Nordic Alliance
for Clinical Genomics

WORKSHOP REPORT

NACG 10th workshop, 31.May – 4. June 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
03072021	1	Bobbie N. Ray-Sannerud

Symbols

Lecture / presentation

Interactive workshop

Abbreviations

FIMM	Institute for Molecular Medicine Finland
HUS	Helsinki University Hospital
IVDR	In-Vitro Diagnostics Medical Device Regulation
LoD	Level of detection
MDR	European Medical Devices
MOMA	Department of Molecular Medicine
NACG	Nordic Alliance for Clinical Genomics
OUS AMG	Oslo University Hospital – Department for Medical Genetics
PoN	panels of normal
RP	Retinitis pigmentosa
SV	Structural variant
SVDB	Structural variant database
WGS	Whole genome sequencing

Contents

Executive summary

This report summarizes the 10th workshop of the Nordic Alliance for Clinical Genomics (NACG). Due to the global covid-pandemic, the workshop was organized as a virtual NACG week, with daily two-hour lunch sessions 31. – 4. June 2021.

Even if we were unable to arrange for a physical meeting, the upside of a virtual event became very clear in that this format attracted an all-time-high audience of more than 165 registered participants[1] from about 55 different organizations in 11 countries, 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).

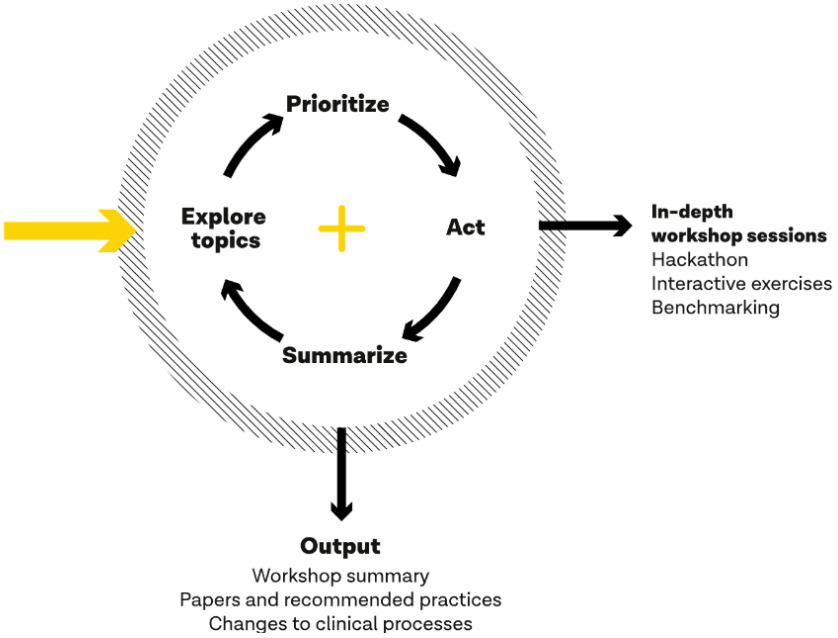


Figure 1 NACG members discuss and explore topics of interest to identify shared challenges and strategies for overcoming them. Prioritized topics are explored in in-depth interactive exercises. Findings and learnings are summarized in workshop summary reports and collaborative papers and contribute to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

[1] Actual workshop attendance: Opening and keynote (60), Automation of sequencing operations and data management (77), Digital dynamic consent (59), IVDR compliance (61), Somatic variant calling - benchmarking (70), Variant interpretation and data sharing (88), Clinical diagnostic unsolved cases (55).

NACG week agenda

The agenda for the NACG week is outlined in Table 1 with further introduction of the workshop sessions detailed in Table 2.

Table 1 NACG virtual week - agenda

Time (Oslo GMT+2)	Monday 31st May	Tuesday 1st June	Wednesday 2nd June	Thursday 3rd June	Friday 4th June
12:00	**Opening and keynote** **Genome sequencing in clinical microbiology** Rasmus L Marvig, Center for Genomic Medicine, Rigshospitalet DK	**Digital dynamic consent in clinical genomics** Sharmini Alagaratnam, and Courtney Nadeau, DNV NO	**Somatic variant calling - benchmarking exercise with in-silico spiked-in variants.** Oleg Agafonov, DNV NO, Valtteri Wirta, Clinical Genomics Stockholm, SciLifeLab SE	**Variant interpretation and data sharing** Dag Undlien, OUS NO; Stephen McAdam, DNV NO; and Sharmini Alagaratnam DNV NO	**Clinical diagnostic unsolved cases** Maria Rossing, Center for Genomic Medicine, Centre of Diagnostic Investigations, Rigshospitalet DK
12:30	**Automation of sequencing operations and data management** Tony Håndstad, Department of Medical Genetics, OUS NO				
13:00		**IVDR compliance progress** Cathrine Høgseth Nordhus, Department of Medical Genetics, OUS NO			**Social networking hour** Sharmini Alagaratnam, DNV NO
13:30			END		
14:00	END	END		END	END

Table 2 Description of workshop sessions

Topic	Description	Contact person
Keynote Genome sequencing in clinical microbiology	The session will provide an introduction to ***Genome sequencing in clinical microbiology***	Rasmus L. Marvig (rasmus.lykke.marvig@regionh.dk)
Automation of sequencing operations and data management	While most labs automate the execution of their variant calling pipeline, the further operational aspects of data management are often handled by manual procedures, cron jobs and many diverse scripts. As data management and analyses increase in complexity, an adequate level of automation is needed. In this session, we learn how some of the larger labs automate their operations and also learn about a specific open-source event-based system that can be particularly suited to the task. Target audience: All bioinformaticians and others interested in operations, automation and data management	Tony Håndstad (tonyha@extern.uio.no)
Digital dynamic consent in clinical genomics	Join us for an interdisciplinary, interactive workshop where we will explore challenges in consent across the clinical genetics landscape, identify common focus areas and share potential approaches to moving towards a more dynamic and digital reality. Target audience: Anyone and everyone working within clinical genetics who collect, manage and/or need consent, either directly or indirectly, for their work.	Sharmini Alagaratnam (Sharmini.Alagaratnam@dnv.com) & Courtney Nadeau (Courtney.David.Nadeau@dnv.com)
IVDR compliance progress	All actors in the field of medical genetics will have to comply with the new European Medical Devices (MDR) and In-Vitro Diagnostics Medical Device Regulation (IVDR) by May 2021 and May 2022 respectively. In this session Nordic laboratories will share the status of their efforts to secure compliance to the new regulations. The goal of the session is to compare the different laboratories' approaches to these regulations and to identify areas where the NACG members can work together to address challenges. Topics to be addressed are formats for collaboration, use of open source code, factory developed test arguments and market surveillance. Target audience: Labs or those interested in developing or deploying their own pipelines or software for genetic diagnostics.	Cathrine Høgseth Nordhus (cahnor@ous-hf.no)
Somatic variant calling - benchmarking exercise with	The adoption of molecular diagnostics based on NGS technologies is challenging from a quality assurance perspective. In contrast to more established assays, the often broad contents and technically complex workflows	Valtteri Wirta (valtteri.wirta@scilifelab.se) & Oleg Agafonov (oleg.agafonov@dnv.com)

