Nordic Alliance for Clinical Genomics

WORKSHOP REPORT

NACG 11th workshop, 25. Nov – 26. Nov 2021

About NACG

The Nordic Alliance for Clinical Genomics (NACG) is an independent, non-governmental, not-for-profit Nordic association. NACG gathers stakeholders in clinical genomics who collaborate to identify and address emerging challenges to the implementation of clinical genomics and precision medicine. NACG partners collaborate to identify and address emerging challenges to the implementation of clinical genomics and precision medicine. Learn more about the Nordic Alliance for Clinical Genomics at https://nordicclinicalgenomics.org/ or contact us at post@nordicclinicalgenomics.org.

Mission

NACG partners work together and learn from each other to lift performance standards. We aim at responsible sharing of trustworthy data for improved diagnosis and treatment, and as a resource for research.

Goals and activities

Facilitate the responsible sharing of genomic data, bioinformatics tools, sequencing methods, and best practices for interpretation of genomic data.

Enhance quality of genomic data and processes and explore methodologies to provide assurance.

Understand legal barriers to the implementation of personalized medicine and to engage with key stakeholders that influence these barriers

Develop demonstration projects that challenge perceived legal barriers that limit responsible and ethical sharing of genomic and health data.

Build bridges between research and clinical communities, technologies, and practices to foster innovation

Date of issue	Rev.	Prepared by
11.01.2022	1	Bobbie N. Ray-Sannerud

Symbols

Lecture / presentation

Interactive workshop

Abbreviations

ADA	Assurance of digital assets
API	Application programming interface
CDS	Clinical decision support
CNV	Copy number variation
DC	Dynamic consent
DV	Deepvariant
FFPE	Formalin fixed paraffin embedded
FIMM	Institute for Molecular Medicine Finland
FP	False positive
GATK	Genome Analysis Toolkit
GA4GH	The Global Alliance for Genomics and Health
GIAB	Genome in a bottle
GMS	Genomic Medicine Sweden
hg38	human genome assembly GRCh38 (hg38) from Genome Reference Consortium
HRD	Homologous recombination deficiency
HUS	Helsinki University Hospital
IRDiRC	The International Rare Diseases Research Consortium
ISO	International Organization for Standardization
IVDR	EU In-Vitro Diagnostics Medical Device Regulation
MDR	EU Medical Device Regulation
MME	Match Maker Exchange
MMR	Mis-match repair
MOMA	Department of Molecular Medicine
MSI	Microsatellite instability
NACG	Nordic Alliance for Clinical Genomics
NGC	National Genome Centre
NGS	Next generation sequencing
OUS AMG	Oslo University Hospital – Department for Medical Genetics
PM	Precision medicine
PoN	panels of normal
RP	Retinitis pigmentosa
SNV	Single nucleotide variation
SOP	Standard operating procedures
SSI	Statens Serum Institut
SV	Structural variant
SVDB	Structural variant database
TMB	Tumour mutational burden
VE	Variant Exchange
VUS	Variant of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing

Contents

Executive summary

This report summarizes the 11th workshop of the Nordic Alliance for Clinical Genomics (NACG). Following recovery from the global pandemic situation, the workshop was organized as a physical two-day workshop 25. – 26. November 2021, in Århus, Denmark.

This workshop was attended by 100 registered participants[1] from 25 organizations, from eight countries (Nordics, Malta, Germany, and Spain), representing healthcare providers, governmental organizations, research, and industry.

The objective of this workshop was to progress NACG work to share experiences, data, and best practices relevant for the clinical implementation of genomics, and to collaboratively explore pain points in producing and using genomic data to the best of the patient (Figure 1).

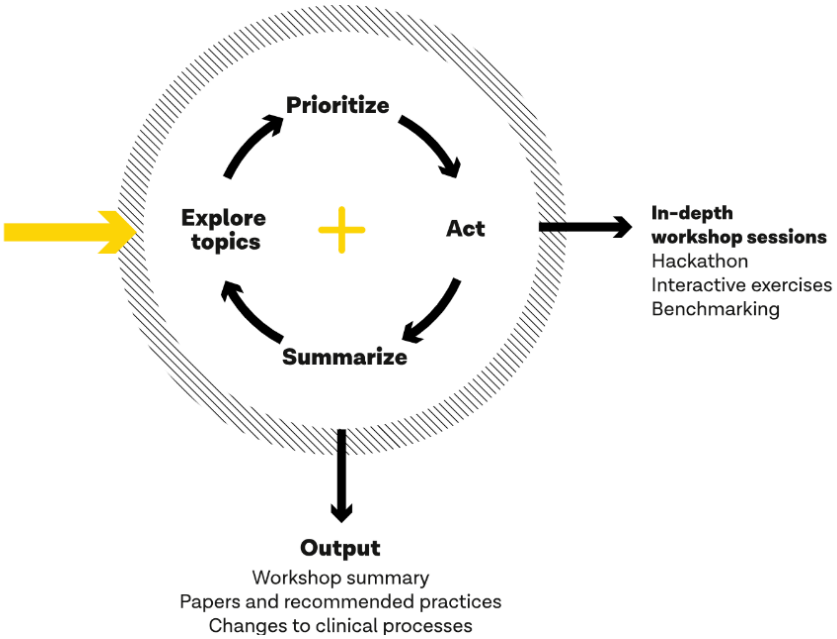


Figure 1 NACG members discuss and explore topics of interest to identify shared challenges and strategies for overcoming them. Prioritized topics are explored in in-depth interactive exercises. Findings and learnings are summarized in workshop summary reports and collaborative papers and contribute to lifting performance standards.

[1] Actual workshop attendance: Keynote: Covid sequencing and monitoring in DK (86), National updates from Nordic representatives (90), Challenging clinical cases (61), IVDR: status update, knowledge sharing and potential Nordic collaboration (61), Overcoming technical and legal barriers to sharing of variant classifications between labs (24), NACG knowledge sharing (82), Transitioning to hg38 reference genome (28), Identification of somatic SVs and complex biomarkers for NGS-based cancer diagnostics (40).

NACG workshop agenda - Day 1

8:30 Coffee – Breakout area	
9:00 Welcome Dag E. Undlien, OUS Department of Medical Genetics, NO, NACG chair Ole Halfdan Larsen, Head of Department of Molecular Medicine/MOMA, AUH, DK	
09:30 Keynote: Covid sequencing and monitoring in DK Morten Rasmussen, Statens Serum Institut, Denmark	
10:15 Coffee – Breakout area	
10:30 National updates from Nordic representatives Dag Undlien (Norway), Valtteri Wirta (Sweden), Cathrine Jespersgaard (Denmark), Jón Jóhannes Jónsson & Eiríkur Briem (Iceland), Anna-Kaisa Anttonen (Finland)	
12:00 Lunch	
12:55 Implementing sharing of variant classifications via the Variant exchange API Tony Håndstad, OUS, NO	**12:55 Challenging clinical cases** Maria Rossing, Center for Genomic Medicine, Centre of Diagnostic Investigations, Rigshospitalet DK *Case contributions from:* Jón Jóhannes Jónsson and Eirny Thorolfsdóttir (Iceland) Kathrine Bjørgo (Norway) Maria Bach Laursen (Denmark, Århus) Elsebeth Østergaard (Denmark, Copenhagen) Morten Dunø (Denmark, Copenhagen)
	14:30 IVDR: status update, knowledge sharing and potential Nordic collaboration Cathrine Høgseth Nordhus, Department of Medical Genetics, OUS NO and Courtney Nadeau, DNV NO
16:00 Coffee	
16:15 Overcoming technical and legal barriers to sharing of variant classifications between labs followed by a NACG ideation exercise Tony Håndstad, Department of Medical Genetics, OUS NO	
16:45 Ideation exercise for next workshop Bobbie Ray-Sannerud, DNV NO	
17:15 Conclusion of day 1	
19:00 Workshop dinner	

NACG workshop agenda - Day 2

8:00 Coffee – Breakout area	
8:25 NACG knowledge sharing *Contributions from:* Ram Neethiraj, SciLifeLab, SE: **Deep variant in a rare disease bioinformatic workflow** Courtney Nadeau, DNV, NO: **Using assurance to accelerate dynamic consent** Mei Wu, SciLifeLab, SE: **New bioinformatic workflow for rare disease diagnostics in Nextflow** Michael Knudsen, MOMA, DK: **Structural Variants at MOMA** Morten Eike, OUS, NO: **ELLA - Recent developments** Henrik Stranneheim, SciLifeLab, SE: **PatientMatcher - A Standalone MatchMaker Exchange Server** Lusine Nazaryan-Petersen, Rigs, DK: **PacBio HiFi sequencing for detection of genetic variants in the hard-to-sequence genomic regions: case report** Tony Håndstad, OUS, NO: **Data storage policy and compression** Anders Jemt, SciLifeLab, SE: **RNAseq in WGS-based RD diagnostics - Experiences from Clinical Genomics Stockholm**	
9:45 Coffee – Breakout area	
10:00 Transitioning to hg38 reference genome Courtney Nadeau, DNV, NO, and Kaisa Kettunen, HUS, FI	**10:00 Identification of somatic SVs and complex biomarkers for NGS-based cancer diagnostics** Valtteri Wirta, SciLifeLab, SE and Oleg Agafonov, DNV, NO
12:00 Lunch	
13:00 Conclusion of workshop	

