DANISH NATIONAL GENOME CENTER

Minutes

Date: 21-06-2022
Unit: NGC
Caseworker: IVB.NGC
Ref.: 2204653
Doc. No.: 2267563

2. Meeting in DNGC's international advisory board

Date: May 18th, 15.00-17.00 (CET)
Location: Teams (Link is in the Outlook invitation)

Abstract: ***Reporting, Storing, and Use of Data***
A central element in fulfilling the vison of personalised medicine is making data accessible to both clinicians and researchers. But how do we build smart and sustainable data infrastructures that enable valuable reporting, storing and use of data?

This is one of the key challenges and tasks that The Danish National Genome Center is currently facing, and a task that we share with colleagues across nations. The Danish National Genome Center would therefore like to invite our members in the International Advisory Board to share experiences with and visions for making data accessible in order to support patient treatment and research.

Agenda

Item	App. time	Activity
1/7	15.00	**Welcome and presentation of the agenda** /Tim Hubbard
2/7	15.05	**Presentation of DNGC's data infrastructure: Reporting, National Genome Database, National Variant Database, and Interpretation Tools** / Cathrine Jespersgaard
3/7	15.20	**Session 1: Making data accessible in the clinic** *Presentation from:* 15.20: Dag Undlien - Oslo University Hospital 15.30: Kym Boycott – University of Ottawa
4/7	15.40	**Shared discussion on key challenges and needs for making data accessible in the clinic** - What are the challenges in reporting, storing, and making data accessible in the clinic? - What needs for data storage and access do you face amongst colleagues or partners? - What are examples of best practice cases?
5/7	16.15	**Session 2: Making data accessible for research** *Presentation from* *16.15*: Ruben Kok - Dutch Techcenter for Life Science *16.25:* Aarno Palotie – University of Helsinki

6/7	16.35	**Shared discussion on key challenges and needs for making data accessible for research** - What are the challenges in reporting, storing, and making data accessible in research? - What needs for data storage and access do you face amongst colleagues or partners? - What are examples of best practice cases?
7/7	16.55	**Concluding remarks** /Tim Hubbard

Participants
Tim Hubbard, Professor, Kings College London (Chair)
Richard Rosenquist Brandell, Professor, Karolinska Institute (Vice Chair)
Valtteri Wirta, Dr., Ph.d., Karolinska Institute
Heidi Rehm, Ph.D, Broad Insitute
Dag Erik Undlien, Professor, M.D., PhD, Oslo University Hospital
Russ Altman, Professor, Stanford University
Jean- François Deleuze, Ph.d., Head of CNRGH
Aarno Palotie, M.D., Ph.d, Institute for Molecular Medicine Finland
Ruben Kok, Ph.d., Director, Dutch Techcentre for Life Science
Kym Boycott, Clinical geneticist, University of Ottawa

DNGC's Secretariat
Bettina Lundgren, Director, DNGC
Ole Lund, Chief Bioinformatics Officer, DNGC
Ali Syed, Head of the HPC Platform Team, DNGC
Lene Cividanes, Head of Research, Clinic and International Relations, DNGC
Ivana Bogicevic, Policy officer, Research, Clinic and International Relations, DNGC

Minutes

1/7 Welcome and presentation of the agenda /Tim Hubbard (Chair)

Tim Hubbard welcomed everyone and introduced the theme shortly.

2/7 Presentation of DNGC's data infrastructure / Bettina Lundgren (CEO)

Bettina Lundgren gave a short presentation of the Danish National Genome Center's data infrastructure, including:

- The Danish healthcare system is decentralized and organized across ministry, regions and municipalities.
- The Danish National Genome Center (DNGC) is a health authority under the Ministry of Health.
- The DNGC's core tasks are to:
    - Collect and store Danish genome data from both clinic and research.

    - Make genome data accessible to clinical personnel, researchers and patients.
    - Promote development of personalised medicine in Denmark.
- The DNGC is currently developing and overseeing the national infrastructure for personalised medicine, which consist of:
    - Two national WGS centers that perform WGS for the healthcare system.
    - A national supercomputer that can load, process and display sequence results to interpreters and clinicians in the healthcare system.
- The DNGC is also developing a National Genome Database and a National Variant Database.
    - The intention is, that data will be reported to the two databases from both the Danish healthcare system and clinical research projects.
    - The vision is to create both a classified and an un-classified variant database. However, the legal framework is still a challenge.
    - The first 60.000 WGS in the National Genome Database will stem from 17 selected patient groups.
- Research projects which are associated with a certified Danish research institution, promote the development of personalised medicine, have a significant societal benefit and are approved by the Danish ethical committee system can gain read-only access to the data in the National Genome Database.
    - These projects can also get access to a secure private cloud on the supercomputer.

3/7: Session 1: Making data accessible in the clinic

Tim Hubbard introduced shortly the two speakers in session 1.

