NORDIC CLINICAL GENOMICS DATA SHARING WORKSHOP 3
18.-19.04.2017

Workshop summary report

Oslo Universitetssykehus
SciLifeLab
Karolinska Universitetssjukhuset
University Hospital Copenhagen (Rigshospitalet)
DNV GL

Report No.: 3, Rev. 01
Date: 2017.08.09

Project name:	Nordic Clinical Genomics Data Sharing workshop 3 18.-19.04.2017
Report title:	Workshop summary report
Participants:	Oslo Universitetssykehus SciLifeLab Karolinska Universitetssjukhuset University Hospital Copenhagen (Rigshospitalet) DNV GL
Date of issue:	2017.08.09
Revision:	01

Objective:

The objective of this report is to summarize topics discussed and agreed actions from the Nordic Clinical Genomics Data Sharing workshop at Rigshospitalet's Kennedy Centre in Copenhagen 18.-19. April 2017.

Prepared by:

Guro Meldre Pedersen
Principlal Researcher
DNV GL

Sharmini Alagaratnam
Principal Researcher
DNV GL

Rev. No.	Date	Reason for Issue
0	20170612	Draft workshop summary report
1	20170809	Final workshop summary report

Table of contents

1 EXECUTIVE SUMMARY

This report summarizes the workshop held in Copenhagen 18.-19. April 2017. The agenda included a session exploring legal barriers to data sharing and the implications of the EU General Data Protection Regulation (GDPR). The bioinformatics tools in use by the participant laboratories where explored and sessions benchmarking laboratories in terms of variant calling and variant classification were performed. Finally issues around data sharing were discussed and a wish list of functions for data sharing were developed by the labs. Key decisions and conclusions from the two days included:

- Variant calling across three pipelines using raw sequence data from NIST reference material NA12878 (HG001) allowed side-by-side comparison of the benchmarking output (true/false positives, true/false negatives) and for the effect of particular parameter/ pipeline choices to be teased out. Based on this evidence and these results, one unit decided to adapt their choice of parameters to improve performance of their production pipeline.
- Studies exploring how useful national and local databases might be for the filtering out of "normal" variants indicated this would be limited. The international databases Exac was able to reduce the number of variants of interest in found in 879 exomes analysed against the BevegForst gene panel from over 300 000 to 11 000, while the SweGen national population database and the OUS in house database were only able to remove a further 30 and 153 variants respectively.
- Two of the laboratories are working to explore, develop and test bioinformatic tools that address structural variants. There are indications that including these in production could improve diagnostic yield by 5-10%
- Benchmarking across four laboratories of the classification of 39 variants identified 12 variants might impact on medical management. While some of the discrepancies may be due to a superficial evaluation not representative of a real clinical setting, others discrepancies indicated more fundamental challenges that should be further explored. These included individual knowledge related to specific genes and competence.
- Although participants in Workshops I and II showed some initial interest in the sharing of variant frequencies this was not now seen to be a priority for clinical labs. Rather there was consensus on the need to identify and manage discrepancies in variant interpretation/classification (see Section 4.2 Benchmarking – variant interpretation). It was agreed to explore how variant classifications could be shared and this should include standardised reporting that provides evidence that supports the classification.
- Discussions from the partners indicated that there was variation in how and when variants were reported to patients, especially related to variants of uncertain significance. It was agreed that this should be explored in an upcoming workshop.
- The participants recognised the continuing value of the workshops and wished to formalize the cooperation. A draft constitution based on the GA4GH model will be prepared by DNV GL and circulated. The formalizing of the collaboration will be part of the Nordforsk funded Nordic Alliance for Sequencing and Personalized Medicine.

1.1 About the workshop – background and objectives

This workshop was the third in a series of workshops between Nordic entities focusing on responsible sharing of trustworthy data for improved diagnosis and treatment, and for research. The previous workshops are summarized in as Clinical Genomics Data Sharing – workshop summary reports No 1 and 2, available at the community Sharepoint or by contacting the authors of this report.

Meeting objectives included:

- Formalise Nordic clinical genomics data sharing collaboration
- Continue sharing of data, tools and methods
- Review progress and contribute to sharing of genomic variants
- Discuss quality of data and processes through benchmarking of variant calling and classification
- Keep each other informed about ongoing projects and processes related to clinical genomics data sharing
- Understand legal barriers to sharing of genomic data

1.2 Workshop participants

The workshop included representatives from the below units. A full list of participants is provided in Appendix 2.

- Oslo University Hospital, Department of Medical Genetics, Norway
- Copenhagen University Hospital (Rigshospitalet) and the University of Copenhagen, Department of Clinical Genetics at The Juliane Marie Centre, Denmark
- Copenhagen University Hospital (Rigshospitalet) and the University of Copenhagen, Center for Genomic Medicine, Denmark
- SciLifeLab, Clinical Genomics Unit SLL, Sweden
- Karolinska Universitetssjukhuset, Centre for Inherited Metabolic Diseases (CMMS), Sweden
- Karolinska Universitetssjukhuset, Clinical genetics, Sweden
- DNV GL

Observing institutions:

- Institute of Clinical Molecular Biology, Christian-Albrechts-University Kiel / University Hospital Schleswig Holstein, Germany
- Institute for Molecular Medicine Finland, FIMM
- Oslo University Hospital, Legal department
- Helsinki University Hospital, Laboratory of Genetics, Finland

Guest speaker:

- Marjut Salokannel, LL.D., Adjunct Professor

1.3 Goals of collaboration

Goal of collaboration were defined in the second workshop and confirmed in this meeting (Figure 1).

Figure 1 Goals of collaboration

1.4 Community Sharepoint

Presentations and other material referred to in this summary report are available at the community SharePoint: https://meet.dnvgl.com/sites/nordic-collaboration/SitePages/Home.aspx.

2 UNDERSTANDING LEGAL BARRIERS AND ENGAGING WITH KEY STAKEHOLDERS

Recognizing that ongoing initiatives and existing regulatory framework has a bearing on opportunities for sharing of clinical genomic data, the workshop included orientations on national initiatives, regulatory framework and relevant projects.

