



Global Alliance for Genomics & Health

Collaborate. Innovate. Accelerate.





Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Opening Remarks

Ewan Birney



Approved Technical Standards

ga4gh.org/genomic-data-toolkit

htsget API V1

Download read data for only the subsections of the genome that you are interested in

refget API V1

Access reference genomic sequences without ambiguity from different databases and servers using a checksum identifier based on the sequence content itself

Data Use Ontology V1

Semantically tag your genomic datasets with usage restrictions, allowing them to become automatically discoverable based on a health, clinical, or biomedical researcher's authorization level or intended use

Beacon API V1

Make your datasets discoverable to the international community by implementing the Beacon API which allows users to determine whether or not a specific allele is present in the dataset

Workflow Execution Service API V1

Run your bioinformatics workflows (defined using CWL or WDL) on multiple platforms, clouds, and environments, with the confidence it will deliver the same results.

File Formats

CRAM V3: genomic data compression
SAM V1/BAM V1: reads data
VCF V4/BCF V2: variant data

Agenda Overview



APRIL 29

- Updates from Work Streams & Driver Projects
- Breakouts by Work Stream and Deliverable
- Work Stream Lead - Driver Project Champion Speed Meetings

APRIL 30

- Breakouts by Work Stream and Deliverable (with breaks!)
- Breakouts Report Back

MAY 1

- Steering Committee Meeting

IP Policy Meeting



Discuss core elements of a draft GA4GH intellectual property policy, which will outline guidelines for collaborating to develop open technical standards

TOMORROW 1:30 PM - 3:00 PM

Chestnut Suite

MEETING CHAIRS

Jorge Contreras, Professor of Law, University of Utah (Co-lead)

Ray Krasinski, Director, Standards, Philips (Co-lead)

Adrian Thorogood, REWS, GA4GH (Coordinator)

Seydina Touré, McGill University (Academic Secretary)

Work Stream / Driver Project Speed Meetings



Series of 20 Minute Meetings

TOMORROW 4:00 PM - 5:40 PM

Rosalind Franklin

WHO

- All Work Stream Leads
- 2 Driver Project Champions (*max 3*)



Tracking the Matrix



Global Alliance
for Genomics & Health

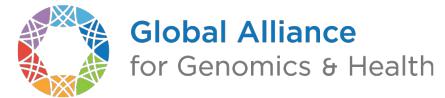
Use comments to update your Driver Project at bit.ly/2019matrix
OR email ashley.hobb@g4gh.org

A	B	C	D	E	F	G	H	I	J
1	<p>A = adopting C = contributing B = contributing & adopting ? = unassigned strong medium weak (white)</p>								
2									
3	Clin/Pheno Phenotype Representation for Genomic Medicine and Research	B?		?				A	
4	Clin/Pheno Clinical Data Exchange (FHIR/Phenopackets)	B		?	?			B	
5	Clin/Pheno Pedigree Representation	C	?	B					
6	Cloud Data Repository Service (DRS) API	C	A		B			A	
7	Cloud Cloud Testbed Interoperability Demonstration	C?	A	?	B				
8	Cloud Workflow Execution Service (WES) API	C?	A	?	B				

Goals

- Identify unmet needs/gaps in genomic data sharing standards.
- Ensure the GA4GH Strategic Roadmap adapts as the field evolves
- Ensure the Strategic Roadmap meets the needs of the new Driver Projects as well as continuing to meet the needs of existing Driver Projects

2020 Strategic Roadmap



Work Streams

Driver Project Consultations

Draft Strategic Roadmap Updates

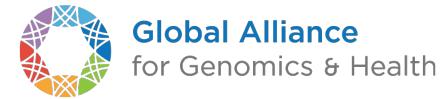
Integrate Findings From Gap Analysis & Finalize Roadmap

Strategic Roadmap Committee

Gap Analysis

May Jun Jul Aug Sep Oct Nov Dec

Attendee Mobile App



- Access the meeting agenda
- Locate breakout rooms
- Connect and communicate with other attendees
- Stay up-to-date on meeting announcements



Once you have downloaded the app, search "GA4GH" to add the event to your dashboard.

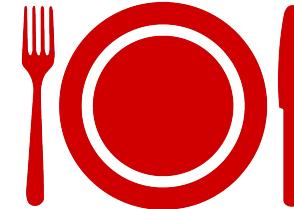
Meals & Transportation



RECEPTION DINNER TONIGHT

Kitchen Garden Bar opens at 18.00

Dinner at the Restaurant from 19.00 - 20.30



TUESDAY EVENING

Bus to Cambridge

Departs Hinxton at 18.00

Returns from Cambridge City Centre at 21.00



Drinks on Campus

Graham Cameron Bar in Hinxton Hall (cash bar)

BUS TO LONDON ON WEDNESDAY

Departs at 13.30 from the Conference Center (reception entrance)

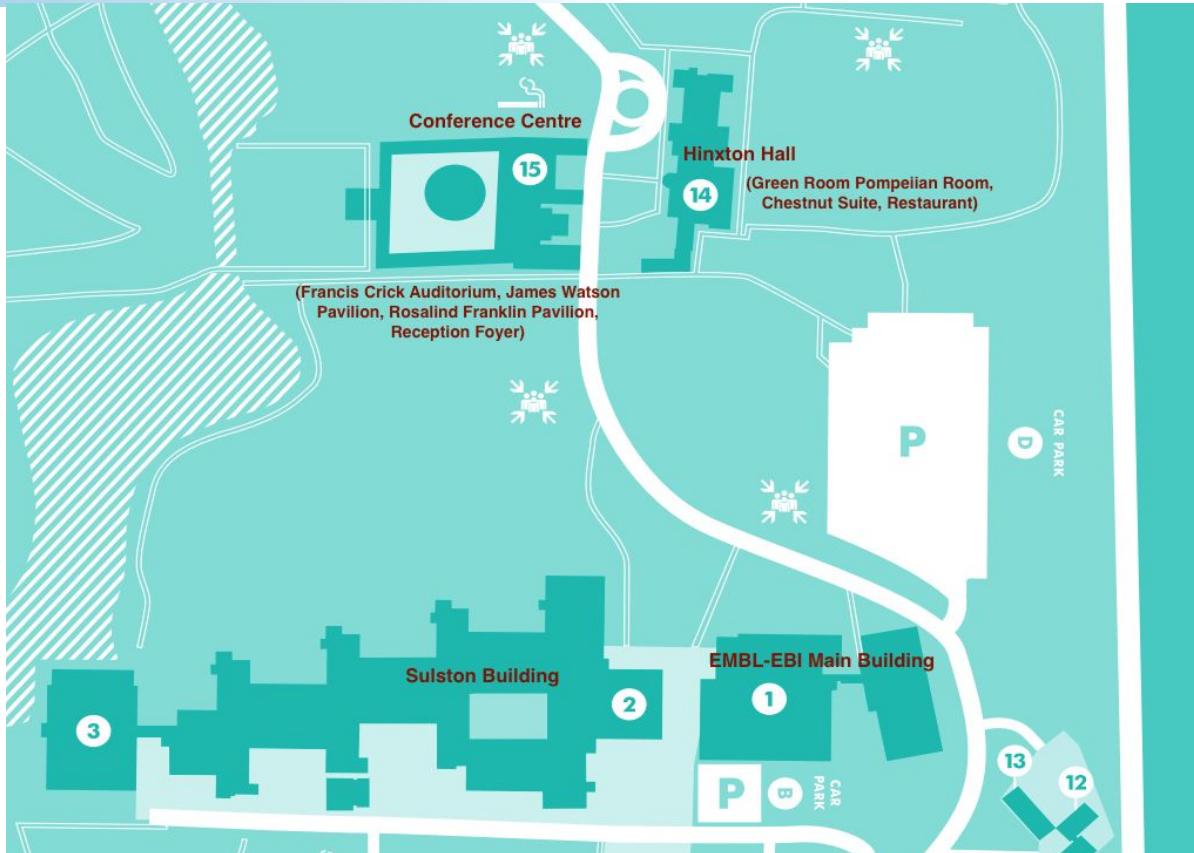
Arrives at Wellcome Collection, 183 Euston Rd, London NW1 2BE at ~15.00

Do you have **dietary
restrictions?**

Talk to the Conference
Centre kitchen staff.



Campus Buildings



For Meetings in:

EMBL-EBI Main Building
Garden Room
Courtyard Room

Sulston Building
Sulston C3-02

meet at reception and walk over together.

Openness & Transparency



All GA4GH meetings and satellite meetings are conducted in an **open** and **transparent** manner to advance interoperable standards and frameworks to enable data sharing.

Decisions made in this meeting will be shared publicly through a link on the GA4GH website.

Expectations for Conduct



GA4GH expects all attendees to help us create an inclusive environment to advance the GA4GH mission.

- Be Respectful
- Be transparent about potential conflicts of interest



Work Stream Updates

Current projects

1. Best practices & Open terminologies

- Delivering an environment for open terminology work - terminology mapping, FHIR server, tooling support - e.g., Redcap

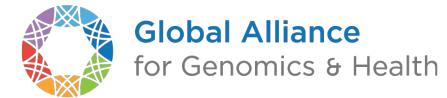
2. Patient-level phenotype exchange

- Phenopackets exchange standard - moving phenotypic info around like genomic info, in partnership with HL7 and ISO/TC215 (Health Informatics) to get data into and out of EHR

3. Family health history capture and interoperability

- Better management of pedigree data and how we use it (long-term goal of propagating info between family members)

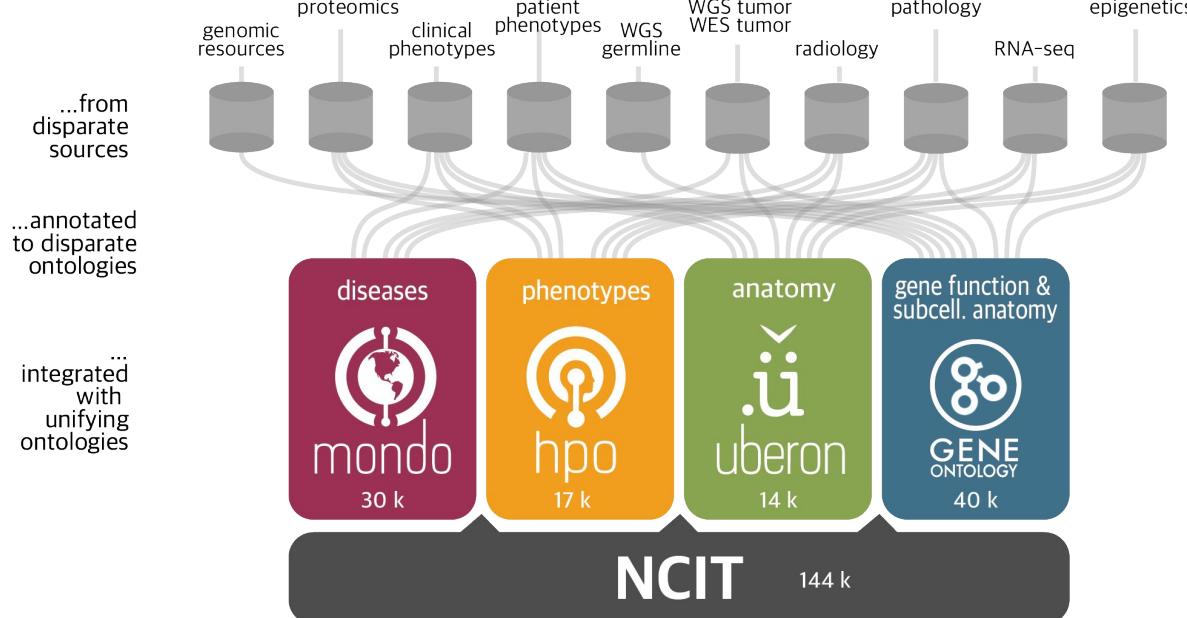
Clinical & Phenotypic Data Capture



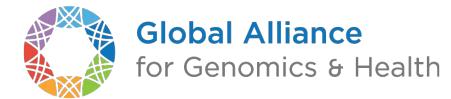
Best practice guidelines for terminology use

Criteria for use in GA4GH

- ✓ Open
- ✓ Good content coverage (cancer, rare, and complex disease)
- ✓ Meets best practices for development (release, documentation, structure, etc.)
- ✓ Interoperable with other standards
- ✓ Relevant for genomic medicine (e.g. variant interpretation)
- ✓ REDCap Ontology support



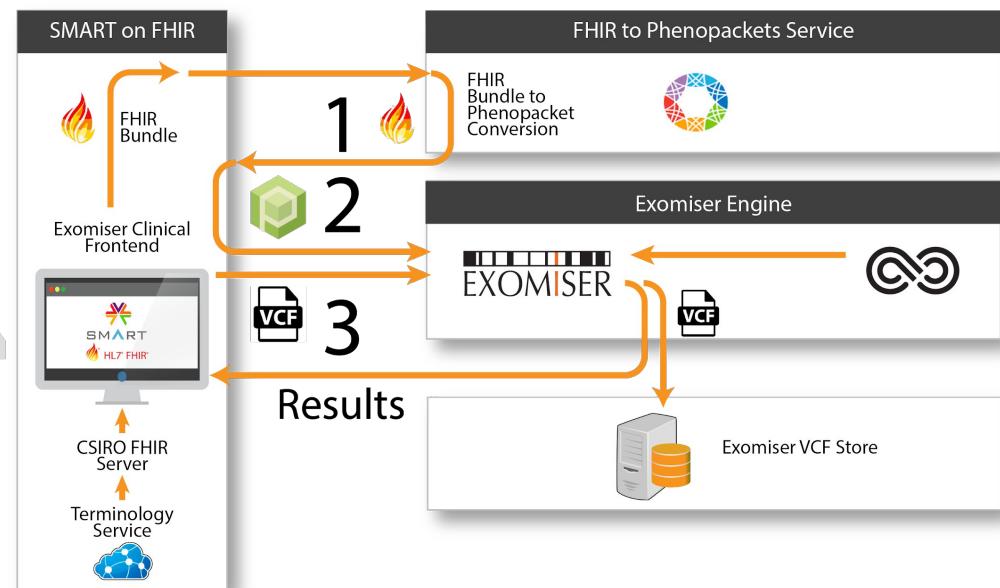
Clinical & Phenotypic Data Capture



Phenotypic data exchange: Phenopackets, an interoperability schema

Criteria for use in GA4GH

- ✓ Provide standardized and modular phenotypic representation for exchange in different contexts and associations (e.g. GKS variant annotation, Discovery, MME, Beacon, GKS annotation, etc.)
- ✓ Interoperability with EHRs via FHIR
- ✓ Terminology agnostic
- ✓



Phenopacket structure



patient:

id: "PROBAND#1"

sex:

id: "PATO:0000384"

label: "male"



who?



Global Alliance
for Genomics & Health

phenotypes:

- type:

id: "HP:0000520"

label: "Proptosis"

what?

severity:

id: "HP:0012828"

label: "Severe"

how?

classOfOnset:

id: "HP: 0003577"

label: "Congenital onset"

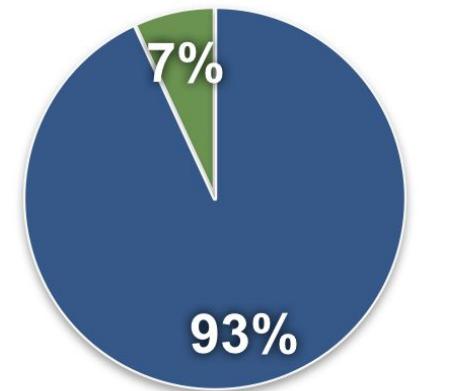
when?

Family Health History: Project goals



- A family health history/pedigree standard for genomic medicine
- Structured, computable information
- Shareable between patients, EHRs, research
- Allows automatic visualization
- Powers ecosystem of CDS tools

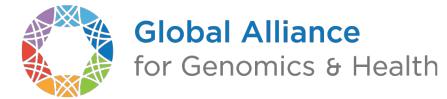
“Do you think family health history clinical decision support would be helpful to clinicians in your country?”



● yes

● no

Cloud Overview

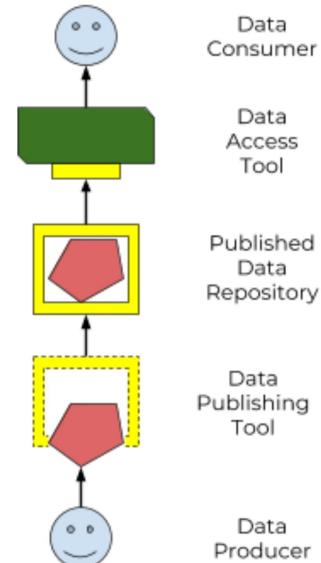


	Champions	Purpose	today → Fall goal
Data Repository Service (DRS)	Brian O'Connor David Glazer	map a logical ID to a physical means of retrieval	v0.1 → v1.0
Workflow Execution Service (WES)	James Eddy	send the algorithm to the data	v1.0 (approved) → v1.1
Tool Repository Service (TRS)	Denis Yuen Susheel Varma	share workflows and tools	2.0.0-beta.2 → v2.0
Task Execution Service (TES)	Angel Pizarro Susheel Varma	run a single job	v0.3 → v1.0

Bonus wish: How might we build an end-to-end story and demo of DRS + WES + Search + DURI?