in-silico spiked-in variants.	commonly seen in NGS diagnostics mean that assay validation and verification is difficult. In comparison to the identification of germ-line variation, somatic variation imposes an extra layer of difficulties due to several factors: (i) most tumour samples are comprised of an unknown fraction of both normal and tumour cells (tumour purity); (ii) the ploidy of cancer cells is unknown; (iii) due to the subclonal evolution the cancer cell population could be heterogeneous. In this session, we aim to gather best practices of somatic variant calling and subsequent variant filtering through a benchmarking exercise, for which we use deep sequenced, highly characterized reference with in-silico spiked-in variants in a set of oncogenes Target audience: Laboratories that perform clinical NGS-based oncology testing (Illumina, WES or gene panels).	
Variant interpretation and data sharing	The results from a pre-workshop variant sharing and variant interpretation benchmarking exercise focusing on hereditary cancer and retinopathy variants will be shared and discussed. Opportunities for extending these efforts to somatic variants will also be discussed. Target audience: Everyone involved in variant classification	Dag Erik Undlien (UXDAUN@ous-hf.no)
Clinical diagnostic unsolved cases	The session will cover case presentations from various clinicians &/or bioinformaticians on clinical diagnostic unsolved cases. If you have an interesting case you would like to share, please contact Maria Rossing and/or register your interest in contributing in the registration form. Target audience: Everyone involved in the processes around clinical diagnoses and solving of cases	Maria Rossing (caroline.maria.rossing@regionh.dk)
Social networking hour	We will be using Spatial.chat for our closing social event, to virtually recreate the bar at the end of a successful meeting. This app allows you to informally meet new participants and catch up with familiar faces. Try it here beforehand, and we're looking forward to seeing you all there!	Sharmini Alagaratnam (Sharmini.Alagaratnam@dnv.com)

NACG opening & keynote

Welcome and opening remarks

Number of participants in this session: 58

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

Key information

Dag welcomed the 10th NACG workshop and introduced the organisation as well as the ambitions for the week. 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. The broad audience and high attendance of over 165 registrations was mentioned and celebrated, confirming NACG's position as an important platform for collaboration in the Nordics where professionals come together to collaborate and share experiences to progress clinical genomics. Tentative dates to the next workshop at Århus Nov 25-26 or 18-19th with the hope that this will be a physical meeting was also mentioned.

Keynote

Number of participants in this session: 60

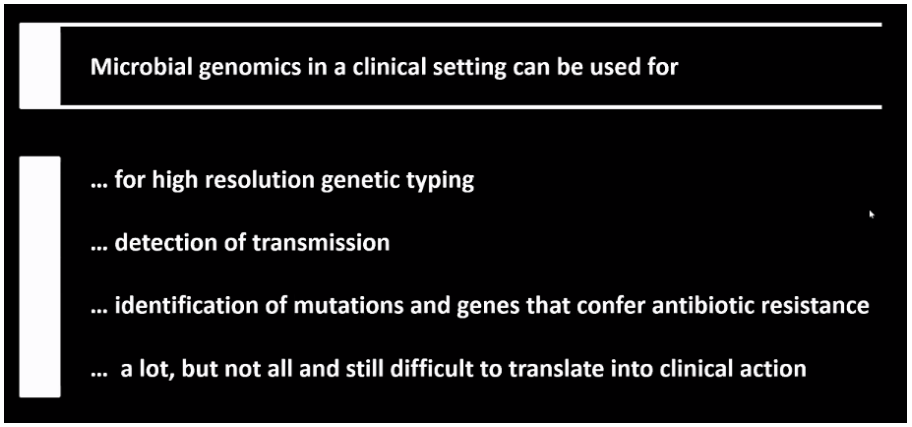
Speaker	Rasmus L Marvig, Center for Genomic Medicine, Rigshospitalet DK
Title	Genome sequencing in clinical microbiology

Key information:

Rasmus presented a keynote on the topic of genome sequencing in clinical microbiology. His presentation covered the following highlights:

- Looking at genome of microbial pathogen as opposed to in host
- Genomes are moving targets which mutates quickly, heterogeneity even within a single sample
- Aim to use NGS to answer 3 main questions: is there something, what is it, what can it do
- Collaboration with Department of Clinical Microbiology at Rigshospitalet
- 2,780 bacterial genome sequenced for routine diagnostics in period 2015-2021
- A total of 106 bacterial species has been genome sequenced as part of routine diagnostics
- Requirements for analysis varies across species
- Why WGS for microbiology? It can identify species, strain & clonal lineage (depending on species)
- WGS of P. aeruginosa from 36 CF patients: compared to find genetic diversity by pairwise SNP instances btw bacterial isolates
- Genotyping offers insight into strain dynamics to determine if patient is chronically infected
- WGS also identified Achromobacter clinical isolates as other than original A. cylosoxidans, and some other new species (previously based on MALDI-TOF of API N20 typing)

- Different gene content and phenotypes (incl more resistant to different antibiotics)
- Build phylogenetic trees of clone types to show evidence of transmission between patients also within hospital
- Sudden need for national surveillance of SARS-CoV-2: All samples from patients and staff at Rigshospitalet sequenced from Jan 1, 2021, and onwards.
- Local SARS-CoV-2 sequencing feeds into Danish Covid-19 Genome Consortium db for local and national surveillance



Q&A

Q: Are you sequencing mostly cultured material, or is it fecal/sputum material, whole-genome amplified, or flow-sorted or dropletted single cells? A: Yes, only cultured, to avoid the host DNA problem

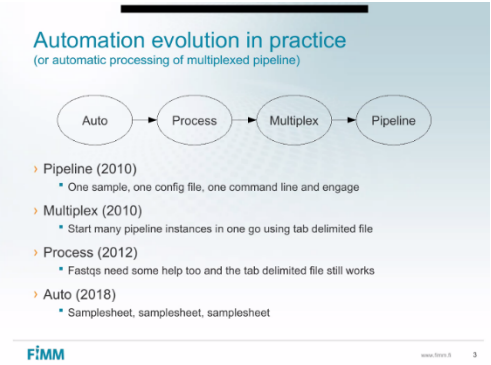
- A: 1. Mostly cultured material (bacteria) or PCR-amplified genomes (SARS-CoV-2). 2. MRD-style would be valuable but not something we do. We have tried metagenomic sequencing of blood samples to follow bacterial sepsis but the human to bacteria DNA ratio troubles us. Karious Dx is the frontrunner on this. 3. We sequence C. diff from transplanted patients to see if their C. diff infection is from self or from transmission. All MRSA are sequenced for national surveillance of this pathogen.
- Q: For another time perhaps, I would be interested to learn how you manage the (meta)data in microbiology, with so many different reference genomes etc. A: Good point and I agree. We have bacterial nomenclatures and reference genome dictionaries that we can rely on. We do not have smooth integration between genomes and patient metadata (except for basic information such as patient ID, ward, date of sample).

