The Norwegian Genomics Consortium
A National Platform for High-Throughput Genomics

NGC NEWSLETTER Nr.01/2015

The Norwegian Genomics Consortium (NGC) is a collaboration between nodes 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 a leading provider of high-throughput genomic analyses in 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 commercial microarrays (Affymetrix, Agilent and Illumina) to study genome structure, dynamics and function. Our services include experimental design, laboratory and bioinformatics analysis.

www.genomics.no

Liquid biopsies: A non-invasive insight into the tumour

By Heidi M. Namløs[1] and Leonardo A. Meza-Zepeda[1,2,3]
Departments of [1]Tumor Biology and for [2]Core Facilities, Institute for Cancer Research, Oslo University Hospital. [3]Norwegian Cancer Genomics Consortium.

Cancers arise from accumulative alterations of the genome. As the patient is treated, the tumour evolves to circumvent treatment, gathering new genomic changes that convey resistance and allow the tumour to further progress and evolve. Traditional examination of the tumour is normally performed through needle biopsies or surgical specimens, and these invasive methods limit the examination of the tumour and only provide a snapshot of a specific tumour in space and time.

In cancer patients, DNA from apoptotic or necrotic cancer cells is shed into the bloodstream and circulates freely together with DNA shed by normal dead cells. The circulating tumour DNA (ctDNA) in plasma has been shown to contain the various tumour-specific alterations seen in the primary and metastatic tumours, and may more accurately represent the genetic profile of the whole tumour. For this reason, ctDNA can be seen as a "liquid biopsy" of the tumour, and provides an alternative, non-invasive method of tumour sampling to traditional tumour biopsies. Several recent reports have demonstrated that sequencing of plasma ctDNA represents a novel paradigm for non-invasive characterization of cancer, and ctDNA has been used for prognostication, prediction and monitoring of therapeutic response, resistance and to study tumour evolution. High-throughput sequencing of ctDNA has been performed for ovarian and breast cancer patients, showing that levels of mutant alleles reflect the clinical course of the disease and its treatment. Also, analysis of serial plasma ctDNA has identified mutations associated with acquired drug resistance in advanced cancers.

Levels of ctDNA vary a lot from person to person and can be hard to detect, especially for small tumours in their early stages, and commonly make up barely 1% and possibly as little as 0.001% of the total DNA. Thus, the blood samples have to be collected using specific procedures to preserve ctDNA and reduce genomic DNA contamination from white blood cells. Technological advancements such as droplet digital PCR and high-throughput sequencing, also combined with allele-specific PCR, have allowed for unprecedented progress in this field, giving up to 0.001% sensitivity.

Our group has started a clinical protocol to evaluate the clinical impact of ctDNA as biomarker for disease monitoring in soft tissue sarcomas (CircSarc). The prospective study has been implemented together with surgeons, oncologists and pathologists, and today samples are being collected at Oslo and Haukeland University Hospitals. For each patient, tumour material is being collected at the time of surgery, as well as blood samples before and after surgery, before and after each treatment cycle and at each routine control. The aim of this study is to show that targeted resequencing of ctDNA from plasma can be used to indicate tumour burden in soft tissue sarcoma. With high sensitivity, we aim to identify recurrent disease prior to changes in clinical symptoms, which increases the possibility of an early intervention. We have established exon and targeted sequencing-based methodologies to identify somatic mutations in the tumour. We are currently evaluating different sequencing approaches to identify the mutated regions in ctDNA from plasma and to follow them over time.