Description of workshop sessions

Topic	Description	Contact person
National updates from Nordic representatives	At this session, representatives from the Nordic countries will provide national updates related to clinical genomics.	NACG Nordic representatives
Implementing sharing of variant classifications via the Variant Exchange API	Variant Exchange is a solution for sharing variant classifications. This workshop will introduce web APIs in general and how to communicate with them using the Python programming language. We will then have a practical session where the purpose is to learn how to use the Variant Exchange API to automatically upload variant classifications. The general introduction should be accessible to everyone, and the practical session will provide bioinformaticians from each lab with the foundation to more easily get started with sharing variant classifications.	Tony Håndstad (tony.handstad@medisin.uio.no), Øyvind Evju, Tor Solli-Nowlan and Marlon Polo de Melo
Challenging clinical cases	The session will cover case presentations from various clinicians &/or bioinformaticians on clinical diagnostic unsolved cases.	Maria Rossing (caroline.maria.rossing@regionh.dk)
IVDR: status update, knowledge sharing and potential Nordic collaboration	This interactive workshop will focus on sharing updates and knowledge around compliance to the requirements for in- house exemption as it applies to a) the documents that might fulfil General Safety and Performance Requirements (procedures, validation reports, etc.) and b) risk assessments of inhouse methods. The workshop will also focus on IVD performance assessment for inhouse methods and compare how various labs plan to monitor performance for their in-house methods.	Cathrine Høgseth Nordhus (cahnor@ous-hf.no) and Courtney Nadeau (Courtney.David.Nadeau@dnv.com)
Overcoming technical and legal barriers to sharing of variant classifications between labs	This wrap-up session will include a summary of the variant sharing hackathon, discussion of barriers and next steps.	Tony Håndstad (tony.handstad@medisin.uio.no),
NACG Knowledge Sharing	An interactive session where NACG members will share projects and learnings.	NACG members
Transitioning to hg38 reference genome	This interactive session will allow participants to share experiences, identify common issues and pitfalls, and discuss strategies and approaches to overcoming these that clinical genomic labs can apply for a smooth transition to the hg38 reference genome. This includes for example issues such as dependencies with external resources and which methods to transition when.	Kaisa Kettunen (kaisa.kettunen@hus.fi) and Courtney Nadeau (Courtney.David.Nadeau@dnv.com)
Identification of somatic SVs and complex biomarkers for NGS-based cancer diagnostics	Structural variants comprise a large fraction of variation in cancer genomes and play a significant role in cancer development. Nevertheless, identification of these variants is challenging due to tumour heterogeneity and sample purity. In addition to identifying SNVs and SVs, NGS technology allows the identification of genome-wide biomarkers, broadly termed mutational patterns, and signatures. This session will allow participants to share experiences and identify common issues of identification of SVs and complex oncomarkers, such as HRD, MMR, TMB and MSI.	Valtteri Wirta (valtteri.wirta@scilifelab.se) & Oleg Agafonov (oleg.agafonov@dnv.com)

NACG opening & keynote

Number of participants: 86

Speaker	Dag E. Undlien, OUS AMG & NACG steering committee chair
Objective	Welcome, opening remarks and present information on status and development of NACG

Key information

Dag welcomed the 11th NACG workshop and introduced the organisation and community, as well as the ambitions for the two days.

The Nordic Alliance for Clinical Genomics (NACG) is an independent, non-governmental, not-for-profit Nordic association. NACG partners work together and learn from each other to lift performance standards. We aim at responsible sharing of trustworthy data for improved diagnosis and treatment, and as a resource for research. Dates for the next physical workshop, potentially to be held in Iceland, are the 28-29th April 2022.

NACG has a large number of members, and non-members are encouraged to get involved if they aren't currently. NACG is open to organisations, institutions, and individual members. NACG success and failure is dependent on us: committed people must help us progress. The 11th workshop is the largest physical workshop in our history, with good/broad representation.

Dag gave an overview of the NACG steering committee, and highlighted that decisions are made by consensus.

Special thanks were also given to our sponsors, MSD and Novogene.

Speaker	Ole Halfdan Larsen, Head of Department of Molecular Medicine/MOMA, AUH, DK
Objective	Welcome

Key information

Ole introduced the workshop participants to Denmark, Århus and the local clinical genetics facilities.

The department of molecular medicine (MOMA), Århus, contains the National Centre for whole genome sequencing and works to develop synergies between the diagnostic and research units.

Keynote

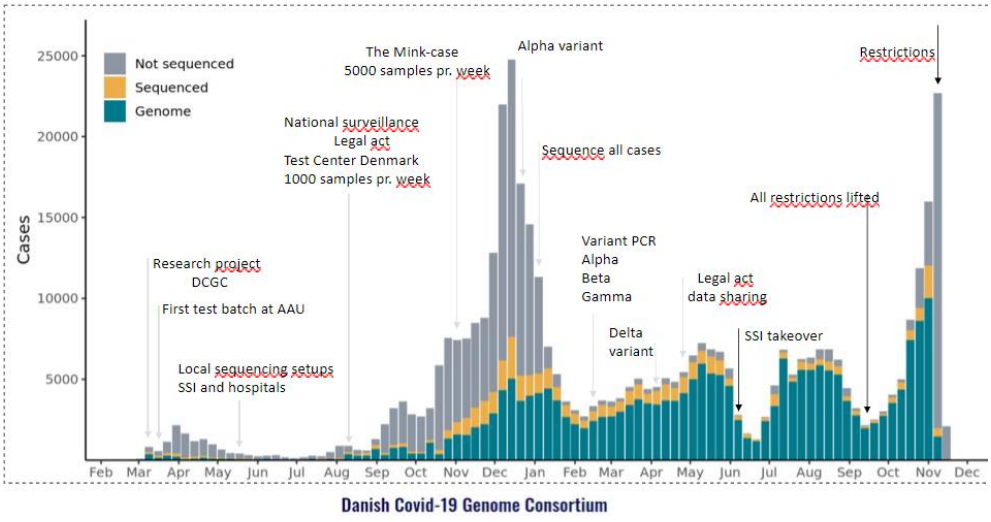
Number of participants: 86

	Speaker	Morten Rasmussen, Statens Serum Institut, Denmark
	Title	Covid sequencing and monitoring in DK

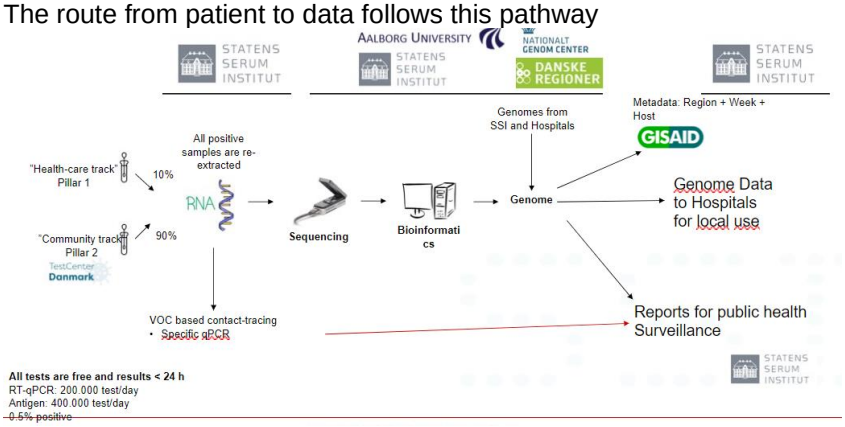
Key information:

Morten presented on the topic of SARS-covid sequencing timeline in DK.

- Confronted with legal and compliance barriers at the start of pandemic to work through the number of samples (1K samples / week).
- Mink cases/outbreaks sequenced, farmers testing positive this past summer. Results of sequencing was part of the data used to evaluate the situation that ended with the decision to exterminate the minks.
- In April 2021 DK data could be legally shared (bek nr777 af 29/04/2021)
- Sample and data flow currently has two tracks. 1) healthcare track and 2) community track all of which are via Illumina sequencing.



The route from patient to data follows this pathway



Data usage is via

- Linelist" data integration
  - Integration of WGS data and metadata
- SUM/STM reports
  - Status of sequencing coverage and VOCs
  - Detecting and monitoring emerging variants

- Signal reports
    - Weekly report covering new signals
    - Weekly reports for WHO variant meeting
- VOC identification-> contact tracing (STPS)
    - Either based on qPCR or WGS data
- Outbreak report
    - In depth reports covering specific outbreaks
- Data to modelling group
    - Modelling R-numbers and reopening scenarios
- Breakthrough infections after vaccination
    - Surveillance of infections post vaccination
- Surveillance of diagnostic assays
    - Primer/probe statistics

Currently, there is a turnaround time of 72-96 hours with more than 45% of cases sequenced and shared. However, there are challenges remaining with the logistics of moving samples and data around, and the necessity to explore legal avenues and support for local and international data sharing.

Morten concluded with a comparison of common variants and upcoming variants with spike mutations (usual to observe up to 15 spike mutations - new one in SA (Omicron) with 30).

Q&A

Q: What drives the request for more sequencing? This is partly driven by politicians, but from a public health perspective sequencing nearly 100% of all samples gives us a unique surveillance system. It allows for ID of new variants at a very early point although less than 100% would also allow this.

Q: What is the recommended number? Modelling done for early/quantity detection by random in general supports the need for not more than 4000 per week.

Q: What protocol is used? The ARTIC protocol is used with primerset V3, plus a few spike ins. All sequencing at SSI is now Illumina.

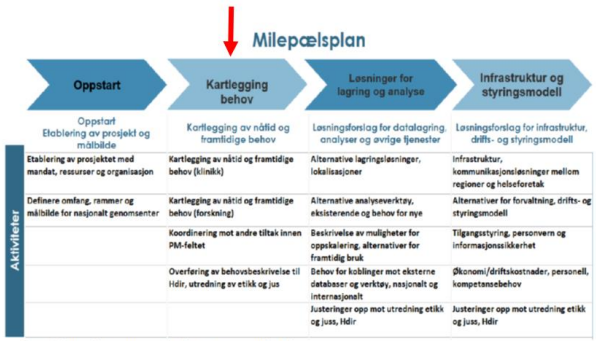
Q: Are there any attempts to link sequence to disease? Immunologists are doing so.