Purpose: map a logical ID to a physical means of retrieval

- Status
 - released v0.1 ([docs](#)) this week as baseline for discussion
 - aim to have v1.0 approved in the fall
- Goals for this week
 - discussion of rough edges / needed work from v0.1 to v1.0
 - lock in a plan for a testbed
 - get “soft commitments” to implement DRS v1.0 within a year
 - frame the product approval document for the fall
- Feedback desired
 - gaps between v0.1 and something you’re happy approving as v1.0



Purpose: send the algorithm to the data

- Status:
 - v1.0 was approved in Oct 2018
 - we are working on a v1.1 to accommodate DRS
- Goals for this week:
 - review of active endpoints, discussion/design of a WES registry, on-boarding, and compliance testing
 - discuss multitenancy, delegated auth
- Feedback desired:
 - WES interaction with DRS
 - WES interaction with auth

Purpose: share workflows and tools

- Status:
 - 2.0.0-beta.2 is our latest release
 - aim to have v2.0 approved in the fall
- Goals for this week:
 - work on product approval documentation
 - spec out integration with testbed and WES
- Feedback desired:
 - discuss direction of API with Drivers Projects
 - surface requests from GEM Japan and other Driver Projects

Cloud: Task Execution Service (TES)



Purpose: run a single job

- Status:
 - v0.3 is latest release
 - aim to have v1.0 approved in the fall
- Goals for this week:
 - identify TES implementations
 - TES Roadmap
- Feedback desired:
 - gaps between v0.3 and something Drivers are happy approving as v1.0

Data Use & Researcher Identities



DURI builds the Data Use Ontology and Researcher Identity standards to accelerate processes of accessing protected health & genomics data



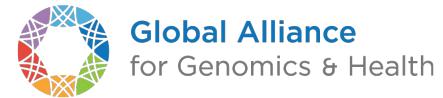
Three main ways of working and feedback request

1. Evolving RI & DUO standard definitions, v1, v2, ...
2. Coordinating leadership connect DURI to the Security, Cloud, Discovery, and driver projects
3. Driver project requirement gathering maximizing understanding and compliance with the existing data access policies and processes of e.g. cohorts to increase standard uptake speed. (drivers from US and Europe and Asia identified)

DURI Future session - Roadmap update

We prepare for the October GA4GH plenary and first half of 2020. RI specification v1 aims for approval in coordination with the Security AAI v1 (*critical*). By October we hope to achieve end-to-end technical implementations to demonstrate benefits of DUO and RI for DACs and Data holders.

Data Use & Researcher Identities



DURI builds the Data Use Ontology and Researcher Identity standards to accelerate processes of accessing protected health & genomics data



Meeting goals

[Researcher Identity specification v1 DRAFT RFC presented](#)

Introduce and discuss the spec v1 with DPs, Cloud, Discovery, Security -> aim to lock for implementations by Google and ELIXIR

Driver Projects are most welcome to attend the RI and DUO sessions please! Technical demonstrations Google in Researcher Identity stream and also a hands on tomorrow

Claims lab deep dives to technology to identify additional claims (for v2).

Connector DUO – RI integration discussion

Discovery



- Ongoing expansion of API for query, response
 - **handover** method for data endpoints w/ format specs. outside Beacon API (e.g. stream VCF, clinical data in authorised environments)
 - moving focus to **clinical use cases**
 - upcoming: more structural variants, **filters** (e.g. phenotype matches), ...
- This meeting:
 - planning Beacon v2
 - +++ access levels, authenticated Beacons
 - “flavours” - engage DP for **Evidence**
Beacon design
 - string it together: example use of Beacon w/ DUO && AAI && handover?



SchemaBlocks {S}[B]

- Cross-WS, cross DP initiative to document & align core data models, objects and conventions used in GA4GH products
- This meeting:
 - present first {S}[B] examples
 - establish procedural principles
 - formalities of engagement with DP & WS?
 - spread the message & get **involvement**
 - sound out the future
 - {S}[B] in GA4GH product approval process?
- FYI please see schemablocks.org:
 - mission statement, testbed examples...

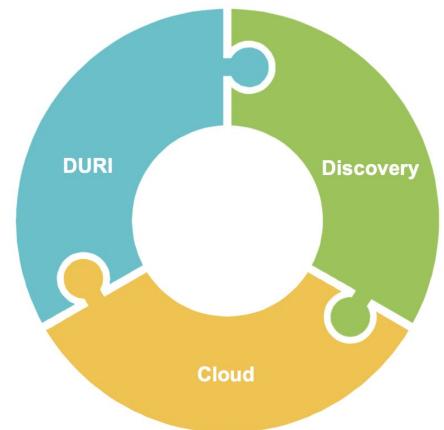


Global Alliance
for Genomics & Health



Search

- A *general purpose framework* for federated multi-modal search across GA4GH (and custom) data models
- **Progress:**
 - Collected use cases and queries from Drivers; implemented “connectors” to adapt diverse datastores of Drivers to a common Query Language
 - One [demonstration](#) illustrates the integration of Discovery and Cloud Work Stream APIs, where the *outputs* of Search are mapped to the *inputs* of Workflow Execution
- **This Meeting:**
 - Review pilots, to inform decision making on the query language (syntax)
 - Independently, discussion of data models (semantics) and their governance
 - Define projects that intersect DURI, Discovery, and Cloud standards, plan their execution



What is GKS?

- Enable the exchange of reference genomic annotation through shared APIs and data models
- Variation is seen as the current key unit of annotation
- Unambiguous representation is key
- Variant representation
 - Larry Babb, Alex Wagner, Reece Hart
 - Describes the low-level atomic types of variation and their model
- Variant annotation
 - Matt Brush and Javi Lopez
 - Builds statements to describe the impact of a variant as described previously

Progress since Basel 2018

- Variant representation
 - Wider focus on somatic variation
 - Developing new tools based on our existing VMC model
 - Pushing data types through VMC
 - <https://tinyurl.com/vr-update-april-2019>
- Variant annotation
 - More landscape analysis from driver projects (VICC, Monarch, BRCA, ClinGen) and annotation modelling
 - Competency questions analysis
 - <https://tinyurl.com/va-update-april-2019>
- We recognise the mutualistic relationship these groups and bridging the gap between them

Planned goals from Hinxton 2019

1. Connecting variant representation to variant annotation
 - Formalising our plans in this area from the document <https://tinyurl.com/var-model-april-2019>
 - Start to enact those plans for further development ahead of October 2018
 - Seeking feedback on these plans to ensure suitability for all Driver Projects (both old and new)
 - Working sessions and hackathons
2. Harmonisation with other work streams within GA4GH
 - Clinical & Phenotypic Data Capture our main target
 - Links into Discovery and Large Scale Genomics
3. Organisational changes
 - Integration of existing GKS GitHub repos back into the GA4GH space

Refget

Andy Yates

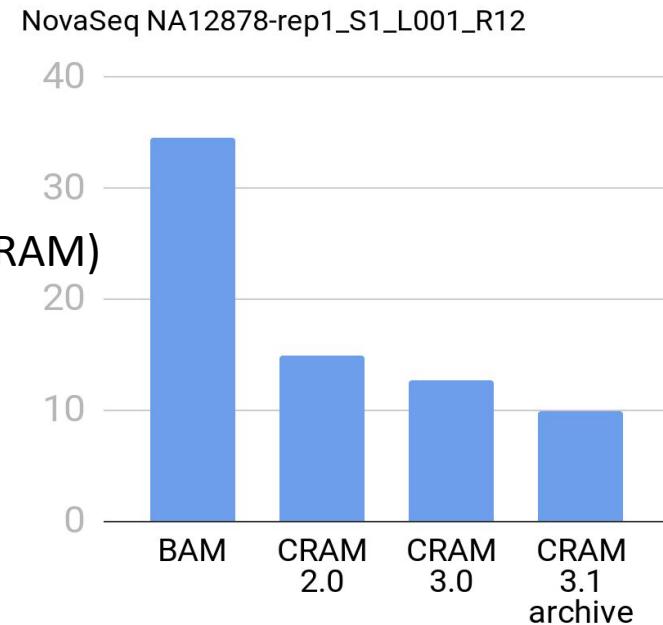
- 2019 Q1: Reference sequences in AWS's Public Data Set Program ✓
- 2019 Q2: AWS implementation
- 2019 Q3: Application note
- 2019 Q4: Draft specifications of next services

Workshop discussions: Alias to checksum, sequence collections and discovery network support (with Discovery WS), serverless implementation

File Formats

James Bonfield, Louis Bergelson

- 2019 Q2-3: Nanopore standardisation (Fast5 <-> CRAM)
 - *Generalisation of any machine-specific data.*
 - ***Initial implementation available for feedback***
- 2019 Q3: CRAM v3.1 codecs for “archive mode”.
- 2019-: Methylation data representation
 - ***Initial proposals - more expert input required***



Future of VCF

Yossi Farjoun, Cristina Gonzalez

- 2019 Q4: VCF successor: Discussion phase
 - Discussing “User stories”
 - Evaluating initial implementations
 - spVCF, Hail, Tachyon
 - **More real data (>50K samples)** would be useful for benchmarking

#1 Non-Superlinear growth	Status	
#2 Improved Phasing for >2 ploidy	Status	
#4 Distinguish No-Data from No-Variation	Status	
Ex	Project	
User Story As an analyst who interprets variants, or the lack thereof, I would like to be able to distinguish between a lack of a variant due to monomorphism, and that due to low coverage so that I would be able to correctly distinguish between no-data and data indicating of a monomorphic site.		
Notes/Comments <ul style="list-style-type: none">• This is currently possible with a gVCF.		

Encrypted Container Format

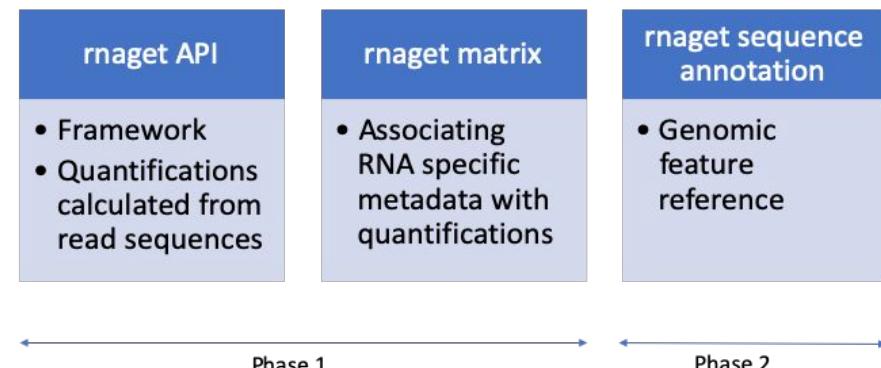
Alexander Senf, Rob Davies

- Crypt4GH initial specifications draft: updated for modern cryptography
- **Discuss** Crypt4GH 1.0 specification and publish
 - Aiming for September steering committee meeting
- Adjusting format for efficient use with GA4GH htsget (**in discussion** for 1.0)
- Crypt4GH interop matrix: Java and python implementations available
- htslib & htsjdk support planned

RNASeq

Roderic Guigo, Barbara Wold, Sean Upchurch

- Product Review Committee review of rnaget API
 - **Discuss PRC feedback; compliance and testing suite**
- Implementations
 - Caltech, CRG (March 2019)
 - CanDIG and IHEC Data Portal
(in development)



Phase 1

Phase 2

htsget

Mike Lin, Jerome Kelleher

- EVA implementations in process
- Protocol support solicited in tenders/RFPs from biobanks & repositories
- Q1/Q2 focus: Documented, open-source building blocks for new servers
- 2019 direction: Integration with FHIR Genomics
- **Explore DRS, Crypt4GH support**
- **How to promote OSS dev without blessing any one implementation?**

Data Security: Main Objectives of the new leadership

- Revision of the security infrastructure and questionnaires, to set-up a more streamlined collaboration with the work streams.
- Increase the awareness and relevance of security-by-design in GA4GH products.
- More targeted and “personalized” security review process, adapted to the products’ scope and scenarios.
- More active role of security in the product design and specification. We welcome questions and more interaction with the active products being developed in each work stream, to ease the assessment and security analysis.



Jean-Pierre Hubaux (JP)
Professor



Juan Troncoso-Pastoriza
Post-doctoral researcher



Mickaël Misbach
Software Engineer



Joao Sa
Software Engineer



Jean-Philippe Bossuat
Software Engineer

Tentative Work plan for 2019

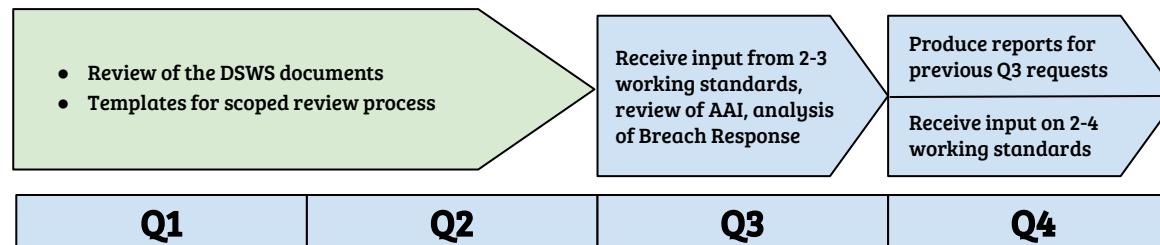
Two phases:

- **1st Phase:**

- Bootstrapping (Q1)
- Revision of the 'Security Technology Infrastructure' document (Q1+Q2)
- Update of the current questionnaires to conform with the new review process (Q1+Q2)

- **2nd Phase:**

- Review the *GA4GH Authentication and Authorization Infrastructure (AAI) and Breach Response* standards
- Security analysis of proposed and submitted products



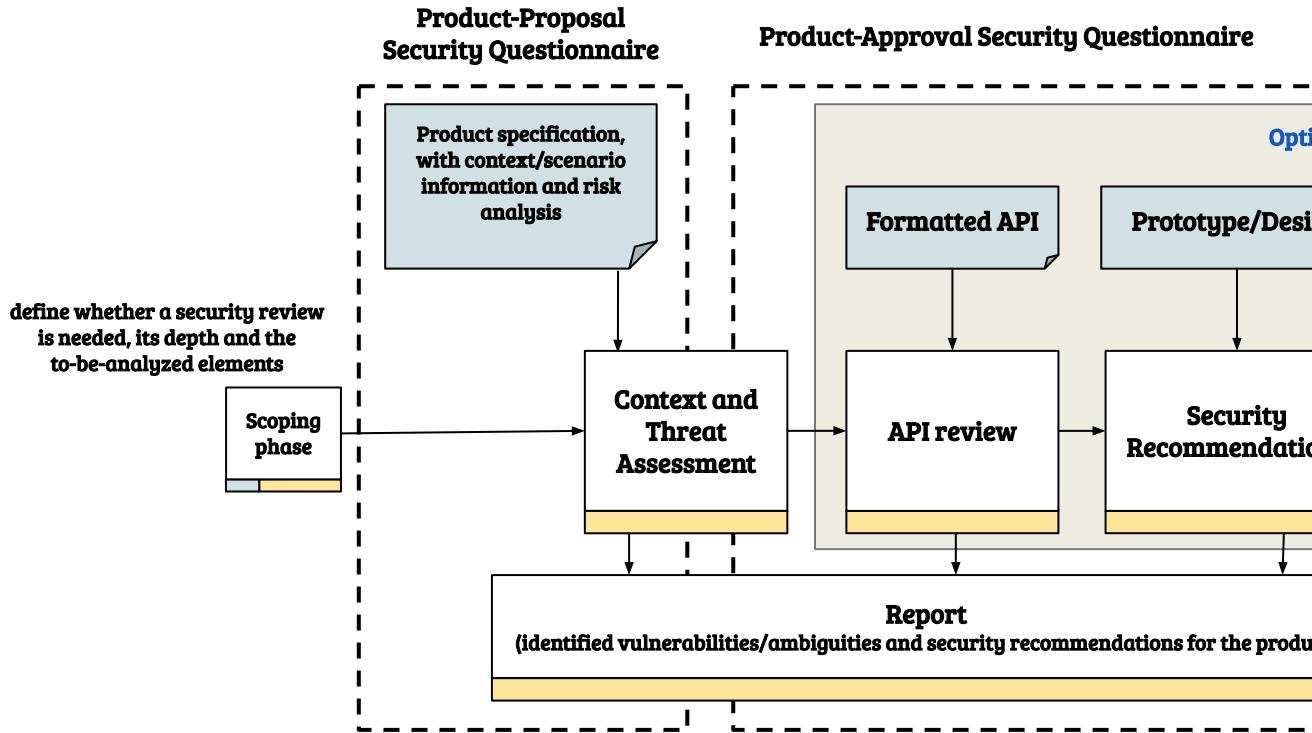
2019

Security Technology Infrastructure

Main changes:

- Updated references to the *GA4GH Authentication and Authorization Infrastructure (AAI)* and the *Breach Response* standards (both yet to be approved).
- Added “Open Web Application Security Project (OWASP) top 10 security risks” to the list of threats to consider during the risk assessment.
- Reviewed the allocation of responsibilities for each stakeholder.
- Added new security primitives to the cryptographic controls:
 - Homomorphic Encryption
 - Secure Multi-party Computation (SMC)
 - Post-Quantum cryptography
- Broadened the scope and increased the generality of the recommendations and guidelines, to avoid binding them to specific cryptographic implementations that might be vulnerable in the future (e.g., specific hash or encryption functions).
- Updated references and added recent attacks to -omics systems.

