Inaugural meeting of the Nordic Alliance for Sequencing and Personalized Medicine

Venue: DNV GL, Høvik, Norway
Time: November 15th – 17th 2017

NASPM NOVEMBER 2018 MEETING

Workshop summary report

Revision No.: 0
Date: 2018-01-22

Nordic Alliance for Sequencing and Personalized Medicine

Project name: NASPM November 2018 meeting
Report title: Workshop summary report

Date of issue: 2018-01-22

Revision No.: Rev. 0

Objective:

This report summarises the inaugural meeting of the Nordic Alliance for Sequencing and Personalised Medicine taking place at Høvik, Norway, 15.-17. November 2017. Following a plenary session focusing on common challenges, two separate workshops were held focusing on clinical and research-specific needs. This emphasis of this summary report is on discussions and agreed actions from the 4th NASPM clinical genomics data sharing workshop.

Rev. No.	Date	Reason for Issue	Prepared by
0	2018-01-22	First issue	Guro Meldre Pedersen (guro.meldre.pedersen@dnvgl.com)

Table of contents

1 EXECUTIVE SUMMARY

This report summarises the inaugural meeting of the Nordic Alliance for Sharing & Precision Medicine (NASPM) held at DNV GL headquarters in Høvik, Norway 15.-17. November 2017. The formalisation of the NASPM follows three informal workshops focusing on sharing of clinical genomics data at a Nordic level.

The meeting was organised in two parts; a NASPM plenary session and two parallel workshops focusing on research and clinical needs.

<table>
<tr><th></th><th>15. Nov</th><th>16. Nov</th><th></th><th>17. Nov</th><th></th></tr>
<tr><td rowspan="2">Morning</td><td rowspan="2"></td><td colspan="2">NASPM plenary meeting</td><td rowspan="2">NASPM 4th clinical genomics data sharing workshop</td><td rowspan="2">NASPM 1st Nordic workshop for National genomic research infrastructures</td></tr>
<tr><td colspan="2">Inaugural general assembly for NASPM</td></tr>
<tr><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Afternoon</td><td>NASPM plenary meeting</td><td>NASPM 4th clinical genomics data sharing workshop</td><td>NASPM 1st Nordic workshop for National genomic research infrastructures</td><td>NASPM 4th clinical genomics data sharing workshop</td><td></td></tr>
</table>

1.1 About the NASPM

The Nordic Alliance for Sequencing and Personalized Medicine is an independent, non-governmental, not-for-profit, Nordic association that has received initial funding from Nordforsk.

The Nordic alliance brings together key national research infrastructures in genomics as well as key clinical environments for implementing genomic medicine in the Nordic countries. The alliance is a result of a realization that the Nordic countries with its high trust in government, transparent societies and similar health care systems based upon the ideal of equal access to care for all members of society, have the potential to become leading countries in the sustainable implementation of personalized medicine. However, this requires Nordic collaboration as the Nordic countries individually are too small.

Following three informal workshops the inaugural meeting of the alliance was planned based on a wish to bring together key stakeholders in the Nordic countries to identify specific action points for the future of the alliance, bringing in international expertise with the aim of identifying best practices and state of the art for the field.

1.2 Mission

The overall mission of the NASPM is to share trustworthy genomics data and technology competence for improved diagnosis and treatment, and as a resource for research.

1.3 Goals of the NASPM

The Nordic Alliance for Sequencing and Personalised Medicine has defined the following goals:

- 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 personalised 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

1.4 NASPM contacts

Table 1 Transitional Steering Committee members

SC Chair	Dag Undlien	Oslo University Hospital	Norway
SC Vice Chair	Valtteri Wirta	SciLifeLab	Sweden
SC Vice Chair	Karin Wadt / Morten Dunø	Department of Clinical Genetics, Rigshospitalet	Denmark
SC Member	Joakim Lundeberg	SciLifeLab	Sweden
SC Member	Jón Jóhannes Jónsson	Dept. of Genetics and Molecular Medicine, Landspitali - National University Hospital / Dept. of Biochemistry and Molecular Biology, Faculty of Medicine, University of Iceland	Iceland
SC Member	Maria Rossing	Center for Genomic Medicine, Rigshospitalet	Denmark
SC Member	Stephen McAdam	DNV GL	Norway
Secretariat	Guro Meldre Pedersen Guro.meldre.pedersen@dnvgl.com	DNV GL	Norway

2 NASPM PLENARY MEETING

The plenary session began 15. Nov 2017 at 12:40 and concluded 16. Nov 2017 at 13:00. The objective of the NASPM plenary session was to gather Nordic and international expert stakeholders within clinical genomics to share knowledge regarding developments within their intuition and country.

The NASPM plenary session consisted of three sessions:

1. State of the art of personalized medicine, chaired by Dag Undlien / OUS;
2. Personalized Medicine – Nordic initiatives, chaired by Valtteri Wirta / SciLifeLab; and
3. Personalized Medicine – data sharing and big data; chaired by Karin Wadt and Morten Dunø / Rigshospitalet.