Automation of sequencing operations and data management

Number of participants in this session: 77

	Speaker	Tony Håndstad, Department of Medical Genetics, OUS NO
	Title	Automation of sequencing operations and data management
Key information	Tony Håndstad introduced the topic of automation of sequencing and data management by saying that automation helps to scale and standardize processes, improving throughput and quality. Automation is very central to bioinformatics but there are many ways to automate and building automation systems and adapting them to change can be laborious. If our needs for automation are increasing, how do we best invest in automation?	
Q&A	N/A	

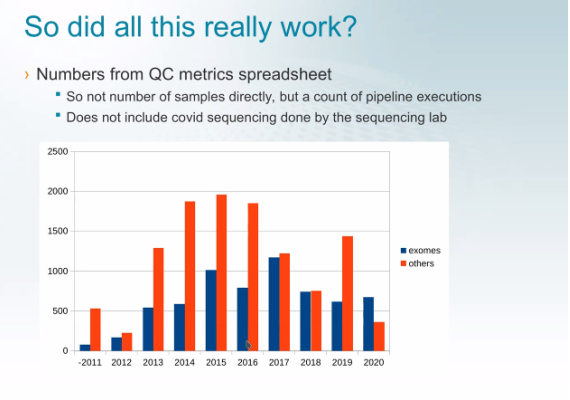
	Speaker	Henriki Almusa
	Title	Bioinformatician, FIMM
Key information	Henriki presented on automation at FIMM (covering only DNA projects with basic germline analysis) with the following highlights: - Henriki has created a simple automated system that starts from Excel-based sample sheet and runs de-multiplexing and variant calling pipelines - With the coming introduction of Dragen, automation is affected, e.g., the pipeline itself will be executed sequentially - The system is simple to start. The lab engineers have the information given by customers (paper → excel → iLab) and information flow always slows things down. The lab engineers at FIMM can manage the command line and start the pipeline. - Tracking execution can be done using a web page, qc is added to a spreadsheet, which allows for monitoring of trends	



- Process over 8-10 years to fully automate
- Disk usage: comments they are working with Finnish national supercomputer center (CSC) on getting the pipeline results go there and possibly move existing

data sets as well. Another disk array holds close to 550TB of sequencing data so total nearly 1PB

- Printout of conflict file for manual review before starting jobs
- Created website to see status of jobs and track execution
- Additionally track 30-40 different QC metrics, e.g. GBs, mean track coverage etc, to track sequencing quality between runs



Main issue for future (apart from transition to Dragen): Disk usage as now have nearly 1PB of data

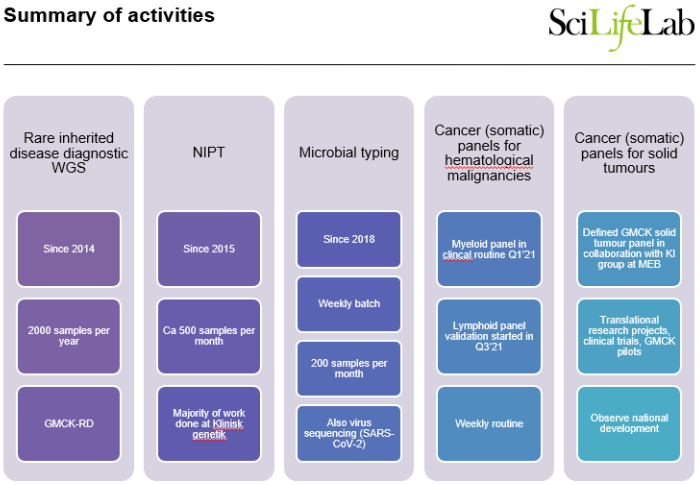
Q&A	Question to Henriki Almusa Q: Do you use any LIMS systems to store what has been run on samples? A: No, only a project management tool that is being used by seq lab. Can be extracted from sample sheets. Q: What will change when you switch to Dragen? A: Single machine, queue and pipe, so no stacking or batch-wise runs. Will probably still keep a compiled set of results from Dragen runs, will need to adapt.

	Speaker	Henrik Stanneheim
	Title	Head of Bioinformatics, Clinical Genomics, SciLifeLab

Key information

Henrik presented on the topic of Automation and data management at Clinical Genomics, SciLifeLab. The highlights of his talk are:

- Infrastructure specifically established for processing samples from clinical routine with personnel of approx. 45 FTE, 1/3 wetlab, 2/3 bioinformatics, SW development etc
- Use of high level of automation in prep lab, 2 Agilent Bravo Option B, 3 Hamilton NGS Star, High sequencing capacity, 3 NovaSeq™ 6000, 2 HiSeq™ 2500 (until end of Q2'21) and all in-house IT systems, HPC and associated Pb-scale storage and browser-based clinical decision support to include ISO17025 accredited analyses

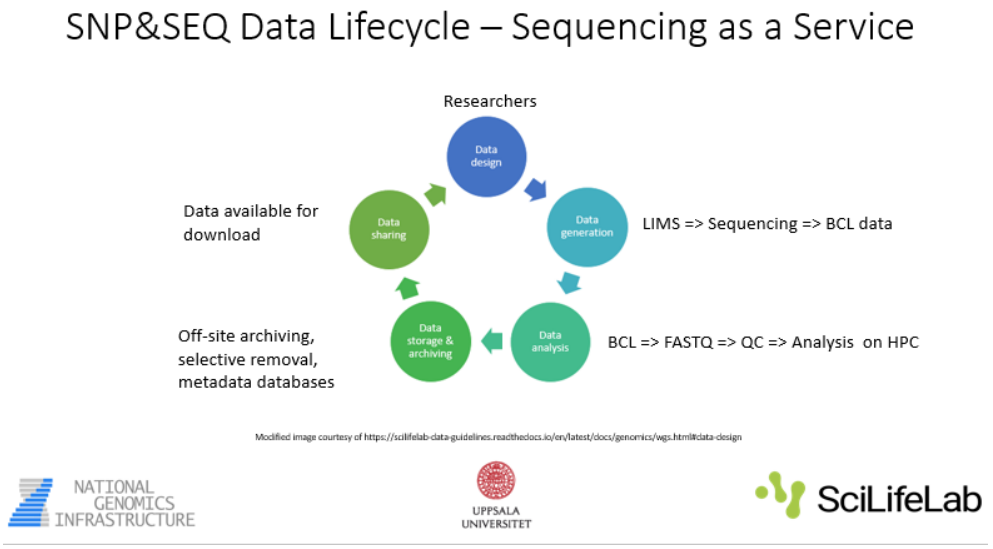


- Current automated activities: Data transfer from sequencing, Start demultiplexing, Start of workflows, Monitor workflows, Store results from workflows, Upload to in-house databases (e.g. chanjo, loqusdb, Scout), Upload to external databases e.g. GSAID, Generate delivery reports, Clean up intermediate data from pipelines, Compress/decompress FASTQ data, Archive BCL data, and Fetch archived BCL data.
- An internally developed system (CG) acts as an orchestrator to control the different activities. The state is stored in a "status database," and a web-based UI, Trailblazer, is used for monitoring. These systems are available as open-source, however they are tailored to the workflow at Clinical Genomics.
- A recent paper gives a good overview of their clinical activity: Stranneheim, H., Lagerstedt-Robinson, K., Magnusson, M. et al. Integration of whole genome sequencing into a healthcare setting: high diagnostic rates across multiple clinical entities in 3219 rare disease patients. Genome Med 13, 40 (2021)

Q&A	Q: When you are running so many different kinds of analysis, what is most important when building automation tools? A: Tight communication with customer, agree on what can be provided. Keep very clean APIs and where information is placed so data is cleanly separated but still connected, and making sure all the connections are good and smooth. Nail down rules for each analysis being done. Q: You seem to have made most of these systems yourselves? A: Historical, balance between freedom to develop as needed but also required efforts to update, maintain and document. Q: Is Trailblazer custom built or ready made software? It looks really nice! Would love to hear more about how it was built and integrated. A: Custom-built, is more integrated not standalone, not easy to install and adapt to run on a different cluster. But application is small so theoretically possible. Q: It seems like you've all had to custom-build everything. I know in other use cases there are orchestrator/master API software, is there any pre-made stuff that can help? A: Yes but it's hard to have APIs to LIMS etc within ecosystem. Difficult to have all APIs and keep it simple and agnostic. There are likely many ways to do this but so far this, no one correct way. Ours has worked well for us. Q: Which technology is used for hemato and solid tumour seq? A: Hyper plus for library prep, followed by twist for enrichment Q: Is Trailblazer custom built or ready made software? It looks really nice! Would love to hear more about how it was built and integrated. A: Was originally intended to be stand alone but has become more integrated (more than I would like). Not easy to install and run, but it could be done. All Clinical Genomics Stockholm tools are available on GitHub. https://github.com/Clinical-Genomics/. Trailblazer is here https://github.com/Clinical-Genomics/trailblazer.