Review Process



Product-Proposal Security Questionnaire

Main changes:

- Changed its focus to encourage security-by-design
- Clarified the original questions and added new ones aimed at:
 - Defining the scenario/context, including the potential use cases that motivate the product
 - Identifying and characterizing the corresponding agents and stakeholders
- Added a new standardized way to perform the risk analysis based on the 'OWASP Risk Rating Methodology'
 - Estimate risk likelihood (based on threat agent and vulnerability factors)
 - Estimate risk impact (based on technical and business impact factors)
 - Estimate severity
 - Propose countermeasures
- Added a new set of questions to assess the compliance with the AAI and 'Breach Response' standards (clarification of vulnerability and breach reporting)

Product-Approval Security Questionnaire

Main changes:

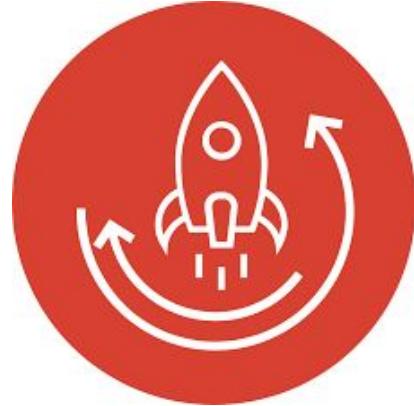
- Simplified and streamlined its contents, aligned with the new proposal questionnaire
- The new set of questions covers 4 main objectives:
 - Validate the initial information provided by the product team (in the proposal questionnaire)
 - Identify and analyse any new or unaddressed risks
 - Detect and remove ambiguities that could lead to vulnerabilities in implementations
 - If applicable, collect useful links to: code repositories, documentation, live demos, etc.

Updates:

- International Survey - public perspectives on data sharing (YOURDNAYOURSAY - Anna Middleton): 35,000+ from 16 countries, incl. Egypt and Iceland. Recruitment done autumn, dataset ready Dec.
- Return of Individual Genomic Results:
 - [Are laws and policies keeping step?](#) [published]
 - [Systematic Review of international views and experiences with RoR](#) - 126 empirical studies of participants, publics, researchers, clinicians, RECs; little on commercial entities. Analysis/ drafting summer.
- Inclusion of Decisionally-Vulnerable Populations in Research and Data-Sharing
 - [Key Implications of Data Sharing in Paediatrics](#) (KIDS) (+ highlighted in paed. cancer [Science editorial](#))
 - [Consent recommendations](#) Dementia Data Sharing (+ commentary in [Lancet Neurology](#))

Goals for Meeting:

- Adapting data sharing to the EU [General Data Protection Regulation](#)
 - The Human Cell Atlas: A Legal Basis for Data Sharing - public interest? consent? already made public?
 - Updates to GA4GH [Privacy and Security, Consent Policies and Templates](#) (by October)
- Intellectual Property and Open Licensing
 - 1) Launch of [IP and Technical Standards](#) (session on Apr 30)
 - 2) Best practices for open licensing of genetic variant databases? See the [ReusableDataProject](#) Paper



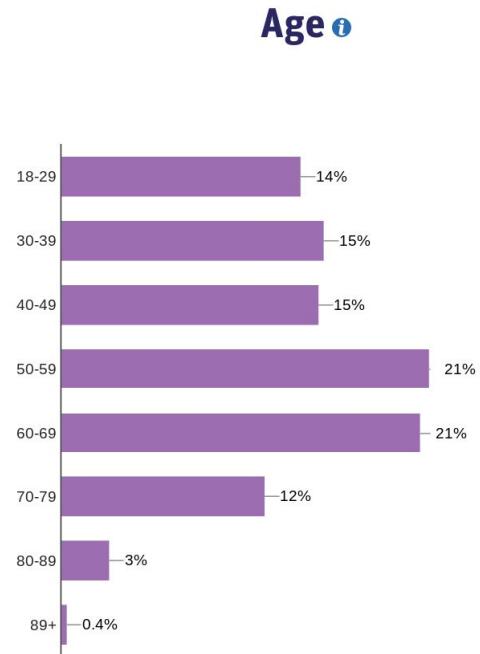
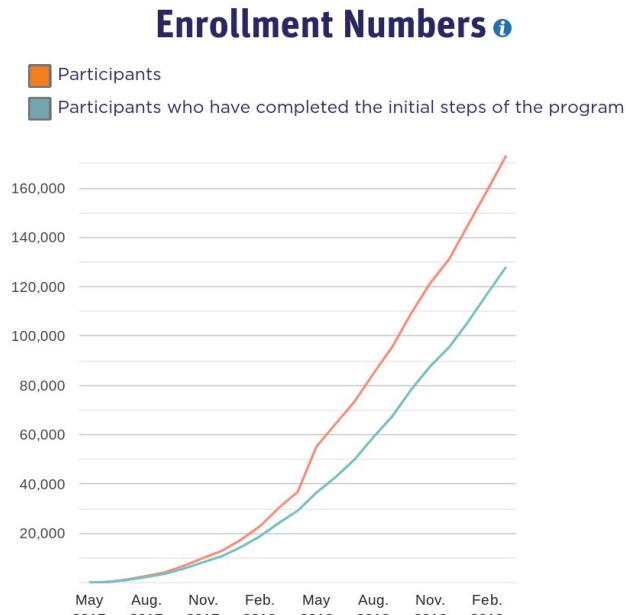
Existing Driver Project Updates

All of Us RESEARCH PROGRAM

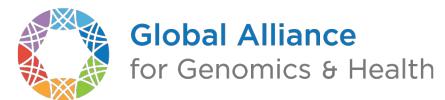
LOCATION: USA

CHAMPIONS: David Glazer,
Anthony Philippakis, (Robert Carroll,
Meg Doerr)

All of Us Research Program



All of Us Research Program

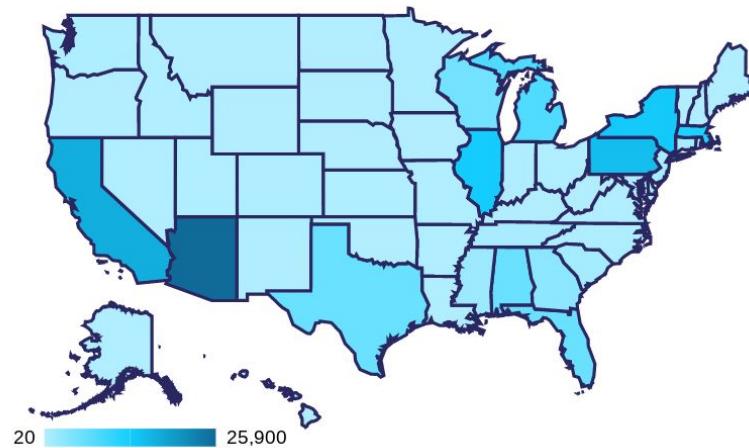


100+

Funded Partner Organizations

300+

Sites Collecting Samples and Measurements



80,000+

Electronic Health Records

132,000+

Biosamples

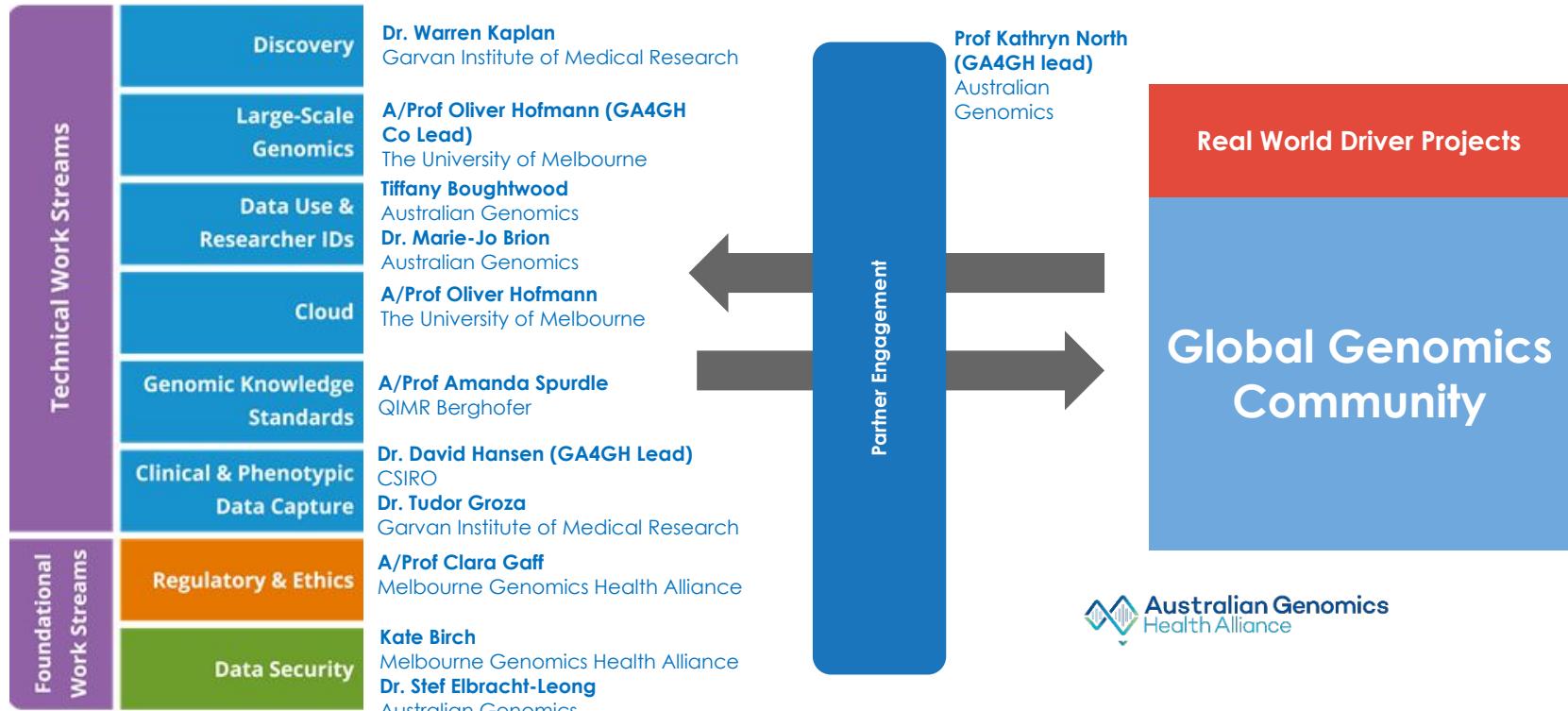
- Timeline
 - May 2017: beta enrollment opens
 - May 2018: public enrollment launch
 - May 2019: public data release (aggregated data) ← next Monday!
 - Winter: first researcher data release (row-level data)
 - 2020: first genomic data release
- GA4GH priorities
 - Cloud WES+DRS integration for genomic data release
 - DURI (depending on access policy evolution)
 - Clin/Pheno integration and interoperability with new data types



LOCATION: Australia

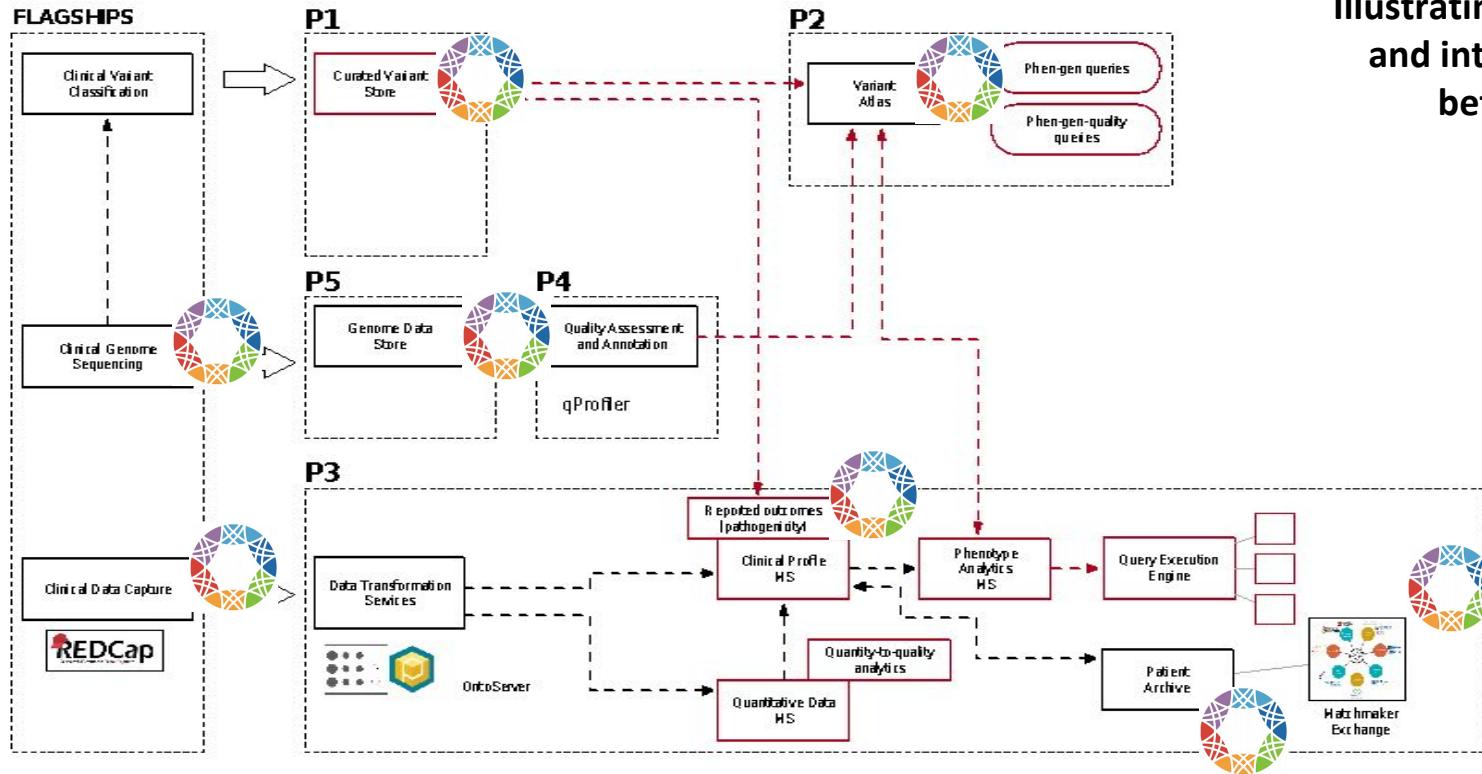
CHAMPIONS: Kathryn North, Clara Gaff, Tiffany Bougwood

Australian Genomics as a 'driver project' of the GA4GH



Data Federation and Analysis

PROGRAM LOGIC



Illustrating the activities
and interrelationships
between our data
projects

Australian Genomics

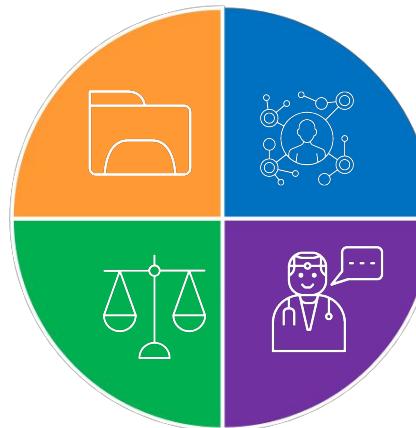
RESOURCES DEVELOPED / STANDARDS IMPLEMENTED

Data

- CRAM
- WES endpoint testing (Cloud)
- Local EGA: Crypt4GH, htsget (Large Scale Genomics)
- Phenotype standards through HPO and SNOMED CT
- Piloting implementation of DUO (ADAM/Consent codes)
- Breach response protocol
- Landscape governing genomic data ownership and sharing (2017)
- FHIR based clinical data models (OntoServer, Shrimp)
- 'Shariant' platform for the sharing of curated variants between Australian diagnostic laboratories, and contributing to international databases (ClinVar)

Evaluation, Policy And Ethics

- Position statements on use of genetic information in insurance (2018)
- Personalised electronic health data capture (2018)
- Understanding public preferences around genomic testing (discrete choice experiments) (2018/2019)
- Catalogue of genomic initiatives with G2MC



Diagnostic Network, Clinical Implementation

- National consent form for genomic testing (2017-2019)
- CTRL 'control' participant platform incorporating dynamic consent (2018/2019)
- **genomicsinfo.org.au** online resources for public and patients –information about genomics, data sharing, testing and support (2018/2019)

Workforce & Education

- Needs analysis of education and training for the genomic workforce (2016).
- 5P health professional education survey for non-genetic health professionals (2018/2019)
- Workforce survey regarding educational needs and implementation readiness for rapid genomic testing in acute paediatrics (2018)
- Program logic for the development of genomic education activities / draft evaluation framework (2018/2019)

National Initiatives Meetings



Value of the network:

- Valuable to compare experiences to avoid "reinventing the wheel"
- Different stages of development. Complementary perspectives.