The plenary session included invited speakers from the following organizations, the full agenda can be found in appendix 1:

- Genomics England 100,000 Genomes Project, Clare Turnbull
- Abbvie, Steinar Thoresen
- The Danish Genome Centre, Kasper Lindegaard-Hjulmann
- Genomic Medicine Sweden initiative, Richard Rosenquist Brandell (Karolinska Institutet)
- Norwegian Ministry of Health and Care Services, Maiken Engelstad
- Ministry of Social Affairs and Health, Finland, Jaakko Yrjö-Koskinen
- Global Alliance for Genomics and Health, Thomas Keane (EMBL-EBI)
- UMC Groningen dept. of Genetics, Joeri van der Velde
- University of Copenhagen, Lars Juhl Jensen
- SITRA, the Finnish Innovation Fund, Pia Heikkurinen
- BigMed Lighthouse Project, Thomas Smedsrud (Oslo University Hospital)

Following the plenary session, an open meeting was held to inform about this Nordic alliance and initiate the process of formalising the further collaboration.

3 NASPM 4TH CLINICAL GENOMICS DATA SHARING WORKSHOP

The agenda and list of participants for this workshop are included in appendix 2 and 3 of this report. Presentations and other material from the workshop that are approved for sharing are available at the community SharePoint: https://meet.dnvgl.com/sites/nordic-collaboration/SitePages/Home.aspx.

Reflecting three of the focus areas agreed in previous workshops (illustration below), this workshop was organised in three sections:

- Section 3.1: Establishing vehicles for sharing, chaired by Tony Håndstad, OUS
- Section 3.2: Enhancing quality of data and processes, chaired by Sharmini Alagaratnam, DNV GL
- Section 3.3: Sharing of data, tools and methods, chaired by Henrik Stranneheim, SciLifeLab / Karolinska

3.1 Establishing vehicles for sharing

chaired by Tony Håndstad, OUS

The introductory presentation to the session discussed how current legal barriers seemingly create an imbalance between the healthcare benefits of sharing patient data on one side, versus the harm to patient privacy on the other side. Further, the presentation looked toward some potential technological solutions in the future and discussed principles for future systems and sharing.

3.1.1 The BigMed Trusted Variant eXchange (TVX) prototype

By Sharmini Alagaratnam, DNV GL

Sharmini presented the Trusted Variant eXchange (TVX), a prototype solution to enable the sharing of variant frequencies and classifications with evidence between trusted partners. The rationale behind this database is two-fold: firstly the volume of genomes sequenced is expected to increase dramatically in the next 15 years, which, secondly, will exacerbate issues with quality and decision making in variant interpretation. As a part of BigMed work package 3, and in collaboration with NASPM participants, requirements and functionalities for a secure database to collate variant classifications from a network of trusted partner labs were defined. An important function of the solution is to flag and report discordances in classification when present. The prototype database was presented at the workshop session, demonstrating its features, followed by an interactive feedback session to collect input on the top 3 features, least important 3 features, and what should be done differently. The result of this session

is summarized in the table below. These will be entered in a design iteration and newer versions made available to partners as relevant.

Top priority features
Feeding in and out of the database /pipe-in and pipe-out
Classification – include what resources were used to get here. Good to have this in one resource. All the bits I need to make the classification and how this was made. It will save people time. (Google Scholar search, link to ClinVar)
Easy to use, fast response time
Free & open
Evidence for classification
Possibility to contact other submitters
Disagreement in classification
Flagging difference and conflicts
What type of cohort the frequency comes from? Is it even allowed to respond to this info?
Mechanism to submit variants to ClinVar
Who submitted the variant
Reporting back conflicts to submitting labs
Common guidelines (ACMG)
Classification / ACMG
Frequency and allele count
Classification with evidence
Opportunity to add more evidence
Standard APIs for upload & download to feed in an out to pipelines
Management of reference genome
Phenotype is valuable, but challenging legally. Rather focus on HPO at this stage.

Less important features
Sleeker frontpage: simpler, cleaner
Front page should have altered focus, maybe the flag
Detailed phenotype info
We do not need favourite variant on front page or recent search
Linkouts/ annotation
Frequency and classification are two separate problems
What is allele frequency based on?
Keep it clean and simple
Good to skip repeating what is already in variant triage / public db tools
Annotation and link to external sources less important
No need for recent searches

Should be done differently
Discussion / comments module
Gather information from other sources (ExAc, gnomAD, ClinVar)
Reclassification when "rules" change
Approach to calculate significancies (case control): valuable for inherited cancer
Support multiple ref. genomes + visualise what reference genome was used
Functionalities to solve conflicts
Classification interface with key information easily available
API – upload / download
ClinVar compatible?
Clinical trial links
References / publications for arguments
Arguments for the variant classifications
Feedback if someone submits a conflicting class of the same variant
'Expert' labs for particular genes?
ACMG codes optional
Contact information for labs
Evidence details

3.1.2 NORVARIOM

by Øyvind Evju, OUS

Øyvind introduced the background, status and scope of NORVARIOM, a database of variants in the Norwegian population. Norvariome is a database of genomes/exomes for use in research and clinical settings. The data will primarily come from consenting patients undergoing sequencing at the Department for Medical Genetics at Oslo University Hospital. In the presentation, proposed solutions for preserving privacy were highlighted, with focus on pseudonymization and data availability. In addition, possible future directions for such databases were outlined, including how it could be used both in clinical and research settings.

Discussion points:

- Similar initiatives in Sweden and Denmark
    - SweFreq available
    - Danish ambition to set up a similar database, orchestrated by the Danish Genome Centre.
    - Finland holds similar resource. Users can apply for access to data (with limitations). Frequencies are released for public use.
- Input data
    - Combining input from different labs could be challenging in terms of pipelines used and annotation of submitted data.
- Discussion regarding additional data to be included, e.g. age of patients.
- Consent management
    - Electronic solutions for consent available, but not tailored to the needs of genomic databases. Consent management could be skewed with electronic solutions, as younger people have lower barrier. It is challenging to secure informed consent; time consuming and costly, and requires consultation. Opportunities for standardization of information, and for check of comprehension.
    - Suggestion to use films for information, Genomics England has some useful examples.

