Workshop Agenda

Rare Disease Community

Date & Time:	September 20, 2023 at 11am PT
Slides:	Rare Disease- GA4GH Connect 2023 slides
Meeting Recording:	https://drive.google.com/file/d/1KrCUvsYvPfua0kc4EhUxyztlOLegA9J1/view?usp=drive_link
Chairs	Nicole Vasilevsky & Tudor Groza
Attendees*:	Charles Steward, John Seavitt, Nara Sobreira, Soichi Ogishima, Robert Davies, Nadia, Eric Sid, Jen Harrow, Renan Martin, Ainslie Tisdale, Joseph Manuel, Mike Becich, Catalina Bustamante, Jan Friedman, Hector, Tony Brookes, Jose M Fernandez, Bryan Laraway, anne O'Donnell-Luria, Adriana Malheiro, Kayla Socarras, Evgeniy Mozheiko, Nicholas Owen, Senkei Umehara, Nicola Mulder, Ada Hamosh, Kym Boycott, Michael Baudis, Monica Munoz-Torres, Sveinung Gunderson, Ashley Hobb, Jordan Lerner-Ellis, Giselle Kerry, Neerjah Skantharajah **Please note that the names of a portion of in-person attendees were not collected. If you attended the meeting, but do not see your name please feel free to add it.*
Aim of the Meeting:	
Actions:	☐ Plan follow-up meetings ☐ Any other business arising

Session Description

The GA4GH Rare Disease Community aims to promote global cooperation, data sharing and collaborative rare disease research through identifying the need for new standards, and/or implementing existing GA4GH standards. This mission is accomplished by collaboratively defining scientific rare disease use cases that may benefit from global cooperation, by a commitment to advancing those use cases through either the development or implementation of

GA4GH standards, and by providing practical examples and guidance for rare disease sharing and analysis. The RD Community Connect session focuses on one of the most critical dimensions of rare disease (RD) diagnosis and care coordination: phenotypes. Deep phenotyping of patients suspected of or diagnosed with a RD - using the Human Phenotype Ontology - has become standard practice in the last ten years. And while the adoption of standardized terminologies and an increasing willingness to share data internationally have laid the foundation for progress in the RD field over the course of the last few years, for many patients the diagnostic odyssey will ultimately be futile: more than 50% of RD patients do not receive a molecular diagnosis. This session aims to discuss some of the challenges underpinning further progress, including the encoding of longitudinal phenotypes, the communication and exchange of phenotypes in human language or the definition and sharing of statistical data on phenotypes in diagnosed cohorts.

Workshop Agenda

	Agenda Item	Chaired by	Item Type	Time
1	Introduction	Nicole / Tudor		5 mins
2	Keynote	Nicole / Tudor		25 mins
	Dr. Charles Steward - Head of Head of Patient and Participant Engagement, Genomics England **Title:** Phenotyping lessons from patients' lived experience with rare diseases			20 mins
	Q&A			5-10 mins
3	Panel Discussion	Tudor / Nicole		30 mins
	Topics: • Phenotype internationalisation efforts - Multilingual phenotypes • Representing and capturing longitudinal phenotypes • Curating layperson terminology • Definition and sharing of statistical data on phenotypes in diagnosed cohorts • How to name and define rare diseases Panelists: • Prof. Ada Hamosh • Prof. Kym Boycott • Prof. Nicola Mulder • A/Prof. Monica Munoz-Torres • Prof. Michael Baudis • Prof. Soichi Ogishima			
4	Group Q&A (sli.do)	Nicole / Tudor		10 mins
5	Next Steps	Nicole / Tudor		10 - 15 mins

Meeting Summary

[General Overview - Challenges and Opportunities Identified](#)
[Detailed Minutes](#)
- [Introduction](#)
- [Keynote](#)
- [Panel Discussion](#)
  - [Topic 1: Language and communication](#)
  - [Topic 2: Disease Naming and Definitions](#)
  - [Topic 3: Data Sharing](#)
- [Next Steps](#)

General Overview - Challenges and Opportunities Identified

Challenges

- **Phenotype Standardization:** Standardizing phenotype descriptions across languages and cultures remains a challenge, but finding synonyms and engaging in mapping efforts will be essential for global collaboration.
- **Disease Naming:** The lack of consensus in disease naming and definitions can lead to confusion in research and healthcare.
- **Data Silos:** Patient phenotype-genotype data is often siloed in institutional databases, hindering collaborative research efforts.

Opportunities for Follow Up**

- **Standardization Efforts:** Continue efforts to standardize phenotyping and engage in mapping efforts across languages and cultures to facilitate global rare disease research.
- **Consensus Building**: Work towards a consensus on disease naming and definitions within the rare disease community.
- **Data Integration:** Explore privacy-preserving methods for integrating patient phenotype-genotype data to foster collaborative research.
- **Patient Involvement:** Develop tools and platforms that enable patients to contribute to phenotype profiles and participate actively in rare disease research.
- **Education:** Educate clinicians and researchers about the importance of standardized language and terminology in rare disease research.
- **Continued Collaboration:** Foster collaboration between the clinical and patient engagement worlds to improve rare disease research and diagnosis.
- **Monitoring Changes:** Establish mechanisms for tracking changes in disease definitions and nomenclature to adapt to evolving knowledge.