Opportunities for sharing:

- **Implementation of GA4GH standards and tools.**
- **Beta testing of tools** e.g. Capturing clinical phenotype.
- Minimal **clinical dataset** to interpret genome.
- Flexible **consent** for data sharing.
- **Educational** resources - including approaches to public engagement.
- Experience with **Indigenous populations**.
- Share **reference populations**.
- **Evidence** to support policy and government funding: cost effectiveness, diagnostic efficacy and changes in patient management.
- **White Paper**: State of the Nations

COMMENTARY | VOLUME 104, ISSUE 1, P13-20, JANUARY 03, 2019

Integrating Genomics into Healthcare: A Global Responsibility

Zornitza Stark • Lena Dolman • Teri A. Manolio • ... Peter Goodhand • Ewan Birney • Kathryn N. North

The 8th Plenary Meeting of the Global Alliance for Genomics and Health will be hosted in Melbourne in September 2020, bringing together international stakeholders in genomics and health.





LOCATION: International
CHAMPIONS: Melissa Cline, Amanda Spurdle, Gunnar Rätsch

We just received a strong score on a grant proposal to the NCI ITCR program, reflecting in part the value of GA4GH engagement

Two aims emphasize GA4GH collaboration

Specific Aims

1. Container-based data collection
2. Integrative analysis
3. Dissemination of results

We just received a strong score on a grant proposal to the NCI ITCR program, reflecting in part the value of GA4GH engagement

Two aims emphasize GA4GH collaboration

Specific Aims (*Key Workstreams*)

1. Container-based data collection

REWS (Open data licensing); Cloud Workstream (WES, TRS); Clinical/Phenotype: (phenopackets)

2. Integrative analysis of phenotype from clinical & other data

3. Dissemination of results

We just received a strong score on a grant proposal to the NCI ITCR program, reflecting in part the value of GA4GH engagement

Two aims emphasize GA4GH collaboration

Specific Aims (*Key Workstreams*)

1. Container-based data collection
2. Integrative analysis of phenotype from clinical & other data
3. Dissemination of results

GKS: Variant Representation and Variant Annotation APIs for sharing data with ClinGen, CIViC.



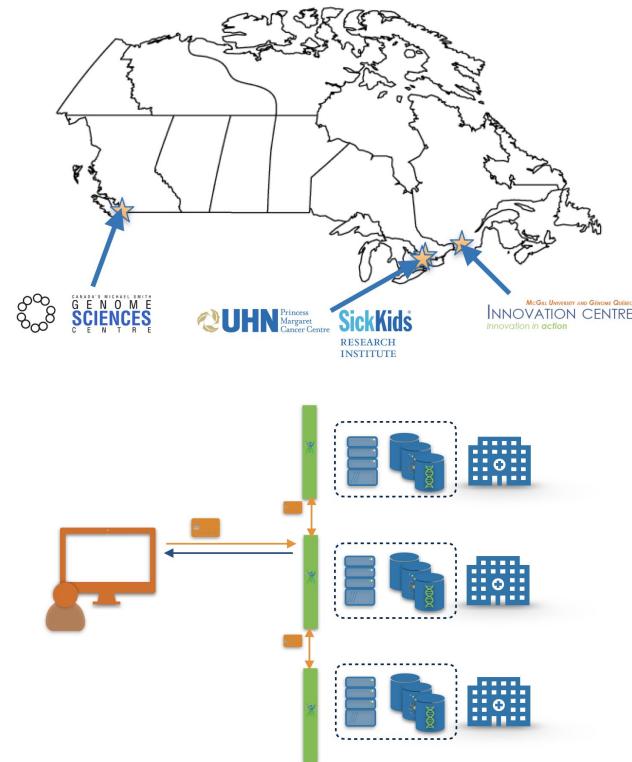
LOCATION: Canada
CHAMPIONS: Mike Brudno,
Steven Jones, Guillaume Bourque,
Jonathan Dursi

CanDIG: What we'd like to talk about this meeting



CanDIG is about **fully decentralized federation**, building complex capabilities by **composing small pieces**.

- **Cloud**
 - WES + DRS: closing the loop
- **Discovery**
 - Search: prototyping w/ PRESTO + GraphQL
 - Networks: lightweight, decentralized federation
- **DURI**
 - Common Claims
 - Alignment w/ OpenID Connect workgroups
 - Lightweight, decentralized AAI
- **LSG/RNA Expression**
 - Implementation
 - Re-use for other features
- **Cross-workstream considerations**
 - Common data models like bioschemas
 - Coordinating better to reduce gaps & overlaps

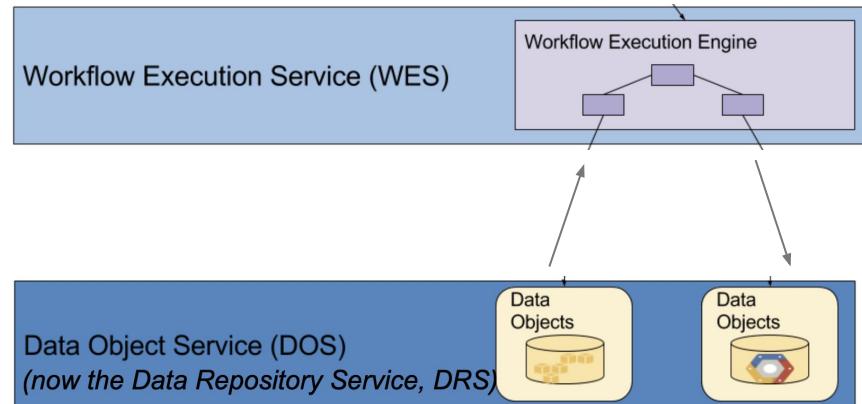


CanDIG: What we'd like to talk about this meeting



CanDIG is about **fully decentralized federation**, building complex capabilities by **composing small pieces**.

- **Cloud**
 - WES + DRS: closing the loop
- **Discovery**
 - Search: prototyping w/ PRESTO + GraphQL
 - Networks: lightweight, decentralized federation
- **DURI**
 - Common Claims
 - Alignment w/ OpenID Connect workgroups
 - Lightweight, decentralized AAI
- **LSG/RNA Expression**
 - Implementation
 - Re-use for other features
- **Cross-workstream considerations**
 - Common data models like bioschemas
 - Coordinating better to reduce gaps & overlaps

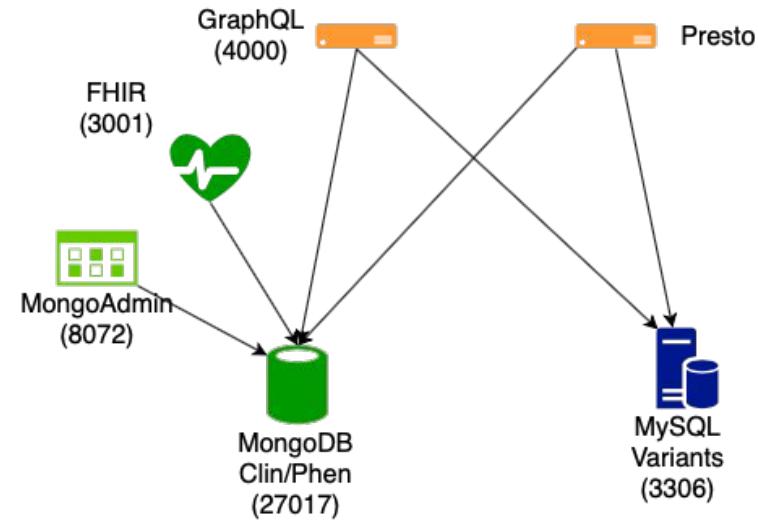


CanDIG: What we'd like to talk about this meeting



CanDIG is about **fully decentralized federation**, building complex capabilities by **composing small pieces**.

- **Cloud**
 - WES + DRS: closing the loop
- **Discovery**
 - Search: prototyping w/ PRESTO + GraphQL
 - Networks: lightweight, decentralized federation
- **DURI**
 - Common Claims
 - Alignment w/ OpenID Connect workgroups
 - Lightweight, decentralized AAI
- **LSG/RNA Expression**
 - Implementation
 - Re-use for other features
- **Cross-workstream considerations**
 - Common data models like bioschemas
 - Coordinating better to reduce gaps & overlaps

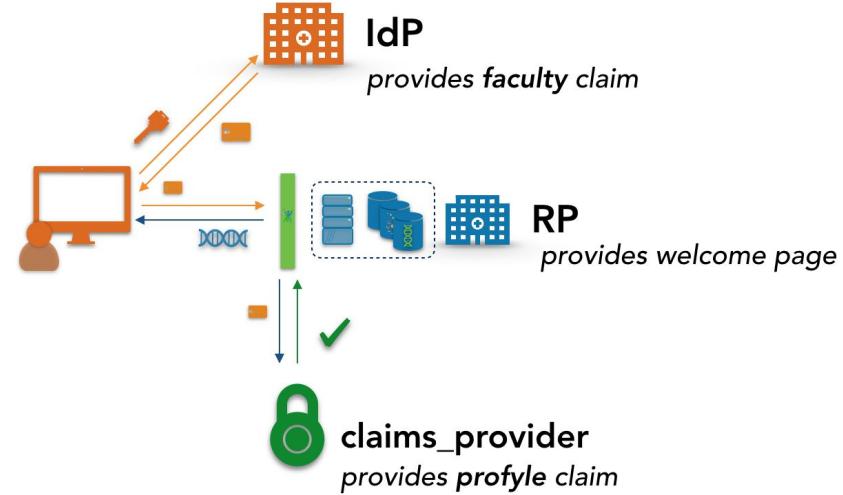


CanDIG: What we'd like to talk about this meeting

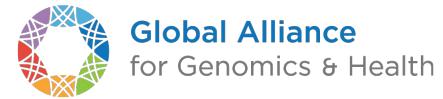


CanDIG is about **fully decentralized federation**, building complex capabilities by **composing small pieces**.

- **Cloud**
 - WES + DRS: closing the loop
- **Discovery**
 - Search: prototyping w/ PRESTO + GraphQL
 - Networks: lightweight, decentralized federation
- **DURI**
 - Common Claims
 - Alignment w/ OpenID Connect workgroups
 - Lightweight, decentralized AAI
- **LSG/RNA Expression**
 - Implementation
 - Re-use for other features
- **Cross-workstream considerations**
 - Common data models like bioschemas
 - Coordinating better to reduce gaps & overlaps



CanDIG: What we'd like to talk about this meeting



CanDIG is about **fully decentralized federation**, building complex capabilities by **composing small pieces**.

- **Cloud**
 - WES + DRS: closing the loop
- **Discovery**
 - Search: prototyping w/ PRESTO + GraphQL
 - Networks: lightweight, decentralized federation
- **DURI**
 - Common Claims
 - Alignment w/ OpenID Connect workgroups
 - Lightweight, decentralized AAI
- **LSG/RNA Expression**
 - Implementation
 - Re-use for other features
- **Cross-workstream considerations**
 - Common data models like bioschemas
 - Coordinating better to reduce gaps & overlaps



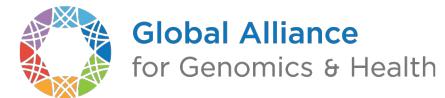
CanDIG: What we'd like to talk about this meeting



CanDIG is about **fully decentralized federation**, building complex capabilities by **composing small pieces**.

- **Cloud**
 - WES + DRS: closing the loop
- **Discovery**
 - Search: prototyping w/ PRESTO + GraphQL
 - Networks: lightweight, decentralized federation
- **DURI**
 - Common Claims
 - Alignment w/ OpenID Connect workgroups
 - Lightweight, decentralized AAI
- **LSG/RNA Expression**
 - Implementation
 - Re-use for other features
- **Cross-workstream considerations**
 - Common data models like Bioschemas
 - Coordinating better to reduce gaps & overlaps

CanDIG: What we'd like to talk about this meeting



CanDIG is about **fully decentralized federation**, building complex capabilities by **composing small pieces**.

We can't be everywhere today and tomorrow - come talk to us!



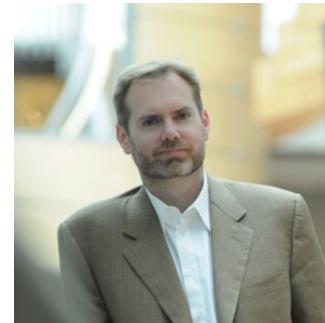
Guillaume Bourque, PI
CanDIG/EpiShare



David Bujold
CanDIG/EpiShare



Jonathan Dursi
CanDIG



Steven Jones, PI
CanDIG



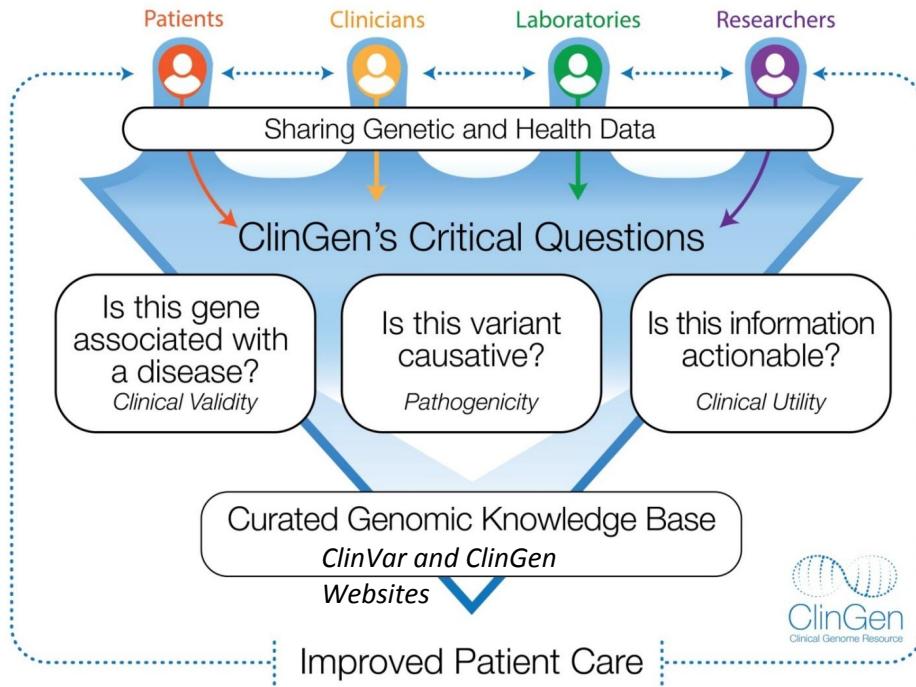
Shaikh Farhan Rashid
CanDIG



LOCATION: USA (funding)
>27 countries participating

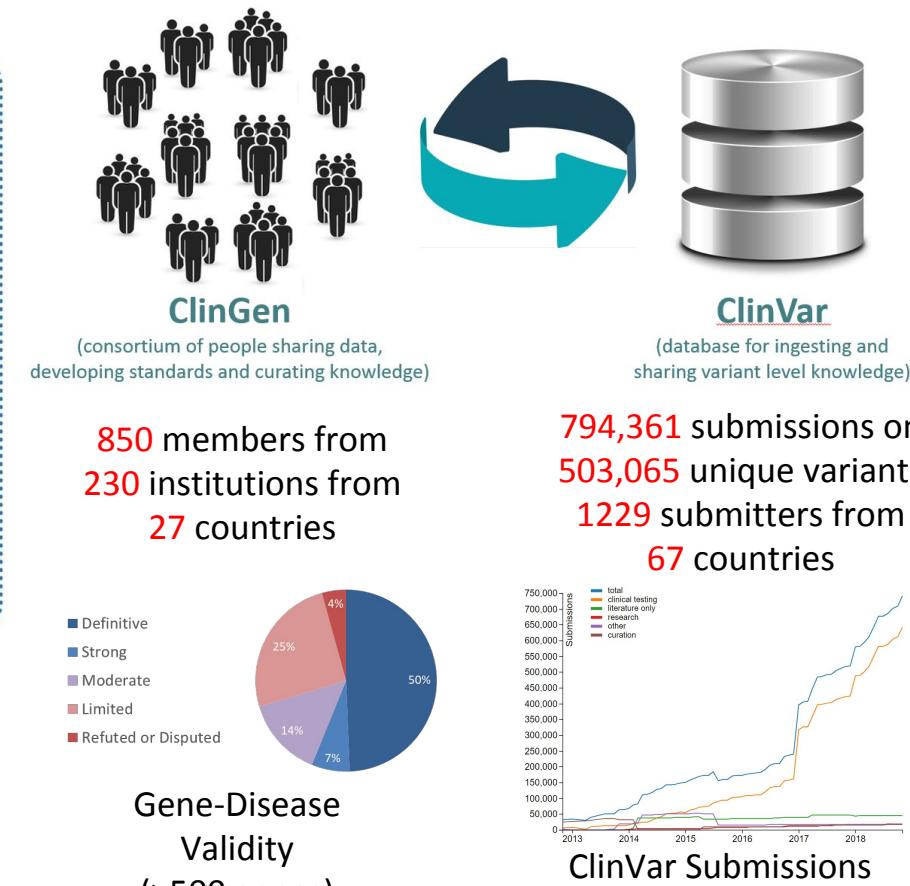
CHAMPION: Heidi Rehm

The Clinical Genome Resource



Purpose: Create an authoritative central resource that defines the clinical relevance of genes and variants for use in precision medicine and research.