Q: How did mink infection spread? There have been many hypotheses, but the reason is still not fully understood

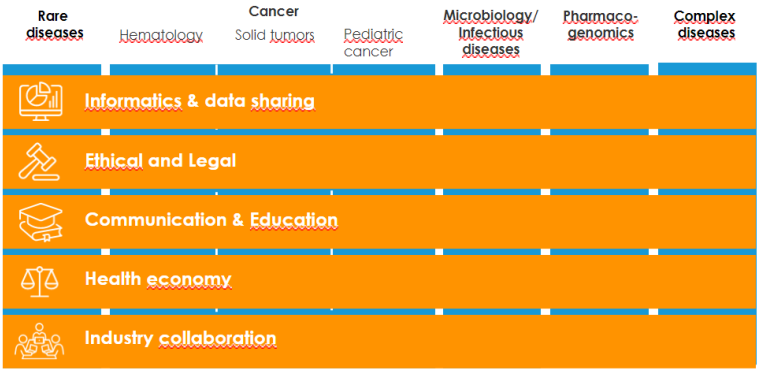
National updates from Nordic representatives

Number of participants: 90

	Speaker	Dag E Undlien
	Title	Updates from Norway
Key information	Dag introduced the following key areas ● A NGC is planned - government initiative - focus will be on ICT infrastructure and ICT analysis tools to provide a place to store and analyse data (sequencing done elsewhere) - mapping - design for storage - governance model - within 2022. ● New legislation: Prop. 112L - allows for using health data to help the next patient (for CDS and AI (data aggregation)) and establishment of register for DNA variants and associated health data - the data from genomic sequencing is covered by the legal right to establish a shared health registry with high level and structured information (incl. phenotype /predictive /diagnostic/prenatal data). ● Establishment of a national competence network for personalized medicine. The aims of the network are to develop a forum for information exchange, knowledge sharing, increase harmonization nationally, increase national collaboration, secure equality of care nationally, and provide a *channel for proposing national projects to promote personalized medicine.* ○ Projects launched through the national competence network include: Digital solution for collecting structured phenotype information as part of ordering genetic tests for rare genetic disorders, Implementing NGS diagnostics in cancer – national harmonization and collaboration, and a database for microbial genetics.
Q&A	Q: In terms of using family data, does this new act provide for that? The question on data use of relatives is not specifically addressed - still rely on consents. Q: In terms of 'helping the next patient', can you opt out and still be analysed? Patients must be informed that they can opt out from the registry - however, in terms of DS/AI development it is less possible - although mechanisms are in place to ensure directly identifiable data is not show



Figur 2 Oversikt over prosjektaktiviteter og milepæler

	Speaker	Valtteri Wirta
	Title	Updates from Sweden
Key information	Valtteri introduced Genomic Medicine Sweden (GMS), which coordinates implementation of precision medicine across Sweden. It offers equitable access in all healthcare, regions and is: • Clinically focused – bottom-up initiative, • National - 7 Genomic Medicine Centres, • Harmonised - national standards, guidelines, and methods, • Cost-effective – coordinated implementation • Innovative – population-based genomic data The new scientific advisory board to GMS was presented **Focus areas - diagnosis and treatment**  National Genomic Platform was presented as a national resource for healthcare, research, and innovation, offering scalable national solutions for information management: standardisation, real-time analysis, and visualization. Currently being prototyped in a pilot implementation setting. Links to national registries, and national data sharing will be possible within research projects with ethical approval. Working on nationally coordinated panels for haematology and solid tumours; also piloting WGS in combination with RNAseq for selected patient groups (acute leukaemia, paediatric cancers). In these projects predisposing germline variants will also be evaluated. Envision making healthcare generated data available for researchers in the future, however currently all work is done under the research umbrella and broader secondary use of data is not possible. Other challenges include access to clinical studies, competence development avenues are unclear, and health economy and long-term financing is problematic. • Cutting edge implementation of precision medicine in Europe will be held 22-23 Sept 2022, by J. of Internal Medicine, GMS and ZPM (German precision medicine initiative)	
Q&A	Q. and A. basket to basket trials difficult to run in small countries Q. With respect to running the same panel across Sweden - is bioinformatic analysis done the same? Partly, but not fully and working towards improved harmonisation.	

	Speaker	Cathrine Jespersgaard
	Title	Updates from Denmark

Key information

Cathrine introduced the DNGC:

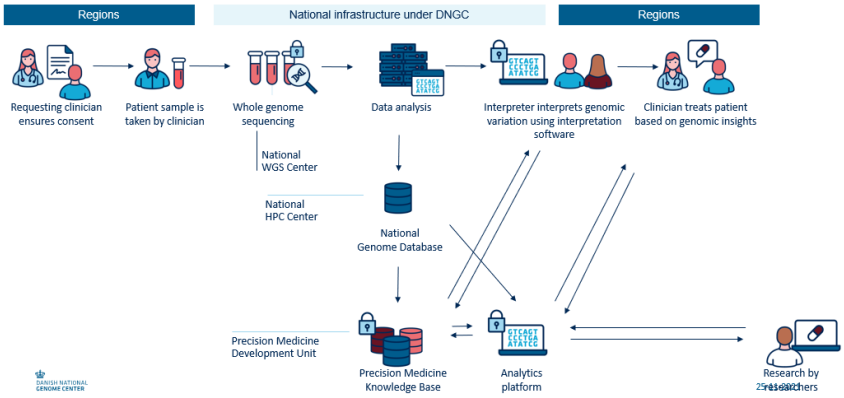
- An institution under the Danish Ministry of Health assisting the Minister with the central administration of issues related to personalised medicine.
- Supports the development of personalised medicine in collaboration with the Danish healthcare system, research institutions, patient organisations etc.
- Develops and runs a joint, national information infrastructure for personalised medicine, including a national infrastructure for performing genome sequencing and storing of information in a national genome database.
- Makes information available for people from the healthcare system and for patients, including information from the joint, national genome database for patient treatment etc.

The unique set up benefits from:

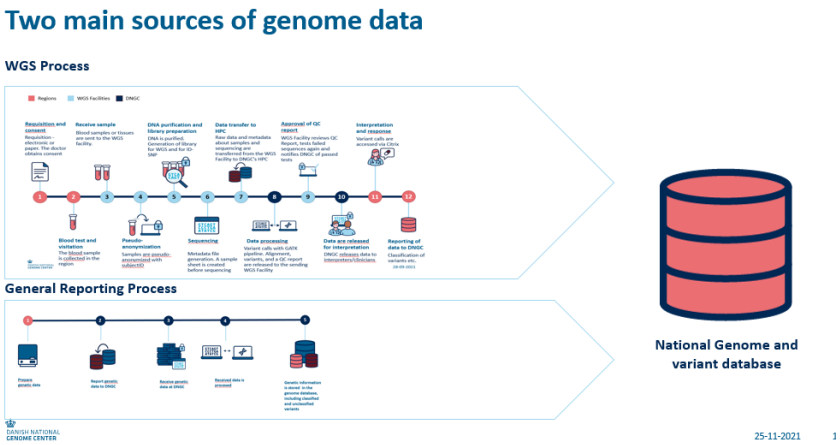
- A focus on diagnosis and treatment of patients.
- Broad political and state support for DNGC (Ministry of Health).
- A national Strategy for Personalised Medicine.
- A law on data reporting of comprehensive genetic analyses to NGC.
- Broad support from the regional healthcare system, The Organisation of Danish Medical Societies, and universities.
- Financial foundation through financing from the state, existing regional funds, the Novo Nordisk Foundation, and others (e.g., research funds in the future).
- Strong tradition for registration of data in health registries and biobanks.

The governance of the DNGC is structured through a steering group, advisory boards, and clinical and technical working groups (advising from universities and regions on infrastructure). There is an emphasis on DNGC providing expertise and value for patients - with better access, broad effect, and socioeconomic considerations. WGS of 60,000 patients is expected by 2024.

Flow of clinical WGS data through the infrastructure



There is an obligation to report all comprehensive genome analysis to DNGC - from DNGC and regional healthcare - data provided for clinical and research purposes, through cloud services, for the development of future solutions.



How to get research access to data and clouds in DNGC:

- DNGC Research Services.
- Research within PM and significant societal interest.
- Certified research institutions.
- Companies can get access through research collaboration.
- Use of data stays under public control.
- Access requires research ethics approval.
- Pseudonymous data.

Q&A

Q: Is there access for international researchers to DNGC? Yes, this is available through collaboration with a DK institution.

Q: Is 60,000 achievable by 2024? Yes, this is ambitious but possible.

Q: Do you have sample collection processes in place to meet sequencing capacity? Yes, this is a challenge, not currently receiving patient group numbers desired but hoping for more.

Speaker Jón Jóhannes Jónsson & Eiríkur Briem

Title Updates from Iceland

Key information

Jon introduced creation of a new clinic building opened Spring 2021

- Presented challenges associated with open landscape working policies: discontent around open-offices and hot desks which don't support the work the department needs to do regarding communications, paperwork and 'clean-desk-policy'.
- Eiríkur presented details of the new building for clinical labs at Landspitali in 2026 both positives and negatives. The new hospital will merge clinical genetics labs (apart from Genetic Counselling) with better space usage but still open office policy, and helicopters are planned to deliver patients on top of clinical labs building
- Jon presented on challenges associated with unsolved cases and new approaches/strategies to heterozygous pathogenic variants in AR disease genes, no pathogenic variants in AD diseases, and with questions to how others approach this - current tendency towards nanopore.

- Presented on a new computer system called GLIMS.

Q&A

	Speaker	Anna-Kaisa Anttonen
	Title	Updates from Finland
Key information	Anna-Kaisa presented on the law on the National Genome Centre and on genetic testing currently under public review following critical comments. The previous version is now split into two parts and raises critical comments from relevant stakeholders. Heads towards variant registry, regulation on a few clinical points. Question to how funding will be allocated. Currently, a major reform is under development for funding for public health care. Within the University hospital clinics, the number of WES going up in clinical testing: Number of WES doubled from 2020 to 2021 while the Number of aCGH, panels reducing
Q&A	Q: Funding for public health: regional or national, how will this be done? Currently: University hospitals funded by district communities; plans are to gain funding from governments for larger districts Q: WES quantity specifics? 100s still Q: aCGH going down - does this capture all? Not entirely convinced yet that it does e.g., phoniatric problems requests for arrays down (probably due to covid keeping patients away?)