Table 1 Understanding legal barriers and engaging with key stakeholders – agenda content

Agenda item	Main content	Responsible	Reference file
What's up – per country - Denmark - Norway - Sweden	Status and ongoing processes	Rigshospitalet - Morten OUS - Dag SciLifeLab – Valtteri	• 20170418 WS III Denmark • 20170418 WS III Norway • 20170418 WS III Sweden
Relevant projects	Nordic Alliance for Sequencing and Personalized Medicine	OUS - Dag	• 20170418 WS III Norway
Relevant projects	Status of the BigMed project	DNV GL – Vibeke	• 20170418 WS III BigMed
Legal	Research across borders: legal perspective on sharing of health-related data for scientific research purposes in the Nordic countries	Marjut Salokannel	• 20170418 WS III Research across borders

Table 2 Actions

Theme	Action	Responsible	Comment
Legal	Share links on WP29	Guro	Working Party 29: The website seems to be under construction. Older opinions can be found in the former website. - http://ec.europa.eu/newsroom/just/item-detail.cfm?item_id=50083 - http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/index_en.htm BBMRI biobank guidelines: nothing has been published yet but they will be sent for comments some time at the end of summer
Understanding legal barriers and engaging with key stakeholders	Progress on national initiatives	• Rigshospitalet - Morten • OUS - Dag • SciLifeLab – Valtteri	Update at next workshop

2.1 National status and ongoing processes

2.1.1 Denmark

The Danish National Strategy for Personalized Medicine[1] 2017-2020 was published December 2016, and 5 million DKK was allocated to two politically appointed working groups who reported from their work in 2016. A funding of 100 million DKK for the strategy period has been indicated, but it is not clear if this is from current regional budgets or fresh money. There are great expectations to industry involvement and funding.

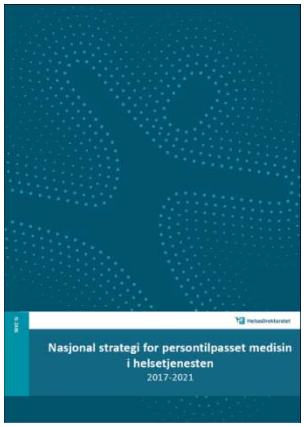
Currently initiated processes include establishment of joint governance and establishment of a National Genome Centre which will be led from the Copenhagen area. Core tasks will include:

- Building a cohesive technological infrastructure, eg. national genome database
- Establishing secure and flexible access to data for researchers and clinicians
- Establishing database - and cooperation on - knowledge to clinical practice
- Informing patients, citizens, health care professionals etc.
- Support the work of the Board for Personalized Medicine and advisory committees

Focus will be on disease areas and risk groups characterised by

- Those posing special challenges to the Danish society, e.g. affecting a lot of patients or relatives.
- Those associated with a significant genetic component, also having a considerable research potential.
- Those in which progress and new results are anticipated in the short term, e.g. in the form of better or new treatment forms.

2.1.2 Norway



The Norwegian National Strategy for Personalized Medicine in Healthcare was published June 2016. The main implementation project is led by the Directorate of Health, and two initiatives are funded in the 2017 national budget:

- Establishment of anonymous national database of genetic variants (5 million NOKs)
- Establishment of national network of regional competence centres (3 million NOKs)

A steering group including representatives from all regional health authorities and working groups for the two initiatives are being established.

[1] http://www.sum.dk/~/media/Filer%20-%20Publikationer_i_pdf/2017/Personalised-Medicine-Summary/SUM_klar_diagnose_summary_UK_web.ashx

2.1.3 Sweden

There is no national strategy for personalized medicine in place in Sweden. Instead a bottom-up initiative including groups from all health regions have proposed the establishing of Genomic Medicine Sweden to focus on rare diseases and cancer. The goal is to establish infrastructure based on current national resources to provide equal and cost-efficient care across the country, and be a unique resource for academia and industry.

2.2 Relevant projects

2.2.1 Nordic Alliance for Sequencing and Personalized Medicine

OUS, SciLifeLab and partners have received start-up funding from Nordforsk for 2017 (450.000 NOKs) to

- facilitate the implementation of personalized medicine in a sustainable way in the Nordic countries and to bring the Nordic countries to the international forefront in this field promising to radically improve healthcare.
- establish a network of NGS expertise in the Nordics to support specific areas of common interest, including basic research and clinical implementation of precision medicine programs

The project could serve as an umbrella for the further collaboration between the workshop partners, see section 8.

2.2.2 BigMed

The Big data medical solutions (BigMed) is an Innovation project financed by the Norwegian Research Council and partners with a budget of 60 + 74 mNOK and 3-4 years duration. The project kicked off Q1 2017 and is seen as highly relevant for the implementation of the National strategy for personalised medicine in Norway with its ambition to lay the foundation for an ICT platform that addresses the analytic bottlenecks for the implementation of precision medicine and pave the way for novel big data analytics. The solutions will provide the patients with an optimized care which takes their unique individual characteristics into proper consideration.

SciLifeLab, Karolinska, OUS and DNV GL are all partners in the BigMed project, and the work on genetic pipelines and accumulation, governance and sharing of genetic variant data is partly overlapping with activities taking place in this Nordic workshop series.

2.3 Legal barriers and opportunities for sharing

An introduction to 'research across borders – legal perspective on sharing of health-related data for scientific research purposes in the Nordic countries' was provided by Marjut Salokannel, invited speaker. The discussion was then based on this introduction on the EU General Data Protection Regulation (GDPR), which is in force and will be applied as of 25th May 2018 in all EEA states.

Table 3 Topics discussed

Topic	Discussion
About the GDPR - Implementation timeline - National interpretations	- Timeline for national interpretations - Potential Nordic alignment
Definitions - Health data - Personal data - Pseudonymous data - Anonymous data	
Legal basis for processing of personal data, such as - Consent - Public interest - Compliance - Protection of vital interests	
Requirements for processing of personal data - Consent - Public information	
Processing of health data for healthcare or administrative purposes	
Processing of health data for public health purposes - Public health interest (Regulation No 1338/2008 (EC) - Safeguard measures	
Processing of health data for scientific research purposes	- Research a wide concept that must be clarified. - Opportunity for harmonisation on nation interpretation of "scientific research"
Further legislative leeway for Member States	- Discussion on opportunities for pooling data across borders - A separate directive on cross border health data focuses on travelling patients and member state responsibilities.
GDPR and consent - Sensitive data require explicit consent for specific purposes - Withdraw and opt out options	- Member States are provided with a possibility to make an exception for the opt-out under Art 89.2 in specific cases. - Up to national laws to deal with how data is managed after a person is diseased. - Mutually recognised consent an opportunity for Nordic collaboration - WP29 draft guidelines on consent,also valid for biobanks (BBMRI) to be released October, implemented May 2018 - Certified of consent at European level
Safeguards for processing data for research purposes Organisational and technological safeguards Anonymised or pseudonumised data	- Data protection impact assessments - Data protection agencies will be obligated to evaluate impact assessment prior to implementation – this is new.
Further processing of health data for research purposes	- Discussion on rare variants and opportunities for reidentification of individuals - Interpretation of "public interest" national responsibility, this must be clarified as it may pave the way for sharing. - In Finland, a landmark case determined that the primary goal (research or healthcare) defines how data is to be managed. But note that this is a fine and often blurry line!
Cross-border processing of health data for scientific research	- Discussion on building research infrastructures based on consent - Interpretation of "broad consent"; future research is not permitted as a term in the consent. - Dynamic consent
Legal basis for research infrastructures	