www.clinicalgenome.org



ClinGen Goal: Use GA4GH VA & VR Products to represent ClinGen & ClinVar knowledgebases



ClinGen activities supporting this goal:

- Variation Representation
 - Evolving VMC to an updated GA4GH VR schema specification
 - Providing a reference implementation and tools to support translation & registration
 - Working with Baylor to support VR schema in the ClinGen Allele Registry
- Variation Annotation
 - Applying the SEPIO model to all ClinGen knowledge and evidence statements
 - Applying the SEPIO model to ClinVar annotations
- Management
 - Supporting staff and contractors to gather requirements, facilitate meetings, develop work products

ClinGen Members Contributing: Bob Freimuth, Larry Babb, Reece Hart, Tristan Nelson, Chris Bizon, Ronak Patel, Christa Martin, Heidi Rehm

Goal: Harmonize standards for gene-level resources to enable collaboration across projects

Gene Curation Consortium (GenCC)

Orphanet/ORDO

Catalog and ontology – presence in a publication is sufficient for entry

OMIM

Descriptive entries for reported gene-disease implications (requires a minimum level of evidence to enter database, includes “?” entries)

G2P/DECIPHER/TGMI

Database of cases with curation of the evidence for implicated genes

Rahman Gene-Disease Map

Rapid review of gene evidence for highest disease implication - Completed for all genes July 2018, will not be updated

ClinGen Gene-Disease Validity

Expert consensus review of evidence for gene-disease implications

GEL PanelApp

Resource for documenting which genes are valid for use in diagnostic panels

NCBI Genetics Home Reference

Accessible educational resource for genes and disease

Comparison of Terms → Delphi Survey → Harmonized

ClinGen	Terms	GDM	G2P	Orphanet	OMIM	PanelApp
Definitive	Red	Confirmed	Present	Yes	Green	
Strong	Red	Confirmed	Present	Yes	Green	
Moderate	Grey-Red	Probable	Absent	Yes	Amber	
Limited	Grey-Blue / Grey-Red?	Possible	Candidate	?Disease	Red	
No evidence	Blue	Absent	Absent	No disease claim	Red	
Disputed	Grey-Blue	Absent	Candidate	?Disease	Red/Amber	
Refuted	Blue	Absent	Absent (Suppressed)	Reclassified-VUS	Red	

→

GenCC
Definitive
Strong
Moderate
Limited
Animal Model Only
No Known Disease Relationship
Disputed Evidence
Refuted Evidence



LOCATION: Europe

CHAMPIONS: Jordi Rambla; Juha Tornroos; Gary Saunders

- Beacon specification:
 - Goal: Extending Beacon to leverage their potential in data discovery
 - Fact: Exploring a Beacon with Proteomics content and a Beacon for the Plant community
- Beacon Network goals:
 - Allowing to organize the Beacons above into coherent thematic networks
 - Allowing for private networks, e.g. in healthcare related consortia
- Deployment goals:
 - Helping ELIXIR Nodes to deploy Beacons, e.g. as a service they can offer internally
 - Providing a Reference Implementation (including web interface/s)
 - Helping deployment of Beacon Network/s across ELIXIR
- Needs:
 - Schemas, schemas and more schemas (or agreeing in “recommended” ontologies)
 - Common schemas for Variant description and annotation
 - Common mechanisms/schemas to describe phenotypes
 - Common schemas to describe artifacts and operations (e.g. “download”, “file format”, “charts”) for the handover functionality
 - Authentication and (flexible) Authorization commons





EUROPEAN
GENOME-PHENOME
ARCHIVE



 ENA
European Nucleotide Archive

LOCATION: Europe

CHAMPIONS: Thomas Keane,
Jordi Rambla, Dylan Spalding,
Cristina González García

Current (in production)

- DUO for EGA submissions and discovery (DURI)
- Htsget for secure streaming reads/variants from EGA/EVA (LSG)
- Refget: CRAM reference registry at ENA (LSG)

Future (pilot in 2019)

- Researcher IDs: ELIXIR/EGA/Sanger collaboration (DURI)
- Crypt4GH: Finalise standard v1.0, interoperability, demonstrate for EGA archiving/distribution (LSG)
- Refget
 - Cloud deployment of CRAM reference registry (LSG/GKS)
 - Sequence alias/accession/checksum resolution service (LSG/GKS)
- Phenopackets: Phenotype submission to EGA (C+P)



LOCATION: UK

CHAMPIONS: Augusto Rendon,
Peter Counter

Current

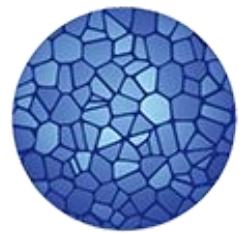
- Genomic Medicine Service (test directory, test ordering system, consent services, WGS sequencing and analysis service, genomic data stores, research environment, ...)
- HTSGET implementation (<https://gitlab.com/genomicsengland/htsget/gel-htsget>)
- CRAM readiness
- GKS (annotation)

In the backlog (adoption)

- DURI products
- Clouds products

Special interests

- Gene curation network (<https://panelapp.genomicsengland.co.uk/> and <https://github.com/genomicsengland/panelapp>)
- Clinical Beacon / Search API
- Variant aggregation



**HUMAN
CELL
ATLAS**

LOCATION: International
CHAMPIONS: Laura Clarke,
Timothy Tickle

Cloud

- Continue to provide use cases and design feedback for DRS and WES, using DRS in handoff to Terra

Data Security

- HCA Data Coordination Platform adopting GA4GH DSWS recommendations for AAI and Breach Response

Large Scale Genomics

- Continued use of standard formats
- Supporting RNASeq group on Matrix Format and API use cases

Data Use and Researcher Identity

- Using GA4GH policies/standards as investigating supporting managed access data

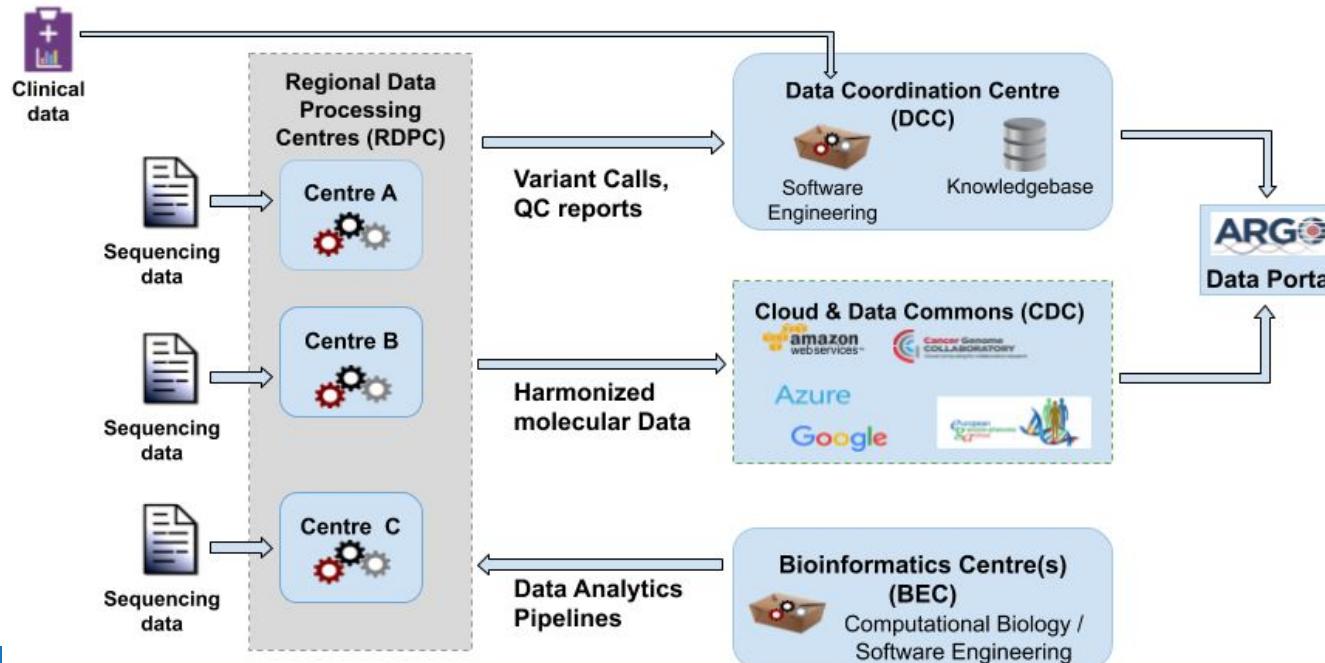
Regulatory and Ethics Workstream

- CGP at McGill is leading the HCA Ethics Working Group (EWG) and using existing GA4GH policies to support creation of HCA ethics toolkit



LOCATION: International
CHAMPIONS: Christina Yung,
Lincoln Stein

- Project goal: catalog the genomic profiles of 100,000 donors with rich clinical data across multiple cancer types.
- 19 Expression of Interests. Development of the ICGC-ARGO Platform has ramped up.



ICGC-ARGO: Priorities with GA4GH



Adoption of GA4GH products: htsget, refget, CRAM, Beacon

Cloud

- WES implementation for regional data processing centres
- Continue to contribute to DRS

DURI

- Adoption of DUO by ICGC Data Compliance Office (DACO)

Search API

- Arranger - our implementation of a search API with GraphQL querying against Elasticsearch
- Sharing our use cases and implementation with Discovery workstream

Clinical & Phenotype / GSK variant representation + annotation / AAI

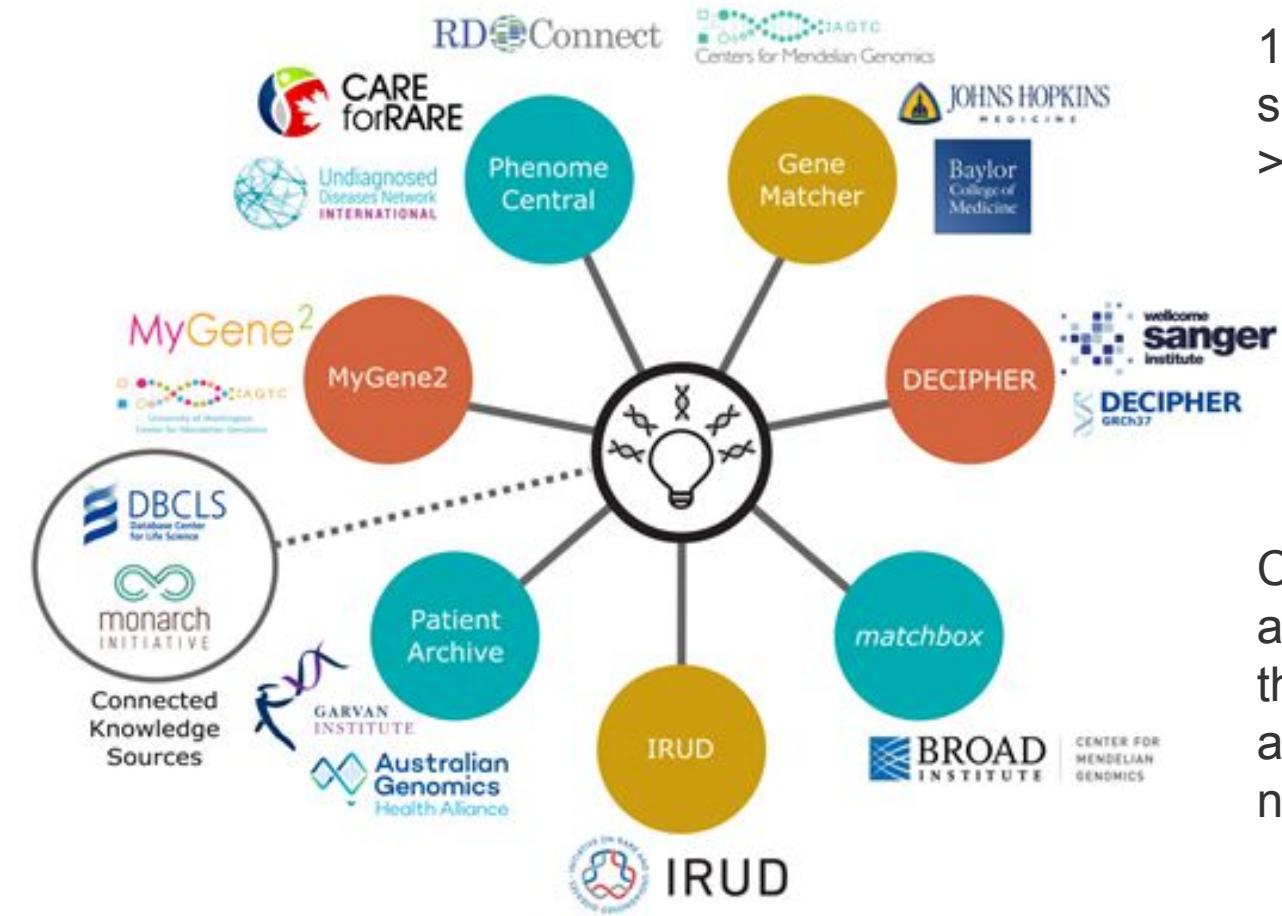
- Will get more involved



Matchmaker Exchange

LOCATION: International

CHAMPIONS: Ada Hamosh, Heidi Rehm



130,000 rare disease cases
submitted from 84 countries
>10,000 unique genes

Current efforts with GA4GH are focused on working with the Discovery WS to define and eventually implement the new Search API

VariantMatcher

VariantMatcher is a database open to search on genomic locations. It harbors genomic data as part of the BHCMG.