	Moderator	Bobbie Ray-Sanneruud
	Title	Panel discussion
Key information	All presenters were invited to answer questions during the panel session 

	Panel speakers (L-R): Valtteri Wirta (*VW*), Eiríkur Briem (*EB)*, Dag Undlien (*DU*), Cathrine Jespersgaard (*CJ*), Anna- Kaisa Anttonen (*AK*), Jón Jóhannes Jónsson (*JJJ*)
Q&A	*Comment: Research supports theory that open offices aren't always beneficial* *Q: Patients want their data, and support this data access/sharing, but what data is relevant for them?* *CJ:* They have not been asked this yet. *VW*: patients have the right but what data they get may not be helpful (e.g., raw). *DU*: OUS have a process in place to copy FASTQ files. *AK:* have experienced this in one instance, where FASTQ was provided on USB. *JJJ:* process in place to supply VFC files - it is becoming more common to give data to patients *Q: Fetal data registration - when is this data stored with mother or moved to self*? *EB:* new GLIMS system gives fetus ID connected to mother, which moves to own ID when born. *AK*: the fetus data is connected to the mother but cannot currently be transferred to its own repository *Q: Is there any laws that prevent multiple data sets from the same patient from being later compiled into a single data lake/source for later secondary use?* *AK:* current consent form offers opt in/out for testing, but one cannot say no to second acts on data (this is required currently but there is hope to remove this). *JJJ:* Secondary use of data from clinical samples is required. *VW:* This inability to opt out seems strange, and in conflict with GDPR requirements. *Q: Is consent given to specific methods e.g., WGS or can other methods be used?* *CJ:* consent for all comprehensive genomic analysis is given, informed data will be stored in DNGC. It is possible to opt out of secondary findings, analysis for testing is broad. *Q What long read is most interesting and with what criteria?* *VW:* We don't know - we must look at different data properties and analysis methods over time

Implementing sharing of variant classifications via the Variant exchange API

Number of participants: 24

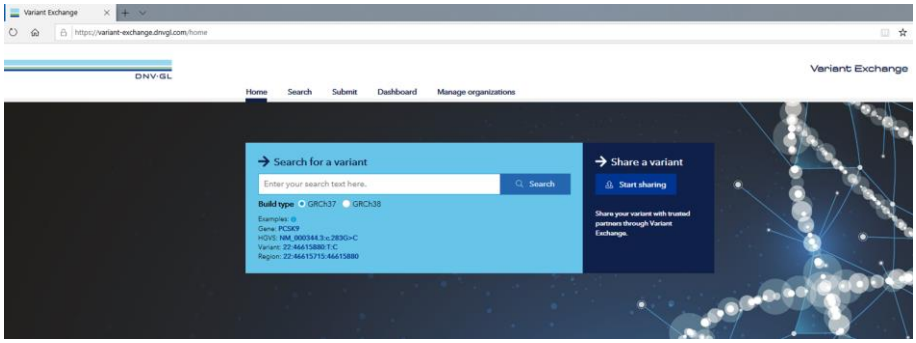
Speaker	Tony Håndstad, Department of Medical Genetics, OUS NO Marlon Polo de Melo, DNV

	Title	Implementing sharing of variant classifications via the Variant Exchange API

Key information

Tony emphasized that the current session fulfils NACG objectives in sharing data. The objectives of the workshop are designed to enable all to learn how to interact with web service APIs via Python, lower the barrier to sharing variant classifications via Variant Exchange, and discuss remaining barriers to sharing variant classifications.

Marlon presented the Variant Exchange by DNV. Variant Exchange (VE) is the result of a collaboration by Norwegian precision medicine project BigMed and NACG with the objective of sharing variants across borders.



The VE has the following features relevant for NACG, 1) quality control and variant classification management, 2) secure sharing, 3) between trusted partners and organisations (only) 4) warns about conflicting classifications between your lab and collaborating labs, and thus helps standardise classifications among collaborators.

An introduction to API was provided as a type of software interface, offering a service to other pieces of software. This was followed by an introduction to authentication and authorization, JSON, and the specifics of the VE API.

VE has a user interface that allows the user to search and browse per gene and variant, manage collaborations, etc, and has an API for upload and download of classifications

Participants of the workshop formed groups and then worked on reading and writing classifications via the API. Perhaps the most challenging part was to successfully get the authentication and authorisation to work. A later debrief included a discussion of their own processes and storage of variant classifications. Currently, many use Excel and are awaiting a national solution / database. Variant interpretation in the Nordics use different software's and infrastructures. For example, in Sweden, variants are stored in SCOUT in their own infrastructure and data is submitted through Clinvar, they also use Matchmaker Exchange. Helsinki (HUS) uses a commercial software (Euformatics) with some issues currently being resolved.

Q&A

Challenging clinical cases

Number of participants: 61

	**Speaker**	Maria Rossing, Centre for Genomic Medicine, Centre of Diagnostic Investigations, Rigshospitalet DK
	Title	Challenging clinical cases
Key information	-	This session is designed to make us better at sharing cases - comparing knowledge improves diagnostic outcome - direct diagnostic results are obtained from these sessions.
	-	Although cases were presented as anonymized with all identifying information removed, in order to further protect the cases, it was agreed that no notes would be taken on this session for publication to the workshop summary.
	-	*Case contributions were presented by:*
	-	Jón Jóhannes Jónsson and Eirny Thorolfsdóttir (Iceland)
	-	Kathrine Bjørgo (Norway)
	-	Maria Bach Laursen (Denmark, Århus)
	-	Elsebeth Østergaard (Denmark, Copenhagen)
	-	Morten Dunø (Denmark, Copenhagen)

IVDR: status update, knowledge sharing and potential

Number of participants: 61

	**Speaker**	Cathrine Høgseth Nordhus, Department of Medical Genetics, OUS NO Courtney Nadeau, Healthcare programme, DNV
	Title	IVDR: status update, knowledge sharing and potential Nordic collaboration
	Objective	Understand where labs are regarding IVDR compliance

	Identify areas for collaboration between NACG members Identify key people and establish an informal working group to address one (or more) of these topics
Key Information	Courtney introduced the structure: ● A show-and-tell from 8 labs ● Discuss 6 topics in groups Updates from Nordic Labs •HUS - All IVDs are mapped in the lab, and a strategy is in preparation for each - many things are covered by validation, accreditation, and SOPs. Lab is accredited to ISO 15189. Scientific validity reports to fulfil IVDR requirements in prep. Currently mapping assays in the lab and collecting specific criteria for in-house exemption. Unsure of how to fulfil the criteria for continuous risk assessment for each in-house method. •GMCK (KI and SciLifeLab) - No ISO 15189 instead ISO 17025. IVDs in use are mapped (WGS and target analysis for germline/somatic analysis). There are problems with time and resources to handle this - would like a roadmap to move forwards. •Danish NGC - Sequencing carried out at two accredited labs and NGC runs the bioinformatic pipeline. The role of NGC here is to accredit the pipeline using in-house exemption and they have begun working (initial stages) towards obtaining ISO 15189. No actions to IVDR have been taken. Biggest challenge is the timeline and uncertainty in interpreting the regulations. •OUS: Cancer Cytogenetics - Very beginning phases - just started mapping some IVDs in the lab (e.g., FISH, cytogenetics, and array), but currently no strategy or work towards accreditation. The biggest challenge is resource and time - there are no employees that can continuously work on this. •Landspitali - Registry of assays but not all IVDs defined in the regulation. No strategy is determined. Will use in-house documents and Johner institute documents. Not accredited to ISO 15189 but preparing concurrently with IVDR. Challenges involve difficulty in understanding regulation, limited help, confusion about compliance definitions or date compliance is required by (postponements). Collaborations ongoing internally to make this happen - four people working on this. Uncertainty about what is needed in 2022 compared to 2026. •MOMA - Accredited to ISO 15189, but no mapping to IVDs, no strategy or implementation in quality systems. Trying to understand rules for evoking article 5.5 to understand in-house exemption. Lacking support from big institutions on how to interpret. *Q on how did the network of DK hospitals begin?* It started with clinical academics working together to solve the same problems, who met 5-6 times virtually. Hypothesise that if no one knows or can advise on how to interpret the regulations then we will interpret them and hope the authorities follow this. •OUS: Medical Genetics - All IVDs mapped, expected majority in in-house exemptions, no plans to CE mark, as assembly is based on conforming products combined. Most 'new' documents required should be covered by existing procedures and validation reports. Accredited to ISO 15189. The challenge is interpreting rules for in-house exemption, as there is a lack of clear guidance from the EU and Norwegian Medical Agency. Collaborations successfully formed with other relevant regions and hospitals in Norway. •DNV - regulatory updates from recent publications*: ● MDCG 2021-24: Guidance on Classification ● MDCG 2021-13 Rev.1: Obligations for EUDAMED Registration ● MDCG 2021-22: Clarification on Expert Panel Consultations ● Performance Evaluation Consultation Procedure (PECP): First opinion issued Regulatory updates upcoming*:

- 2 New SARS-CoV-2 PECP opinions expected soon
- MDCG Guidance on:
- Performance evaluation: 2022
- Summary of Safety and Performance Template: 2022
- Qualification of assays used in clinical trials for medicinal products: 2021
- In-house devices: 2022
- Legal status of app providers: 2021
- Market surveillance for In-house Manufacturers: 2021
- Clinical Investigation Report Summary Template: 2021
- Proposed amendment to IVDR

*not-limited

Discussion: A big part of genetic analysis is bioinformatic pipelines with integrated internal software development, which makes CE marking unfeasible. Switching to a company with a CE-marked equivalent product will be required if one becomes available. However, experience with a NIPT test that is CE marked and regularly used has not been without problems - which require consultation with the company manufacturing it (timely etc). With the GDPR, allowing access to sequencing data and pipelines is also problematic. More competence is required to address these issues.

Workshop and Brainstorming

Participants broke off into smaller groups to discuss selected topics relevant for clinical labs. Groups were asked to discuss their approaches, challenges, and potential common activities for NACG, and to summarize their discussions. To support these discussions, relevant legal definitions were presented, and excerpts from the IVDR were available for each topic.

	How can a lab address this topic?	**Potential NACG activities**	**Main challenges and unanswered questions**
Intended Purpose	-Read art 2(12) -Fill out documents based on annex II 1.1.c		-Most important -Do first
Analytical Performance		-Agree on common interpretation for measuring analytical sensitivity for non-quantitative assays -NACG can be proactive in proposing an interpretation of the IVDR	-Many assays are not quantitative -Cost of running many samples -Access to samples with relevant mutations: Finding samples, Ethical and legal approvals to use existing samples
Clinical Performance	-Public/private partnerships with industry	-Support industry/hospital matchmaking for clinical validation studies	-How are manufacturers going to get access to clinical samples?