3.1.3 Lighting the SciLifeLab Beacon: Experiences so far and future plans

by Chiara Rasi, Scilifelab.

Chiara discussed the necessity of data sharing for rare diseases. A Beacon is a public data discovery web-service; variants are uploaded to a server available to others for queries. If you query and find a variant in the network of Beacons available, it is possible to follow up and contact the laboratory for more information. Beacon is powerful as it has a network all over the world, and the query of a variant can be performed by anyone in the network. Beacons is based on the willingness of laboratories to share data variants with the community. The Beacon APIs standardize what is shared and how. Today there are three types of access; Open, registered and controlled, the latter allowing more freedom in data to share.

Chiara introduced Beacons for genetic data sharing, general tips to avoid problems when you design your data sharing system, available software and experiences at Clinical Genomics, SciLiifeLab Stockholm, Sweden.

Known security issues include the possibility to reidentify individuals through multiple queries and use of probability models. This requires previous knowledge of the individual's DNA sequence. Mitigating actions include increasing the Beacon sample size, avoid naming datasets after phenotypes and sharing only small portions of the genome (genes, exomes, gene panels).

SciLifeLab has chosen to connect to the Elixir Beacon. Sweden already has a node, the SweFreq node in Uppsala. SciLifeLab is currently sharing 22000 variants in an open Beacon setup as a proof-of-principle demonstration.

3.1.4 Structured Phenotyping

by Yngve Sejersted, OUS

Structured phenotype is not collected, stored or shared at OUS today. Precise and detailed phenotyping will aid diagnostic genetic analysis. The Phenomizer and other tools could be applied in diagnostics today.

Discussion points

- Phenotyping and reporting; reporting of search terms used.
- Input from requisitioner key, close collaboration needed.
- Filtering on OMIM and HPO terms.
- Process depends on panel- or exome sequencing.

3.1.5 Automated phenotyping

by Ketil Heimdal, OUS

Ketil represented the clinical perspective, and presented the gap between doctors' work in free text to algorithms' need for structured data for diagnostic workup. Resources include EHR (Electronic health records), lab requisitions and local IT software, quality registries, as well as resources outside of OUS. Doctors are trained to use the London dysmorphology terms for syndrome searching.

Phenotypic data collection at OUS AMG is not done systematically, only collection of indications for testing. Projects have demonstrated that this is possible but there are challenges related to data quality, consistency and completeness.

Hurdles mentioned were

- Phenotyping must be of high granularity to be clinically useful
- Efficient phenotyping takes training and practice, which is time-consuming
- It takes time to change doctors' mode of operation, and motivation may need to be a combination of carrots and sticks.
- Fully developed phenotyping tools have not been integrated into clinical systems
- Privacy concerns slows the process considerably

Alternative ways forward could be standalone phenotyping systems or integrating HPO-terms in EHR. The BigMed project would like to implement an HPO browser and e-requisition in EHR and establish in-house phenotype DB as quality registry.

Questions from the audience were related to:

- Why HPO should be integrated in the EHR? This would reduce having a doctor perform it twice. We would like the doctor just enter data once with an automated requisition form, or any other technological solution.
- streamlining of the interface between the technologies and the human (doctors).
- Difficult vs easy cases: with difficult cases, it requires a multidisciplinary discussion (Denmark). Ketil responds that this information is retrieved from the lab and should be properly integrated and reward those with permissions to make use of secondary data.

3.2 Enhancing quality of data and processes

chaired by Sharmini Alagaratnam, DNV GL

This session focused on exercises and tests run, specifically on variant calling with an extension of the benchmarking with additional samples (3.2.2), and an evaluation of variant callers for a benchmarked dataset (3.2.3). In the second half of this session, a facilitated discussion on challenges and approaches to reporting difficult or problematic variants was held. Workshop participants were divided into five groups, each assigned a topic. The results of these discussions are summarized below (3.2.4). In general reporting as a theme for subsequent workshop activities was stressed.

3.2.1 Benchmarking of variant calling Trio HG002-004: All variants and PU panel evaluation of caller performance

by Courtney Nadeau, DNVGL

Courtney presented the results of an ongoing SNP and indel benchmarking initiative. In this experiment, reference DNA from a trio of well-described individuals was used to generate WGS libraries, which were sequenced at one site using their standard protocols. Data (.fastq files) were then distributed to three partner laboratories, which used their production SNP/small indel pipelines to quality control, align, de-duplicate, and call variants from each dataset individually. Data were then transferred to DNV GL for benchmarking.

The benchmarking protocol was based on the precisionFDA truth challenge. For each of the three samples, regions with consistent SNP and indel calls have been previously identified by the GIAB consortium using 12 different sequencing technologies. The variants called by each of the three pipelines were compared to the gold standard set using hap.py v0.3.7.

Summary of Results:

- All pipelines called a substantial number of SNPs and indels outside of high-quality regions, and called false positive SNPs and indels within the high confidence regions.
- SNP recall and precision varied by pipeline. Some pipelines matched other external benchmarks, while others scored lower. One of the low-precision pipelines seemed to be miscalling zygosity, while correctly identifying the variant. Recall and precision were lower for indels than SNPs, and a greater number of false positives were produced. All pipelines greatly overestimated the number of heterozygous indels.
- Concordance between pipelines in the variants detected was high. Not only were the same gold-standard calls shared, but false positives were also shared to a great degree, possibly due to similarities in the pipelines or shared artefacts in the primary data.