Preliminary work shows that ctDNA can be detected in plasma from soft tissue sarcoma. In a proof-of-concept study, we have analysed the primary tumour and normal material as well as serial plasma samples from a patient with high-grade malignant spindle cell sarcoma. The tumour was sequenced using a Norwegian Cancer Genomics Consortium (NCGC, cancergenomics.no) custom cancer gene panel enriching for exons of 900 cancer-related genes, as well as selected promoters and introns frequently involved in fusions. Eight somatic mutations were identified in the primary tumour, of which six were detected in ctDNA from plasma using the same method. One of the mutations, KRAS c.35G>T (p.G12V), was also confirmed in plasma using the ThunderBolts Cancer panel from Raindance technologies (Figure 1). The level of ctDNA in plasma was represented by the allele frequency of the mutated genes, and we see that the level in plasma at the time of surgery was comparable with the primary tumour. Three days after surgery, low levels of mutations could still be detected in plasma, reflecting persistent microscopic disease. After six weeks, the allele frequency in plasma had increased to a similar level as before surgery, consistent with the presence of widespread metastatic disease in the patient. These findings show that levels of tumour-specific mutations in liquid biopsies can be correlated to clinical manifestation of metastatic disease in sarcomas, and have the potential to detect disease progression at an early stage.

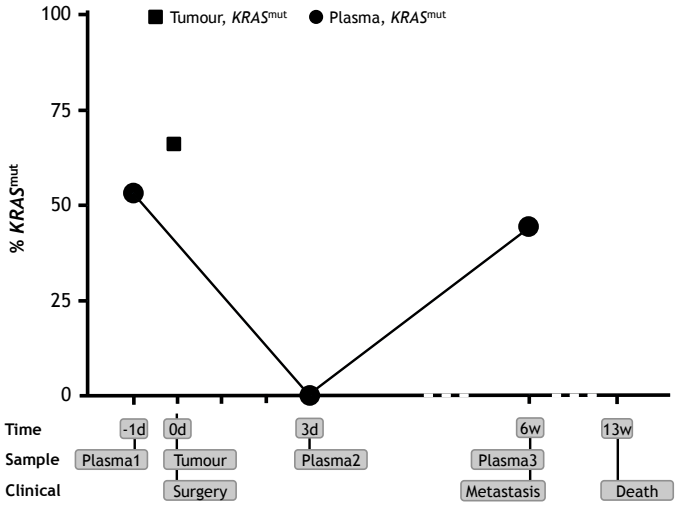


Figure 1: Levels of dtDNA in serial samples during disease progression.

The technology established by the CircSarc project constitutes a universal platform that can be easily expanded and adapted to other tumour forms. At present, we are establishing the use of this technology for other cancer types, including lung and ovarian cancer. The CircSarc project is part of NCGC and financed by the Norwegian Cancer Society and NCGC grants from the Research Council of Norway.

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

New funding to research infrastructure

The National consortium for sequencing and personalized medicine (NCS-PM) has been granted 60 MNOK for large-scale research infrastructure from the Research Council of Norway. The consortium is a collaboration between the Norwegian Sequencing Center and the Norwegian Genomics Consortium aiming to bring genome sequencing towards personalised medicine.

www.genomics.no

RNA SEQUENCING OF LASER MICRODISSECTED TISSUE – TOWARDS RESOLVING THE COMPLEX TISSUE PROBLEM

By Atle van Beelen Granlund[1,2]
[1]Institute for Cancer Research and Molecular Medicine, NTNU, Trondheim, Norway
[2]Centre of Molecular Inflammation Research, NTNU, Trondheim, Norway

Most gene expression studies of human and animal specimen are done on samples consisting of several cell and tissue types, i.e. complex tissue. This poses challenges when interpreting data from the analysis, as differences in sample cell type composition often introduces data variance unrelated to the subject you are trying to study. For instance, when contrasting the gene expression analysis of inflamed colonic tissue with healthy, the dominating difference in expression will come from the fact that mononuclear cells infiltrate the inflamed tissue, while largely lack in the healthy controls. The result is an analysis ideally suited for the study of the inflammatory processes in action. But if your motivation was to understand the epithelial processes that might have allowed for initiation of the inflammation, this analysis would be poorly suited. The epithelial contribution to gene expression levels would drown in the clutter of thousands of differentially expressed genes arising as a result of differences in mononuclear infiltrate. Furthermore, if we wanted to identify which cell or tissue type that contribute to the expression, we would have to do some type of histological evaluation of the sample, a process that can be time-consuming and expensive when faced with thousands of differentially expressed genes.