	Risk Management System	-Same assessment for one method (NGS, FISH) -External Quality Assessment	-Risk assessment scheme -Share what different departments are using	-Continuous evaluation: Resources and time required -How often to update?
	Post-market surveillance	-Trending performance metrics using a reference sample or panel on a repeated basis		
	Market monitoring			
Discussion / questions / comments				

Wrap up day 1: Overcoming technical and legal barriers to sharing of variant classifications between labs and ideation exercise for next workshop

Number of participants: approx. 80

	Speaker	Tony Håndstad, Department of Medical Genetics, OUS NO
	Title	Overcoming technical and legal barriers to sharing of variant classifications between labs

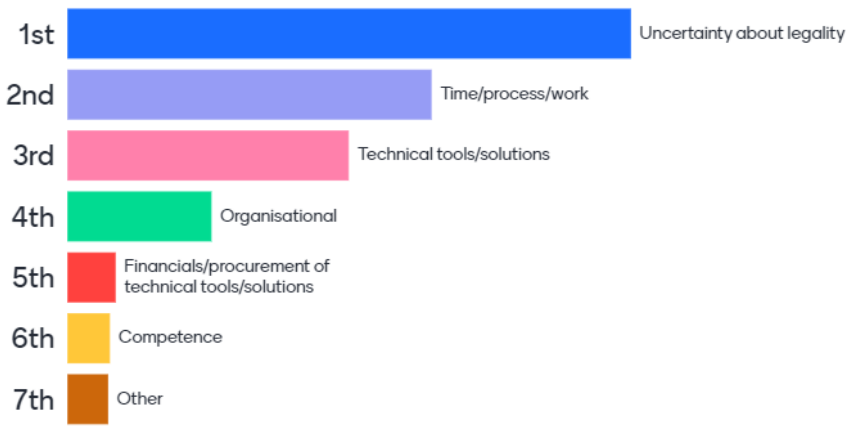
Key information

Tony provided an introduction on the Variant Exchange (VE) by DNV which included an overview of main features such as sharing of classification variants, notification of discordance, graphical user interface (search for variants) and inclusion of an API to upload and download classified variants. The VE API was used in the hackathon earlier in the day. Prior to the hackathon, a survey indicated that only half of the participants had used an API before and therefore the session included an introduction to using APIs. At the hackathon, they additionally read from the VE database via API and uploaded variant classifications via the API.

Reflections were that the hackathon went well and was “fun” and all took away learnings.

Tony then facilitated a Menti session on overcoming technical and legal barriers to sharing of variant classifications.

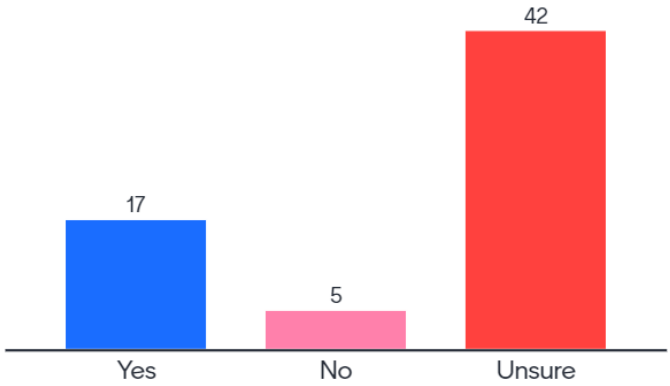
1. What is the greatest barrier for sharing variant classifications at your institution? (n=65)



Comments from participants relating to this include:

- Concerns that there is a re-identification risk from multiple reanalysis procedures, also from identification of really rare variants.
- If sharing was more automated, then it would be more feasible - manual seeking results is a deterrent.
- Interesting variants to share are the rare ones which are most valuable - incl. VUS.
- If you don't have time, then you need a really good solution to make it easier to share (which should be integrated into the pipeline if possible).

2. Are there legal barriers against sharing variant classifications in your country? (n=64)



Comments from participants relating to this include:

- A significant number of participants are unsure if sharing is legally allowed - help is needed here and the 12th NACG WS could lean on expertise from NordicPerMedLaw to advise on this.

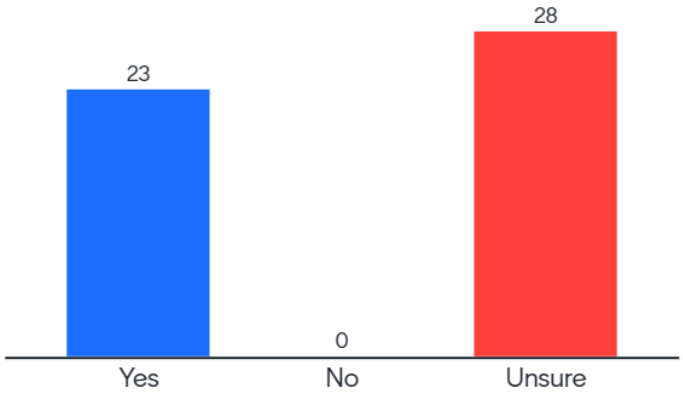
The next question followed on from this

3. Which questions would you like the legal teams in NordicPerMedLaw to answer?
    - What can you do legally in terms of variant exchange?
    - Can we all agree that classified variants are medical information and not personal data?
    - To what extent can a genetic result/variant classification/genotype-phenotype correlation be anonymized for sharing?
    - Is variant exchange health care or research?
    - Are there significant differences between national legislation in the Nordics lying on top of GDPR which is common?
    - IVDR exploration

Comments from participants relating to this include:

- Asking legal questions around variant sharing is limiting, whilst the need to share a much broader data set is required (e.g., matched phenotypes). Perhaps to accommodate this we should be asking what we are not allowed to do as opposed to what we are allowed to do.
- A research project is not constrained by the same ethical approval as clinical projects.
- There is disagreement on legal interpretations even amongst lawyers in the same department, therefore it may be even harder to get consensus internationally.

4. Is Variant Exchange the right solution for sharing variant classifications? (n=51)



Comments from participants relating to this include:

- Participants that were unsure lacked knowledge about variant exchange, participants that said 'yes' highlighted that approval is already in place and competing products are not available.

Q&A	Additional questions with some answers raised by both Tony and the audience included: • Is VE for research or healthcare? • Does consent need to play a role for use? • Is VE an application under IVDR? (legal status is data processing: storage, retrieval, messaging, and simple search so assumed not to be) • Are there any alternatives to VE? Examples of these include ClinVar, ClinGen, MME, Beacon, email, phone calls...

	Speaker	Bobbie N. Ray-Sannerud
	Title	Ideation exercise for next NACG Workshop
Key information	Bobbie asked the audience to give NACG their wish list and volunteer to lead new sessions - to share with the steering committee for development of the 12th workshop. Topics where suggestions for leaders were given included: • Long reads - which technology, how to implement, comparison to short reads and what benefit will we get (MOMA - Ebbe) Long-Read sequencing (Valtteri Wirta, SciLifeLab) • Liquid biopsies and CT DNA, Meza (Leonardo Zepeda, OUS) • LIMS system implementation use, problems (Pia from NCG) • Nordic comparison to fetal analysis and how results are handled (Peter from DK NGC)	

- Part II hackathon with use of Variant Exchange (Dorte from MOMA)
- Something on CT DNA - assays and use (Thomas Reinert from MOMA)
- What are new research methods to go into clinical practice AI/ML, DepSeq etc - then how do we implement them ASAP? Henriki
- What are barriers to using cloud in HC? How can we use what is there? Moving to the cloud (Henrik?)
- More case studies, Challenging cases. Continue with unsolved cases - what do we do and for how long? (Dorte Lildballe, MOMA)
- Joint Nordic interpretation of IVDR. More about IVDR topic specifically on how labs handle post market surveillance for cytogenetics (Nelli Karhu from Finnish Medical Agency) - IVDR (Courtney)
- Clinical aspect - how to handle issues with pathogenic variants that are heterozygous in recessive diseases (Maria)

Topics involving contributions from all labs include:

- Which are prediction tools that can be used to automate e.g., ACMG classification - on a lab basis - rules and thresholds (all labs)
- Hackathon from challenges from current workshop - long read sequencing or common and converging tools pipelines. Phase II of 11th workshop
- Practical exercise between the workshops - e.g., variant interpretation in RD or cancer, or SV
- Pipeline focussed talks - SM and RD

Topics where suggestions for leaders were not given included:

- Multiomics - how to combine layers and analyse - understand library preparation, DNA/RNA/Methylation.
- Focus discussion on fetal analysis and results (from clinical point of view)
- Pros and cons of various variant sharing tools and repositories.
- Implementation of new research methods (e.g., AI/ML, mRNA seq, deep variant) into clinical practice. How can we translate research to the clinic faster without sacrificing safety?
- What are available funding opportunities - follow up on successes with these after 6-12 months?
- Variant interpretation.
- CNVs for WGS.
- Clinical trials and similar in genomics field including evaluations.
- Best practices for bioinformatics analysis
- Quality assurance
- Emerging technologies (longread seq, bioname)
- Polygenic risk scores
- Data driven identification of SVs - how to communicate with clinicians more efficiently

Summary	Participants of the 11th workshop were asked to vote, in an online survey provided to them after the event. on selected topics from these to determine which of them they found most interesting and would plan to attend.

NACG knowledge sharing

Number of participants: 82

Moderator	Bobbie Ray-Sannerud, Healthcare programme, DNV
Title	NACG knowledge sharing
Objective	An interactive session where NACG members share projects and learnings.

Nine presenters from five NACG organisations gave short (7 minute) presentations on diverse topics.

Ramprasad Neethiraj, Clinical Genomics-Stockholm: Deepvariant in a rare disease bioinformatic workflow

- Ram introduced challenges in exchanging statistics analysis across platforms, leading to quality issues and a necessity for deepvariant technology. This looks at a broader region, base and mapping quality, strand bias, base differing from the reference and the read supporting the variant.
- Ram chose to investigate DV workflow in comparison to current SciLifeLab workflow (GATK), using the GIAB data set. DV performed much better for performance, sensitivity, F-measure, and false/negative positives and did so faster than GATK.
- DV is currently in production to populate databases and they expect a full switch in a couple of weeks.

Q: Did you compare it to other comparable technologies such as Sentieon? No, but we use some of the other ones in other sequencing applications e.g., cancer

Q: Retraining of DV? No, we can't match their level, so we use the default model

Q: Why did you drop base recalibration and how much memory does this use? This was dropped as it affects quality of calls, and 180/190 GB

Courtney Nadeau, DNV: Using assurance to accelerate dynamic consent

- Courtney introduced the concept of DC for managing preferences in data management, and the ADA assurance framework - 1: profile asset - understand, context, stakeholders, risks, and opportunities etc, 2: Iterative management - strategy, uncertainty, substantiate, assess.
- Courtney introduced a pilot project run with Australia Genomics on CTRL that identifies and mitigates risks and prioritizes and capitalizes on opportunities.
- The process involved document review and stakeholder interview (both technical and conceptual), analysis of design documents - building of entity ecosystem map, and categorisation of risk. Three events to mitigate - data breach, ethics/trust breach (identification by participants data used in an unexpected way) and bias/unidentified breach. Security and data breach were most interesting to Australian Genomics - through identification of these we prove the need for mitigating actions and continuous assurance.

Q: How frequently is it used and how often do patients change their minds? CTRL is rarely used for clinical studies; it has been ethically approved as they have a complimentary paper option. The number of people that use it is variable and not as much as hoped.