Email :

Password :

VariantMatcher (VM) created by:

- Nara Sobreira
- François Schiettecatte
- Ada Hamosh
- BHCMG Center for Mendelian Genomics

Your search included the following features:

Hypotonia, Microcephaly, Global Developmental delay, Esotropia

A submission match notification, for **your search: '6:34004293:T>C'**, was sent to the following:

BHXXXX - Patient - Affected - 6:34004293:T>C

Salmo Raskin - genetika@genetika.com.br - PUC Brazil

Bilateral Cleft

BHXXXX - Patient - Affected - 6:34004293:T>C

Hamza Aziz - haziz2@jhmi.edu - JHU

Bicuspid Aortic valve, Aneurysm, ascending aortic

BHXXXX - Patient - Affected - 6:34004293:T>C

Samantha Penney - penney@bcm.edu - Baylor College of Medicine

Encephalopathy, Ataxia, Hypotonia

BHXXXX - Patient - Affected - 6:34004293:T>C

Samantha Penney - penney@bcm.edu - Baylor College of Medicine

Ataxia, Spasticity, adult onset spinocerebellar ataxia

BHXXXX - Mother - Unaffected - 6:34004293:T>C

Filippo Vairo - fvairo@hcpa.edu.br - Hospital de Clinicas de Porto Alegre

BHXXXX - Father - 6:34004293:T>C

Daryl Scott - dscott@bcm.edu - Baylor College of Medicine

BHXXXX - Mother - 6:34004293:T>C

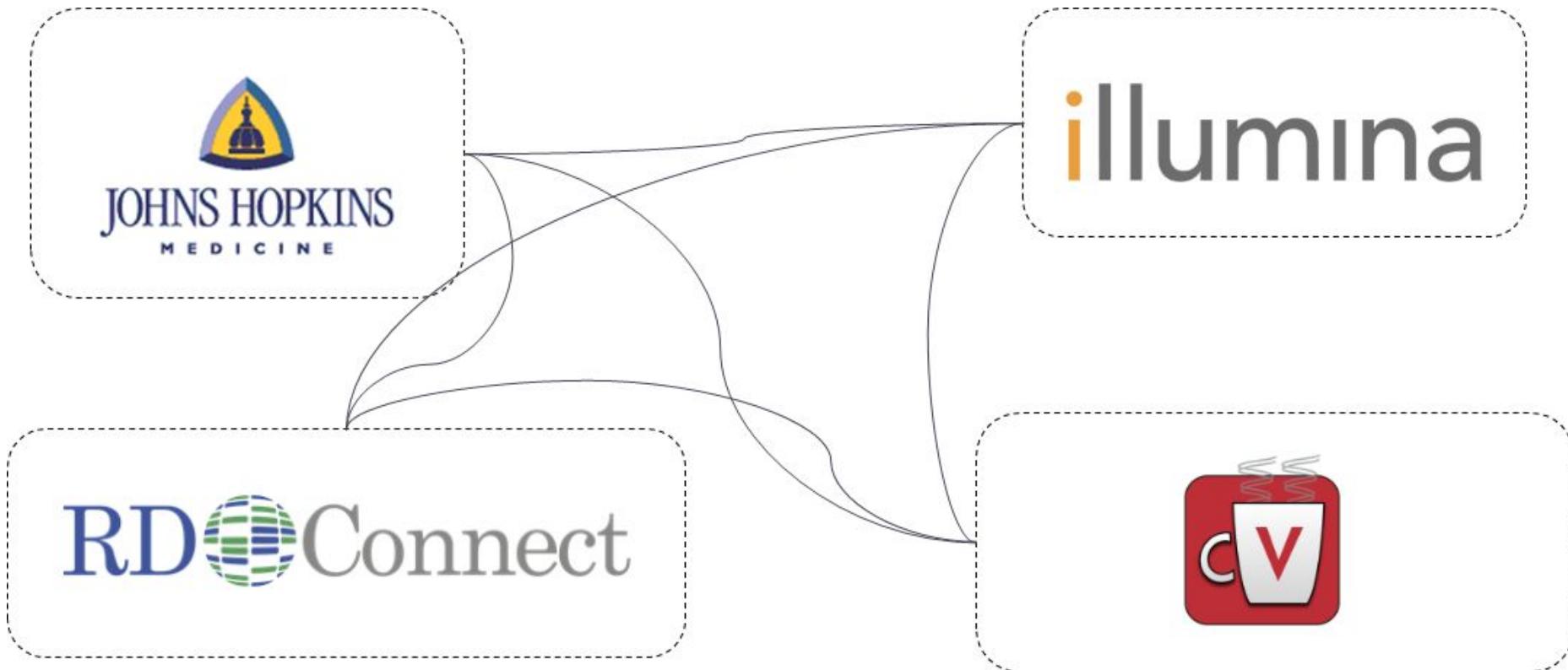
Samantha Penney - penney@bcm.edu - Baylor College of Medicine

BHXXXX - Father - 6:34004293:T>C

Samantha Penney - penney@bcm.edu - Baylor College of Medicine

Please do not reply to this email, it was sent from an unattended email address; however, you can email us at variantmatcher@jhmi.edu or use the [contact form](#).

Federated Network for Variant Matching – Test Stage



Test Version of “MME Variant”

[VariantMatcher](#) Home About Contact Us Editor Admin Help

Search Discovery Search Submission List

Discovery Search

Variants : 5:118860953 T>C 5:118792051 C>T 17:42929130 G>A

Servers : Café Variome Illumina PhenoDB Search Staging Int. RD-Connect

Response : Exists Contact Phenotype Subject Variant Gene

Require match on all variants



Discovery Work Stream Search API

Not yet live!

Nara Sobreira
François Schiettecatte
Ada Hamosh

Anthony Brookes

GRCh37 chr5 118860953 Chr env Tg.2016000
GRCh37 chr5 118792051 Chr env Ig.C16000
GRCh37 chr17 42929130 Chr env Ig.2016000

PHENOTYPE

(D) HP:0000985

Add Remove

HP:0000110
HP:000158
HP:000160
HP:000175
HP:000179
HP:000202
HP:000204
HP:000215
HP:000218
HP:000230

NCBI36 GRCh37

CV

(A OR B) AND C AND D More information

Reset Build Query

Source VariantMatcher test data Counts 1

Kyle Farh

illumina Case Log

lchr14:30095714:C:T (PRKD1)

Search

Phenotype	Mutations observed	Mutations expected	P-value (Poisson exact test)
Developmental delay	4 de novo mutations	0.038	$P < 8 \times 10^{-10}$
Congenital heart disease	3 de novo mutations	0.009	$P < 3 \times 10^{-6}$



LOCATION: International
CHAMPIONS: Melissa Haendel,
Monica Munoz-Torres, Peter Robinson,
Tudor Groza

A bit different than most driver projects in that we are about **DATA HARMONIZATION** rather than about producing data (most similar to VICC)

Monarch provides: standards & ontologies, best practices, and semantic tools

- **Human Phenotype Ontology**; contributions to other GA4GH recommended standards e.g., NCI Thesaurus, LOINC, ...
- **Disease-to-phenotype** data models and annotations
- **Phenotype profile matching** and variant prioritization algorithms
- **Exomiser** variant analysis tool, used especially for **rare and Mendelian disease** diagnostics
- **SEPIO** ontology and modeling framework for evidence and provenance information

Main products we've contributed to within GA4GH to date:

- **Phenopackets** (Haendel co-leads Clinical & Phenotypic Data Capture WS)
- **Variant Annotation** standard (Brush co-leads in Genome Knowledge Standards WS)
- **VICC manuscript** and variant alignment
- **DUO ontology**, data licensing

What we'd like from GA4GH at THIS MEETING:

- Finalize **integration** between Variant Annotation and Phenopackets and **release Phenopackets 1.0!!!**
- Better understanding of **semantic data representation and integration needs** from driver projects
- Coordinate & advance **data & knowledge licensing** standards w/ Regulatory and Ethics ([ReusableDataProject](#))



LOCATION: USA

CHAMPIONS: Robert Grossman,
Zachary Flamig, Michael Fitzsimons

Areas of collaboration:

- Cloud workstream
 - DOS/DRS
 - WES/TES/TRS
- DURI workstream
- Large Scale Genomics workstream

Goals for meeting:

- Understand where DCFS can help to augment current APIs and standards

Expected next steps:

- DRS implementation and testing
- Providing DCFS use cases and proposing new APIs to the community



National Heart, Lung,
and Blood Institute

LOCATION: USA

CHAMPIONS: Albert Vernon Smith,
Alastair Thomson, Goncalo Abecasis

Areas of collaboration and interest

- Cloud
- Large Scale Genomics
- Data Security
- Data Use & Researcher Identities
- Regulatory & Ethics
- Clinical & Phenotypic Data Capture

Priorities

- Standards and implementation protocols for genomic data access and analysis across continents
- Standards to cope with evergrowing WGS and other omics datasets
- Standards relevant to the capture of molecular phenotypes from WGS and other omics data

Goals for meeting

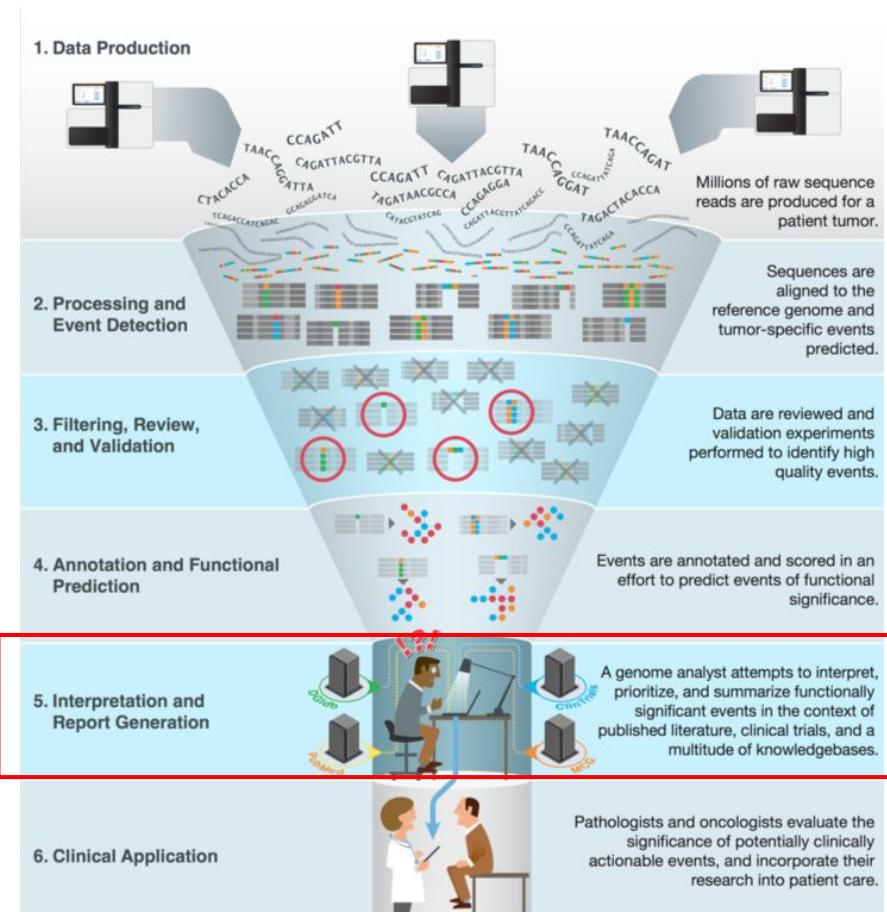
- Understand where TOPMed can contribute and interact in areas of interest



Variant Interpretation
for Cancer Consortium

LOCATION: International
CHAMPIONS: Alex Wagner, Obi Griffith, David Tamborero, Malachi Griffith

Addressing the Interpretation Bottleneck

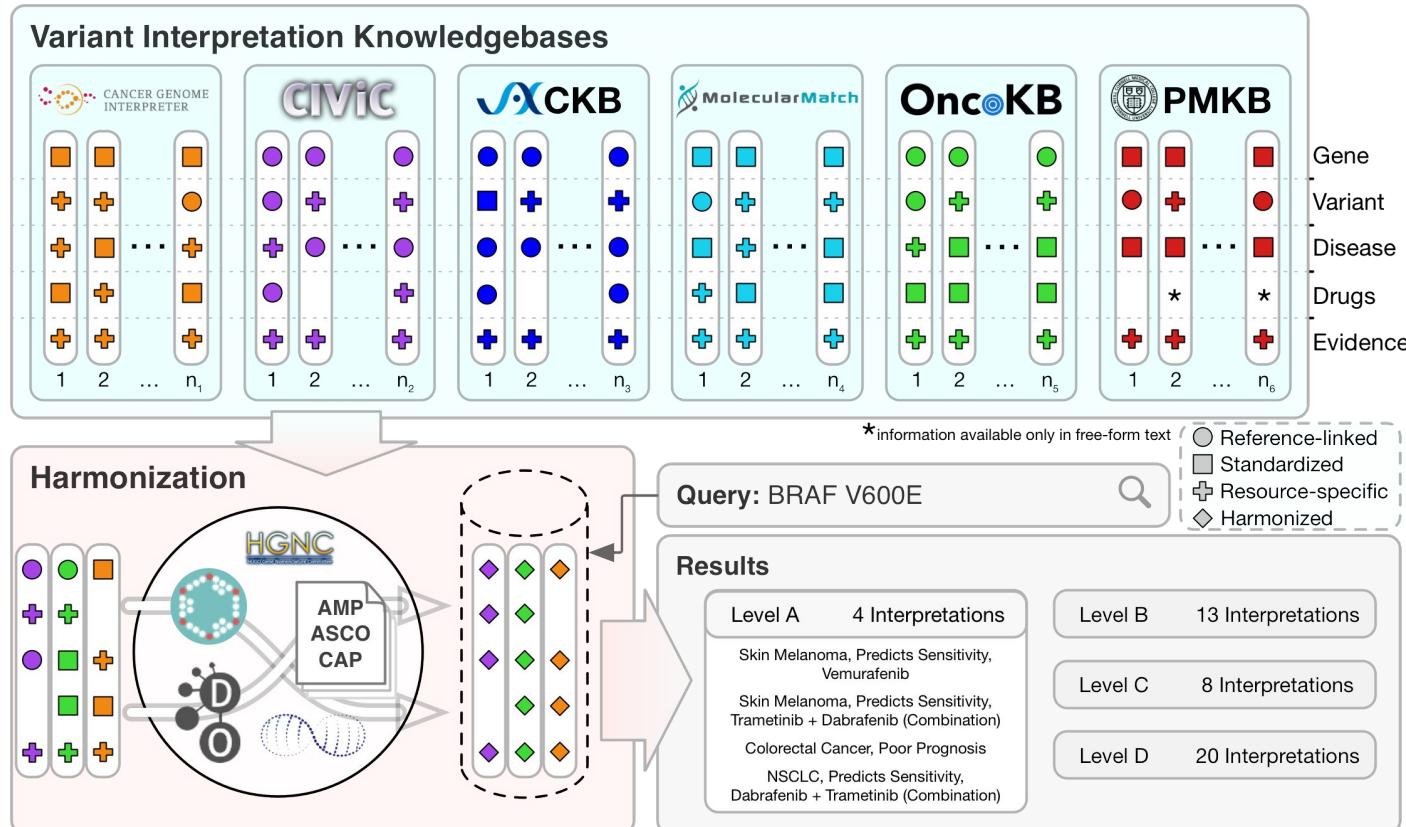


Good BM, Ainscough BJ, McMichael JF, Su AI†, Griffith OL†. 2014. Genome Biology. 15(8):438.

An International Consortium from GA4GH

- Year Started: 2016 (AACR GA4GH Meeting)
- Global: Leadership from USA, Canada, UK, Spain, Germany
- Mission
 - Global integration of knowledgebases for clinical interpretation of cancer variants
- Clinical Focus
 - Ultimate Goal: expert curated interpretations integrated into clinical reports

A Harmonized Meta-Knowledgebase



Wagner AH, et al. bioRxiv 2018

VICC Goal: Genomic Knowledge Standards

VICC activities supporting this goal:

- Variant Representation
 - Updating the VR model to contain rule-based variations for somatic knowledge representation
 - Updating VMC to create computable representations of these
- Variant Annotation
 - Informing Therapeutic Response, Diagnostic, Prognostic, Pathogenicity Annotation types

VICC Goal: Data Usage

VICC activities supporting this goal:

- Licensing:
 - Contributions to the Reusable Data Project (Monarch Initiative)
- Data Use Ontology
 - Examining Contributions for Knowledge Licensing to DUO

VICC Goal: Discovery

- Evidence beacon project
 - Focus on using VICC as a harmonized data source

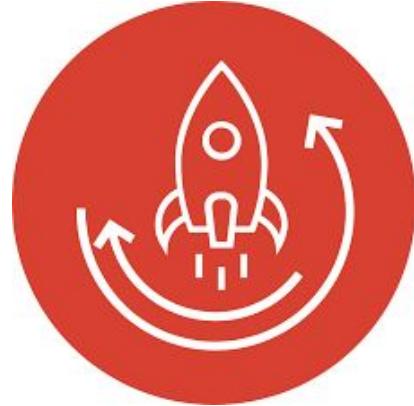


Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

BREAK





New Driver Project Introductions



LOCATION: International
CHAMPIONS: Dean M. Hartley,
Thomas Frazier

New 2019 Driver Project

Autism Sharing Initiative - Autism Speaks



Autism Advocacy Organization

- Sponsors autism research
- Conducts awareness - WAM
- Outreach activities
 - families
 - governments
 - public
- Based primarily in US and Canada
- Public Health programs world-wide

Impact Statement:

To enhance lives today and accelerating a spectrum of solutions for tomorrow

Autism Sharing Initiative



Global Alliance
for Genomics & Health



Dean Hartley



verily



**Cloud-Based Genetics Database to
Understand Molecular Basis of Autism**



Stephen Scherer



ASI - to enabling research world-wide



**179 Investigators
58 Institutions
16 Countries**

ASI - open science to understand autism



Global Alliance
for Genomics & Health



verily



DNASTACK

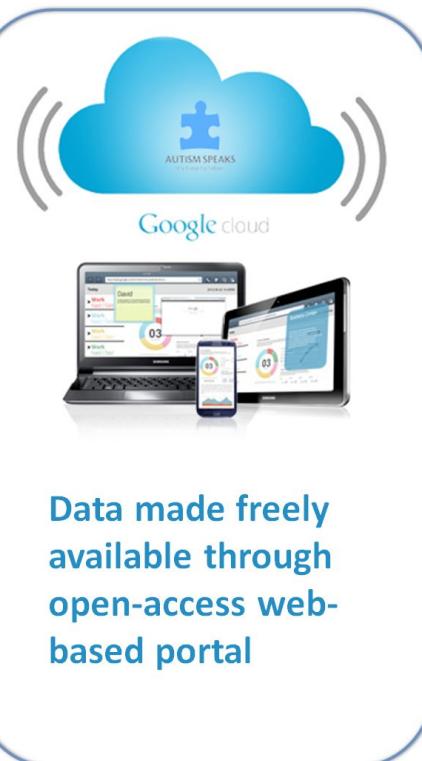
Google™

Private-Public
partnership to improve
understanding and
treatment of autism

2019 reached 11,100+
whole genome
sequences with
phenotype data from
families with autism



Data made freely
available through
open-access web-
based portal



ASI Work Streams – Current Focus

Regulatory & Ethics

- Trust (to participate)/ Risk (to execute)

Clinical & Phenotypic Data Capture

- Deep phenotyping to identify autism subgroups

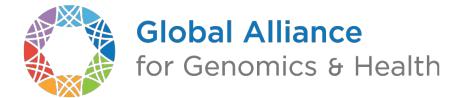
Data Use & Researcher Identities (DURI)

- Federated credentials to expedite sharing

Discovery

- Registered access to Beacon and Search

Autism Sharing Initiative - Summary



Organization & Driver Project – Autism Speaks/Autism Sharing Initiative

Vision – Sharing autism data across platforms

Mission/Goals – looking for consensus, responsible sharing, and reduce risk

Current Status – launched ASI website & membership agreement, working an identity concentrator, initiated conversation with autism cohorts

GA4GH Plans/Ideas Goal – Build-out MSSNG platform into a federated network to seamlessly share genetic and phenotypic data to understand Autism

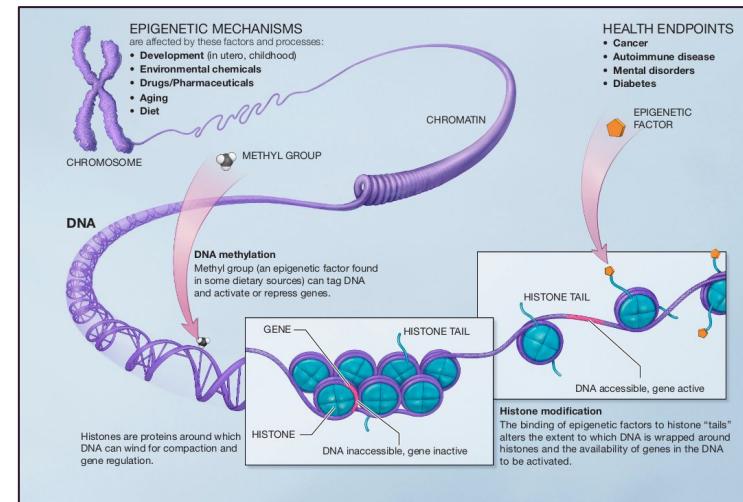


LOCATION: International
CHAMPIONS: Yann Joly, Mike Cherry,
Guillaume Bourque

New 2019 Driver Project

What is epigenomics?