Dag Undlien gave a presentation on the diagnostic data flow at Oslo University Hospital, including:

- Data infrastructures are always a compromise between functionality, data security and privacy.
- The diagnostic data flow at Oslo University Hospital consists of a secure infrastructure for sequencing and another secure infrastructure for storing and making data available for healthcare personnel.
- Due to security measures ICT-infrastructures do not have access to internet, this increases workload and makes most tasks cumbersome e.g.:
    - Transferring data across different ICT-infrastructures – often done with a stick or disk.
    - Maintaining the infrastructure and different environments.
- Software considerations are important.
    - Using commercial software might put some restrains on getting data in and out of the system, or on importing data from other sources. Often LIMS vendors need to help with even simple statistics.
    - In-house software gives an advantage in terms of full control of data, controlling export and import, and it gives you full control

over statistical jobs e.g it becomes possible to measure how much time is used on interpreting each variant.

- Making effective and valuable infrastructures for data sharing requires one to be attentive to even small details, especially the way data is reported and disclosed. Valuable sharing requires use of data standards – e.g. the GA4GH standards.
- Matchmaker Exchange might be a solution for sharing across borders.

Kym Boycott gave a presentation on the Canadian "All 4 one"-project, including:

- The "All 4 One" health data ecosystem is in its development phase.
- It has two goals:
    - Facilitate high quality Clinical GWS as a standard-of-care.
    - Facilitate precision health research.
- Two-pronged data solution:
    - A Canadian variant database for QA/QI of clinical GWS.
    - Opt-in re-contact registry for REB-approved research projects.
- Engagement of GAPP Projects across the country.
- Clinical Use Case: Rare Disease Knowledge Network
    - Stakeholders: Data custodians, payers, and regulators.
    - Ensuring hospitals report in a standardized fashion:
        - HPO Terms
        - Diagnosis
        - Variants
        - Pathogenicity
        - Supporting evidence
    - Creating infrastructural links between data custodians in the healthcare system.
- Research Use Case: Re-contact registry
    - Enable the identification of eligible research participants.
    - Send out invitations to participate in REB-approved research.
    - Canadian legislation for using clinical data in research is difficult, the patient/family/citizen needs to free their data for research. Thus, when the patient/family/citizen registers in the "All for One Connect"-registry, they consent to being contacted and they make their data available for research.
        - Initial consent is to all types of sharing including international sharing.
    - The "All for One Connect"-registry has a governance, overseen by national REB-boards, and is held by a non-for-profit.
    - The registry holds a minimum of patient information:
        - Participant ID
        - Contact information
        - Information on referring provider
        - Variant file (this can flow into the "Genomics4RD"-infrastructure)
- Genomics4RD-infrastrucutre has three access levels (at the moment it is mainly level 2 that is available):
    - Fully open
    - Controlled access (read only)
    - Restricted access

4/7 Shared discussion on key challenges and needs for making data accessible in the clinic

Following the two presentations, Tim Hubbard opened the floor for a shared discussion on infrastructures for the clinic.

The following three main themes were discussed:

Theme 1: Secondary findings

The shared discussion underscored, that there are multiple ways of defining "secondary findings" or "incidental findings"; and that not all countries are using the ACMG-list. There are also multiple ways of dealing with reporting secondary findings:

- Genomics England has found that 85% of patients consent to having secondary findings reported back. They have decided on a narrow reporting, so only actionable mutations are reported back to consenting patients.
- The US has also found that the majority of patients consent to having secondary findings reported back. However, secondary findings place an enormous burden on patients, clinicians and interpreters. Thus, the US is looking into a new approach – especially relating to material from biobanks – here an expert panel decides if a secondary finding meets the threshold for being medical significant or not. If it meets the threshold, the patient or family will be informed, and get the opportunity to receive the information or not. In that way, the patient still has autonomy.
- In Canada patients must consent. Based on the ACMG-list, interpreters and clinicians evaluate the found variants, and assess whether they are medical valuable for the patients or family, before reporting them back. They have found, that they only report secondary findings in 5% of clinical cases.
    - In a research setting, they do not actively look for secondary findings. They only report back in case a medical significant variant is found by accident.
- In France they have not been allowed to report secondary findings. However a new legislation has just been approved allowing them to report back based on the ACMG-list.
- It was highlighted that it is important to ensure that reports on secondary findings are informative.

Theme 2: Quality control

- All agreed that conducting benchmark and blind tests for the entire flow is important.
- All countries experienced that getting pipelines ISO-certified and validated is a high priority.
- Quality and accuracy control is more than conducting benchmark test. Two additional questions were highlighted as central:
    - How accurate is the interpretation of pathogenicity?
        - Here national and international knowledge sharing and sharing of variant classification is important.
    - How accurate is the causality interpreted?

- Here clinical data is crucial.