In general, pipelines did well at detecting true positives, though there is room for improvement particularly for small indels. Pipelines did however generate a large number of false positive calls, indicating that care must be taken when filtering, prioritizing, and interpreting these data.

3.2.2 Benchmarking variant calling pipelines. false negatives

by Ying Sheng, OUS

The aim of this study was to investigate the reasons for false negative variants, only focusing on SNPs. Ying introduced how the whole genome sequenced FASTQ files from NIST reference material NA12878 (HG001) were examined through two different variant calling pipelines (U1 pipeline: using multiple variant callers including GATK3; U2 pipeline: only GATK3 variant caller). The output VCFs were compared pairwise to the truth VCF made available by the GIAB consortium using the python tool hap.py and the results were reported. Different categories of false negative variants were extracted from hap.py comparison results from both variant calling pipelines.

Three categories of false negative variants are defined by the tool hap.py:

- Category 1: The variant is not called by the pipeline.
- Category 2: The genotype of the variant is assigned wrong.
- Category 3: The variant is in the neighbour region of a true variant.

Four features were investigated on different categories of false negative variants from both pipelines:

- whether the false negatives are variants in the truth VCF sequenced from sequencers other than Illumina sequencers, and were called by variant callers not in U1 and U2 pipelines?
- whether the false negatives are located in the difficult DNA sequencing regions (RepeatMasker defined regions)?
- whether the false negatives are in regions with low coverage (< 10 reads) or the alternative alleles were covered with less reads (reads mapped to alternative allele accounts for less than 30% of reads covered the region)?
- what is the genotype composition in different false negative categories?

One control variant set (all variants called from chromosome 15) was also investigated on the above four features to be compared with.

Conclusions:

False negative category 1 and category 3: the variants in these two categories share the same features. The main reasons for variants missing in these categories are:

- variants are located in difficult DNA sequencing regions
- variants are in regions with low coverage
- Half of the false negative variants are heterozygosities in truth VCF.

The U1 pipeline generated less false negative variants in these two categories, compared to those generated from U2 pipeline.

False negative category 2: 96% of false negative variants from U1 pipeline were in this category, which was also the reason why U1 pipeline showed lower sensitivity than that of U2 pipeline. The main feature difference was that the genotype errors of the U1 pipeline unique false negatives were wrongly assigned as heterozygosity to a homozygosity, whereas the errors of common and U2 pipeline unique false negatives were wrongly assigned a homozygosity to a heterozygosity.

Possible extensions of the investigation:

- Extend the investigations to other NIST reference materials (HG002, HG003 and HG004).
- Investigate which pipeline steps can be improved to get better performance.
- Extend the investigations to the results from other pipelines (e.g. Dragen, pipelines from other partners in the project).

3.2.3 Management and reporting of problematic variants – group work

Facilitated by Sharmini Alagaratnam, DNV GL

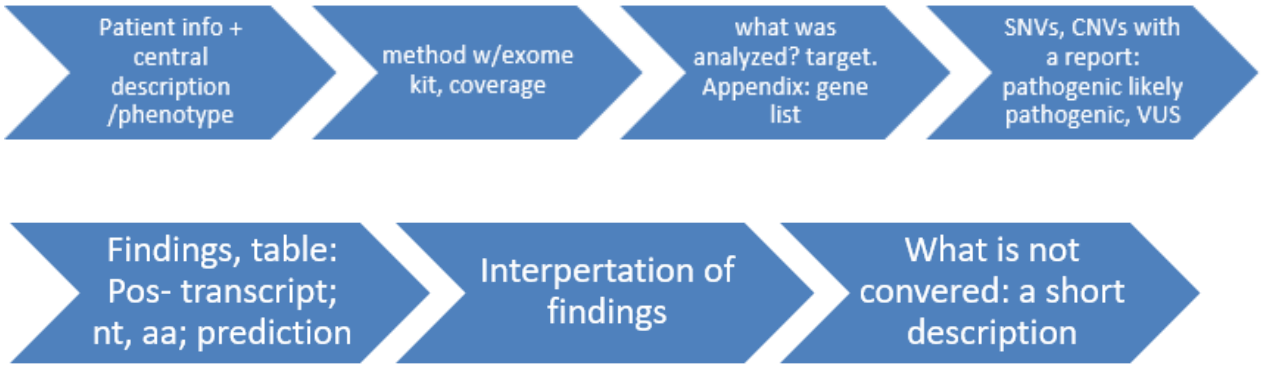
Group discussion topic	Note taker	Summary of discussions
Variants of Uncertain Significance (VUS) - Perform functional analyses/ RNA - Segregation /family analysis: de novo/ phasing	Courtney Nadeau, DNV GL	- Different indications trigger different thresholds for reporting - Class 3 variants are reported; it is then the responsibility of the requisitioner to request additional analyses - Internal / external references - Follow-up functional / family studies: new requisitions (after report). - Reporting will depend on who will use the information and expected competence. Education of other physicians on the differences between class 2 and 3 is important.
Pleiotropy - Variants with documented pathogenicity in other phenotypes - How close should phenotypes be? - Additional phenotypes - Interpretation and reporting	Daniel Nilsson, Dept of Clinical Genomics, Karolinska	- Differences between clinical and research (more open). - Pleiotropic gene; check genotype-phenotype correlation - Re-evaluation of patients - Novel findings require research to match with phenotypes. - What are trusted sources of information (e.g. OMIM, PubMed)? - Appropriate contact with referring physician required before results are delivered.
Technically challenging variants - Low coverage - Repeat regions - Calling and validation of copy number aberrations -	Chiari Rasi, Scifilab	Definition of low coverage depends on type of data you are working on. Some labs report all variants identified using a specific threshold, others process differently based on the technology available. Thus, "Low coverage" is an uncertain term. IP cov > min cov → validate (Sanger) - Validate if not seen by the hospital
Reporting	Eidi Nafstad, OUS	This group discussion performed an initial mapping of report contents between three different labs as outline in the illustrations below. - Structure of report are similar on main topics. - Details of report sections and contents need further investigations. - Need to distinguish between reporting of panels and WES; the latter being more exploratory. - **Further work:** suggest more thorough mapping of reporting at next workshop