Mei Wu, SciLifeLab: New bioinformatic workflow for rare disease diagnostics in Nextflow

- Mei introduced Nextflow - an open-source workflow management software that enables fast prototyping with existing scripts, portability, and reproducibility, and ease of integration with HPC and cloud servers in the lab.
- Mutation Identification Pipeline (MIP; https://github.com/Clinical-Genomics/MIP) is being ported to Nextflow using nfcore framework - nfcore is an organisation that sets the standard on bioinformatic pipeline development - allowing harmonisation.
- The pipeline is currently hosted on nfcore, which has been in development since summer 2021 across six nodes in Sweden.
- Conclusion - moving to Nextflow with nfcore improves usability, scalability, and supports continuous development and efficiency of the pipeline. Working with other nodes in Sweden allow cross use/testing of node pipelines,

Q Is it on premise? Can be both on the cloud and on the premises depending on licence purchase.

Michael Knudsen, MOMA: Structural variants

- Michael introduced problems associated with calling of germline SVs - easy to do but often call too many and there is difficulty in identifying truth. How is effective prioritisation and removing the false calls of variants done? Can look at reports in VCF files (discordant reads) and ignore those without any reads spanning breakpoints. Recurring noise and common variants can be filtered out - e.g., Delly and Manta VCF files include confidence intervals for breakpoints - can analyse frequency of breakpoints in background sets.
- For somatic samples it is still a work in progress - currently annotating variants with gene names, prioritizing based on actionable genes and combining them with RNA-seq.

Q: Have you used panel of noise to eliminate false positives e.g., in later filtering steps? This is achieved using the breakpoint frequencies described above. *Have you tried anything other than Delly which is noisy?* Currently testing Manta. Using ExomeDepth for CNAs.

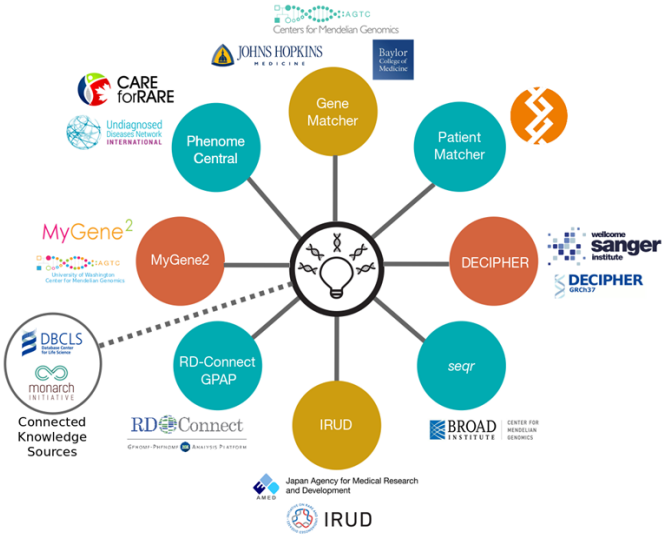
Morten Eike, OUS: ELLA - Recent developments

- Morten presented recent developments in ELLA, an inhouse clinical variant interpretation method, with a focus on documentation and reproducibility, and faster interpretability.
- Currently *bare bones* support for CNV interpretation - ELLA shows CNVs in separate variant list, allows documentation and reuse of CNV interpretations, and offers annotation tracks in embedded genome browser. Current limitations include:
    - Only DEL, DUP and DUP:TANDEM.
    - No CNV filtering in ELLA (handled upstream) (to be developed Q3 2022).

    - No CNV-specific annotation except in tracks (to be developed Q1 2022).
    - No CNV-specific ACMG criteria (to be developed Q2 2022).
    - Gene interpretation (to be developed Q4 2022).
    - Current limitations mean this version is recommended for smaller gene panels only
- Also refactoring the front end - lifting code from one (inefficient/outdated/messy) framework to another for increased efficiency.
- Website https://allel.es ** Open source ** and demo site available http://demo.allel.es (testuser1-8, password: demo)

Q. Do you have a front-end developer? One hired on limited contract, others are full stack

Henrik Stranneheim, SciLifeLab: PatientMatcher - A Standalone MatchMaker Exchange Server

- Henrik introduced MME which consists of several nodes supported by GA4GH and IRDiRC.
    - 

- MME works through the introduction of patient information by a clinician into their node, and resultant notification of matches by email.
- Participants can choose to join an existing MME network or create their own node. SciLifeLab chose to create their own as advantages include No need to deposit data outside facility, ad hoc database, and ability to define own matching algorithms and ranking of returned results. Although this is complicated and can require more time investment.
- There are limitations to the network, one can only ask for similar patients e.g., a phenotype linked to a genotype.
- PatientMatcher is opensource https://github.com/Clinical-Genomics/patientMatcher
- SciLifeLab is currently connected to 3 other nodes including GeneMatcher, and connection is carried out through Scout to PatientMatcher and then into MME.
- Limitations - the matching algorithm could be improved, Orphanet is used but it's not currently compatible with Scout, not yet able to calculate disorders similarity between patients as distance of terms in the ontology.

- Future plans: want to join phenome central and Decipher and need to be more integrated into 'routine use' to increase the number of shared patient cases.

Q: The use of data shared is limited to three categories - Do you have to share the whole genome? No, just a little bit of data is shared (either specific variants or candidate genes) and then matching happens - the more specific the better it works.

Q: Are you not allowed to use gene matcher? Gene matcher is too manual, MME is more integrated into the process and provides a better overview.

Lusine Nazaryan-Petersen, Rigshospitalet: PacBio HiFi sequencing for detection of genetic variants in the hard-to-sequence genomic regions: case report

- Lusine introduced a pilot project to improve upon the current status of only 40% variants identified with gaps missing
- The case is a balanced inversion - homologous segmental duplications that cannot be identified by array. The inversion was first identified by FISH. Test to see if PacBio (2x) can identify inversion at the breakpoints - it can also show indels and SNVs with good calling.
- PBSV cannot call inversion, only observable by manual inspection. This case was more complex than just an inversion - inserted repeats were present. Another SV caller was compared - SVIM (used by GATK) was better but identified it as translocation as opposed to an inversion.

Q: What was the cost? Per smart cell 30000SKK x 2 = 60000SKK

Q: Was there any difference observed between the two runs? The results were confirmed with the second smart cell (PacBio recommend using 2, for indels 3)

Tony Håndstad, OUS: Data storage policy and compression

- Tony requested audience feedback on how data is stored - FASTQ, BAM and VCF storage costs escalate - e.g., estimation of 1 million NOK per year if sequencing around 96 genomes per week.
- To reduce costs could sequence less, compress data, remove redundancy, delete old data, use cheaper storage
- OUS currently compresses BAM files to Crumble CRAM, which can cut storage needs to 1/10th without affecting variant calling. Would like to delete FASTQ files as the data they contain can be found in BAM/CRAM files but have not done so yet.
- One possibility is to remove old data when new data becomes available - currently reanalysing against GRCh38 reference genome - so should old data against GRCh37 be deleted (and can it be?)

Tony asks: What do you do - store indefinitely, long term, delete, any alternative suggestions for reducing costs?

Denmark found that CRAM had positive effects, and they also delete FASTQ and intermediates.

Should we keep BAM for the patient journal? Consensus seems generally yes, even in solved cases.

Anders Jemt, SciLifeLab: RNAseq in WGS-based RD diagnostics - Experiences from Clinical Genomics Stockholm

- Anders introduced the current production pipeline: can get splicing and fusions represented in Scout and monoallelic expression with complimentary text files. Fusions come in pdf output
- Developing a new research pipeline - Anomaly - that will be using an RNA first strategy, using background datasets with the aim of identifying candidate variants without DNA results, with normalisation to find the most important findings in an html report.
- They still need background data - so sequencing 300 RNA samples for database (that DNA samples exist for) with the protocol etc being defined.

Q: In Nextflow are you using nfcore modules? Yes, where we can, if not we will add them. OUS is potentially interested in collaborating.

Q: Where is the RNA from? It's mainly from blood, in selected cases from fibroblasts.

Discussion

Transitioning to hg38 reference genome

Number of participants: 28

Speaker	Courtney Nadeau, Healthcare programme, DNV Kaisa Kettunen, HUS
Title	Transitioning to hg38 reference genome

Courtney introduced the necessity to cover centromeres and regions of complexity previously missed in GRCh37 = hg18 which was released in 2009 with GRCh38 = hg38 released in 2013.

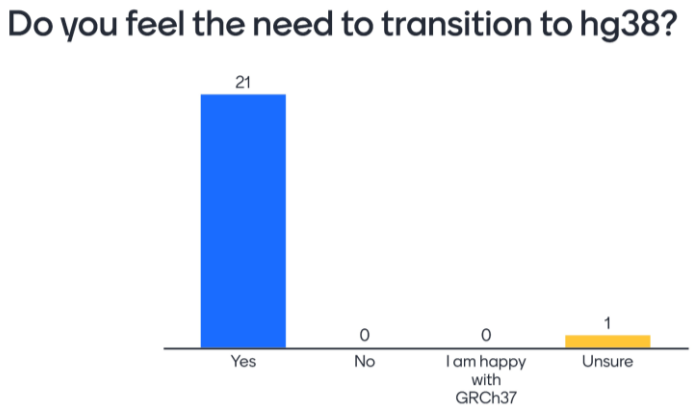
Three main updates in hg38:

- Repair of incorrect reads
- Inclusion of model centromere sequences
- Addition of alternate loci

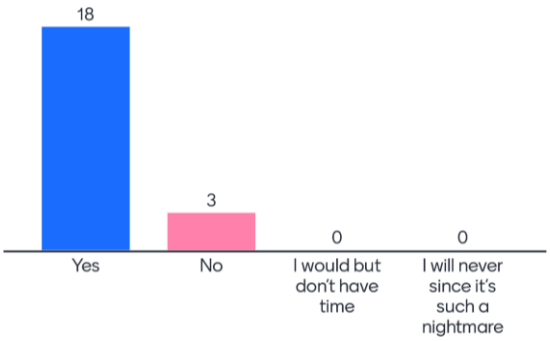
Kaisa presented examples of issues with two genome builds from Helsinki:

- Not all databases (e.g., GnomAD v2) support current versions of hg38, with discrepancies in the ref/Alt alleles.
- Not all software producers support hg38 (despite their guarantees e.g., Alissa Interpret - annotations missing - to some extent solved)
- In-house generated databases still in hg37 - so what should be done with old versions/data - should it be re-run?
- Some genes cannot be found in GRCH37 at all

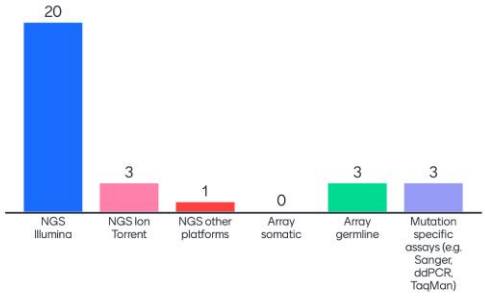
A Menti poll was carried out, with the following results:



Has your lab started working with hg38?