Since the emergence of Laser Capture Microdissection (LCM) it has been suggested that this method could solve some of the problems associated with the analysis of complex tissue. LCM lets the user isolate minute parts of tissue sections for subsequent RNA, DNA or protein isolation, allowing for the focused analysis of sample material such as microcarcinoma, renal structures and even single cells. Although LCM solves the initial challenge of separating the different cell and tissue types in a heterogeneous sample, it is not trivial to analyse the isolated tissue. Formalin fixation prior to LCM will considerably hamper the efficacy and accuracy of downstream analysis of proteins, DNA and RNA, and special precautions have to be made during isolation of biomolecules from such samples. If isolation from frozen tissue is an option, you still have to grapple with rapid degradation of nucleic acids in the conditions imposed on the sample during LCM. The RNA integrity of samples isolated from LCM target varies widely as a result of the variation of nucleases activity in the tissue. For whole genome gene expression (WGGE) array analysis an RNA integrity number (RIN) >7 would generally be needed. And while RNA-Seq is less sensitive to RNA degradation, you will still need a way to select mRNAs from the samples as total RNA samples consist of >90% rRNA. Both specific rRNA removal and mRNA capture by poly-T priming are viable alternatives for mRNA selection.

With the introduction of the RNA Access TruSeq library preparation kit by Illumina, a third option for selection of mRNA for RNA-Seq is available. Originally intended for use on RNA isolated from formalin fixed samples, the RNA Access protocol includes a selection of mRNA through the use of exon-specific capture sequences. The sequencing of a pull-down using these probes will according to the manufacturer cover expression analysis of 98.2% of RefSeq exons. In our initial tests of this system, we have established a protocol using RNA Access to perform RNA sequencing of samples isolated using LCM on frozen sections of colonic epithelium. In this analysis, the epithelial monolayer of IBD patients was contrasted with normal controls. Approximately 10 000 cells were isolated from each frozen section using LCM (CellCut plus LCM microscope), and total RNA was isolated. Although RIN values were poor, a quality assessment focusing on determining the fraction of RNA fragments longer than 200 nucleotides (DV200) showed a satisfactory result fully compatible with the RNA Access protocol. While still in the initial parts of analysis, it is clear that the analysis identified > 4000 differentially expressed genes when contrasting IBD with normal controls. Furthermore, the results show an enrichment of epithelial-related expression and a lack of expression related to infiltrating blood cells, suggesting that the microdissection successfully minimizes contribution from the mononuclear infiltrate to the expression analysis. We hope that this focused analysis will help us understand more about the epithelial processes associated with IBD initiation and perpetuation.

Although LCM has been around for over ten years, it has been difficult to couple its use with whole genome analysis. With the broader use of sequencing technologies, and the introduction of target preparation kits suitable for the analysis of degraded samples, we believe that we might be closer than ever to a viable solution to the complex tissue problem.

RareVariantVis: identification of causitative variants in rare monogenic disorders from whole genome sequencing data

By Tomasz Stokowy[1] and Rita Holdhus[1,2]: [1]Department of Clinical Science, University of Bergen and [2]NGC

The search for causative genetic variants in rare diseases with unknown etiology of presumed monogenic inheritance has been boosted by the implementation of whole exome (WES) and whole genome (WGS) sequencing. WGS seems to be superior to WES thanks to equally distributed coverage and possibilities of non-coding variant analysis, but the analysis and visualization of the vast amounts of data is demanding. To meet this challenge, we have developed a new tool – RareVariantVis – for analysis of genome sequence data The tool visualizes variants along their respective chromosomes, providing information about exact chromosomal position, zygosity (i.e. percentage of variant reads) and frequency, with click-able information regarding dbSNP IDs, gene association and whether the variant is inherited from the mother, father or both. Rare variants with no dbSNP ID or frequency below a certain user-defined threshold as well as de-novo variants can be flagged and visualized in different colors.