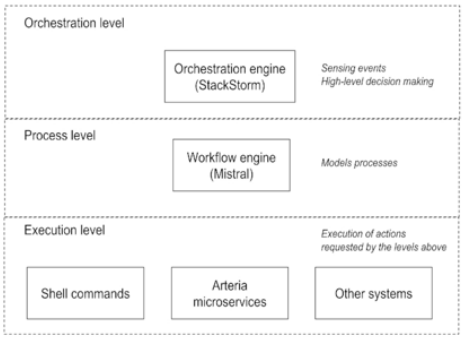
	**Speaker**	Mariya Lysenkova
	Title	Systems Developer and PhD student, National Genomics Infrastructure, SciLifeLab, Uppsala
Key information	Mariya Lysenkova presented on Arteria – an automation system for a sequencing core facility. The highlights of her talk included: - Presented info on NGI, https://ngisweden.scilifelab.se/ - Presented a need for automation to deal with increasing data complexity, seq tech, lab infrastructure, other systems integration, data processing pipelines, archiving procedures etc. - SNP&SEQ Data life cycle – Sequencing as a service. Presented how things were *before* Arteria project: Moving between stages of the cycle there was a lot	

of manual work, including shuffling data, checking for completed sequencing runs, scripts run manually, restarting of processes, error prone



- Designing automation: flexibility is key. Sequencing as a service means multiple workflows, where some of the processing is not simple I/O, e.g: Updating a database, Emailing reports, Third-party system integrations. Must account for differences between labs, ISO accreditation => everything must be trackable. Therefore, a flexible, extensible, and robust system is key
- A solution: event-driven architecture: Instead of having to manually trigger various steps, sensors automatically detect events and trigger actions. Manage WHEN and HOW actions are taken
- Arteria is an event-based framework for process automation at sequencing core facilities. It is a flexible, convenient interface for monitoring, helps to scale the number of samples processed, reduces human error => higher quality data. 200 WGS a month, now up to 9000 in total. Easier to track what has been done to a sample, good for compliance.
- Arteria packs contain re-usable units for the StackStorm event-driven platform. Then there are the microservices that do the actual work. The microservices of community interest are in the Arteria project on Github.
- Arteria implemented at 3 sequencing core facilities: SNP&SEQ Technology Platform at Science For Life Laboratory, Clinical Genomics Uppsala at Science For Life Laboratory, The University of Melbourne Center for Cancer Research.

Arteria Architecture



Arteria is built on existing open source technologies, with a modular design allowing for a community-driven effort to create plug-and-play micro-services

Gigascience, Volume 8, Issue 12, December 2019, giz135, https://doi.org/10.1093/gigascience/giz135

NATIONAL GENOMICS INFRASTRUCTURE

UPPSALA UNIVERSITET

SciLifeLab

- Open source and available on Github. Can run in Docker. Paper published

Q&A	Questions to Mariya Lysenkova (NGI SLL Uppsala) Q: Question around off the shelf Lims + Clarity A: Found Lims to be limiting in terms of integration (required many resources). Plan to replace this part of their system. Q: Is Arteria actively developed today? A: The core is not under active development but the microservices are. Microservices are not difficult, but understanding the entire infrastructure can be. Q: How does one use this? A: Available on GitHub and happy for more collaboration Q: How does this integrate with LIMS and Clarity for example? A: Using off the shelf LIMS is very limiting. Currently building a framework for exchanging to open source LIMS. Q: Maybe we need a bioinformatic LIMS, does Arteria do that in some way? A: Yes it does manage dataflows for bioinformaticians. Open for trying out and contributing? Q: Is it difficult to build microservices/executors/tasks yourself? A: No, the biggest hurdle is understanding the entire infrastructure, But even with a small instance of these running, it should be easy to import bioinfo logic into this service. Henrik: Our system is quite custom designed. A challenge when to break off a service as a microservice and when to keep it in the monolith. Mariya: If the organization is relatively limited, perhaps Arteira is overkill? High start up costs to implement. Good to have, but resource heavy to get there. Henrikki: We (at FIMM) have grown and expanded organically, Created tools as needed to ensure enough resources for maintenance and use.

Dynamic consent

Number of participants: 59

	Speaker	Sharmini Alagaratnam, and Courtney Nadeau, DNV NO
	Title	Digital dynamic consent in clinical genomics

Key information

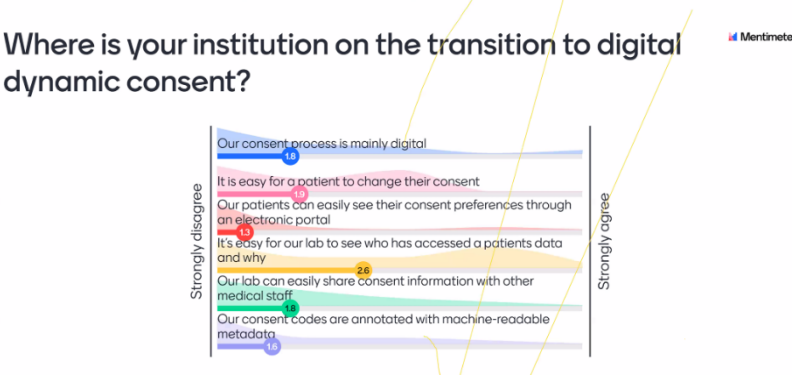
Workshop participants were polled to describe what the features of dynamic consent are:

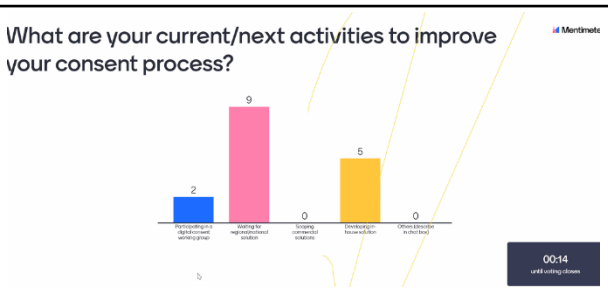
The hallmarks for DC were presented as including 1) options for consent can be presented and set through sustainable two-way communication between patients, healthcare professionals and researchers; 2) Patients have the option to modify their preferences over time if desired and when relevant; and 3) The preferences set the basis for dynamic downstream clinical and data management actions.

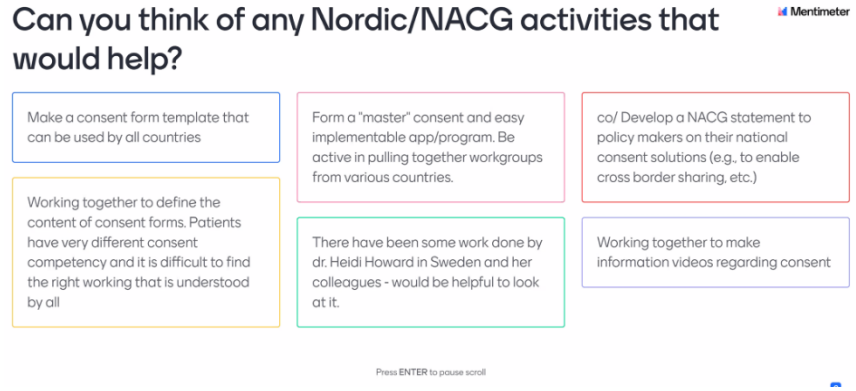
The benefits of DC to participants were control over data processing preferences, a reduction in laborious manual consent processing using static forms, patient empowerment, trust building, increased willingness to participate in research, patient engagement, and easier recontact.