Which methods have you started using/plan to transition hg38 on?



What do you think are the benefits of starting to use hg38?

- Better build
- Including chr6 contigs
- Better variant calling x 2
- Some genes are not in hg37
- Improved regions
- Newer, better data x 4
- A better resolution/improved analyses and interpretation/calling x 3
- Better and more correct interpretation of the human genome x 5
- DK-NGC is using it - easier with one refseq
- Updated alignment of some genes

The first group work asked participants to identify: what are your problems? Challenges, issues, and pitfalls then consequential solutions

- Old data needs to be handled; however, it is costly to rerun the data. Production, development, and software/database changes (liftover) requires investment of time, money, and personnel to support this, which is not readily available.
- The current system works so it can be difficult to justify prioritisation of this change. In addition, hg38 has been around for eight years, new updates are anticipated soon, due to faults that must be accounted for and in many labs, change has only recently begun to be actioned.
- Rerunning old background data sets is costly in terms of computation, personnel, and storage - but it is a requirement for reference

genome/background in-house datasets. Tweaking of pipelines may be required, which requires subsequent re-running of validation sets.

- Uncertainty about how validation can be done. Is ISO15189 compliance, GIAB only, all panels/methods separately or selected samples the way forward? How can this be done in SVs?
- The transition from hg37 to hg38 can lead to lots of variance that is difficult to resolve due to genomic coordinates mismatching. A second opinion variant caller could be used to identify these variations.
- It has been noted at MOMA that it is faster to run pipelines with hg38 - this could be software related, updates all at once etc.
- Experience with liftover is variable or minimal, and there is some hesitancy to do this, as complex variants for example can be problematic.
- It is a big change with factors around the pipeline impacted: QC etc.

Finally, a necessity for an awareness of the anticipated problems was highlighted to inform each labs strategy for the transition, which led well into the next group work:

The second group work asked participants to identify: what is your hg38 transition strategy and how can you prioritise and order this?

- Group 1: Transition to hg38 - success criteria defined - it should do the same as current, ideally better. Then develop pipelines - select variants that are difficult to call, ask interpreters to find test sets, when it can be run (even when not complete), run it in shadow mode to allow correction of basic problems. When pipeline is ready for validation run the set and look for variants, (run validation report) when happy then place the pipeline into production
- Group 2: Define acceptance criteria - what we need will determine transition, which samples to use for external quality, precision, and sensitivity (including GIAB). Use in house samples for methods running e.g., regarding trios (inheritance) and panels to ensure previously detected variants are found (not just numbers). Build pipeline and do a lot of tests to ensure capture of mean coverage/all variants, then lock pipeline, update documentation, QC, and validation report, inform clinicians, in-house allele database etc, update other software to default to hg38.
- Discuss between interpreters and developers the build - what would you like? It is a big change so worth investing the time to try and align priorities and needs.
- Group 3: Infrastructure must be in place to parallelise runs, the pipeline for RD will be the same, just re-write the references, but interdependent of each other. Validation and reporting are required, but this is standard procedure. Requires time and personnel.
- Group 4: Existing databases want to be kept, so find new software that aligns, then do small incremental changes to build, then re-run some samples to identify issues and validate. Re-run all data to hg38 to fill databases - validate end result with end-user and accept that not all data will go to new format - most useful data is maintained. Make the switch.

Comments and questions

Identification of somatic SVs and complex biomarkers for NGS-based cancer diagnostics

Number of participants: 40

	Speaker	Valtteri Wirta, SciLifeLab Oleg Agafonov, Healthcare programme, DNV
	Title	Identification of somatic SVs and complex biomarkers for NGS-based cancer diagnostics
	Objective	This session will allow participants to share experiences and identify common issues of identification of SVs and complex oncomarkers, such as HRD, MMR, TMB and MSI.

Khurram Maqbool, PhD, Karolinska Institute	Introduced somatic SVs. Take home message: ○ Calling somatic SVs is a multi-layered biological and bioinformatic pieces of puzzle ○ There are few group(s) working on samples from homogenous population with several cancer-types ○ Undefined/random choice of SV-callers, strategy for SV-merging ○ Lack of consensus for use of tools/databases for sharing
Britt Elmedal Laursen, consultant, associate professor, PhD, MOMA /Department of Oncology, Aarhus University Hospital	There are several predictive biomarkers for immune checkpoint inhibitor treatment: PD-L1, MSI, dMMR, TMB. In 2017, FDA has approved usage of pembrolizumab (humanized antibody) for patients with MSI-H or dMMR is solid tumours, while in the EU its approval is limited to patients with colorectal cancer. Several analyses have shown that there is a correlation between high TMB, and the clinical benefit of antibodies targeted against CTLA-4, PD-1 and PD-L1. TMB-high status identified a subgroup of patients who could have a robust tumour response to pembrolizumab monotherapy (response 29%). In 2020 FDA has approved use of pembrolizumab for patients with solid tumours and TMB-high status. HRR is required to accurately repair DNA breaks. In HRD cells repair is imprecise and leads to DNA damage accumulation. To target HRD cells, PARP inhibitors are used that cause inhibition of DNA repair within cells leading to apoptosis. At present, four PARP1/2 inhibitors have received marketing authorization in Europe and/or the US: Olaparib/Lynparza®, niraparib/zejula®, rucaparib/rubraca®, talazoparib/talzenna®. Detection of genomic instability HRD-associated genomic alterations that have been investigated in PAOLA-1 include genome wide loss of heterozygosity, telomeric allelic imbalance and large-scale transition, which are continuous measures with predefined criteria and score. Composite genomic instability score (GIS, also called HRD score) is determined when the combined measures and respective scores are used to assess the extent of specific genomic aberrations accumulated in tumour cells. Lower score defines lower likelihood of hr deficiency of tumour cells, and higher score determines higher likelihood of HRD of tumour cells at the time of the sample collection relative to exposure to DNA damaging agents. Validated cut-offs should be used to determine GIS positive status. HRD positive status can be defined by a composite GIS score for HR deficiency-associated

	genomic alterations tested by an experienced laboratory using a validated test. Because HRD is observed in many tumour types other than ovarian and breast cancer, PARP inhibition is also being explored in several other tumour types including prostate, pancreatic, lung, gastric and several other cancers. Traditionally, no companion diagnostics approved by EMA, the European Council has initiated a revision of the legislation on pharmaceutical compounds. EMSO recommends that clinical research centers develop multi-gene NGS as a tool to: use in NSCLC, cholangiocarcinoma, prostate, and ovarian cancers; test TMB in well- and moderately differentiated neuroendocrine tumours, cervical, salivary, thyroid, and vulvar cancers; screen patients eligible for clinical trials and to accelerate drug development; prospectively capture the data that could further inform how to optimize the use of this technology. Challenges: DNA, epigenome, RNA, tumour microenvironment, relatively small number of participants in clinical trials.
Updates from Nordic labs on SV detection and NGS identification of complex biomarkers	
MOMA somatic SV detection	•Paired tumour-normal. Fresh-frozen tissue, but occasionally FFPE. •Wide range of different cancers. •WES and mRNA-seq •100X for blood and 200X for tissue. •Deletions, duplications, inversions, and translocations (e.g., fusion). •Delly, SvABA, and CNVkit for DNA-seq. STAR-Fusion (with Fusion-Inspector) and Arriba for RNA-seq. •Calls are merged. Clinically relevant calls are evaluated manually by interpreters. •Variants are prioritized by whether they involve clinically actionable genes. Working on further scoring of variants based on frequency in background set and number of supporting reads. •The biggest challenge is effectively filtering false positives while maintaining high specificity. Not a huge problem for WES, but it will become more challenging with WGS.
MOMA NGS-based identification of complex biomarkers:	•TMB, MSI, and currently working on HRD. •TMB is calculated simply as number of somatic mutations per megabase. Only variants with allele frequency >2% in targeted regions are included. MSI is estimated using MSIsensor. For HRD we have evaluated CHORD, scarHRD, and shallowHRD. We are currently testing Illumina Dragen. •TMB and MSI may allow for inclusion in trials. This is also expected for HRD. •TMB and MSI are in production based on WES. HRD is to be based on WGS. •The biggest challenge is the lack of consensus on how TMB and HRD should be calculated.
Cancer cytogenetics, OUS NGS-based identification of complex biomarkers	***•Which types of biomarkers does your lab identify (e.g., HRD, MMR, TMB, MSI, or other similar biomarkers)?*** Our laboratory specializes in searching chromosomal aberration in various types of haematological malignancies and solid cancers. We combine G-banding analysis with next-generation sequencing technologies in search of genes involved behind these aberrations. Since a biomarker is an indicator of a biological state or condition, the identification of cancer-specific chromosomal rearrangements in specific tissue contribute to identify not only the presence of a disease but also can give diagnostic, prognostic, and in some cases, information related to choice of treatment. Whenever we identify «new biomarkers/such aberrations» our goal is to characterize them as good as possible in a process