Thanks to the new method we discovered and successfully confirmed causes of 3 rare diseases for which causative mutation was not known. Some of these diseases were studied in Bergen for years, without success. We have also promising candidates for other disease samples, however the most recent results still need verification.

Our package supports both germ line and somatic variant analysis, also in non-coding regions of genome. The RareVariantVis tool accepts vcf files and annotated variant tables. It can be efficiently run on a desktop computer.

The tool with its documentation is available for download under the following link: http://bioconductor.jp/packages/3.2/bioc/html/RareVariantVis.html

Figure1

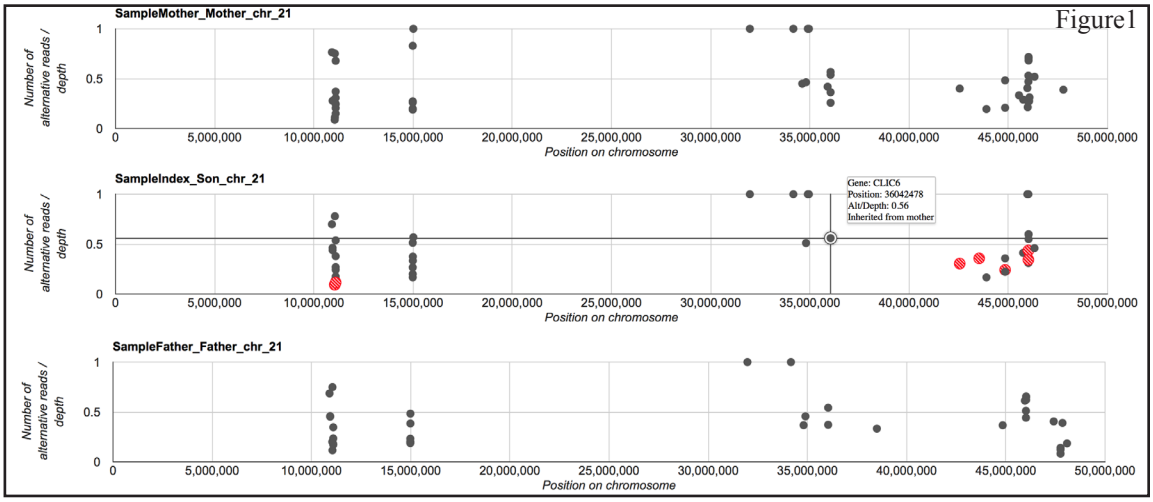
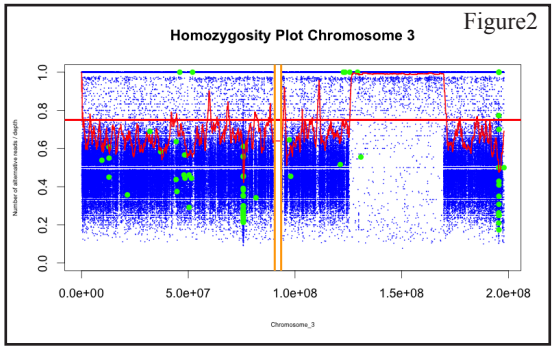


Figure2



Two of the plots produced by the RareVariantVis tool.
Figure 1: TrioVis plot. Rare variants plotted on chromosome 21 for sick child and healthy parents. Grey dots represent variants, whereas red dots represent possible de-novo variants.
Figure 2: Homozygosity plot. For recessive monogenic diseases finding homozygous regions is important. Here all variants are plotted and a moving average is calculated. This plots clearly shows a homozygous region on chromosome 3 for this sample.

www.genomics.no

Norwegian Genomics Consortium www.genomics.no
Bergen +47 55975326 bergen@genomics.no
Oslo +47 22781769 oslo@genomics.no
Trondheim +47 72825345 trondheim@genomics.no
NGC Newsletter Editor (Rita Holdhus) Rita.Holdhus@uib.no