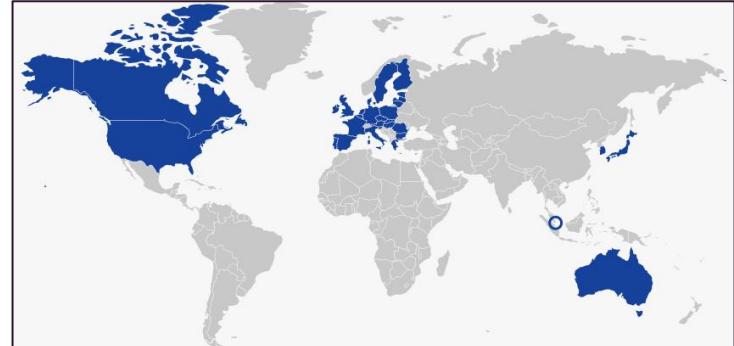
- Study of epigenetic modifications on genetic material of cells
 - Reversible modifications on cell DNA or histones
 - Affect gene expression without altering DNA sequence
 - Partly inherited, partly imputable to environment



<http://commonfund.nih.gov/epigenomics/figure>

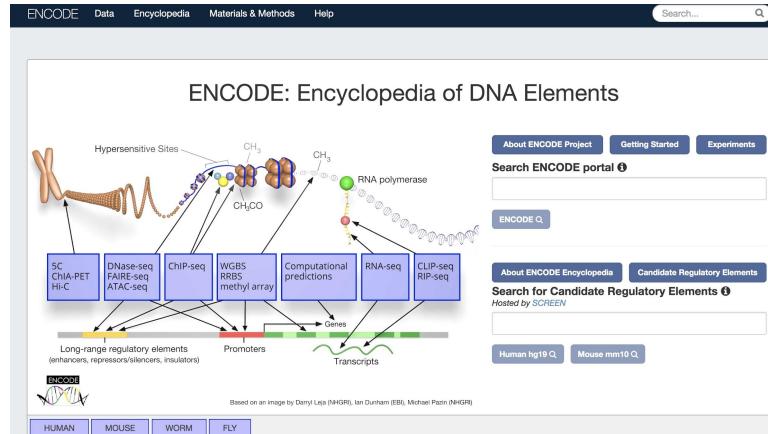
What is ENCODE/IHEC?

- Goal: Provide reference epigenomes for a variety of normal and disease tissues
- International epigenomic datasets production effort with several funding agencies
 - Workgroups develop standards & toolboxes (assays, data ecosystem, integrative analysis, ethics...)
- All IHEC standards and tools are freely available online
 - <https://github.com/IHEC>



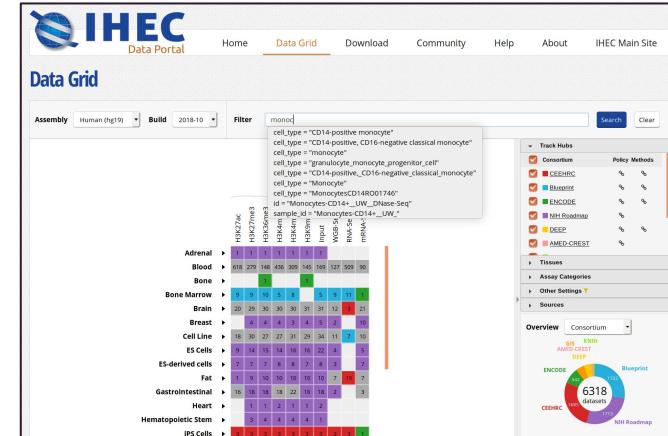
ENCODE and IHEC Data Portals

- Goal: Making epigenomics data discoverable and easily obtainable.



The ENCODE Encyclopedia of DNA Elements homepage features a central diagram illustrating the relationship between genomic features and gene expression. At the top, a ribbon-like DNA strand is shown with 'Hypersensitive Sites' indicated by arrows. Below the DNA, various experimental techniques are represented by blue boxes: SC ChIP-seq, DNase-seq, FAIRE-seq, ATAC-seq, ChIP-seq, WGBS, RRBS methyl array, Computational predictions, RNA-seq, CLIP-seq, and RIP-seq. Arrows from these boxes point to 'Promoters' and 'Genes', which then lead to 'Transcripts'. A green bar at the bottom represents the 'Genome'. A legend at the bottom left defines the color coding for HUMAN, MOUSE, WORM, and FLY. The top navigation bar includes links for ENCODE, Data, Encyclopedia, Materials & Methods, Help, and a search bar.

- ✓ Metadata
- ✓ Processed data
- ✓ Raw data



- ✓ Metadata
- ✓ Processed data
- ✗ Raw data



Challenges

There are multiple challenges bound to using sensitive metadata and raw datasets produced by this decentralized effort, even before getting to the bulk of the analysis!

- Obtaining (meta)data access
 - Application to a Data Access Committee (DAC)
- Downloading
 - Getting the data from controlled access repositories
- Comparing datasets across subgroups
 - Metadata is often hard to collate
- Analysing the data
 - Heavy use of resources (space, memory, processing power)
 - Standardization of codes
- GDPR compliance inside / outside of Europe

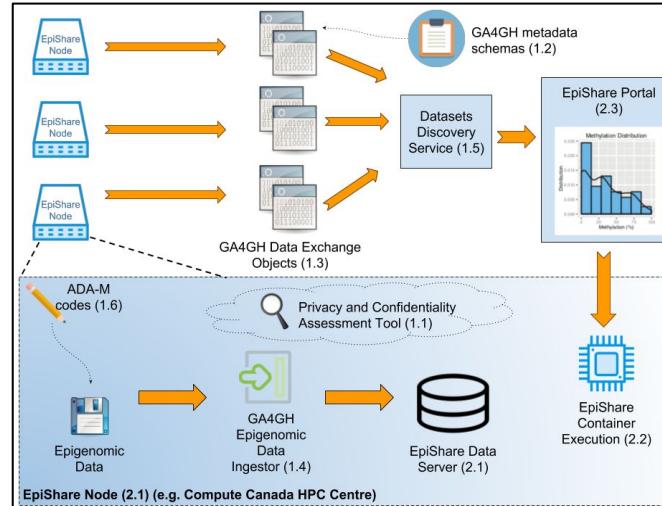
EpiShare

- Goal: Extend GA4GH APIs, standards and tools for epigenomic data
- Also a Genome Canada-funded project (2018-2021)
- Website: <http://epishare-project.org/>
- Who?
 - IHEC: Guillaume Bourque, Yann Joly, David Bujold, Roxanne Caron et al.
 - ENCODE: Mike Cherry, Barbara Wold, et al.



(Meta)data sharing / processing

- Platform enabling launch of multi-omics analyses on controlled-access datasets at their storage location
 - Goal: **Implement resources and a portal to make epigenomic data more easily discoverable, accessible and analyzable**



EpiShare - Ethics and policies

- ELSI – Literature, law & policy review
 - Privacy and sensitivity issues specific to epigenomic data
 - Consensus position and best practices for data sharing in genetic and omics research
 - Key privacy legislations (GDPR in particular)
 - International standards on data sharing and protection
 - Security issues specific to cloud computing
 - Best practices in IT security
 - Goal: **Mapping the privacy and security obligations**



EpiShare - Ethics and policies

- ELSI – Concrete tools
 - Development of a privacy and confidentiality assessment tool
 - Questionnaires that researchers can answer to help them navigate confidentiality and security obligations
 - Updates of the tool as needed
 - Contribution to ADA-M and DUO
 - Gather information from IHEC projects and other volunteers from the epigenetic community for including in ADA-M
 - **Goal: Harmonizing and streamlining data access agreements**

Conclusion

- IHEC has adopted different strategies to make epigenomic data more FAIR
- Still, making restricted access data truly available to the research community has been challenging
- New emerging standards within GA4GH will be adopted to further improve the ways IHEC and ENCODE share genomic data
- EpiShare = Extend GA4GH to epigenomics data
- In which workstream should we discuss experimental meta-data?

Thank you!



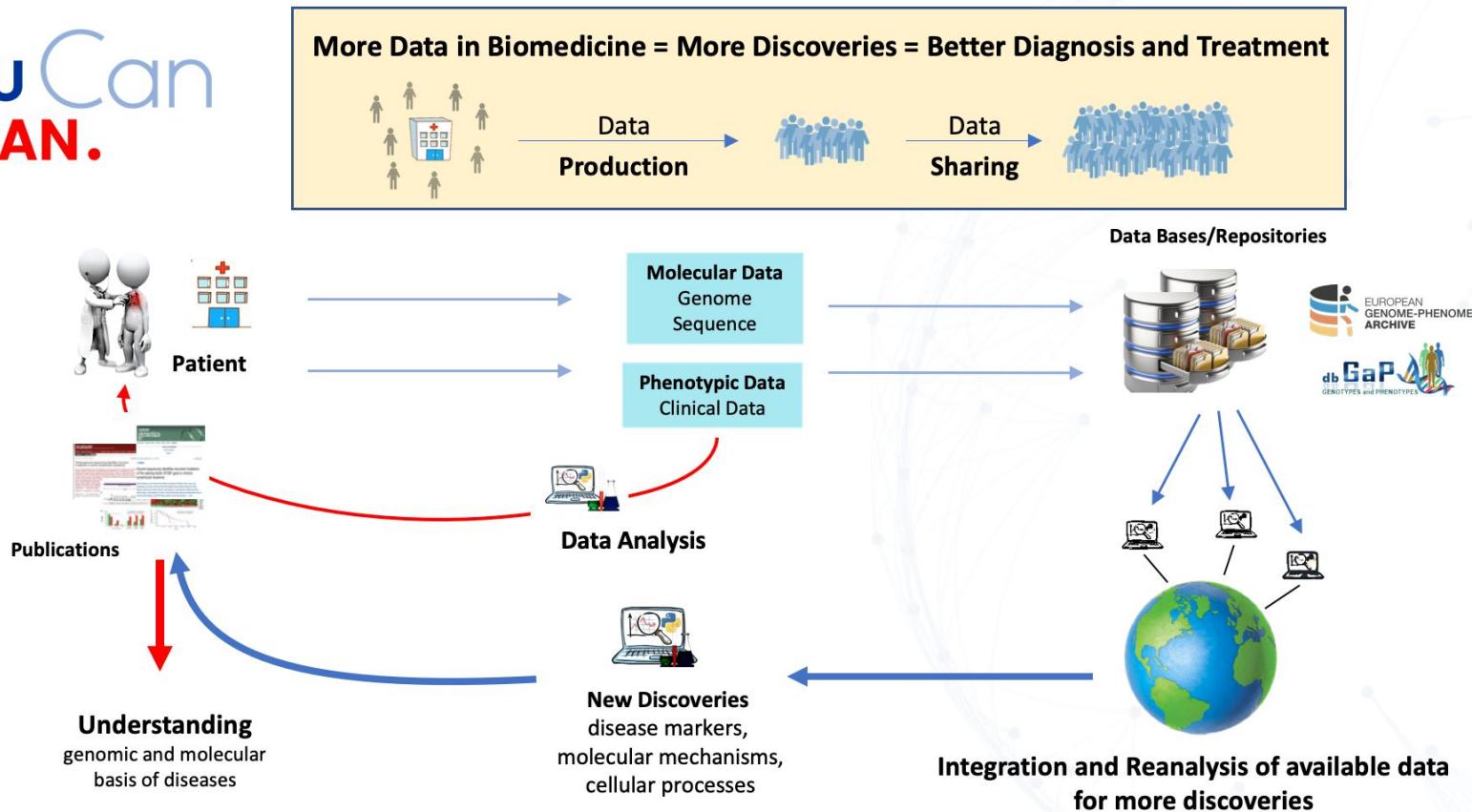


LOCATION: Europe-Canada

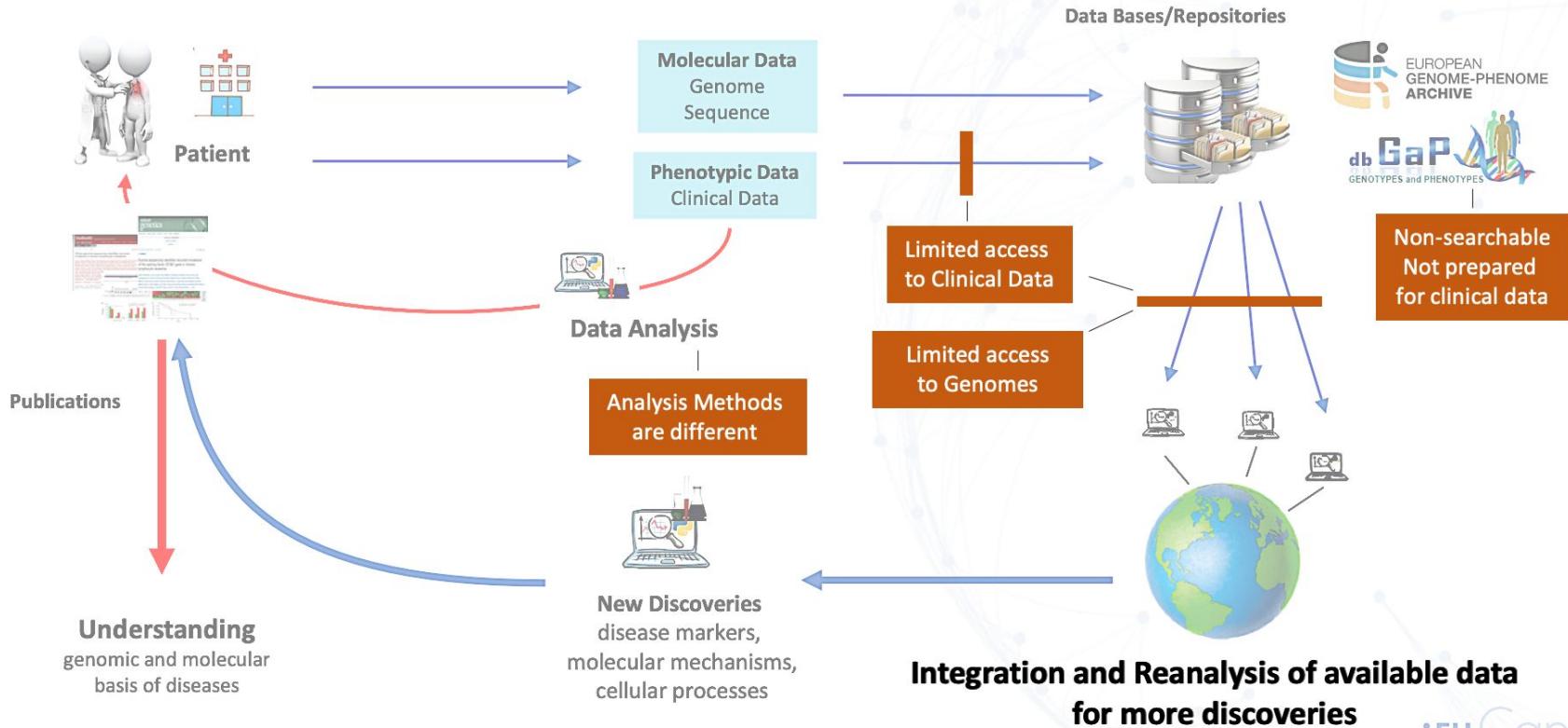
CHAMPIONS: David Torrents,
Alfonso Valencia

New 2019 Driver Project

The Rationale



Current Limitations for data sharing are Ethico-Legal, Technical/Methodological and Cultural



• EU Can
CAN.