The limitations of DC were described as technical cost to implement and maintain, the potential for consent fatigue, potential for data fragmentation, overcoming institutional inertia to change entrenched systems, and still weak evidence of benefits.

Participants were polled on the status of DC at their institutions:







A panel discussion was held to reflect on the above topics consisting of 1) Elsebet Østergaard, Clinical geneticist & assoc. prof, Rigshospitalet, Copenhagen, DK, 2) Mahsa Shabani, Assistant prof in privacy law, Univ Ghent, BE, 3) Wenche Sjursen, Clinical Laboratory Geneticist, St Olavs Hospital, Trondheim, NO, Professor, NTNU, NO. Highlights of panel discussion were:

- Elsebet comments that in Denmark they still use a paper form which is scanned and uploaded to their patient journal. Refers to this as “old fashioned” and desired a digital consent format with features that allow patients to change preferences. This would be valuable as it will reduce admin burden and for patients it will be easier for them to be more hands on. Currently, it is not possible to develop an in-house solution but they are making comments to the Danish authorities on how this consent solution should look. Also working on developing educational videos to patients to explain genetic testing and sequencing to help reduce time in the genetic counseling and to focus more on their questions.
- Mahsa discussed topics related to the secondary use of genomic data. A challenge is to identify what is the best model of consent for reuse for secondary purposes. She is specifically working on the EU health data space and the topic of secondary use of data. At a recent meeting, it was discussed a challenge in understanding the expectations of patients, for regulators what is required for consent, and these questions drive frustrations in blocking progress. Commented that the first step is to bring more clarity for a model in enabling a use of secondary data and then, establish a harmony across member states (e.g. across border data sharing). Also mentioned was the EU Data Governance Act proposal where the topic of consent is introduced and the pillar on data altruism which has implications for consent. The question is if it will imply harmonize consent or for member states to figure this out on their own.
- Wenche’s comments from a medical genetics perspective in Trondheim, Norway. They currently use several paper forms that are scanned and put into the patient’s EHR. They have no visibility from the lab side to these forms, even though patient preferences may influence the analysis. She is in a group developing consent for research purposes, currently now only within inherited

	cancer. She finds the paper forms are too many and there is a desire for a digital solution to engage with patients on research opportunities and to share information on research projects. There currently is no digital platform for this. They are waiting on regional solutions in the short term but ideally it should be a national solution. Currently the health platform in development in Norway (to be released in 2022) may have a consent solution but she has not received clarity on this despite efforts to inquire.
Q&A	Comments from the audience were: - A comment from a Norway perspective, that requisitions can be done in the lab, but consent is best handled nationally. - A comment from a Danish perspective, not looking forward to a national/regional solution but it is the only route to go as in-house solutions will not be accepted by the authorities. - A comment from a Finnish perspective, eagerly awaiting a digital solution to consent but feel this is taking a long time to be realized. Currently, test requisition is electronic but lab systems are disconnected from the clinic which is problematic. - Comments around how often patients change their consent preferences, panel responded that this is quite rare but it does happen. This is also dependent on whether or not this would be a research vs. clinical context. In a clinical context, one use case is when a child receives a diagnosis, but an adult relative does not want to know their carrier status.

IVDR compliance

Speaker	Cathrine Høgseth Nordhus, Section Manager for Quality at the Department of Medical Genetics at OUS
Title	IVDR compliance progress[2]
Objective	To compare the different laboratories' approaches to IVDR regulations and to identify areas where the NACG members can work together to address challenges.

Number of participants: 61

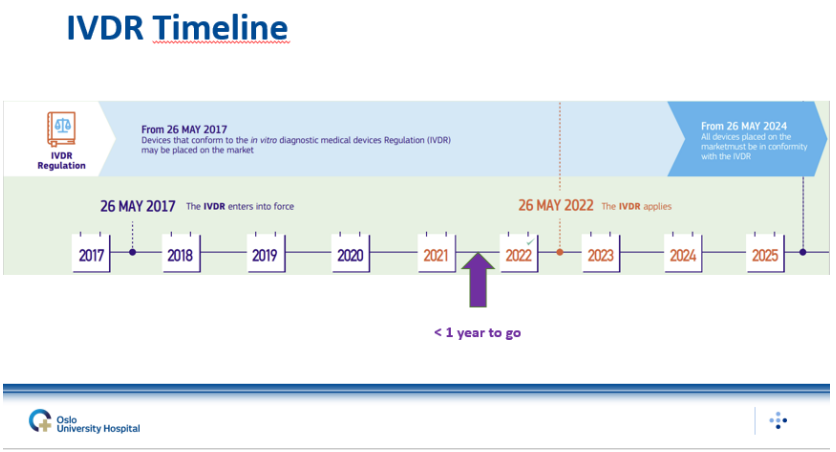
Key Information

IVDR has been a topic at three previous NACG workshops as described in the workshop reports[3]:

- In November 2018, Courtney Nadeau (DNV) presented an introduction to the IVDR requirements.
- In November 2019, Alexey Shiryaev and Nick Baker (both DNV) gave an overview of the regulation and discussed the applicability and requirements for transition.
- In November 2020 Cathrine presented risks and challenges, and sought opportunities to collaborate.

Relevant papers are also available through the BigMed project[4].

The IVDR timeline is as follows, with less than one year ahead:



Cathrine presented a status of the IVDR compliance process at OUS and in the health regions in Norway. Progress to date has involved collaboration with other

[2] For this session, slides and other resources are made available at https://nordicclinicalgenomics.org/projects/preparing-for-ivdr.
[3] https://nordicclinicalgenomics.org/resources#report
[4] https://bigmed.no/assets/Reports/clinical_decision_support_software.pdf and https://bigmed.no/assets/Reports/clinical-sequencin-g_regulatory-frameworks-and-quality-assurance-for-ngs-based-diagnostics.pdf

	clinics in OUS and with other hospitals. Current activities involve mapping out in-house diagnostics in the various laboratory medicine departments, as these will be regulated as lab-developed tests under IVDR. After this mapping, departments will need to implement procedures to check whether equivalent CE-marked tests are available on the market, a process which is currently difficult due to the unavailability of certain modules of EUDAMED. Additionally, after this inventory, a process will be initiated to determine the further course of action for these tests: either to adopt a CE-approved test (if available), to offer these as lab-developed tests under IVDR, or to develop these tests into CE-approved diagnostics. She reported that in-house templates have been useful in the process due to the large number of tests provided, and that these have been used by other units. The approach to self-declaration by OUS for IVDR was described as: • The department of medical genetics is accredited to ISO 15189, and in discussions within the work group in Health Region South East the medical genetics work group the medical genetics lab at OUS has decided to set up the initial self-declaration document based on the accreditation scope (Part B of the self-declaration) • The various method descriptions are more or less identical, e.g. for the NGS methods it is only the panel filtering criteria that differs from panel to panel. • We have established an excel template where we record compliance with each general safety and performance requirement for every method. • If later, we decide that we have to prepare one declaration per method, it will be easy to expand the excel spreadsheet with one row per panel, including specific details on indication per panel under intended use, and adding details on the list of genes for each panel in the “Electronically programmable systems” column in part C of the Self Declaration • An overview of Northern Ireland Health Inst exemption was presented. • Mention that ISO 62304 has guidance on software of unknown provenance and Classification of software under MDCG 2019-11
Discussion / questions / comments	- Q: Did you mention that you are aiming to survey every 3rd year the market for available IVDR solutions on the market? What action will be taken if an IVDR solution becomes available during the 3 year period and you are informed about the availability at that time point? - A: Believe that these guidelines will become more clear.