Theme 3: Re-processing of data

The shared discussion underscored, that there are multiple ways of dealing with re-processing of data, and that most countries are still in the process of figuring out how much value re-processing adds to patient trajectories.

- Genomics England has no systematic flow for re-processing data.
- In Canada, they have guidelines for re-processing:
  - Samples are re-processed every 18 months, if there has been changes to the pipeline.
  - Samples can be re-processed earlier if there is any progress or change in the patient's health status.
- In France the vision is to re-process every six month but only for non-conclusive tests.
- In the US there is an automated pipeline for re-processing. To ensure that the same variants are not interpreted twice, the original analysis is annotated and all variants are given a time stamp.

<u>Besides the two main themes, the following was also discussed:</u>

- It is important to ensure that reports with results are informative and easy to use for healthcare personnel.
  - In Denmark the interpreters are located in the healthcare system, and the results are discussed with clinicians in multidisciplinary teams.
  - Genomics England label their reports with tiers.
  - In Norway the lab sends out the report, however they mostly do in silico panels.
- High quality interpretation requires good reference data and data on the allele frequency of the population.
- Data storage and computer power is a central element in personalised medicine. However, it is expensive.
  - Some countries use cloud solutions, other use on-premise solutions. However, it was highlighted that most countries refrain from storing raw data on cloud solutions due to regulations and data security.
  - The DNGC was advice to always use on-premise solutions to store raw data.
- It was discussed that in the future it might not be necessary to store the raw data. However, at the moment it is highly valuable to store raw data, since it is a central piece in the further development of interpretation.
  - In the US they no longer store exome results, since it has become cheap and fast to conduct an exome sequencing. They do however store WGS, since it is too expensive to re-do.
- Members discussed different ways of storing data e.g. using matrix tables, storing data on variants separately from raw data. Data is then put together in real-time when you need it.

5/7: Session 2: Making data accessible for research

Tim Hubbard introduced shortly the two speakers in session 2.

Ruben Kok presented the vision for the Dutch national data infrastructure, including:

- The vision is to create a federated structure where multiple stakeholders (public and private) can access rich data-sets consisting of health data collected along the life cycle.
- The Heath RI should support both:
    - Data-driven healthcare
    - Research and innovation
    - Learning environments
- They have received a 69 million Euro fund from the Dutch Growth fund. The first focus is:
    - Securing a shared voice and ELSI aspects on data use and re-use across all actors e.g. consent, data linkage.
    - Create a federated health data infrastructure with FAIR data and possibility for distributed analysis and learning
    - Create a "One stop shop" for both public and private users, who need access to data and a service portal.
- At the moment health data sources are not FAIR enough, so there is a huge task in creating and implementing standards for data reporting and disclosure.
    - Further, they do not have a social security number (like Denmark), so it is a challenge linking data across multiple registries and sources.
- They have identified a list of currently 18 legal obstacles for re-using health data. With support from ministries, the project is addressing these obstacles with central stakeholders and working towards changing regulations.
- During COVID, they managed to set-up a decentralized system, where data from all hospitals were reported (based on FAIR-principles) and then made available. They are taking inspiration from this set-up.

Aarno Palotie presented the FinnGen project, including:

- FinnGen is a research project based on a public-private partnership between 13 private companies and all universities in Finland.
- The vision is to have data on 10% of the population.
    - Plan to conduct Axiom QWA Array on 500.000 individuals. Phenotype data will be collected from health registries.
- The project began in 2017 and should run for 10 years.
- The first five years have mainly focused on building the infrastructure.
- In Finland they have a social security number, and this is key in linking data from different registries.
- The biobanks or hospital provide the sample, FinGenn then returns the genotype data. They are not allowed to return phenotype data.
- The infrastructure is based on the google cloud system.
- Every partner has their own sandbox with data (a green box with aggregated data, and red box with raw data). They can bring their own tools and bring their own data e.g. if they need a control group.

- There are strict regulations for taking data out of the infrastructure, and partners can only take data out after special review.

6/7 Shared discussion on key challenges and needs for making data accessible for research

Main points from the discussion were:

- If data should be FAIR, it is crucial to set-up and follow standards for reporting and disclosing data.
    - The Ga4GH were highlighted as a good international standard.
- All countries are looking into creating infrastructures and solutions to share data internationally within the current legislations.
    - Beacon-2 was discussed as a possible solution.
    - The 1+M Genomes federated infrastructure was highlighted as holding great potential.
- There is a movement towards “personal data folders”
    - In the Netherlands they are creating a digital solution, were citizens can store their health data. Thus, citizens have control and provide physician access to data.
    - This is also intended in “European health data spaces”.
    - Finland has the KANTA-system that citizens can use to see their health data.
    - Denmark also has electronic journals, so citizens can see their own health data.

7/7 Concluding remarks /Tim Hubbard

Tim Hubbard and Bettina Lundgren closed the meeting. All members expressed an interest in participating in a meeting in Denmark during spring 2023.