Topic	Discussion
Big data and the GDPR	
Statistical research under GDPR	
Safeguards for processing health data in the big data environment	Liability: 4% or 20 mill Euro – whichever higher
Possibilities for Nordic collaboration • with regard to the safeguards for processing personal data in a cross-border setting • secure technological framework for cross-border research use of sensitive data; e.g. the Tryggve –pilot project (part of Nordforsk funded Nordic E-Infrastructure Collaboration) • mutual regonisance or synchronisation of each others requirements in order to build Nordic research networks and infrastructures • => Nordic research area	
General discussion	Discussion on cooperation across borders to find "the second patient"; whether this is healthcare or research; quality/safety purposes. Safeguards, what data is shared, built-in restrictions.
General discussion	Discussion on progress in national implementation of the GDPR and preparedness - N, S, FI: proposal to be published May / June on leeway path - Guidelines will come, e.g. on consent (oct 2017 from WP 29), code of conduct for biobank consent (summer / fall 2017) - Nordforsk facilitating discussions - Biobank regulations
General discussion	Discussion on sharing of single variants - If anonymous data, the GDPR does not apply - Must processing of data to be aggregated be based on consent. - New technologies to ensure anonymity, such as block-chain? - Relevant to argue for "public interest" in sharing of aggregated data - BigMed an arena to test this case.
General discussion	Discussion on database for sharing of interpreted variants - Safeguards to be specified; structure must be in place before approaching the data protection agencies - A strategy could be to develop arguments for "public interest" rather than base on informed consent
General discussion	The Nordic collaboration could collectively propose to authorities wording for interpretations of definitions and Nordic harmonization.

3 ENHANCING QUALITY OF DATA AND PROCESSES

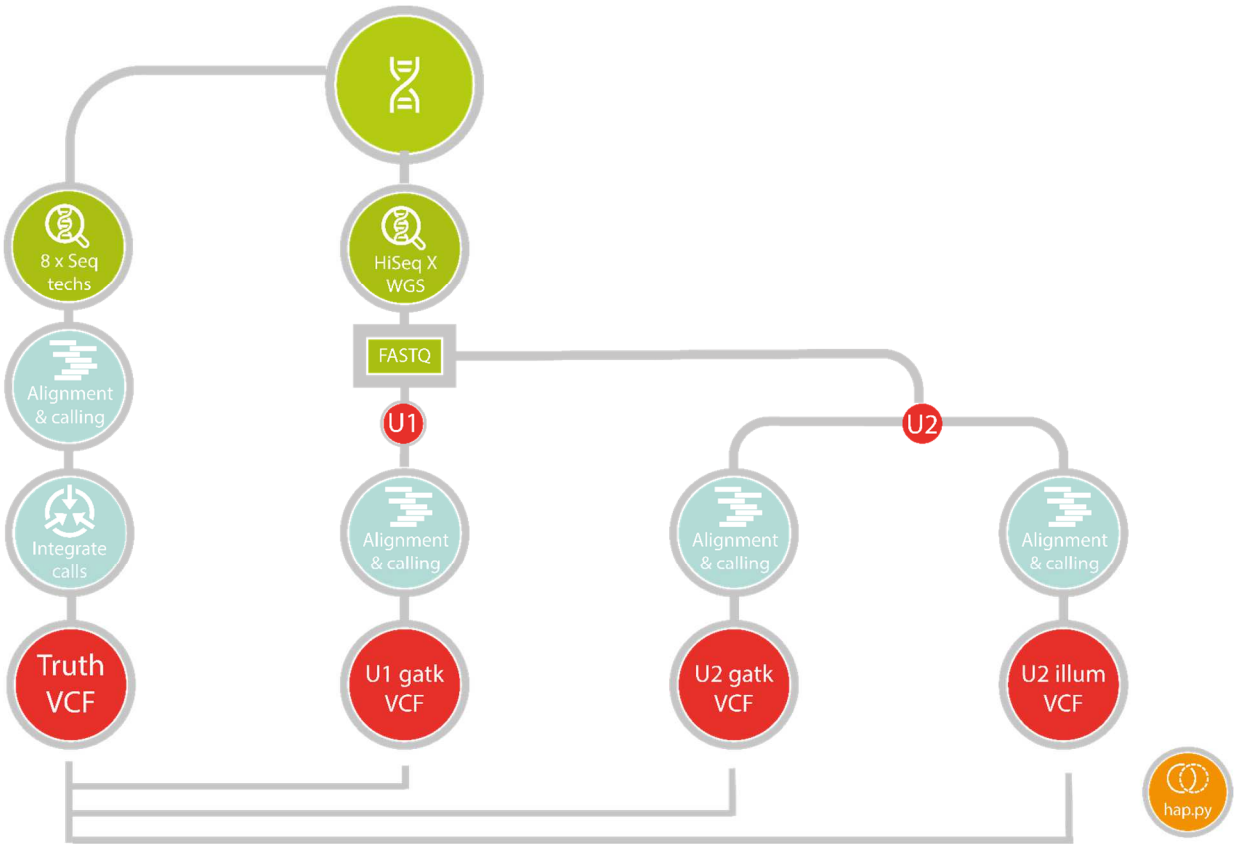
Table 4 Enhancing quality of data and processes – agenda content

Agenda item	Main content	Responsible	Reference file
Benchmarking 1 – variant calling	Benchmarking of sequence alignment, variant calling and variant annotation	DNV GL – Sharm	20170418 WS III Benchmarking 1 variant calling.pdf
Benchmarking 2 – variant interpretation	Benchmarking of variant interpretation per ACMG guidelines	OUS - Morten	20170418 WS III Benchmarking 2 variant interpretation

3.1 Benchmarking 1 – variant calling

Aims and methods

A benchmarking exercise on variant calling was run with two network partners with DNV GL performing the analysis on the output. The design of this exercise is illustrated in the figure below. The NIST reference material NA12878 (HG001) was whole genome sequenced on a HiSeqX instrument by one partner, and the output FASTQ files shared with a second partner. Both partners then subjected the FASTQ files to alignment and variant calling pipelines. The output VCFs were compared pairwise to the truth VCF made available by the GIAB consortium using the python tool hap.py[2].