Somatic variant calling benchmarking

Number of workshop participants in this session: 70

Speaker	Valtteri Wirta (SciLifeLab) and Oleg Agafonov (DNV)
Title	Somatic variant calling - benchmarking exercise with in-silico spiked-in variants.

- Background to the workshop was described: During the NACG event in 2020 benchmarking results for HD832 and HD789 were presented, results indicated that HD samples can't be used to calculate precision HD samples are created by mixing several cell lines, and only a subset of variants is characterized.Therefore, the benchmark in this workshop addresses this by using in silico engineered variants in exome and simulated small targeted panel data sets. For the exercise, a NA12878 dataset sequenced at SciLifeLab was used:
    - Prep: KAPA HyperPlus
    - Target enrichment: Twist Bioscience (8-plex)
    - Indices: IDT Duplex Seq adapter + indexing primers (UDI + UMI)
    - Sequencing: NovaSeq 6000
    - Sample source: NA12878 (250 ng), 673 million read pairs (approx 5500x coverage)
- Reads were mapped to a reference genome GRCh37 and downsampled to 500x for WES and to 2000X for simulation of targeted gene panels. Variants (SNVs, INDELs, and CNVs) were bioinformatically spiked-in into cancer relevant genes with BAMSurgeon.Two types of datasets were distributed:
    - WES - FASTQ files for tumour and matching normal
    - FASTQ files for 7 sets of small gene panels, where the same variants were introduced to each set at different VAF levels (only tumour).

- Overview of variant calling across the Nordic labs was presented:
    - **FIMM, Finland**
      Genome hg38 was used as reference, used dragen 3.8 to align both tumor and normal sample, Truseq adapters were given for trimming options as well as poly-g and minimum quality 3, UMI information was not used. Variant calling was done on the tumor sample using both samples with joint detection turned on, Germline on normal sample and somatic t+n for tumor.
    - **HUS, Finland**
      GRCh38 was used as reference genome.Two strategies were used: in Strategy 1, trimmed UMIs were combined into separate UMI sequence, Strategy 2: UMIs were trimmed off, not used. Burrows-Wheeler Alignment Tool (BWA-MEM) with GRCh38 (strategy 1: GRCh38.p12 EnsEMBL, strategy 2: GRCh38.p12 v0 GATK).Variant calling with Mutect2 (strategy 1: GATK 3.8, strategy 2: GATK 4.1.0.0). Annotation with ANNOVAR v. 2018-04-16 (strategy 1) and Variant Effect Predictor v 100.4 (strategy 2).
    - **Oncodia AB, Sweden**

Proprietary Oncodia pipeline was used: data was preprocessed with Oncodia's own trimmer, aligner and VARify mutation caller. Variant calling was done with default settings. VARify is a general solution that fits any workflow and outputs only raw mutation calls (no BED-file assumptions and no post-filtering). As an exception for the NACG benchmarking, the VCFs were annotated with ANNOVAR, but not filtered in any way.

- **OUS, DMG, Norway**
  Did not use UMI-aware pipeline. Both normal and tumor fastq files were first trimmed with adaptors by Trimmomatic v0.39. The trimmed FASTQ files were then analyzed by both Dragen somatic pipeline with matched tumor-normal pairs and GATK somatic short variant discovery pipeline with tumor-normal pairs. Default parameters for both pipelines were used. In the GATK somatic pipeline, the trimmed reads were mapped to the hg19 reference genome by using BWA MEM v0.7.12 and samtools v1.9. The duplicated reads were marked by using Picard tools v2.22.4. The resulting BAM files from both tumor and normal samples were then used as inputs for GATK (v4.1.0) Mutect2 for calling short variants. CNVs were not analyzed.
- **SciLifeLab, Sweden**
  BALSAMIC pipeline v7.2.2 was used to analyze each of the FASTQ files.
  - Quality control and trimming: FastQC v0.11.5 and fastp v0.20.0, MultiQC v1.7.
  - Genome alignment and post processing: BWA MEM v0.7.15, samtools v1.6 5, Picard tools MarkDuplicate v2.17.0
  - Variant calling: VarDict v2019.06.04.
  - Annotation: Ensembl VEP v100 and vcfanno v0.3.2.
  - Variant filtering: DP > 100, AD > 5, AF > 0.01, MQ >= 40, and GNOMAD AF_popmax < 0.005
- **MOMA, Denmark**
  - Mapping to hg38 without alternate contigs and decoys was done with bwa mem v0.7.17 with default settings.
  - Adapter trimming was performed with cutadapt v3.0 with settings --minimum-length=20, --error-rate=0.1, --quality-cutoff=20, and --overlap=1 (the ones used in the Trim Galore! wrapper).
  - UMI information was not used. By mistake, UMIs were not trimmed before the analysis, which led to a lower sensitivity due to filtered calls in noisy regions.
  - Variants were called with Mutect2 (GATK v4.1.9.0) and subsequently filtered with FilterMutectCalls with the non-default settings --max-events-in-region=3, --min-slippage-length=8, --normal-p-value-threshold=0.0001.
  - Strelka v2.9.10 was used for a second opinion. If a variant was filtered by FilterMutectCalls but called as PASS by Strelka, the call was rescued. This proved helpful for hard-to-call variants while still maintaining a high precision.
  - VCFs were filtered using standard hard-filters used in diagnostic settings: variants with allele frequency < 2% or fewer than five supporting reads were filtered.
  - Copy numbers are called with CNVkit v0.9.8.

- **Center for Genomic Medicine, Denmark**
  GATK pipeline:
  - FASTQs were trimmed using bbduk (v. 38.26).
  - Used pipeline was not UMI-aware and they were not trimmed.
  - Alignment to reference genome hg38 by bwa mem (v. 0.7.15).
  - Duplicates were marked with bamsormadup (v. 2.0.95) from package