Monogenic vs digenic vs other disorders	Savvas Kinalis, Rigshospitalet	International research indicates that about 5% digenic disorder diagnoses. - Nobody in the group experiences 5% digenic disorder diagnosis; often the investigation is stopped after an initial finding, which may lead to overlooking other relevant variants. Open genetic databases have shortcomings, perhaps also the internal databases do? Discussion on how to reduce errors. - Bias – personal gene discovery when looking for phenotypes, disregarded variants when they don't fit. - Genotype-phenotype links might be insufficient. - Positive controls **Further work:** - Suggestion to gather samples for joint research projects - References to existing guidelines on phenotype classification from Baylor college[1], which may impact the lab's diagnostic approach.

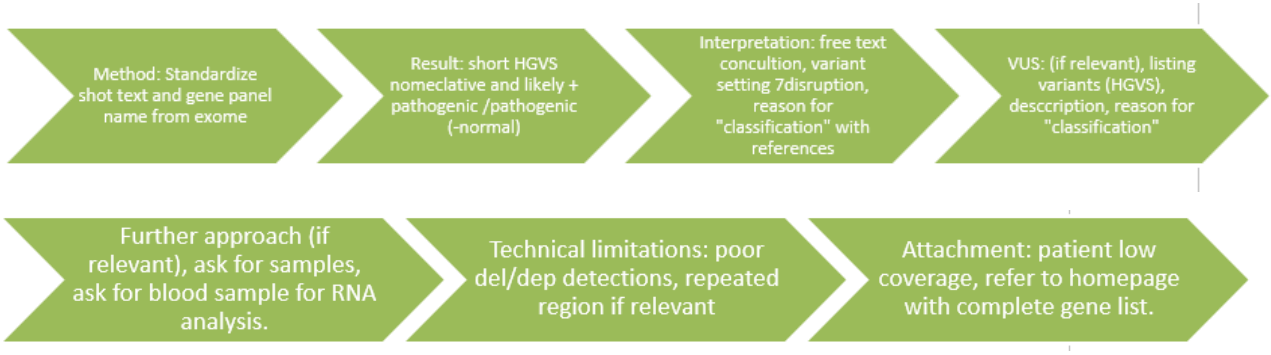
[1] Morten Dunø to provide reference?

Table 2 Initial mapping of reporting between labs

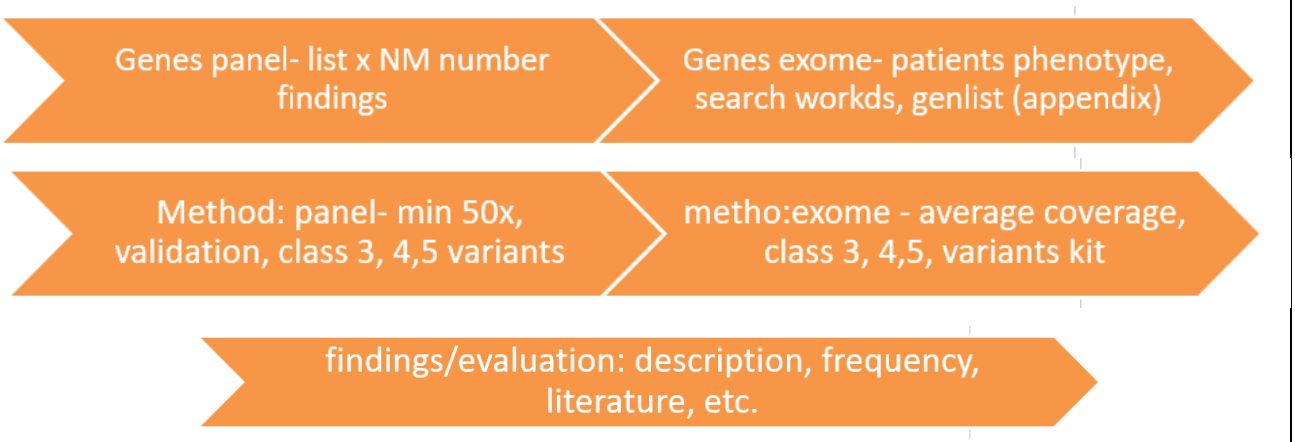
Finland: Mapping:



Norway, OUS, Mapping:



Genomic, Medicine, Denmark:



3.3 Sharing of data, tools and methods,

chaired by Henrik Stranneheim, SciLifeLab / Karolinska

Henrik introduced this section by reflecting on the need for sharing on both Nordic and global level, as a community effort to improve genetic variant analysis.