[2] https://github.com/Illumina/hap.py

Table 5 Topics discussed

Topic	Discussion
True positives	- Only marginal differences between partners and pipelines
False negatives	- Marked difference between FN rate between the two partners both running GATK pipeline, especially for ALL (less so for PASS) variants. Requires some tweaking of parameters defining ALL vs PASS? - Are higher FN rates for site U1 indicative of too harsh filtering?
False positives	- Partner U2 has much fewer false positives than U1 - This is most probably due to U1 using three variant callers, to maximize calls. However, from these results it seems that many of the additional variants called are false positives. U1 verifies findings with Sanger sequencing, to minimize risk of reporting a FP.
GATK3 pipeline parameters	- Very similar pipelines, so surprising that such small differences can give such different results - Devil is in the details! U1 to drop stand_X_conf parameter - U2 GATK3 pipeline without BQSR on a WGS sample: better sensitivity compared to with BQSR, but with more false positives
Performance	- U1 run times significantly shorter than U2 – why? - Dragen is a potential game changer.
General discussion	Coverage is probably defining factor in performance. Consider simulating a lower coverage dataset using a subset of this data and repeat benchmarking to test this hypothesis. - Is automated ranking of variants a good approach? U1: solution to managing large numbers of variants without removing any results U2: will need to consider such approaches as numbers of genes/size of panels increases

Table 6 Possible extensions of benchmarking 1

Action	Question answered
Slice up U1 VCF according to caller, compare per caller to U2	Which variants are missed? Caller performance
Simulate a lower coverage (20x) dataset from this one?	Is coverage the main determinant of performance?
Repeat with U2 generated FASTQ	Benchmark wetlab
U3 and others to also participate	Benchmark drylab analysis
Extend with other GIAB reference materials (HG002-005) on WGS	More data to see if results are similar for more individuals
Extend with GIAB trio	Test more accurate variant calling at all positions, and if ability to make variant calls in low-coverage regions is enhanced
Compare identity of variants called in specific gene panels (eg. 44 gene connective tissue panel from U2)	Variant-level comparisons
Structural variants reference dataset	Check when available from GIAB
Benchmark Dragen if more units have it	Benchmark new pipeline
Benchmarking of raw data generation	Benchmark of wetlab

Conclusions

The results of the benchmarking results in general reflected those obtained by the units individually, however the individual units did not previously have the ability to compare their results with other similar units. As such the facilitation of an inter-unit comparison created value for the partners, where newer partners and observer units both expressed an interest in participating with their own data.

Additionally, side-by-side comparison of pipelines and detailed inspection of the benchmarking output (true/false positives, true/false negatives) allowed the effect of particular parameter/pipeline choices to be teased out. Based on this evidence and these results, one unit decided to adapt their choice of parameters to improve performance of their production pipeline. This concrete example illustrates the benefits and utility of inter-unit performance comparisons for driving quality assurance and improvement of pipelines, which are even more critical in the clinical setting.

It was agreed that unit leaders would discuss and agree on how to extend the benchmarking exercise, with the aim of reporting results at Workshop IV.

3.2 Benchmarking 2 – variant interpretation

Aims and methods

The aim of this session was to identify any differences in variant interpretation strategies, particularly with regard to the application of the ACMG guidelines. 10 variants from each unit were gathered, for a final total of 39 variants. These were then interpreted independently by each unit, without knowledge of previous classifications. The results were summarised in an Excel sheet, highlighting differences between units. At the workshop, conflicts in classification which would have "affected medical management" (see Amendola et. al 2016, Am J Hum Genet 98:1067-76) were discussed (n= 12).

Differences: Preliminary explanations

A total of 12 differences that "affect clinical management" were identified:

3 due to missed literature references

3 due to not having access to family data

2 due to different interpretations of literature references

1 due to "non-matching patient phenotype"

1 due to special characteristic of gene (LOF with no effect)

1 was a typing error

1 with no immediately apparent explanation

Table 7 Topics discussed

Topic	Discussion
Information to accompany variant interpretations	- Need to share evidence to evaluate basis for interpretation - Contact information for further queries
Curation of variant interpretation	Need to include - rationale for change in interpretation - traceability of different interpretations
Tools for handling variant interpretation	Ella could be a possible choice for those who do not have a tool available

Table 8 Possible extensions of benchmarking 2

Action	Comment
Results: summarize reasons for conflicts in classification	
Results: follow-up of identified differences: can consensus be reached?	
Sharing of evidence for classification	Possible to simulate sharing of evidence in Nordic variant database (paper based)?
Include family data	
Results: share arguments	
Results: remaining 10 variants; identify causes	
Benchmarking of variant reporting	Including routines for reclassifications of variants
Variant ranking	E.g. comparison of Scout ranking and exomizer

Conclusions

Any differences in classification that might impact on medical management are clearly worrying although it was suggested that some differences might stem from a "superficial" evaluation not representative of real clinical settings. Trivial explanations include literature references not found and a typographical error. In addition, 3 cases were classified differently due to not having access to important family data.

Other differences were due to differences in experience with and knowledge about how a particular gene works and different interpretations of the relevance of functional evidence, which is also likely due to specific experience with particular genes. Finally, the selection of variants by the different labs were not random and likely to affect particularly challenging variants.

The partners felt that the variant interpretation benchmarking was a valuable exercise and should be continued. Next time a more in-depth analysis of the causes of different classifications should be performed.

4 SHARING OF DATA, TOOLS AND METHODS

Table 9 Sharing of data, tools and methods – agenda content

Agenda item	Main content	Responsible	Reference file
Tools	Scout tutorial - setup and testing of Scout in participant laptops.	SciLifeLab – Henrik, Måns, Robin	NA
Tools	Ella & Scout - Status of testing - Status of communication between Ella and Scout	SciLifeLab – Henrik, Måns OUS – Svein Tore	NA
Methods	Scout - Introduction to variant ranking process	SciLifeLab - Henrik & Måns	NA
Methods	Structural variants – how do we deal with them?	SciLifeLab – Henrik	20170419 WS III Intro to variant ranking process incl structural variants
Data	Comparison of population specific variants - comparison of non-ExAC variants between OUS and SciLifeLab	OUS – Eidi	20170316 Saturation of variant database with increasing number

4.1 Scout tutorial

This session was to include setup and testing of Scout in participant laptops, but due to technical issues installations were not successful. It was agreed that SciLifeLab would set up and distribute a Scout demo version.

Table 10 Actions

Theme	Action	Responsible	Comment
Sharing of data, tools and methods	Set up and send out Scout demo	SciLifeLab/ Robin & Henrik	This has now been done and can be accessed at http://trailblazer.clinicalgenomics.se:8085/.

4.2 Ella & Scout

The session included a brief introduction on how the ACMG classification is performed in Ella and discussion on how to facilitate sharing of classifications, data formats and tools between Scout and Ella.

Table 11 Actions

Theme	Action	Responsible
Sharing of data, tools and methods	Make Ella a standalone solution	OUS/ Svein Tore

4.3 Scout – introduction to variant ranking process

An introduction to the Scout tool and ranking scheme was provided by SciLifeLab. The ranking algorithm uses weighted sums to create a model for prioritizing variants per disease causing potential. This brings the analysis time for analysing millions of variants down to minutes for a clinical test.