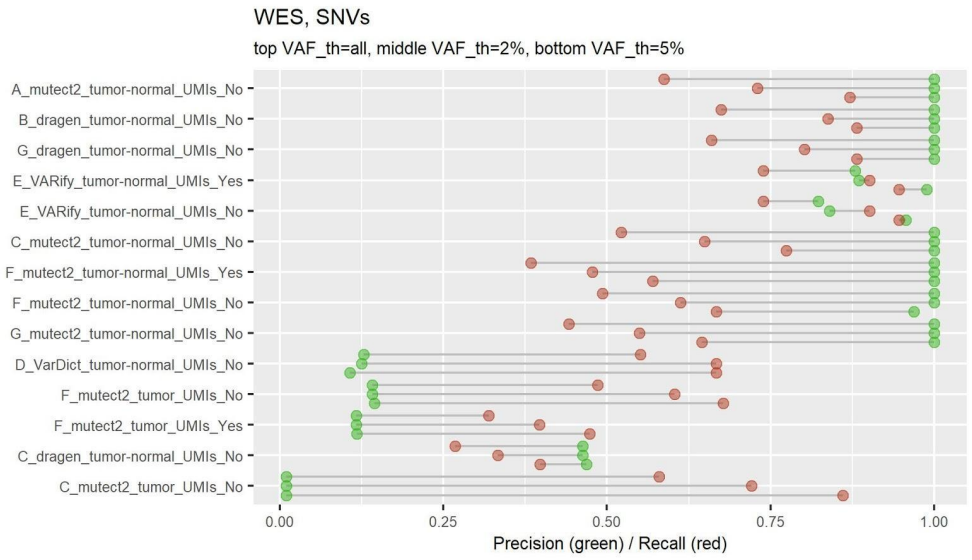
biobambam2

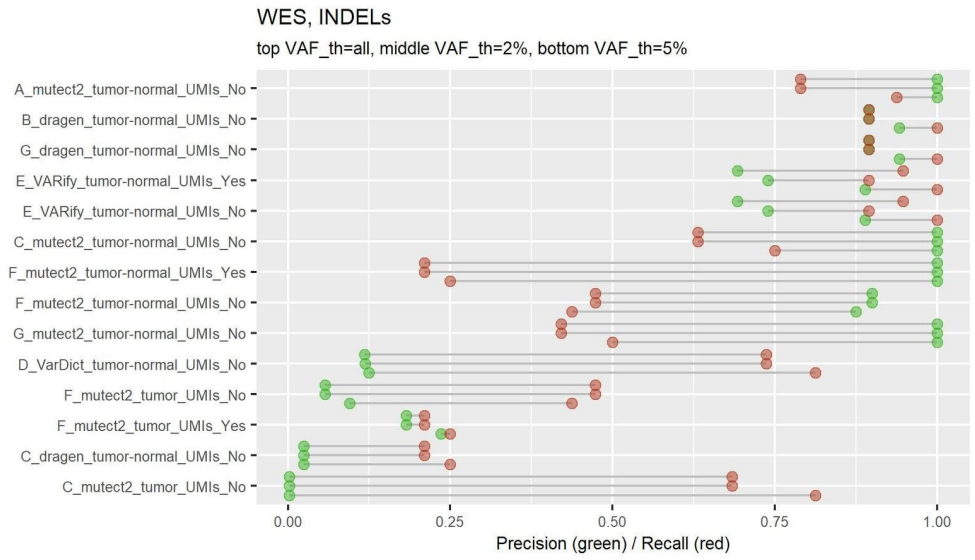
- Somatic variant calling by Mutect2 (v. 4.1.9.0). WES and WES-panels were run as both paired tumor-normal, and tumor-only, with the employment of an internal Panel of Normals.
- Variants were annotated by VEP (v. 100.4) and hard-filtered by allele frequency with cutoff set to 5% (GnomAD)
- CNV calling was not performed.

DRAGEN pipeline:

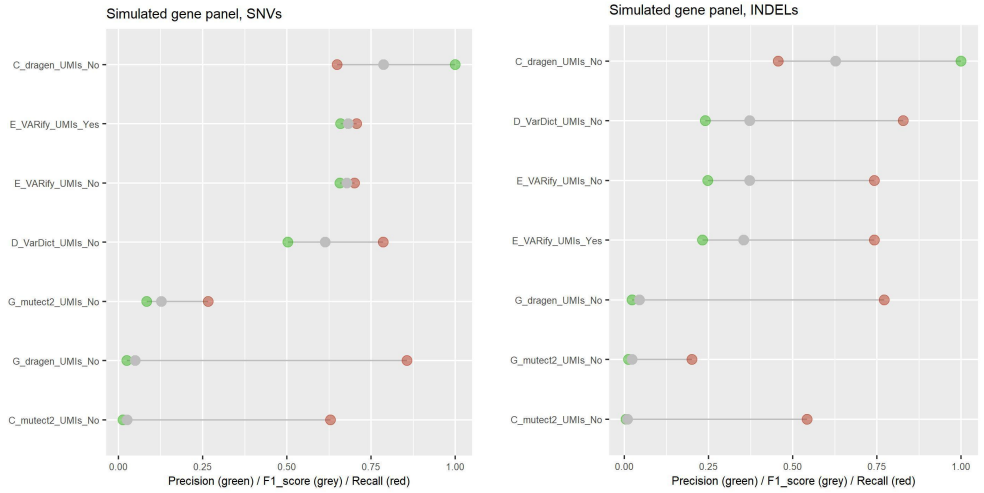
- Dragen v. 3.4
- WES and WES-panels run only as paired tumor-normal
- UMI-aware pipeline
- Reference genome hg38 (no contigs)

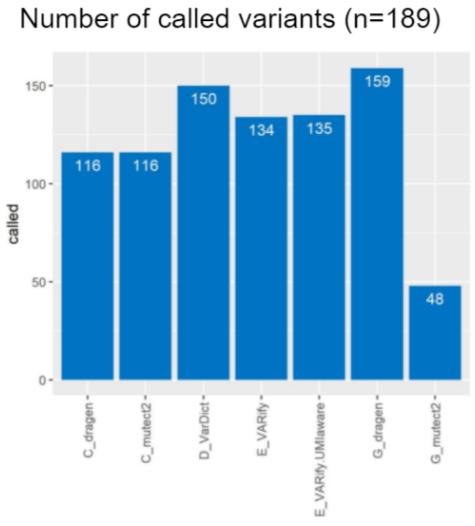
- Overview of the pipelines performance for WES dataset for 3 sets of VAF thresholds - all variants, 2% and 5%.

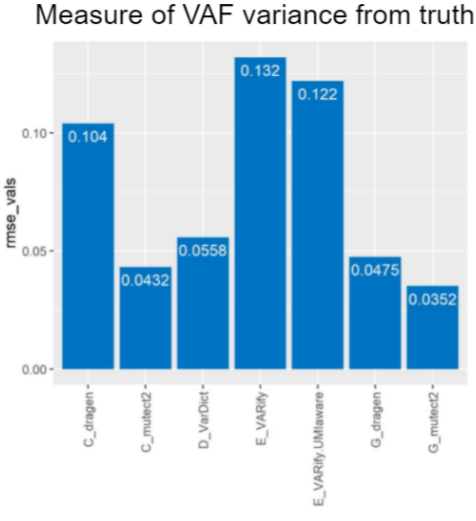




- Overview of the results for simulation of targeted gene panel dataset, presented as the following:







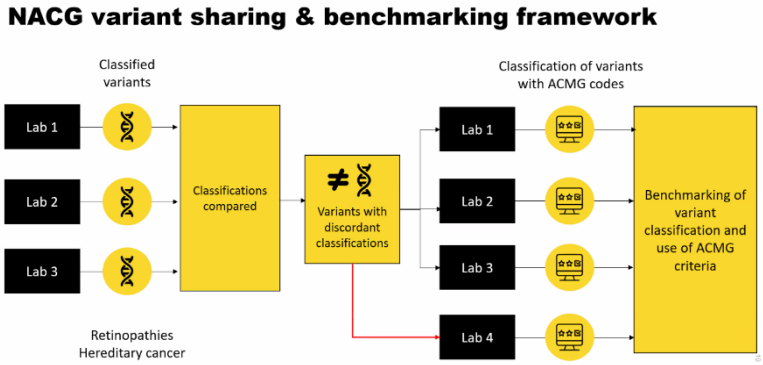
	- The second half of the workshop focused on a discussion around the following topics: Is Recall at the expected level for your lab? If it is not, could you elaborate on why? Is Precision at the expected level for your lab? If it is not, could you elaborate on why? What are good practices to deal with low VAF variants/ variants close to germline VAFs?
Future actions	With use of the MIRO board, ideas for the next workshop were proposed and discussed: - **DRAGEN-related:** - Work together on understanding and improving DRAGEN - Nordic DRAGEN user peer support group/channel/slack/monthly zoom - **Follow-up:** - Participating labs to look in more detail into results - From this benchmark: Is there a summary of what pipes were missing? Always the same usual variants? - Establish working group on use of UMIs for detection of very low VAF variants - **New Benchmarks:** - Benchmarking for germline variants - WGS Somatic Benchmarking - Somatic Benchmarking for RNA Fusions - CNV truth set for benchmarking CNV calls - Interpretation of variants (structural, CNV, SNV, Indels) in a somatic setting - Somatic benchmarking of structural variants - **New Topics:** - Tumour only filtering strategy/variant filtering strategy (2) - Liquid biopsy/ctDNA - TMB - HRD from WES (2) Mutational signatures - Cancer evolution/subclone reconstruction