3.3.1 Mitochondrial variants - biology

by Nicole Lesko, Karolinska University Hospital

Nicole introduced the mitochondrial biology and functions. The respiratory chain enzyme complexes are coded by a combination of mtNDA encoded subunits and nuclear DNA encoded subunits. The circular mtDNA is small, only 16.5 kbp. Some commonly known mtDNA mutations were mentioned, including in transfer RNA genes. Deletions, single or multiple, where also discussed.

One cell contains multiple copies of mtDNA. Mutation thresholds are reached when enough of the copies are affected so that errors are not compensated for by unmutated copies. Mitochondrial DNA are inherited maternally. The "mtDNA bottleneck" describes how a mother could have only a few mtDNA mutation copies, but if these are the ones transferred to the child, they will be multiplied and take effect.

Molecular analysis of mtDNA include Sanger sequencing, southern blot and MLPA to localize large deletions and duplications.

Investigations for a suspected mitochondrial disease include muscle biopsy, skin biopsy and DNA sampling from blood and urine. MITOMAP is a human mitochondrial genome database recommended resource.

3.3.2 Mitochondrial variants - data, tools, analysis

By Henrik Stranneheim, SciLifeLab / Karolinska (data, tools, analysis)

Henrik referred to mitochondrial DNA as "the neglected genome". There are a number of diseases associated with mtDNA, but this is often overlooked in analysis and critical information is missed. There is missing annotation information on mitochondrial variants in major databases such as ClinVar.

Mitochondrial variants are analysed only through WGS, 3k-50k x coverage. Massive coverage chokes alignment viewer and affect structure variant calling. Currently analysis is based on blood samples, but analysis on muscle tissue is being explored.

Discussion points:

- All variants are documented, on average there will be 15 mutations
- A1:5000 risk is estimated for developing mitochondrial DNA associated condition, but it is assumed to be underdiagnosed.
- mtDNA duplications have not been observed / detected.
- Coverage needed to diagnose mtDNA mutations

3.3.3 Structural variants (SV) - biology

By Henrik Stranneheim, Karolinska/ SciLifeLab

Henrik discussed different types of structural variants (SV), such as Deletions (DEL), Duplications (DUP), Tandem DUP, Inversions (INV), Translocations (TRA) and Complex.

An overview of different callers' application to the different SV types was provided, including specification of min var size, seq type, joint calling applicability, processing capacity requirements, installation

complexity, method (paired, split-read, read-depth) and publication (PUBMEDID). He also provided a walk-through of the SV pipeline.

Challenges include

- standardization of variant caller data - vcf format too loose, should be stricter
- SV analysis is a new field, expected to develop rapidly. Currently there is a lack of information / features (genotype, confidence intervals, joint calling, mitochondrial SV calling, annotations)
- Multiple vcf records represent 1 underlying biological representation.
- No GiAB golden standard set

3.3.4 Structural variants – data, tools, analysis,

by Daniel Nilsson, Karolinska - Clinical Genomics

Daniel introduced work on structural variants.

3.3.5 Open cases

The participating labs shared patient cases as basis for discussion.

Case #1: Ying Sheng, OUS: APC c.423-3_454del – known disease causing variant

> Ying shared two difficult cases. Both variants are well known disease causing variants. The regions contain variants are also well covered. However, the variants are missed from the standard variant calling pipeline.
>
> *MSH2 c.942+3A>T*: located in a region with a long stretch of "A"s. The standard pipeline only calls indels at the position. We implemented a method which using read counts to manually call the SNP at the position. The called SNP will always need Sanger verification.
>
> *APC c.423-3_454del*: located in a region with repeats and a stretch of "A"s . The variant is clear by viewing IGV browser, however it is missed from the standard pipeline. We also tested other variant callers (samtools and freebays) and did not predict the deletion. We plan to implement a method to calculate mismatch ratios in the region. When the mismatch ratio is over the threshold, a warning will be sent to the variant interpreter to check the region with IGV browser.

Case #2: Daniel Nilsson, Clinical genomics Karolinska: Structural variant cases

Case #3: Morten Dunø, Rigshospitalet, exome sequencing cases

> Through the exome sequencing cases shared, Morten highlighted the importance of a multidisciplinary effort for qualifying WES (WGS) data, and the awareness of potential mosaic parents to children with "de novo" variants (which are at risk of being filtered out if you solely filter for de novo variants).

Case #4: Maria Rossing, Rigshospitalet, open case

Case #5: Nicole Lesko, Karolinska, open case

3.4 Next workshop

The next workshop will take place in Stockholm in the period 23.-26. April 2018.

4 NASPM 1ST NORDIC WORKSHOP FOR NATIONAL GENOMIC RESEARCH INFRASTRUCTURES

The regular meeting between the Nordic National Genomic Research infrastructures was now also included under the NASPM umbrella. The agenda and list of participants for this workshop are included in appendix 4 and 5 of this report.

APPENDIX 1: NASPM PLENARY MEETING AGENDA

Agenda
15th november

12:00 – 12:40	Registration
12:40 – 12:50	Welcome: Remi Eriksen, Group CEO, DNV GL
12:50 – 13:00	Opening address: Cathrine Meland, Director General, The Department of Specialist Health Care Services, The Ministry of Health and Care Services'.