**these topics may be filtered into an opinion paper led by Tudor Groza and contributed to by the RD community*

Detailed Minutes

Introduction

- Welcome to our new chairs!
    - Nicole Vasilevsky
        - Associate Director of Data Science @ Critical Path Institute, Tucson, Arizona. (Based in Portland, OR, USA)
    - Tudor Groza
        - Data and Technical Coordinator @ Rare Care Centre, Perth
        - Honorary Research Fellow @ Telethon Kids Institute, Perth
        - Visiting Scientist @ SingHealth Duke-NUS Institute of Precision Medicine, Singapore
- The Rare Disease working group has been recently formed and aims to promote global cooperation, data sharing and collaborative rare disease research through identifying the need for new standards, and/or implementing existing GA4GH standards
- Main goal: incubator to gather current challenges and propose and advance potential solutions

Keynote

- Charles Steward
    - Head of Participant and Patient Engagement at Genomics England

Case Study #2: Daisy and Lola Clatworthy

- Lola presented with intractable seizures at 3 months
- Enrolled into 100,000 Genomes Project , and nothing was found
- Father met with a parent from North America who said Lola looked like their child who also had the same seizure phenotype and had a genetic diagnosis for CRELD1. This was a chance encounter via a Facebook patient support group for people with epilepsy.
- Spoke to a clinical geneticist because that child had a diagnosis of CRELD1
- Geneticist said they would not look further into CRELD1 as there was no evidence it was expressed in the brain - it is associated with heart phenotype
- CRELD1 is expressed ubiquitously in 6 different life stages in the brain (Jaffe et al Nat. Neurosci. 18, 154–161 (2015)). These data can be viewed in UCSC and Ensembl genome browsers.
- If you look in Ensembl and look at rodent models for CRELD1 there is data that this gene is associated with abnormal brain development. CRELD1 also has bimode inheritance. AD for heart and AR for brain

- Both of the above facts, suggest that if you look a little deeper, there is data that you can use
- Lola died from sudden unexpected death in epilepsy (SUDEP)
- Alfie, younger brother developed similar seizures to Lola after she died, and again geneticist was asked to look into CRELD1 gene
- Sequenced Alfie and found a bi-allelic variant in CRELD1 and that was the cause of both Alfie's and Lola's epilepsy. This information opens up potential reproductive opportunities.
- Further Reading/Resources:
    - https://rarerevolutionmagazine.com/creld1-warriors-bringing-the-scientific-and-parent-community-together/
    - https://www.creld1.com/

Summary

- Phenotyping is as important to get right as generating the genome sequence
- Phenotyping should not be considered static as disorders may change over time
- The ability to re-phenotype patients for genome reanalysis is essential for suspected and unresolved genetic disorders
- A holistic approach to all family phenotypes, even those that are seemingly unrelated
- Patient concerns should be treated extremely seriously, even if they are imperfectly expressed
- Patients are often inspiring and can help drive research into their disorders and we need to support them!

Q: How do we develop a dynamic process that consists of a phenotype profile contributed to by patients? How do you see this partnership between the clinical world and the patient engagement world evolving?
A: Has to be some sort of automated process where EHR can feed into a genomic database and use that for genomic analysis or re-analysis. For phenotypes it can either be patient-driven, but that's very inefficient so it needs to be automated somehow. Constant reanalysis of all unresolved genomic data is not a feasible/sustainable approach as genomic databases grow rapidly. So the second challenge is that there is a lot of noise in the system/diagnosis coding which makes things difficult and can decrease the efficacy of re-phenotyping.
Q: Where would the patient be able to intervene in this system and contribute to data points?
A: Some institutions do conduct self-phenotyping for patients, but some families are exceptional at that and some are not. Need to be able to facilitate it for families who are health literate.

Panel Discussion

Topic 1: Language and communication

The ability to both convey as well as to understand and retain medical information is critical for medical professionals and patients alike. Two challenges associated with language and rare

disease phenotype terminologies are multilinguality (i.e., the ability to use standardized concepts in a native language - including indigenous languages (the languages of the “first people of the land”) and layperson expressions (i.e., the ability to use standardized concepts in a form comprehensible by individuals with no medical training)

Q: What challenges and solutions do you see with respect to curating and maintaining multi-lingual phenotype standards?