Table 12 Topics discussed

Topic	Discussion
Versions	Changes in Scout versions are communicated through GitHub to users to ensure transparency and traceability.
Processing and reporting of variants	- Variants are handled by geneticist and clinician - List of candidate variants are presented to the medical doctor in Scout. Multiple variants can then be marked as causative. - Variants called are scored and ranked; rank score, with basis in the ACMG guidelines - It is up to each clinic how they use the tool; rank models (panels, tools versions, etc are stored on GitHub)
Reporting of variants	It would be interesting to discuss reporting of variants in this forum - Liabilities - Standardization and harmonisation - Comparison / benchmarking – how do different labs write the reports - Variation in reporting according to type of analysis - Variations in understanding / interpretation by clinicians

4.4 Structural variants

The inclusion of structural variant data has the potential to improve diagnostic yield by 5-10% according to some reports[3] [4] [5]. OUS and SciLife are in the process of exploring, developing and testing SV discovery and annotation pipelines for inclusion in production.

Table 13 Topics discussed

Topic	Discussion
SciLife status	- Pipeline svrank_modelv1.0 for discovering, annotating and ranking structural variants for pathogenicity - For WGS mainly, poor for WES - 4 callers using different tools give separate reports from each caller - Gives approx. 20 k SNVs , merge using SvDb into 1 VCF - Annotate then rank using Genmod_score - Adapt rank model using criterion & weights
SciLife future plans	- Aim to move into Scout - Validation and testing - Test using NIST SV reference dataset (to come online soon)
OUS status	- CNV calling is done by in-house developed software as part of the exome and target pipelines. exCopyDepth calls any exon whose median coverage differs significantly, cnvScan then annotates each call with various databases and counts occurrences in an in-house database. Filtered on gene panel.

Table 14 Actions

Theme	Action	Responsible	Deadline
Structural and mitochondrial variants	Update on work to handle structural and mitochondrial variants	SciLifeLab/ Henrik	Next workshop

[3] P. S. Samarakoon, H. S. Sorte, B. E. Kristiansen, T. Skodje, Y. Sheng, G. E. Tjønnfjord, B. Stadheim, A. Stray-Pedersen, O. K. Rødningen, and R. Lyle (2014) Identification of copy number variants from exome sequence data., BMC Genomics, vol. 15, no. 1, p. 661, Jan. 2014. https://doi.org/10.1186/1471-2164-15-661

[4] P. S. Samarakoon, H. S. Sorte, A. Stray-Pedersen, O. K. Rødningen, T. Rognes, and R. Lyle, (2106). cnvScan: a CNV screening and annotation tool to improve the clinical utility of computational CNV prediction from exome sequencing data., BMC Genomics, vol. 17, no. 1, p. 51. https://doi.org/10.1186/s12864-016-2374-2

[5] Personal communication: Daniel Nilsson Klinisk genetic Karolinska sjukhuset

4.5 Comparison of population-specific variants

Comparison of non-ExAC variants between OUS and SciLifeLab

Population specific databases are used to identify variants found at high enough frequencies that they can be assumed to be benign and thus filtered out during analysis. To find out whether exchanging internal frequency data between SciLifeLab and OUS (which includes specific Swedish and Norwegian variants) would improve the filtering of Norwegian patients the variants found in 879 exomes at OUS were filtered against specific gene panels and different frequency databases as listed in Table 15. Results are provided in Figure 2.

Table 15 List of gene panels and frequency databases used for filtering

Gene panel / frequency database	Number of genes / description of content
Genepanel Iktyose (ichtyosis)	40 genes
Genepanel EEogPU (epileptic encephalopathy and intellectual disability)	57 genes
Genpanel BevegForst (movement disorders like hereditary spastic paraplegia (HSP), hereditary ataxia (HA) and others)	240 genes
SweGen (https://swefreq.nbis.se/#/)	consist of whole-genome variant frequencies for 1000 Swedish individuals generated within the SweGen project and variants with allele frequency >1% was filtered
ExAC (http://exac.broadinstitute.org/)	spans 60,706 unrelated individuals sequenced as part of various disease-specific and population genetic studies and variants with allele frequency >1% was filtered.
ExAC (subpop)	the separate ExAC subpopulations and variants with allele frequency >1% was filtered, but only for the subpopulations with allele count >2000 (1000 individuals) for each specific variant.
InDB	the OUS internal variant frequency database (patient data) and consisted at the time of approximately 1000 individuals (2000 alleles) and variants with allele frequency >5% was filtered.

Figure 2 variants found in 879 exomes at OUS was filtered against specific gene panels and different frequency databases

Table 1: Database filtering with the Iktyose genepanel

Database	Not filtered	Filtered only by this database	Filtered by one other database	Filtered by two other databases	Filtered by all databases
SweGen	2036	0	5	475	79945
InDB	2036	0	6	433	79945
ExAC	2036	0	2730	436	79945
ExAC(subpop)	2036	1175	2741	672	79945

Table 2: Database filtering with the EEogPU genepanel

Database	Not filtered	Filtered only by this database	Filtered by one other database	Filtered by two other databases	Filtered by all databases
SweGen	3023	0	0	462	64231
InDB	3023	0	0	229	64231
ExAC	3023	0	3476	691	64231
ExAC(subpop)	3023	1597	3476	691	64231

Table 3: Database filtering with the BevegForst genepanel

Database	Not filtered	Filtered only by this database	Filtered by one other database	Filtered by two other databases	Filtered by all databases
SweGen	10623	30	292	1776	302672
InDB	10623	153	291	3186	302672
ExAC	10623	0	17677	4216	302672
ExAC(subpop)	10623	5972	17678	4589	302672

Conclusions

The Norwegian national strategy for Precision Medicine recommends developing a national anonymous variant frequency database partly because of an assumption that the international databases available such as EXAC lack normal variants found that are unique for the Norwegian population. However, the analysis discussed at the workshop showed that for the BevegForst genpanel (the largest panel in this comparison), SweGen and InDB (OUS) only contributes with filtering 30 and 153 variants respectively. With almost 11 000 (out of totally over 300 000) variants still not filtered, the contribution of SweGen and InDB (OUS) is quite small.

Considering that ExAC also contains a Swedish population within the European (Non-Finnish) subpopulation and that gnomAD (http://gnomad.broadinstitute.org/ that includes all exomes from ExAC)

with variant frequencies from 123,136 exomes and 15,496 genomes is available now, we concluded in the meeting that exchanging internal frequency databases should not be a high clinical priority. The work to benchmark variant interpretation and share data with classified variants is more urgent and should be prioritized.

5 ESTABLISHING VEHICLES FOR SHARING

Table 16 Establishing vehicles for sharing – agenda content

Agenda item	Main content	Responsible	Reference file
ClinVar	Experiences in submitting variants to ClinVar	OUS - Tony / Discussion	20170419 WS III ClinVar experiences OUS
Sharing of variants	Nordic variant frequency database - 1-pager & personas / user cases - Input data - testing of SweFreq aggregation tool Status and demonstration of prototype	DNV GL – Sharm DNV GL - Brede	20170419 WS III Nordic variant database.pdf
Sharing of variants	Identifying steps towards sharing of patient cases via MatchMaker Exchange	OUS – Tony / Discussion	20170419 WS III Steps toward MME

5.1 Experiences in submitting variants to ClinVar

OUS summarized recent experiences with submitting variants to ClinVar, including user registration of the department in the database, posting templates and process, evaluation process and overall experiences.

