists, genome biologists and bioinformaticians across Norway, and is now extended to include university hospitals and universities in all health regions, the Norwegian Cancer Registry, the Biotechnology Advisory Board and Oslo Cancer Cluster. The consortium has acquired two large grants of totally 75 Mkr from the Research Council of Norway to develop their "National research and innovation platform for personalized cancer medicine". The consortium is collaborating with the Cancer Society, as a representative of the patients, the bioinformatics company PubGene, and the biotech-based pharma company BerGenBio. The Clinical trial units at Haukeland and Oslo University Hospitals are also involved.

The NCGC has based their infrastructure and competence for large-scale analysis of tumour genome data on the Norwegian Genomics Consortium facilities, and has currently on-going or planned research projects where tumour mutation profiles are determined for carefully selected cohorts including, melanoma, colorectal, lymphoma, multiple myeloma, sarcoma, leukemia and prostate cancer. For the major cancer types various international studies have counted mutations already, but the quality of our clinical annotation and better selected and followed up patients give us an advantage. We focus on studies that can provide answers to important biological or clinical questions by selecting case groups, e.g. responders vs. nonresponders, time series, well-defined subtypes etc, or prospective, population-based cohorts for sarcomas, which have not yet been systematically profiled internationally.

We are still open for further studies, provided funding is available, and can collaborate on the search for biomarkers in clinical trials, provide mutation spectra of targets across our tumour types, or other sequence-based studies. We also have a work package on

preclinical validation of targets in models of the various cancer types, to validate known target mutations in new tissue contexts, or new mutations in known targets, and their therapeutic potential. Our focus is on kinases, and we are keen to collaborate with industry to develop new targets and repurpose existing therapies for additional cancers.

We participate in the international cancer genomics benchmarking study as partners of the International Cancer Genomics Consortium (ICGC.org), which has shared among 15 international teams, including ours, two data sets which have to a large extent been validated, to enable comparison and optimisation of bioinformatic analysis pipelines. Our Bioinformatics Team, headed by Prof. Eivind Hovig, is one of the partners in this effort, to ensure our methods are state-of-the-art, and we focus on decision systems for clinical implementation of tumour mutation profiling. We also work with the cancer pathologists to establish streamlined diagnostic procedures for implementation of the moderate number of target mutations that already can guide treatments, the so-called "action-ome".

For more information about project please visit our webpage cancer-genomics.no.



NGC ACHIEVES ILLUMINA CSPRO STATUS

With this NGC announces that we have completed Illumina CSp^{ro}™ certification for bead array (gene expression and SNP genotyping) and high throughput sequencing analysis, gaining entry to an elite group of Illumina genomics service providers globally.

Illumina CSp^{ro} is the collaborative service provider partnership dedicated to ensure the delivery of the highest-quality data available for genetic analysis applications. Illumina CSp^{ro} participants undergo a rigorous two-phase certification process that includes minimum data generation, data certification, and an on-site audit of the facility and processes. Illumina CSp^{ro} certifica-

tion provides a competitive advantage for service providers and ensures that customers who use Illumina CSp^{ro} services receive the industry-leading data quality and service they have come to expect from Illumina. For updated information about our services, please have a look at our web page: www.genomics.no



NOVEL GENE-FINDING APPROACH YIELDS A NEW GENE LINKED TO KEY HEART ATTACK RISK FACTOR

By Kristian Hveem

HUNT Research Centre, Department of Public Health and General Practice, Norwegian University of Science and Technology

Norwegian researchers at Norwegian University of Science and Technology (NTNU) and the Arctic University of Tromsø (UiT) together with researchers at the University of Michigan (UoM) make discovery by scanning genome for variations in genetic code that change proteins.

The Norwegian (PI K Hveem) and US-group (PI Cristen Willer) have discovered a previously unrecognized gene variation that makes humans have healthier blood lipid levels and reduced risk of heart attacks - a finding that opens the door to using this knowledge in testing or treatment of high cholesterol and other lipid disorders. But even more significant is how they found the gene, which had been hiding in plain sight in previous hunts for genes that influence cardiovascular risk. This region of DNA where it was found had been implicated as being important in controlling blood lipid levels in a previous report from several members of the same research group. But although this DNA region contains many genes, none of them had any obvious link to blood lipid levels. The promise of an entirely new lipid-related gene took another six years and a new approach to find. In a recent number of Nature Genetics¹ the team from NTNU, UiT and UoM report that they zeroed in on the gene in an entirely new way.

Genetic information available from a biobank of thousands of Norwegians, the HUNT biobank ("European Research Biobank of the Year" in 2013.), was genotyped at the Genomics Core Facility (NTNU) focusing on variations in genes based on the Exome Chip (Illumina®).

Most of the findings turned out to be already known to affect cholesterol levels and other blood lipids. But one gene, dubbed *TM6SF2*, wasn't on the radar at all. In a minority of the Norwegians who carried a particular change in the gene, blood lipid levels were

much healthier and they had a lower rate of heart attack. And when the researchers boosted or suppressed the gene in mice, they saw the same effect on the animals' blood lipid levels.

Intriguingly, OL Holmen et al suggested that the same gene may also be involved in regulating lipid levels in the liver - a finding confirmed by the observations of a team led by Jonathan Cohen and Helen Hobbs who propose a role for the gene in liver disease in the same issue of Nature Genetics². More research into the exact function of this protein will be needed to understand the role it plays in these two diseases, and whether it can be targeted with new drug therapies to reduce risk -- or treat -- one or both diseases.

The success of the scientific experiment was also due to efficient screening of thousands of Norwegian samples and clinical information amassed over a 30-year period by the The Nord-Trøndelag Health Study (HUNT) and the Tromsø Study, where the initial findings were replicated. The combination of large population-based studies and the rapid development in genotyping technologies will probably be very helpful in understanding a great deal more about cardiovascular disease, and other diseases, in the next few years.

Reference:

- 1) Holmen OL, Zhang H, Fan Y, Hovelson DH, Schmidt EM, Zhou W, Guo Y, Zhang J, Langhammer A, Løchen ML, Ganesh SK, Vatlen L, Skorpen F, Dalen H, Zhang J, Pennathur S, Chen J, Platou C, Mathiesen EB, Wilsgaard T, Njølstad I, Boehnke M, Chen YE, Abecasis GR, Hveem K, Willer CJ. Systematic evaluation of coding variation identifies a candidate causal variant in *TM6SF2* influencing total cholesterol and myocardial infarction risk. Nat Genet. 2014 Apr;46(4):345-51
- 2) Kozlitina J, Smagris E, Stender S, Nordestgaard BG, Zhou HH, Tybjaerg-Hansen A, Vogt TF, Hobbs HH, Cohen JC. Exome-wide association study identifies a *TM6SF2* variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2014 Apr;46(4):352-6

www.genomics.no



Norwegian Genomics Consortium
 Bergen +47 55975326
 Oslo +47 22781769
 Trondheim +47 72825345
 NGC Newsletter Editor (Vidar Beisvåg)

www.genomics.no
 bergen@genomics.no
 oslo@genomics.no
 trondheim@genomics.no
 vidar.beisvag@ntnu.no