Session 1: State of the art of personalized medicine
Chair: Dag Undlien

13:00 – 14:00	Genomics England 100,000 Genomes Project: Delivering genomic medicine in cancer and rare diseases. Clare Turnbull, Clinical Lead, Genomics England
14:00 – 14:30	Nordic collaboration for Personalized Medicine – why this matters for industry. Steinar Thoresen, Medical Director Abbvie.
14:30 – 15:00	Coffee break

Session 2: Personalized Medicine – Nordic initiatives
Chair: Valtteri Wirta

15:00-15:20	The Danish Genome Centre. Status and plans. Kasper Lindegaard-Hjulmann, Adviser, Danish National Genome Centre.
15:20-15:40	Genomic Medicine Sweden initiative. Richard Rosenquist Brandell, Karolinska Institutet.
15:40-16:00	Norwegian strategy for personalized medicine. Status and plans. Maiken Engelstad, Norwegian Ministry of Health and Care Services.
16:00-16:20	Finnish plans for personalized medicine – Implementation status. Jaakko Yrjö-Koskinen, Ministry of Social Affairs and Health, Finland
16:20-16:40	The Nordic Alliance for Sequencing and Personalized Medicine
16:40-17:00	Break
17:00 – 17:40	Panel discussion on Nordic collaboration in Personalized Medicine. Chair: Stephen McAdam Panelists: Clare Turnbull Gert Sørensen Steinar Thoresen Richard Rosenquist Brandell Maiken Engelstad Jaakko Yrjö-Koskinen
17:45	Reception and Dinner at DNV GL (Buses leave for Sandvika, Lysaker, Skøyen and Oslo S at 21:00)

Agenda
16th november

Session 1: Personalized Medicine – data sharing and big data
Chair: Karin Wadt/Morten Dunø

8:30 – 9:00	Data sharing and privacy. Thomas Keane, Global Alliance for Genomics and Health
9:00-9:30	Sharing variant classifications at a National level – experiences from the Netherlands, Joeri van der Velde, UMC Groningen dept. of Genetics
9:30 – 10:00	Big Data and Personalized Medicine, Lars Juhl Jensen, University of Copenhagen
10:00 – 10:30	Discussion
10:30 - 11:00	Break
11:00 – 11:30	P5.fi: Precision Medicine Finally Implemented, Pia Heikkurinen, SITRA, the Finnish Innovation Fund
11:30 – 12:00	BigMED Lighthouse Project, Thomas Smedsrud, Oslo University Hospital
12:00	Lunch
13:00	Formalizing the Nordic Alliance for Sequencing and Personalized Medicine. (by invitation only, contact Guro Pedersen for more info: guro.meldre.pedersen@dnvgl.com)
13:45	Start 4th Nordic Workshop on implementation of genomic medicine in health care. Standardization and data sharing. (by invitation only, contact Guro Pedersen for more info: guro.meldre.pedersen@dnvgl.com)
13:45	Start 1st Nordic workshop for National Genomic research infrastructures (by invitation only, contact Guro Pedersen for more info: guro.meldre.pedersen@dnvgl.com)

APPENDIX 2: NASPM - 4TH CLINICAL WORKSHOP - AGENDA

16. November			
Establishing vehicles for sharing			
Coordination: Tony Håndstad / OUS			
Time	Duration	Theme	Responsible
13:45	5 min	Introduction	Tony Håndstad / OUS
13:50	45 min	BigMed variant database demonstrator (demonstration & feedback)	Sharmini Alagaratnam / DNV GL
14:35	30 min	Norvariom, a database of variants in the Norwegian population	Øyvind Evju & Tony Håndstad / OUS
15:05		BREAK	
15:30	30 min	Lighting the SciLifeLab Beacon - experiences so far and future plans	Måns, Henrik or Chiara / SciLifeLab –
16:00	45 min	Structured phenotyping - Intro - Experiences/challenges with structured phenotyping - Usability of structured phenotype data for variant interpretation and classification - Discussion	Yngve Sejersted / OUS Ketil Heimdal / OUS
17:30		END OF SESSION	

17. November			
Enhancing quality of data and processes			
Coordination: Sharmini Alagaratnam / DNV GL			
Time	Duration	Theme	Responsible
9:00	5 min	Introduction	Sharmini Alagaratnam / DNV GL
9:05	55 min	Benchmarking of variant calling Trio HG002-004: All variants and PU panel Evaluation of caller performance	Courtney Nadeau / DNV GL Ying Sheng/ OUS AMG Sharmini Alagaratnam / DNV GL
10:00	60 min	Reporting of problematic variants.	Group work
11:00		END OF SESSION	

17. November			
Sharing of data, tools and methods			
Coordination: Henrik Stranneheim / SciLifeLab			
Time	Duration	Theme	Responsible
11:10	5 min	Intro	Henrik
11:15	60 min	Open cases: Walk-through of difficult cases or non-standard case	Nicole / Daniel - Karoliska Morten Maria
12:15	45 min	Lunch	
13:00	45 min	Structural variants – data, tools, analysis	Daniel Nilsson / Klinisk genetik (Stockholm) & Henrik Stranneheim / SciLifeLab
13:45	30 min	Mitochondrial variants - Biology - Data, tools, analysis	Nicole Lesko / Karolinska & Henrik Stranneheim / SciLifeLab
14:15	15 min	Wrap up	
14:30		END OF SESSION	