NM

- Creation of Sickle Cell Disease Ontology - combination of existing ontologies with additional terms (standards of care, genetic modifiers)
    - Translated into french, and now into Swahili
- Patients can self-report with what the are feeling when they have an episode at the hospital
- Reason we needed to develop the ontology is because in sub-saharan africa it’s not rare at all but patients are from different educational background and terms used across Africa will be different based on where they are from and their culture
- The way we manage those differences is through language tags (e.g. english to layperson, or english to french)
- Think about what you need it for. Is it for the patient to understand the disease better and report it better? Is it for the clinician/science side? Work out the application and then be able to connect the dots
- 2072 terms for one very well understood disease
- Read more here: https://scdontology.h3abionet.org/

MMT

- Ensuring that there is a source of truth
- Having timely sync up with source of truth
- Make sure there is a regular release cycle of updating
- Ensuring that translation is done in a way that makes it easy to sync up in a regular manner
- You will need volunteers and crowdsourcing - done with HPO

SO

- Experience of HPO translation into Japanese
- Medical terminology is maintained by Medical [..] Committee by the Japan Medical Association
- There are over 100 subcommittees in each disease area
- Very time consuming and costly process, but also very important to address discrepancies in disease areas
- Hopefully AI can assist with this process in the future

Q: What is needed to make phenotyping standards more accessible to patients?

MMT

- Helpful to access groups, both researchers and patients, that are willing to provide feedback and encourage that process - Ask questions like “What’s easy to understand? What isn’t? What would improve your understanding?” Even teaching clinicians to file tickets in GitHub is helpful.

Topic 2: Disease Naming and Definitions

Diseases are named via many different criteria and there is not one agreed upon standard.

Q: How should we best handle these multiple approaches towards disease naming? Is use of synonyms in standards sufficient?

Q: There are many competing definitions of what constitutes a rare disease and the criteria to define these (like the frequency in the population, what is rare in one area may not be rare in another, like Sickle Cell Disease is not rare in Africa but it is in the US.) What kind of criteria do you think should be applied to defining rare diseases? Should there only be one approach?

MMT

- There has to be a conversation that allows all different players to agree on classification strategies, synonyms, curation rules, nomenclatures (Mondo, WHO, ClinGen etc.) have been trying to crystallize this
- Important to agree on a governance to agree on nomenclature

MB

- The community itself defines the rare disease and the somatic genetics is not so much on the horizon when talking about rare disease
- Even in talking about what a rare disease is, that concept changes constantly due to the evolving knowledge
    - Examples - rare disease, syndromic rare disease, cancer type rare disease - how do you go back and address/incorporate the longitudinal changes of knowledge concepts
- Temporal changes - how different changes diffuse through communities

MMT

- Imperative to find a computational solution and one that integrates a lot of resources into all the knowledge we have about rare disease, must be an aggregate effort, not picking one that is the best
- Must also keep track of the changes

AH

- You have to accept that whatever you think is correct might not be correct later. Names and concepts will change. Rare Disease definition ➔ A population prevalence of less than 1 in 2000 - this is a good definition and everyone agrees except the US

- Rare is context specific
- Need to accept something as rare even if it's common somewhere else

MB+MMT+AH

- Rare is multidimensional e.g. rare in the global sense where there is no information globally about the disease, rare in the research context, rare but not genetic (cancer) etc.

AH

- What's important is to not boil the ocean. Synonyms should be cross referenced. If you are using an ontology you must label where you are in that hierarchy so we can all have a conversation and know what each other is talking about. Opposed to symbols, a fan of numbers so people can know where they are in the grander scheme of things.

Topic 3: Data Sharing

Data sharing, enabled by standards, has been one of the major success factors underpinning the rare disease community and domain. Yet, most of the patient phenotype-genotype data remains siloed in institutional repositories / databases. Privacy-preserving federated querying seems to be an emerging solution for a more seamless knowledge integration / exchange. Is there, perhaps, a smaller step that can be performed towards a richer data sharing experience - e.g., by sharing phenotype-level statistics on top of diagnosed cohorts?

Q: Is the information currently curated in the existing standards enough to quantify, for example, the frequency of phenotypes?

Q: Are there better, privacy-preserving, ways to capture phenotype penetration in a disease population? (Including perhaps a way to retain the patient background if this is relevant for this context.) Are there lessons learned we can adopt from existing non-phenotype focused efforts?

KB

- Thinking about it in the context of zero sided match making
- Both datasets need to have overlapping phenotypes to be able to play in the sandbox
- Mapping them across, will characterize who was sequenced
- Working with database in kansas city have the same phenotypic spectrum between them
- Making diagnosis of novel diagnoses higher
- Characterize what patients are sitting in what datasets

MMT

- NC3 gave us a good example when we tried to integrate across seemingly unconnectable places
- General [encalves…?] provided federated access taking into consideration deidentification of patients

- Allowed something in the vicinity of 80 hospitals to share tons of phenotypic characteristics
- We were able to ask questions we didn't know we could ask

MB

- You can build connected and protected systems only to a certain size
- In cancer communities sometimes it's okay and the scale is big enough
- But if youre area of interest is influence by super rare events or different environmental factor affected by certain behaviors then you have these biases, or locally rare events you won't find if you connect only locally and that's why we have this drive for federated data discovery
- That's why it's important to make data discoverable and explain, for example, the type of phenotype profile that your database you have

Next Steps

- Monthly RD community meeting
    - Use these meetings to identify potential opportunities for collaboration
    - Identify the next few big areas that this work wants to work on and break off into subgroups as appropriate
    - After 6 months we will assess the effectiveness of this meeting format and adjust as necessary
    - Other suggestions? Email neerjah.skantharajah@ga4gh.org