Table 17 Topics discussed

Topic	Discussion
Overall experience	Experience shows that even though it took some time to register, submitting variants was fairly easy to do. Several formats could be used to submit Evidence of classification was kept at a minimum level (hardly anything)
Integration in work process	Submitting variants to external databases must be an integrated part of the lab process to ensure that it is done without adding to the workload.

Table 18 Actions

Theme	Action	Responsible	Deadline
Legal	Clarification of legal basis for submitting to ClinVar and clarification of information that can be attached.	BigMed WP5	

5.2 Nordic variant exchange

Although initially in workshops I and II there has been some interest in the sharing of variant frequencies between laboratories, the developments of SweFreq and international databases such as ExAC and gnomAD frequency databases have become available and/or increased in size. As a result, the need for a Nordic frequency database is diminishing (see Section 4.5, Comparison of population-specific variants). At the same time, awareness is growing around the need to identify and manage discrepancies in variant interpretation classification (see Section 4.2 Benchmarking – variant interpretation). In light of this, consensus was that a variant classification database including evidence support would be of more value to the clinical labs.

Ideally a combined variant frequency classification database should be developed but only if including the variant frequencies did not require significant additional resources. A brainstorming session was held with the purpose of identifying desired content and functionalities of such a database. The output of these two separate discussions are collected in Table 19 and Table 21.

Table 19 Variant classification database – content parameters identified

Parameter group	Content parameter	Must have (MH) vs Nice to have (NtH)	Note
Variant	Ref allele, alt allele	MH	
Variant	Chromosome position	MH	
Variant	Reference genome	MH	
Variant	Free text	MH	
Variant	Genotype	?	Homozygote vs heterozygote
Variant	HGVS nomenclature	NtH	http://varnomen.hgvs.org/
Variant	Variant frequencies	NtH	
Variant	Somatic or germline?	NtH	
Variant	Contact info of submitting lab	NtH	
Variant	Quality control of data entered	NtH	
Classification	Detailed annotations	MH	
Classification	Variant classification with details, including low risk variants	MH	
Classification	ACMG codes	MH	
Classification	Evidence, comments	MH	
Classification	Version of guidelines applied	MH	
Classification	Lab ID	MH	
Classification	Date of classification	MH	
Classification	Segregation / family data	NtH	
Classification	Phenotype	NtH	Can this be solved by looking at Matchmaker Exchange?
Classification	HPO / OMIM	NtH	Condition as ClinVar
Classification	Voting system for conflicting classifications		Alternative approach for handling conflicting classifications

Table 20 Variant classification database – functional requirements identified

Step	Functional requirement
Variant input & quality assessment	Guidelines for data input/output
Cross dataset integration and annotation	Classification with evidence
Cross dataset integration and annotation	effect of variant on transcript – is protein reading frame /function affected
Infrastructure & security	Location(Where is the data storage center located)
Infrastructure & security	Governance structure
Infrastructure & security	Logging of access functionality
Infrastructure & security	Use & access guidelines / policies
Infrastructure & security	Security – access, storing, transfer
Infrastructure & security	Open source?
Quality/conflict reporting	Flagging of conflicting classifications; push back to submitters for continuous curation
Quality/conflict reporting	Collaboration tool
Querying	Nice user interface
Querying	API
Querying	Specific variants/specific genes
Querying	Find nearby variants
Querying	Variant extraction per genomic region
Querying	Region-wise graphical representation of variant classification
Querying	Classification
Querying	Phenotype (if in database)
Download for local use	Global database and individual database subsets
Download for local use	To be used for annotation in the pipeline
Further dissemination	Possibility to push data further on to international databases (e.g. ClinVar)
Further dissemination	Compatible with Matchmaker Exchange

It was agreed that this work should continue and the following actions related to this were identified:

Table 21 Actions

Theme	Action	Responsible
Understanding legal barriers and engaging with main stakeholders	Clarification of legal basis for variant database (Nordic variant exchange)	BigMed WP5
Establishing vehicles for sharing – variant database	Agree on use case for variant database	DNV GL/ Sharm
Establishing vehicles for sharing – variant database	Check if Ella can be used as basis for database requirements.	OUS/ Svein-Tore
Establishing vehicles for sharing – variant database	Include harmonisation and quality improvement as benefits of frequency database	DNV GL/ Sharm
Establishing vehicles for sharing – variant database	Safeguards for database to be described	DNV GL/ Sharm
Establishing vehicles for sharing – variant database	Variant database case to be presented to health ministries and data protection agencies	DNV GL/ Sharm

5.3 Identifying steps towards sharing of patient cases via Matchmaker Exchange

A man with a dream (Tony!) introduced Matchmaker Exchange (MME), a federated platform facilitating the identification of cases with similar phenotypic and genotypic profiles (matchmaking) through a standardized application programming interface (API) and procedural conventions. The platform is supported by IRDiRC, ClinGen and GA4GH. MME enables queries containing phenotype (HPO) and gene/variant info to be sent to other MME services that then evaluate and return info about any matching similar cases. The MME requirements are stated in the MME Service requirements. In the BigMed project it is a goal to establish MME at OUS within 2019, and the discussion addressed development of tools and clarifications on data for use in matchmaking.

Table 22 Topics discussed

Topic	Discussion	OUS status & plans	SciLifeLab comments

Tools	- **Database** of variants and phenotypes
	- **Infrastructure**
	  o API for sharing of genotype and phenotype data between trusted partners.
	  o GA4GH Security Technology Infrastructure: Standards and implementation practices for protecting the privacy and security of shared genomic and clinical data.
	- **MME Service** implementation
		- **Database**: Can use Norvariom version 2 technical backend with consent, usage monitoring etc.
		- **Infrastructure**: TSD, but need secure Internet access
		- **MME Service**: Implement ”standard” in BigMed
			- Scout could be connected, but need for secure infrastructure .
			- The Tryggve project could link.

Data	- **Consent**
	  o Querying on broad phenotype descriptions or using standardized terms / codes such as HPO do not require consent.
	  o Querying on more detailed phenotype descriptions and specific variants / genomic datasets may require consent from patients
	- **Legal clarification**; approval from data protection authorities needed?
	- **Phenotypes** preferably based on HPO
		- **Consents**: Norvariom data? Additional consent necessary? We need an OUS solution for dynamic/electronic consent.
		- **Legal go-ahead**: Support from BigMed legal team?
		- **HPO phenotypes**: DIPS (Norwegian EHR vendor) will help with implementation of HPO-based requisition
			- Data consent and legal go-ahead would be a clinical decision.
			- Work ongoing to include option to search for phenotype within own database.

