Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.

Global Initiatives in Genomics and Health

Peter Goodhand
CEO • GA4GH
Outline

1. Why do we need standards?
2. GA4GH
3. Global Collaborations
4. Japanese Example
5. Getting Involved
Since data is distributed globally, we need interoperable standards to answer research questions.
Standards are essential

Interoperability is key to data sharing and enabling the necessary ecosystem

Findable, Accessible, Interoperable, and Reusable Datasets

Interoperable tools

Interoperable compute

Interoperable data access methods
What makes a good standard?

- Clear and specific specification
- Meets the needs of everyone in the system
- Easy to adopt
- Works as part of a broader ecosystem
For standards to help... they need to be deployed.

To be deployed... they need to be built with adoption in mind.

To be adopted... we need to check what people need (everywhere).
How do we advance genomics globally?
The 1000 Genomes Project ecosystem

Common consent across all 26 populations and 5 continents

Common resources: Cell lines, arrays, exomes and genomic sequences

File formats (SAM and VCF) and data sharing methods

Software: samtools, FreeBayes, GATK, VCFtools, tabix, etc.

Interoperable
Together, able to address research questions and lay foundation for future projects
To fully realize the benefits of genomics in health, it is necessary to **work globally**

**Collaboration** is the only way to build standards and achieve global data sharing
GA4GH is a global forum
THE GA4GH MISSION...

The **Global Alliance for Genomics and Health** aims to accelerate progress in genomic science and human health by developing standards and framing policy for responsible genomic and health-related data sharing.
The GA4GH Ecosystem

- 90+ Countries
- 20+ Technical Standards
- 40+ Implementations & Deployments
- 650+ Organizational Members
- 450+ Active Participants
- 8 Work Streams
- 9 Regulatory Policies & Frameworks
- 4000+ Subscribers
- 24 Driver Projects

Enabling the Global Learning Health System
<table>
<thead>
<tr>
<th>GA4GH aims to...</th>
<th>GA4GH achieves this by...</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enable <strong>international</strong> data sharing</td>
<td>✓ <strong>Convening</strong> stakeholders</td>
</tr>
<tr>
<td>Promote sharing across the <strong>translational continuum</strong></td>
<td>✓ <strong>Catalyzing</strong> sharing of data</td>
</tr>
<tr>
<td>Encourage technology-enabled <strong>federated approaches</strong></td>
<td>✓ <strong>Creating</strong> harmonized approaches</td>
</tr>
<tr>
<td>Promote <strong>interoperability</strong></td>
<td>✓ <strong>Acting</strong> as a clearinghouse</td>
</tr>
<tr>
<td></td>
<td>✓ <strong>Fostering</strong> innovation</td>
</tr>
<tr>
<td></td>
<td>✓ <strong>Commiting</strong> to responsible data sharing</td>
</tr>
</tbody>
</table>
Use cases: what are we trying to do?

We have many sources of real world scientific and health use cases...and many more
Australian Genomics: Data Management Work Flow and Capabilities

PanelApp Australia
Curated gene-disease associations and virtual gene panels

'Shariant' Platform
Classified variants and curation evidence shared across laboratories

Contribute to international databases

Genomic Data Repository
Phase 1 genomic data store
Phase 2 comprehensive genomic data catalogue

Data Quality Assessment
Quality reports on BAM + VCF data produced by qProfiler & NGScheck software

Gen-Phen Platform
'Variant Atlas'
Genotypes and Phenotypes available for interactive summary-level queries and visualisations

Genotypes

Phenotypes

Data Access and Approvals System
Approvals issued automatically (low-sensitivity) or via data access review (high-sensitivity)

Access to individual-level data
Access to summary-level data

Phenotype Tools and Resources
Phenotype Analytics
Phenotype Queries

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Data Access and Approvals System

Approvals issued automatically (low-sensitivity) or via data access review (high-sensitivity)

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

Australian Genomics: Data Management Work Flow and Capabilities

Clinical Flagships
Acute Care
Milo
Epilepsy
Brain Malf
Leukodys
Kidgen
Hidden
Cardiac
Gen Immun
Int Disab
Int Lung
Neuromusc
ICCon

Clinical Consult (32 clinical sites)

9 Sequencing or Research Laboratories (4 uploading)

Genomic Data
VCF
BAM
FASTQ

Metadata
laboratory, sequencer, library preparation etc

Accredited Pathology Reports

Genotypes and Phenotypes available for interactive summary-level queries and visualisations

CTRL Consent Portal

Biospecimen

Minimal Clinical Datasets

Australian Genomics Study Database

Clinical Phenotypes
Patient phenotype data coded in SNOMED / HPO terms and represented in FHIR format

Flagship Clinical Phenotypes
Patient phenotype data

Upload Genomic Data + Associated Metadata

Data Access Agreements
Data Governance Policies
Data Access Committee

Access to individual-level data
Access to summary-level data

Genome Phenotype (Gen-Phen) Platform

Genetic Disease Genomes

International Genomics Mini-总局

australiangenomics.org.au

ga4gh.org
H3Africa genomic research ecosystem
Current GHIF Members

Foundation members:

- Australian Genomics
- Genomics England

International HundredK+ Cohorts Consortium (IHCC)

Member of Qatar Foundation

Swiss Personalized Health Network

Canada’s Precision Health Partnership

Global Genomic Medicine Collaborative

Danish National Genome Center

Estonian Genome Center University of Tartu
Collaborations

- CINECA: Common Infrastructure for National Cohorts in Europe, Canada, and Africa
- B1MG: Beyond One Million Genomes Project
- elixir: Human Pangenome Project (HPP)
- Science Council World Health Organization
- Public Health Alliance for Genomic Epidemiology
- International 100K Cohort Consortium
- Global Genomic Medicine Collaborative
- International COVID-19 Data Alliance
- HDR UK
- Human Genome Organisation

ga4gh.org
Alignment with Other Standards Organizations

Research

Clinical Research

Clinical Care

ISO
TC215/SC1

CDISC

SNOMED

HL7 FHIR
Recent implementation updates

Watch these implementer talks from 9th Plenary: bit.ly/9thplenary-videos
Example
Melissa Cline & Japan
CIHR COVID-19 Interoperability Initiative
BRCA variation can also lead to ovarian, prostate, and pancreatic cancers
Allele Frequencies

Proof of Concept with BioBank Japan

- **Rationale:** Disease-causing cancer variants are rare by definition
- If a variant is frequent in some continental population, it can be classified as benign with no further evidence

Slide credits: Melissa Cline, UCSC
Federated analysis of BioBank Japan Data

- RIKEN holds a large cohort of cancer patients and controls from BioBank Japan, which they cannot share directly.
- We shared a Docker container with them to analyze their patient cohort for variant co-occurrences and allele frequencies.
- The container generated variant-level data, which we are now using together with the ENIGMA Consortium to interpret BRCA variants!

Slide credits: Melissa Cline, UCSC
Acknowledgements

GA4GH Host Institutions

GA4GH Core Funders

GA4GH Supporting Funders

GA4GH Staff and Contributors
Contact us

Join a Work Stream!
Contact secretariat@ga4gh.org

Become an Organizational Member
ga4gh.org/members

Subscribe to GA4GH Updates
ga4gh.org/subscribe

Join the GA4GH Slack Workspace
ga4gh.slack.com