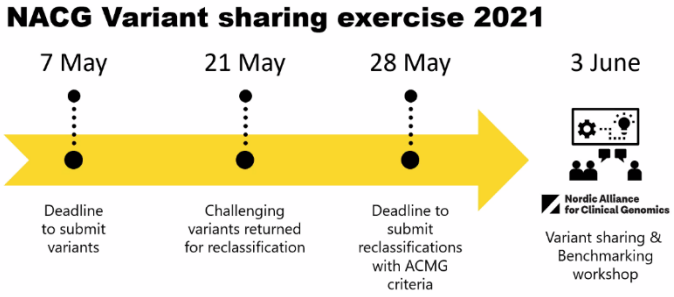
Variant interpretation and data sharing

Number of participants in workshop: 88

Speaker	Dag Undlien, OUS NO and Sharmini Alagaratnam DNV NO
Title	Variant interpretation and data sharing

- Dag opened the session with justification around the value and importance of data sharing, which is not only about technical issues, but also legal and organizational infrastructure issues. More recent work has been focused on the sharing of variant classifications, and in March 2021 a variant classification sharing and benchmarking exercise was planned and executed with NACG members and other parties who indicated their interest. The figures below show the design and timeline of this exercise, which aimed to address the question of if ACMG criteria for classification were being applied in the same way across participating units.



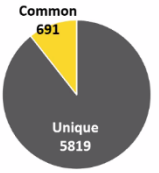


6 organizations from all 5 Nordic countries participated in this exercise. A total of 7058 variant classification in 404 genes were submitted, mainly in hereditary cancer genes but also a handful of retinopathy variants.

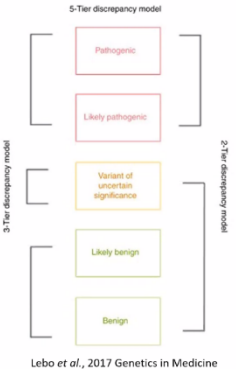
Variant sharing and comparison was achieved through the Variant Exchange software in beta testing mode from DNV and produced the following results:

Results – variant sharing

- **6510/7058 were successfully validated for use of HGVS and ACMG classification**
- **Classifications compared using Variant Exchange in beta test mode: 2 vs. 3 vs. 5 tier comparison**



- 2-tier Discordant: 6 Concordant: 336
- 3-tier Discordant: 47 Concordant: 295
- 5-tier Discordant: 109 Concordant: 233



Lebo *et al.*, 2017 Genetics in Medicine

- Comparison of variant classifications identified 6 variants identified as discordant between two or more labs. These variants were returned to all participating labs for (re)classification with ACMG criteria and other evidence if not already submitted, and also made available to one additional unit which wanted to participate. Although the classifications agreed largely, criteria used vary.
- Teresia Wangensteen (OUS, NO) and Christa Schmidt (HUS, FI) facilitated a discussion of the classifications and criteria that were submitted for the 6 discordant variants. In brief, the discussion around the following variants included:

  BRCA1 variant NM_007294.3: c.4096+3A>G: Previously known discordance in classification, with additional familial evidence for pathogenicity observed in Iceland in particular. Why are criteria applied so differently? Many different guidelines and recommendations on how to use them exist, in addition to different perceptions on how they should be used. Useful to know which classification system was used.
  Annika - Discrepancies in ACMG criteria use are observed every time we participate in external QC exercises, even if all come to the same conclusion on class, which is ultimately most important, in addition to being efficient (do not exhaust search for criteria once final class has been reached).
  Sarah – External QC makes us more aware of how we use the different evidence. May result in more similar use over time, but still a way to go.
  Dag. When we don't reach the same classification, is this because we apply criteria differently?
  Annika – Depends also on the unit performing classification and their experience. Do they work mainly with a few genes they know very well, or large panels where they need to work with all genes? There are some gene specific applications to criteria!
  Need to listen carefully to expert panels, for example for Lynch genes.

  HOXB13 variant NM_006361.5:c.251G>A: initially classified as benign by one unit, but this was due to variation in use of the class for risk variants.
  How to apply AMCG on variants with low penetrance? Jon: People with high and low penetrance variants in Iceland are offered the same surveillance programme, but when to use pathogenic and when to use risk variant?
  Sarah: ENIGMA has tried to develop recommendations on this.
  Wenche: At St Olav we test for this and report it as a risk variant, but its significance is explained more carefully during genetic counselling. More and more important for labs and clinicians to communicate on how to report low penetrance variants. At gene level can reach LP level, but need to keep

	penetrance in mind. Are we moving into polygenic risk scores with genes like these?
	
	MLH1 variant NM_000249.3:c.191A>G: Large variation in criteria applied Pathogenic in LOVD in the past. IN OUS only seen once, and downgraded to VUS. Any other experiences to share?
	Dag: Cancer has more supplementary guidelines compared to rare diseases, but criteria do give us the possibility of clarifying differences in opinion. Better done IRL!
	- *TP53 variant NM_001126112.2:c.394A>G* Two units used ClinGen expert group gene-specific evidence, reaching pathogenic using supporting and moderate evidence. Noted that use of PP5 is no longer recommended.
	Jon: different criteria but same class. Lots of data here! Possible to calculate some kind of average?
	
	- Potential next steps suggested were:
	  - Agree on a process for how to process discordances and document discussions.
	  - It would be beneficial to add phenotype to Variant Exchange.
	  - Test real-world sharing for a period
	  - API implementation would facilitate submission of variants
	  - Alert function similar to ClinVar
	  - Questionnaire to workshop participants on the way forward and interest in participating?

Comments and questions	- Q:Is it possible to compare ACMG criteria used by different labs in Variant Exchange ? A: Yes, when labs have submitted criteria and are sharing partners - Q: What is the view on expert panel classifications? E.g. ENIGMA (expert panel on BRCA1/2) has classified the variant pathogenic A: Participants reported taking expert panels advice seriously, especially for breast cancer genes. - Q: We see discrepancies in ACMG criteria use each time we participate in external quality control even if all come to the same conclusion on class. A gene is a gene. A: Yes but phenotypes can be different! - Q: When to use the classification pathogenic vs the risk variant? A: ENIGMA has published on this issue. Definition used by ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/docs/clinsig/ Risk factor: For variants that are interpreted not to cause a disorder but to increase the risk. Key Word(s) in Title of OMIM Entry or description in OMIM's genemap file: "SUSCEPTIBILITY TO"; "SUSCEPTIBILITY" followed by a numeral "MODIFIER OF" not "RESPONSE"

Clinical diagnostic unsolved cases

Number of workshop participants: 55

	**Speaker**	Maria Rossing, Center for Genomic Medicine, Centre of Diagnostic Investigations, Rigshospitalet DK
	Title	Clinical diagnostic unsolved cases

Key information	-	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.

Discussion / questions / comments	-	N/A

The workshop concluded with a social hour with use of the virtual platform spatial.chat.

Next NACG workshop

The next NACG workshop will be arranged physically 25-26 November 2021 in Århus, Denmark.

Nordic Alliance for Clinical Genomics