Table 23 Actions

Theme	Action	Responsible
Establishing vehicles for sharing	Update on progress with MME from participating labs	OUS Tony
Understanding legal barriers and engaging with key stakeholders	Legal clarifications on MME with data protection authorities	BigMed WP5

6 COLLABORATION

During the second day of the workshop, representatives from the different workshop participant partners gathered to discuss the further organisation of the collaboration / workshop series, which was then discussed further in a plenary session.

Table 24 Collaboration – agenda content

Agenda item	Main content	Responsible	Reference file
Formalisation of collaboration	Establish principles for collaboration	DNV GL – Stephen / discussion	
Naming of collaboration	Feedback from formalisation forum Naming the baby	DNV GL – Vibeke / discussion	
Planning	Rounding up and next steps	DNV GL – Vibeke / discussion	

Table 25 Discussions and actions

Theme	Discussion	Action	Responsible
Confidentiality	Minutes and slides: labs should not be identifiable. Summary reports to be reviewed by heads of sections before distribution	- Slides: to be reviewed before sharing - Summary report to be reviewed by heads of sections of contributing labs before distribution	- Slides: per author - Summary report: review to be organized by Guro
Formalisation	Formulate constitution	Draft constitution for collaboration	DNV GL/ Stephen
Formalisation	Constitution agreement and adoption	Agreement on constitution between heads of sections	DNV GL/ Stephen
Name of collaboration	Alternative names were suggested and voted over.	Agree with partners and funding organisation on name for collaboration	OUS/ Dag

6.1 Formalisation of collaboration

The cooperation to date has been ad-hoc and based on partners self-financing their own activities and covering the cost of the hosting of the workshops. The group has grown for each of the workshops, interest from other parties is growing quickly and we do not wish to exclude relevant individuals and institutions from contributing. These factors as well as the fact that NordForsk funding would allow a more ambitious agenda for workshops led the participants to agree that the cooperation should be formalized.

An organization based on a similar model to GA4GH was proposed and this should incorporate the goals that have been developed in WS II. The forum should be open to promote transparency and sharing, and although the strength of the forum so far has been the intimate setting of a smaller forum, mechanisms for inclusion of new members must be developed. The discussion provided input on parameters to be covered in a constitution for the collaboration, including

- Principles for collaboration
- Funding source(s)
- Partners & responsibilities;
- Mechanisms for including new partners
- Secretariat
- Mission and scope:
- Working groups on specific topics
- Liabilities
- IPR

6.2 Naming of collaboration

A group brainstorming resulted in the below ranked list of suggestions for alternative names to the name under which funding was received from Norforsk. The name used in the Norforsk application was the Nordic Alliance for Sequencing and Precision Medicine, while the workshop identified these additional names to be considered by Dag:

1. Northern clinical genomics alliance
2. Nordic alliance for clinical genomics
3. Nordic alliance for genomics and personalised medicine
4. Nordic alliance for genomics
5. Scandinavian Genomics

7 FOLLOW-UP ACTIONS AND NEXT STEPS

7.1 Summary of actions identified during the workshop

Theme	Action	Responsible
Collaboration	Agree with partners and funding organisation on name for collaboration	OUS/ Dag
Collaboration	Confidentiality • Slides: to be reviewed before sharing • Summary report to be reviewed by heads of sections of contributing labs before distribution	Slides: per author Summary report: review to be organized by Guro
Collaboration	Constitution agreement and adoption: Agreement on constitution between heads of sections	DNV GL/ Stephen
Collaboration	Draft constitution for collaboration	DNV GL/ Stephen
Sharing of data, tools and methods	Make Ella a standalone solution	OUS/ Svein Tore
Sharing of data, tools and methods	Set up and send out Scout demo	SciLifeLab/ Robin & Henrik
Understanding legal barriers and engaging with key stakeholders	Clarification of legal basis for submitting to ClinVar and clarification of information that can be attached.	BigMed WP5
Understanding legal barriers and engaging with key stakeholders	Legal clarifications on MME with data protection authorities	BigMed WP5
Understanding legal barriers and engaging with key stakeholders	Share links on WP29	DNV GL/ Guro

7.2 Next workshop

There was consensus to arrange the next workshop in Oslo late fall 2017.

Tentative dates: 16. & 17. November 2017

7.3 Topics for next workshop

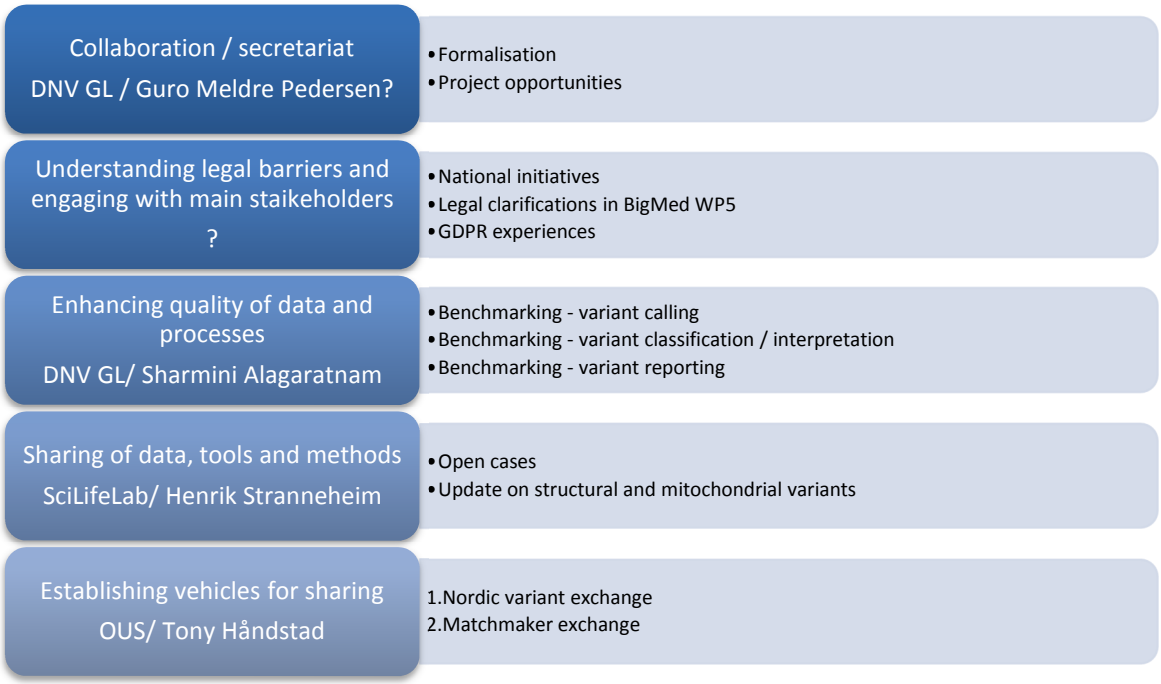
During the workshop, potential follow-up activities and new topics for the next workshop were identified as listed in Table 26.