APPENDIX 3: NASPM - 4TH CLINICAL WORKSHOP – LIST OF PARTICIPANTS

First name	Last name	Organisation
Ane Yde	Schmidt	Center for Genomic Medicine, Rigshospitalet
Birgitte	Bertelsen	Center for Genomic Medicine, Rigshospitalet
Bobbie	Ray-Sannerud	DNV GL
Chiara	Rasi	Clinical Genomics, SciLifeLab
Clare	Turnbull	Genomics England
Courtney	Nadeau	DNV GL
Dag	Undlien	Oslo University Hospital
Daniel	Nilsson	Dept of Clinical genetics, Karolinska Univ Hospital
Eidi	Nafstad	Dep of Medical Genetics, Oslo University Hospital
Guro Meldre	Pedersen	DNV GL
Henrik	Stranneheim	Scilifelab/CMMS
Kaisa	Kettunen	Institute for molecular medicine FIMM
Karin A. W.	Wadt	University Hospital of Copenhagen, Department of Clinical Genetics
Knut Erik	Berge	Dep of Medical Genetics, Oslo University Hospital
Kristina	Lagerstedt Robinson	Dept of Clinical genetics, Karolinska Univ Hospital
Maria	Rossing	Center for Genomic Medicine, Rigshospitalet
Matti	Kankainen	HUSLAB laboratories & Helsinki University Hospital
Morten	Duno	Department of clincal genetics, Rigshospitalet
Morten C.	Eike	Oslo University Hospital
Måns	Magnusson	Scilife lab
Nicole	Lesko	Karolinska University Hospital
Peter	Johansen	Department of clincal genetics, Rigshospitalet
Robin	Andeer	Clinical Genomics, SciLifeLab
Savvas	Kinalis	Genomic Medicine, Rigshospitalet
Sharmini	Alagaratnam	DNV GL
stephen	McAdam	DNV GL
Svein Tore	Seljebotn	Oslo University Hospital
Tony	Håndstad	Oslo University Hospital
Valtteri	Wirta	SciLifeLab
Vibeke	Binz Vallevik	DNV GL
Ying	Sheng	Oslo University Hospital
Yngve	Sejersted	Dep of Medical Genetics, Oslo University Hospital
Øyvind	Evju	Oslo University Hospital

APPENDIX 4 NASPM - 1ST RESEARCH WORKSHOP - AGENDA

1st Nordic workshop for National Genomic research infrastructures

16.11.2017 Day 1

Time	Topic	Speaker
13:45	Welcome and Introduction	Kjetill S Jakobsen
13:50	Sequencing services and experience with 10XGenomics at SciLifeLab	Phil Ewels
14:20	Personalized medicine, services and experience at FIMM	Pekka Ellonen –
15:10	Human genomics, diagnostics and future developments	Nicola Cahill, Pacific Biosciences
15:40	Update on faster workflow, improved flexibility and sensitivity	Håkon Velde, Illumina
16:10	BioCenter Helsinki	BioCenter Helsinki
16.50	Short intro to NorSeq for the SAB, Nordic invitees and all	Kjetill S Jakobsen/ Dag Undlien
17:00	Presentation of the Scientific Advisory Board and discussions with SAB	

17.11.2017 Day 2

Time	Topic	Speaker
09:00	Group presentations – Trondheim, Bergen, Tromsø, Radiumh., Ullevål, CEES	
10:50	Updates from Elixir Norway	Kjell Petersen
11:10	Presentations from sites or Group meetings	lab/ bioinformatics/ leader

APPENDIX 5 NASPM - 1ST RESEARCH WORKSHOP – LIST OF PARTICIPANTS

First name	Last name	Organisation
Anne	Mortensen	Illumina
Hakon	Velde	Illumina
Ana	Lid	OUS Radiumhospitalet
Arnar	Flatberg	NTNU
Arne	Sandvik	NTNU
Arvind	Sundaram	Norwegian Sequencing Centre / Dept. of Medical Genetics, UiO
Ave	Tooming-Klunderud	NSC at CEES
Christofer	Flood	Illumina
Christopher	Fenton	University of Tromsø
Endre	Anderssen	UiT - Norges Arktiske Universitet
Guðrið	Andorsdóttir	Genetic Biobank of the Faroe Islands
Hagar	Taman	UiT
Hans-Richard	Brattbakk	UiB GCF
Jon J.	Jonsson	Landspitali - University of Iceland
Kjell	Petersen	University of Bergen
Kjetill S	Jakobsen	University of Oslo
Lars Birger	Aasheim	OUS
Leonardo A.	Meza-Zepeda	Oslo University Hospital
Marianne	Dalland	NSC, AMG
Marius	Bjørnstad	Oslo University Hospital, NSC
Michael	Sonnestad	Pac Bio
Morten	Skage	Norwegian Seqineing Centre
nicola	cahill	Pac Bio
Noomi Oddmarsdóttir	Gregersen	FarGen
Oleg	Agafonov	Bioinformatics Core Facility, Radium Hospital
Pekka	Ellonen	Institute for molecular medicine FIMM
Philip	Ewels	NGI Stockholm, SciLifeLab
Ragnhild	Aaløkken	NSC, MedGen OUS
Rita	Holdhus	NCS-PM
Ruth	Paulssen	UiT - Norges Arktiske Universitet
Sten Even	Erlandsen	NTNU GCF
Ståle	Nygård	University of Oslo
Thea Charlotte	Smedsrud	The Norwegian Radium Hospital
Tom	Wheeler	NTNU
Tomasz	Stokowy	University of Bergen
Tone	Christensen	NTNU
Tonje	Borch	Norwegian Directorate of Health
Vidar M	Steen	University of Bergen & Haukeland Univ Hospital
Jinchang	Sun	Genomics Core Facility, Radiumhospitalet