	that possibly will lead to the so-called “target molecular therapy”. We first identify the genes involved in the rearrangements, being them fusion genes or deregulated genes, identification of gene variants, mRNA expression, and detection of DNA methylation. ***•What methods do you use for calculating these scores? Are there any common pitfalls?*** We use RNA sequencing to select the candidate fusion genes, to find only those that map to chromosomal breakpoints. The actual involvement of these genes is then validated by PCR and Sanger sequencing analysis. For gene mutation analysis, we use the Ion Torrent technology for NGS. For mRNA expression analysis, we use the NanoString technology, and for the detection of DNA methylation, we use pyrosequencing (PSQ) analysis. Significant variability and reproducibility pitfalls can arise and lead to a false conclusion: -The availability of material is a fundamental factor in research: 1) to allow all investigations planned in a specific moment; 2) to have the possibility to perform additional analysis in a different moment. -Clinical samples frequently contain a mixture of cancer and normal cells. A large amount of normal cells may skew the results. -The comparison between normal and tumour samples from the same patient can be crucial: this is still a challenge. It is not always easy/allow to get the normal tissue from the same patient. -Low or bad quality of DNA or RNA can sometimes lead to non-informative results. -Library preparation is a crucial step in NGS workflows: the available amount of input DNA is key determinant. ***•Are any of these biomarkers used as a companion diagnostic for specific treatments? Inclusion into possible trials?*** No, they are not. ***•Which tools/assays do you use?*** To detect possible fusion genes, we use a different bioinformatic pipeline based on the use of multiple algorithms (e.g., FusionCatcher, TopHat, and FusionMap). To identify multiple somatic mutations in 50 oncogenes and tumour suppressor genes, we use the Ion Torrent AmpliSeq Cancer Hotspot Panelv2 (CHPv2) (ThermoFisher). For the gene expression analysis in 770 human genes and 830 microRNAs based on the nCounter system, we use PanCancer Pathway, PanCancer Immune Profiling, and microRNA assays (NanoString). PSQ analysis is performed using the Therascreen MGMT Pyro kit and PyroMark Q24 system (QIAGEN). ***What are your biggest challenges with identification of biomarkers?*** Intra and inter-tumour heterogeneity is a major challenge for the identification and development of potential biomarkers in solid cancers (e.g., ovarian cancer); Analysis of numerous RNA-seq datasets can reveal false-positive results; The high-resolution data achieved with NGS allow the identification of a large number of genomic alterations, with a considerable increase in computational analysis and bioinformatics support that are now needed for the production and interpretation of these data.
Clinical Genomics, Stockholm, Sweden -	***What types of samples do you use?*** •Fresh frozen / blood for all WGS and certain panels. •FFPE for solid tumour panels.

Somatic SV detection	***Specific cancers or patient populations?*** •Panels: Myeloid malignancies (diagnostics), lymphoid malignancies (research/pilot), breast and lung cancers (research / pilot). •WGS: Paediatric cancers (ongoing), acute leukaemia (ongoing), sarcoma (starting Jan 2022). ***Which technology do you use (amplicon panel, single-primer extension, WES, WGS, RNA-seq)?*** •Panel/WES or WGS: Tumour/Normal. •Panel/WES or WGS: Tumour only. •For all WGS cases we try to also do RNAseq on tumour. ***What is your sequencing depth and coverage?*** •WGS: ~30x (N), 90x (T) •Panel: >200x, but typically >1000x ***Which types and sizes of somatic SVs do you call?*** •CNA, duplication, deletion, insertion, inversion, translocation. ***Which callers do you use? Have you tested other callers, and decided against them? Which? Why?*** •SV: manta (WGS, panels), Delly (WGS) (BRASS (WGS), TIDDIT (WGS)). •CNA: cnvkit (panels), ascatNGS (WGS) (CNVnator/CNVpytor (WGS)). ***If more than a single caller is used for the same variant types, how do you decide on consensus?*** •Merge using SVDB (investigating loqusdb, svtools). ***What criteria do you use to filter SVs, and what do you do with FP calls?*** •Somatic, low-qual (investigating tool specific filters) ***What are your biggest challenges with identification of somatic SVs?*** •False positives. •Sensitivity. •Validation. •Genotyping across callers. •Format/db. •Comparison of tools/methods.
Clinical Genomics, Stockholm, Sweden Somatic NGS-based identification of complex biomarkers updates	***Which types of biomarkers does you lab identify?*** •TMB, MSI for some of the research projects (not in production). ***What methods do you use for calculating these scores?*** •Tumour Mutation Burden (TMB) defined as the number of somatic, coding, base substitution, and indel mutations per megabase in the tumour genome. •To calculate the TMB per megabase: The total number of mutations counted is divided by the size of the coding region of the targeted territory. •Non-coding alterations were not counted. •Alterations listed as known somatic alterations in COSMIC and truncations in tumour suppressor genes were not counted.

	•Alterations predicted to be germline by the somatic-germline-zygosity algorithm were not counted. •Alterations that were recurrently predicted to be germline in our cohort of clinical specimens were not counted. •Known germline alterations in dbSNP were not counted. •Microsatellite Instability (MSI) is a genomic alteration in which microsatellites, usually of 1-4 nucleotide repeats, accumulate mutations corresponding to deletions/insertions of a few nucleotides. ***Which tools/assays do you use?*** •MSIsensor-pro: A multinomial distribution model to quantify polymerase slippages for each tumour sample and a discriminative sites selection method to enable MSI detection without matched normal samples. •The MSIsensor-pro reports all detected microsatellites, the unstable(somatic) microsatellites and the MSI scores. •Are any of these biomarkers used as a companion diagnostic for specific treatments? Inclusion into possible trials? •Specific Cancers with high TMB (and MSI scores) can be treated with immune-modulating drugs. ***What are your biggest challenges with identification of biomarkers?*** TMB: •There is no consensus as to which mutations should be included in the TMB calculation. •TMB threshold used to define "TMB high" (which effectively is proposed to predict response to immune checkpoint blockade) has not yet been established for different cancer types.
Clinical Genomics, Uppsala, Sweden Somatic SV detection	•FFPE, blood, bone marrow; mostly T only and T/N for WGS. •Haematological malignancies, lung/colon/breast-ovarian/melanoma/GIST/CUP cancer. •Panel: 500-1000x (Capture), > (Amplicon); WGS: T/N 90x/30x; RNA: ~50M reads, smaller for panels. •CNV, translocations (fusions), ITD. •Panel: GATK CNV, CNVkit, ONCOCNV, Pindel; WGS: CNVkit, Manta, TIDDIT, in development; RNA-Seq: Arriba, STAR-Fusion, FusionCatcher, Illumina's own (TSO500). •Panel: overlap (CNVkit as base – ONCOCNV/GATK CNV for verification); WGS: No strategy yet; RNA-Seq: manual. •Panel: artefact based, clinically relevant; WGS: No strategy yet/clinically relevant; RNA-Seq: clinically relevant. •Choice of tools, consensus, filtering, FP, validation.
Clinical Genomics, Uppsala, Sweden Somatic NGS	***TMB:*** •In use: Illumina's app TSO500 •In dev: own algorithm •Use case: more general

based identification of complex biomarkers:	***MSI:*** •In use: Illumina's app TSO500 •In dev: MSI-sensor •Use case: colorectal cancer ***HRD:*** •in dev: own analysis based on CNVkit output •Use case: primarily breast/ovarian and lung cancer •Pitfalls: FFPE material -> stringent filtering, artefact filtering based on own data •Markers probably used in MEGALiT-study •Panels TSO500, Twist GMS560 •Lack of open-source software and analysis standards
Oncodia, Uppsala, Sweden	**A decade of development at Uppsala University resulting in several CE-IVD products, provided an update to their solution:** •Oncodia collaborates with different partners in research projects for HRD and LOH evaluation in cancer •There is no global consensus on the use of complex biomarkers at the clinical practise •Guided by the clinical needs of our users we will enhance the somatic mutation pipelines with the necessary metrics •Oncodia algorithms span from FASTQ to VCF, and incorporation of new mathematical models is easy
Centre for Genomic Medicine, Rigshospitalet, Denmark Somatic SV detection:	***What types of samples do you use?*** •FFPE (tumour only), Tumour/Normal (fresh), ctDNA, FNA, Bone Marrow ***Specific cancers or patient populations?*** •All - primarily solid cancers ***Which technology do you use (amplicon panel, single-primer extension, WES, WGS, RNA-seq)?*** •WGS, RNA-seq, TSO500 **What is your sequencing depth and coverage?** •WGS lower cut-off: 25x/50x blood/tumour, TSO500: median > 1000x ***Which types and sizes of somatic SVs do you call?*** •Currently answers are based only Oncoscan/Cytoscan. ***Which callers do you use? Have you tested other callers, and decided against them? Which? Why?*** •Calling from sequencing reads is form Manta, Delly, Lumpy, CNVator. ***If more than a single caller is used for the same variant types, how do decide on consensus?*** •Is based on germline experience, See Gabrielaites et al. ***What criteria do you use to filter SVs, and what do you do with FP calls?***

	•Background panel (PoN) ***What are your biggest challenges with identification of somatic SVs?*** •Very time consuming.
Centre for Genomic Medicine, Rigshospitalet, Denmark NGS-based NGS based identification of complex biomarkers:	***Which types of biomarkers does you lab identify (e.g., HRD, MMR, TMB, MSI, or other similar biomarkers)?*** •HRD (Arrays, CHORD), TMB (WGS), MSI (eyeballing, TSO500 score) assisted by COSMIC signature scores. ***What methods do you use for calculating these scores? Are there any common pitfalls?*** Yes. ***Are any of these biomarkers used as a companion diagnostic for specific treatments? Inclusion into possible trials?*** Yes, all. ***What are your biggest challenges with identification of biomarkers?*** •Calibration and value-distribution
HUS, Finland Somatic SV detection	•Blood, bone marrow, FFPE samples. We do not use tumour/normal pairs. •Haematological malignancies, lung cancer. •Comparative Genomic Hybridization, Targeted panels (Ion Torrent), amplicon panels, FISH, WES and WGS under development. •For haematological malignancies 4000X, other cancers 1000X. •Types: Fusions, translocations, CNVs.Sizes: in targeted panels whole gene deletions, duplications. •Callers for WGS: Manta, Svaba, ControlFreec, lumpy. WES: Codex & ExomeDepth. •Union + filtering. •Filtering against population databases. •Challenges: bad quality of old FFPE samples.
HUS, Finland NGS identification of complex biomarkers	•We are planning to start using TMB. •We are not using the other biomarkers at the moment.
Group discussion	
	Group discussions centred around the following questions and discussed in a plenary session: 1. What does your lab do on the topic? 2. What are the specific challenges? 3. What can we do in NACG to address these challenges?
Highlights from the group discussions:	Data sharing challenges were discussed but acknowledged as possibilities via a Nordic collaboration to address these. Approaches to address challenges to HRD and the value of sharing this knowledge was presented.

	Structural variants: • Benchmarking of SV callers is needed • Establish best practices for merging output of callers • Sharing SVs is important • There is a need to share validation datasets/samples • Need to stay in close contact with clinicians to better understand needs **Complex biomarkers:** • Harmonisation is needed for HRD • Arrays are not ideal, takes too much time • There is a need to benchmark various targeted solutions • Most present labs are interested in WGS-based solutions
Comments and questions	

Next NACG workshop

The next NACG workshop was announced, tentatively to take place in Iceland, between the 28-29th April 2022.

Nordic Alliance
for Clinical
Genomics