Table 26 Potential follow-up activities and topics for next workshop

Theme	Action	Responsible	Comment / input to content
Collaboration	Opportunities for joint research projects		
Enhancing quality of data and processes	Benchmarking - reporting of variants	Rigshospitalet/ Maria or OUS/ Morten?	Reporting of variants - BRCA / exome / genome - Liabilities - Standardization and harmonisation - Comparison / benchmarking – how do different labs write the reports Interested: - OUS/ Yngve, - Rigshospitalet/ Ane, Morten, - Karolinska/ Nicole
Enhancing quality of data and processes	Benchmarking - variant calling	DNV GL/ Sharm	Table 6
Enhancing quality of data and processes	Benchmarking - variant interpretation	OUS/ Morten	Follow-up of identified differences, with aim of reaching consensus about the causes. New test dataset for next workshop: Include essential information, such as family data? Table 8
Establishing vehicles for sharing	Nordic variant exchange - Update on variant database progress	DNV GL / Sharm	Table 21
Establishing vehicles for sharing	Update on progress with MME from participating labs	OUS/ Tony	
Sharing of data, tools and methods	Sharing of open / non-solved cases (consent needed)	Rigshospitalet/ Maria	
Sharing of data, tools and methods	Update on work to handle structural and mitochondrial variants	SciLifeLab/ Henrik	
Understanding legal barriers and engaging with key stakeholders	Progress on national initiatives	Rigshospitalet/ Morten OUS/ Dag SciLifeLab/ Valtteri	
Understanding legal barriers and engaging with key stakeholders	Session on GDPR; data storage, data sharing – sharing of experiences with national legal bodies	DNV GL/ Guro	Sharing of questions to authorities Experiences with impact assessment (to be addressed in 2018 workshop?)

7.4 Organization of activities

Based on goal structure and suggested content for next workshop, the following task forces and coordinators are proposed. The list of subtasks could be subject to further delegation and expansion.



APPENDIX 1: WORKSHOP AGENDA

Table 27 Agenda day 1 – 18. April 2017

Time	Topic	Content	Responsible
12:00	Welcome	Workshop goals, agenda, practicalities, Introduction of participants	DNV GL - Vibeke
13:00	What's up – 10 min per country - Denmark - Norway - Sweden	Status and ongoing processes	Rigshospitalet - Morten OUS - Dag SciLifeLab – Valtteri
13:30	BigMed	Status of the BigMed project	DNV GL – Vibeke
14:00	Legal	Research across borders: legal perspective on sharing of health related data for scientific research purposes in the Nordic countries	Marjut Salokannel
14:00	Benchmarking 1 – variant calling	Benchmarking of sequence alignment, variant calling and variant annotation	DNV GL – Sharm
15:00	Benchmarking 2 – variant interpretation	Benchmarking of variant interpretation per ACMG guidelines	OUS - Morten

Table 28 Agenda day 2 – 19. April 2017

Time	Topic	Content	Responsible
8:30	Formalisation of collaboration	Establish principles for collaboration	DNV GL – Stephen / discussion
8:00	Tools	Scout tutorial Setup and testing of Scout in participant laptops.	SciLifeLab - Henrik & Måns
9:30	Tools	Ella & Scout - Status of testing Status of communication between Ella and Scout	SciLifeLab - Henrik & Måns
10:30	Methods	Scout - Introduction to variant ranking process	SciLifeLab - Henrik & Måns
11:00	Methods	Structural variants – how do we deal with them?	SciLifeLab – Henrik / discussion
11:30	Data	Comparison of population specific variants - comparison of non-ExAC variants between OUS and SciLifeLab	OUS – Eidi
13:00	ClinVar	Experiences in submitting variants to ClinVar	DNV GL – Vibeke / Discussion
13:30	Sharing of variants	Nordic variant frequency database - 1-pager & personas / user cases - Input data - testing of SweFreq aggregation tool Status and demonstration of prototype?	DNV GL – Sharm
14:30	Sharing of variants	Identifying steps towards sharing of patient cases via MatchmakerExchange	OUS – Tony / Discussion
15:10	Collaboration	Feedback from formalisation forum Naming the baby	DNV GL – Vibeke / discussion
16:00	Collaboration	Rounding up and next steps	DNV GL – Vibeke / discussion

APPENDIX 2: LIST OF PARTICIPANTS

Table 29 List of participants

Country	Organisation	Department	First Name	Last name	Role
Denmark	Rigshospitalet	Center for Genomisk Medicine	Birgitte	Bertelsen	Participant
			Frederik	Otzen Bagger	Participant
			Caroline Maria	Rossing	Participant
			Ane	Yde Schmidt	Participant
		Dep. of Clinical Genetics at The Juliane Marie Centre	Jack	Cowland	Participant
			Morten	Dunø	Participant
			Peter	Johansen	Participant
			Lotte	Risom	Participant
			Karin	Wadt	Participant
Finland	Institute for Molecular Medicine Finland (FIMM)		Kaisa	Kettunen	Observer
Finland	Helsinki University Hospital	Laboratory of Genetics	Anna-Kaisa	Anttonen	Observer
Finland	Independent consultant		Marjut	Salokannel	Invited speaker
Germany	Christian-Albrechts-Uni. of Kiel and Uni. Hospital Schleswig Holstein	Institute of Clinical Molecular Biology	Michael	Forster	Observer
			Georg	Hemmrich-Stanisak	Observer
Norway	DNV GL	Analytic Innovation Center	Brede	Børhaug	Participant
		Business Assurance	Stephen	McAdam	Participant
		Global Technology and Research	Sharmini	Alagaratnam	Participant
			Guro Meldre	Pedersen	Participant
			Vibeke Binz	Vallevik	Participant
Norway	Oslo University Hospital (OUS)	Department of Medical Genetics	Morten	Eike	Participant
			Tony	Håndstad	Participant
			Eidi	Nafstad	Participant
			Yngve	Sejerstad	Participant
			Svein Tore	Seljebotn	Participant
			Dag	Undlien	Participant
Norway	Oslo University Hospital (OUS)	Legal department	Randi	Borgen	Observer
			Ingunn	Myklebust	Observer
Sweden	Karolinska University Hospital	Clinical genetics	Anna	Hammersjö	Participant
		CMMS	Nicole	Lesko	Participant
Sweden	SciLifeLab	Clinical Genomics	Robin	Andeer	Participant
			Måns	Magnusson	Participant
			Henrik	Stranneheim	Participant
			Valtteri	Wirta	Participant