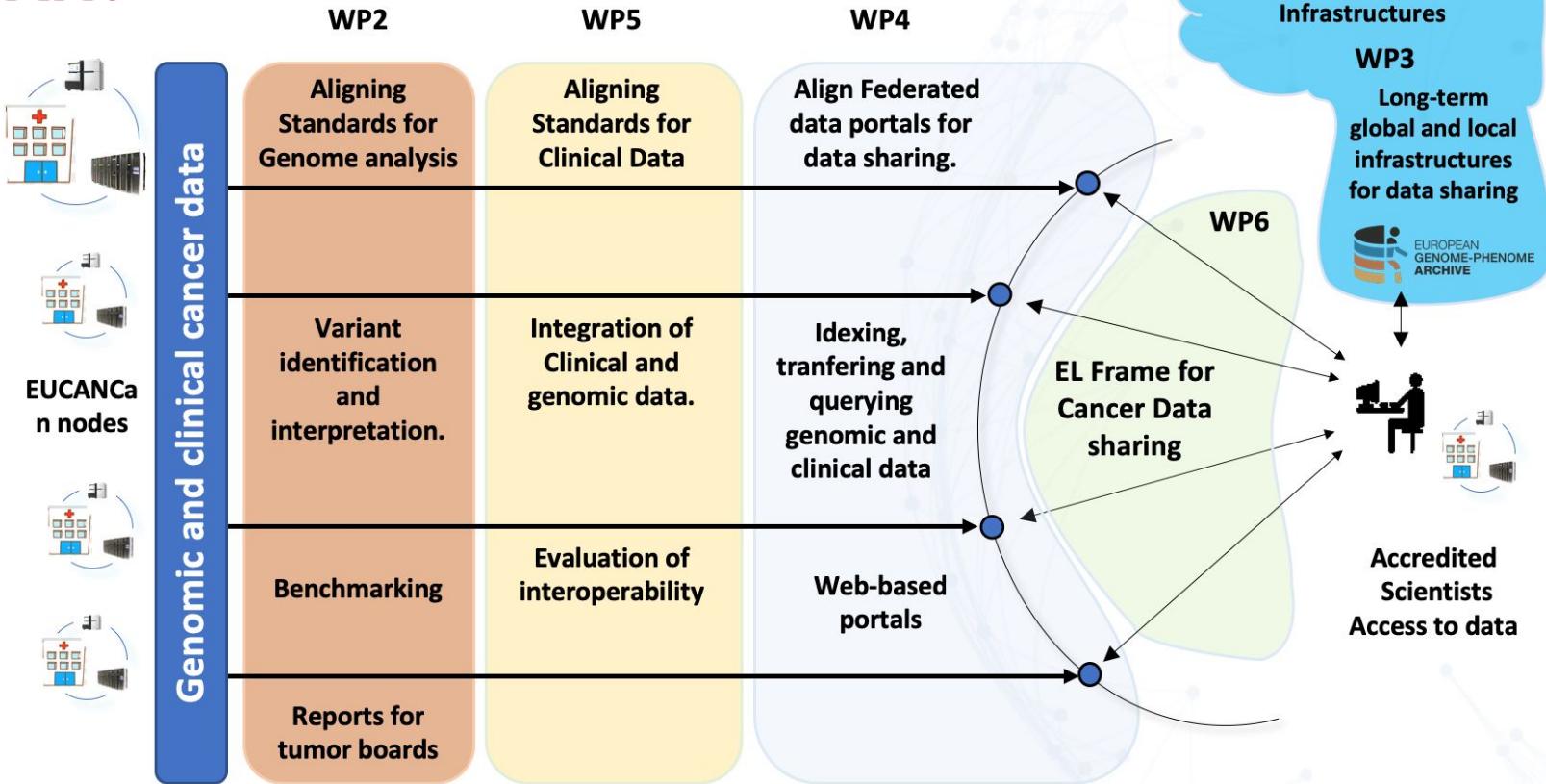
www.eucancan.com

EUCANCan (2019-2023)

A Canada-Europe Federated network for the
globalization of Genomic Oncology and
Personalized Medicine



Data Flow and Work Packages





The Team



Global Alliance
for Genomics & Health



GERMAN
CANCER RESEARCH CENTER
IN THE Helmholtz Association

Research for a Life without Cancer



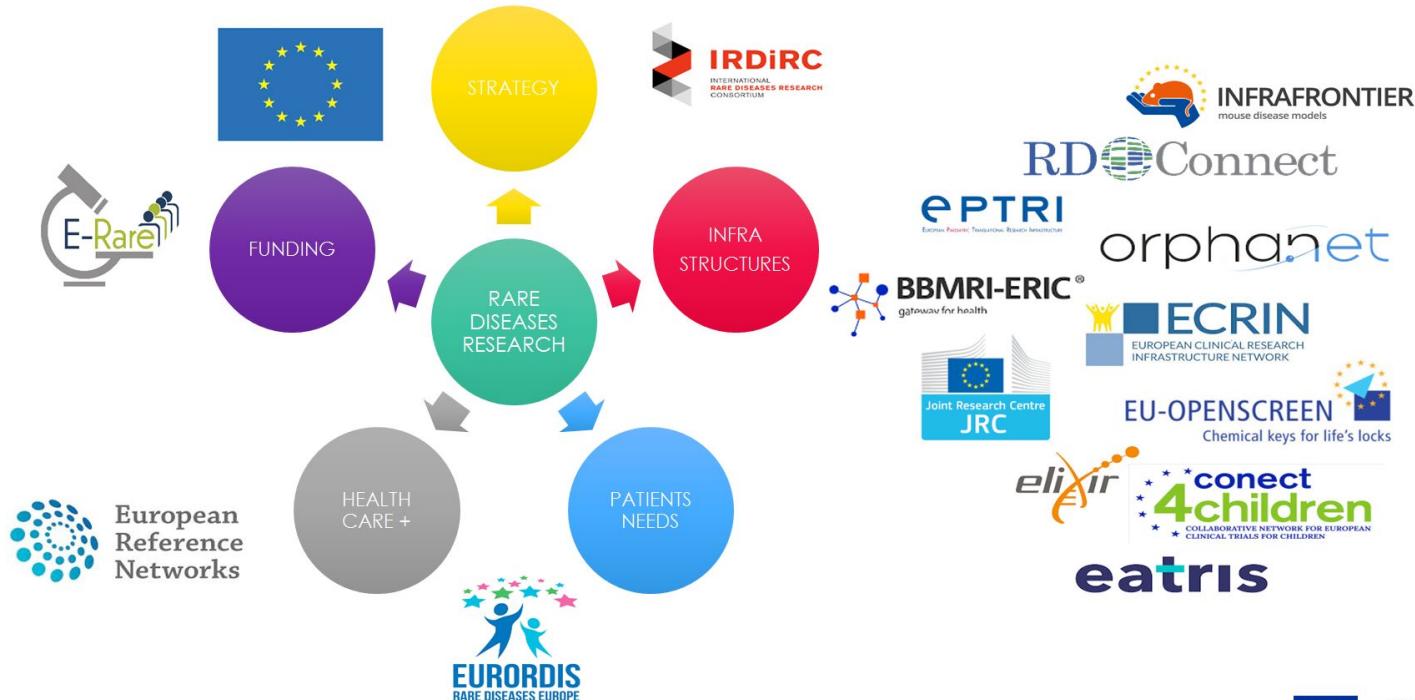


LOCATION: Europe

CHAMPIONS: Sergi Beltran,
Anthony Brookes

New 2019 Driver Project

Rare Diseases Landscape in Europe





Jan 2019

Dec 2023

Total budget (min. submitted): **101 M€** (→ expected > 110 M€)

Union contribution: **55 M€** (70% reimbursement rate)



27 EU MS
7 associated
1 third country



88 beneficiaries

- ☒ 31 research funding bodies/ministries
- ☒ 12 research institutes
- ☒ 22 universities/hospital universities
- ☒ 11 hospitals
- ☒ 5 EU infrastructures (BBMRI, EATRIS, ECRIN, ELIXIR, INFRAFRONTIER) + EORTC
- ☒ EURORDIS & ePAGs
- ☒ 5 charities/foundations (FTELE, AFM, FFRD, FGB, BSF)
- + 50 Linked Third Parties



Funded by the
European Union
GA #825575

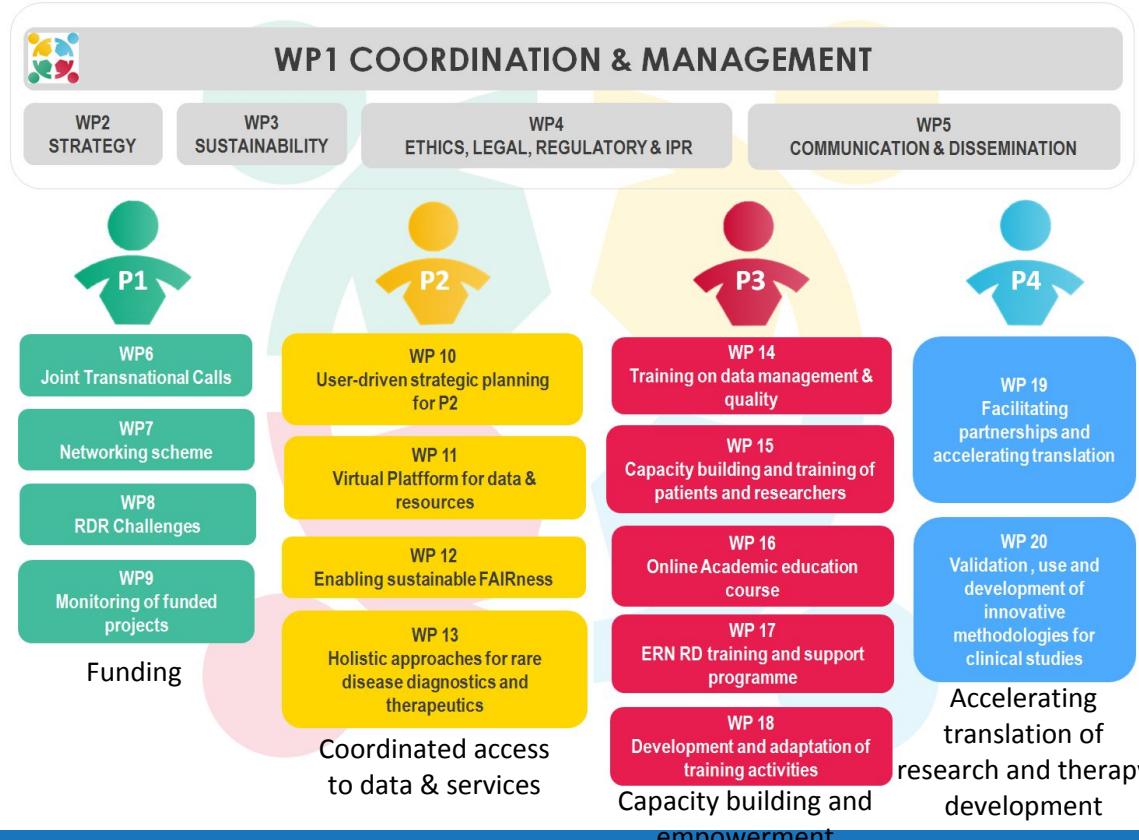


Main objective:

Create a research and innovation pipeline "from bench to bedside" ensuring rapid translation of research results into clinical applications and uptake in healthcare for the benefit of patients

Specific objective:

Improve integration, efficacy, production and social impact of research on rare diseases through the development, demonstration and promotion of sharing of research and clinical data, materials, processes, knowledge and know-how, and an efficient model of financial support for research on rare diseases





First quick expression of interest from selected groups

	1) CLOUD					2) DISCOVERY				3) DATA USE AND RESEARCHER IDENTITIES			4) GENOMIC KNOWLEDGE STANDARDS			5) LARGE SCALE GENOMICS					6) DATA SECURITY		7) REGULATORY AND ETHICS				8) CLINICAL AND PHENOTYPIC DATA CAPTURE				
# PIs showed interest	Task Execution Service	Workflow Execution Service	Data Repository Service	Task Registry Service	Interoperability Testbed	Search API	Beacon API	Service Registry Prototype	MatchMaker Exchange	Data Use	Researcher IDs	Variant Representation	Variant Annotation	HISTOMIC STANDARDS (digital pathology ontologies, annotations etc.)	File Formats	htsget API	RNASeq	CryptidGH	refget API	LARGE SCALE HISTOMICCS (WS file-formats, segmentation, archives)	Breach Response	Authorization and Authentication Infrastructure	Data Protection	Return of Results	International Participant Values Survey	Intellectual Property Licensing Policy	Paediatric Pharmacogenomic Initiative	Data Representations / Ontologies / Interoperability	Data Exchange / Interoperability	AI DRIVEN MEDICINE (reference data sets, federated learning, validations, ...)	EDUCATION AND TRAINING: [non-genomic health professionals / bioinformaticians / IT professionals health related]
4	5	3	3	4	8	4	5	4	4	4	4	1	3	1	2	3	4	3	0	2	4	3	1	1	1	1	8	7	1	1	

Examples:

- Regulatory and Ethics activities are transversal across the EJP-RD (e.g. electronic consents, patient data sharing)
- Clinical and Phenotypic Data Capture standards and ontologies are essential for RD interoperability and data exchange (e.g. HPO, PhenoPackets, etc.)
- DURI and AAI key to the EJP-RD Virtual Platform (Pillar 2)
- Some resources already participating in Discovery, Genomic KS and LS Genomics WGs and using its standards (e.g. MME, Beacon, variant annotation, file formats, htsget, etc.)
- Interest in collaborating towards a FAIR based data infrastructure



H3Africa

Human Heredity & Health in Africa

LOCATION: Pan-Africa
CHAMPIONS: Nicola Mulder,
Mogomotsi Matshaba

New 2019 Driver Project

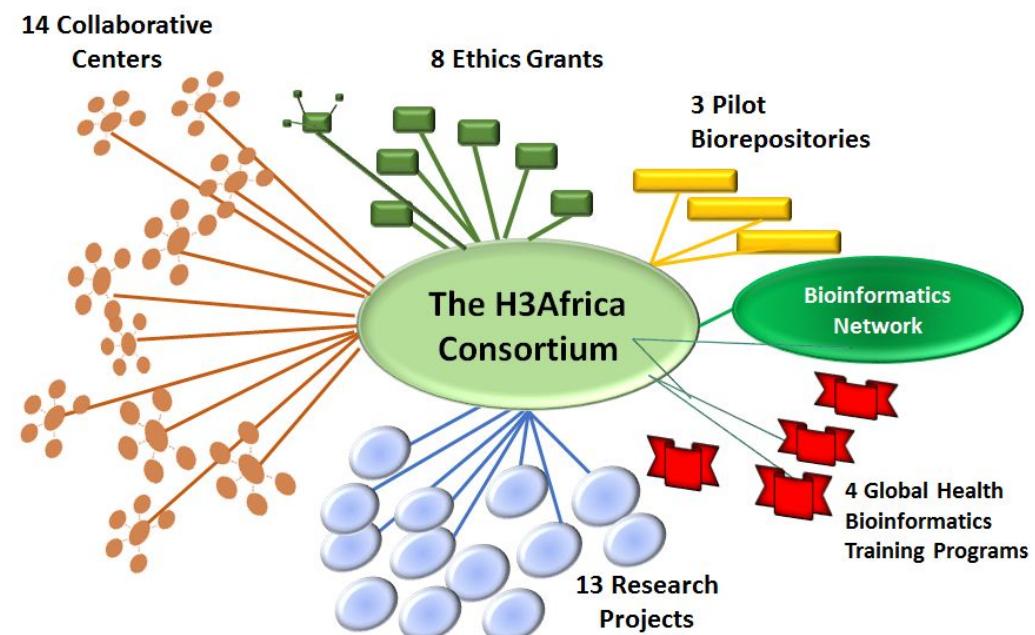
Vision: To facilitate an African-based research approach to the study of genomic and environmental determinants of common diseases with the goal of improving the health of African populations

Additional goals:

Capacity development –human and research infrastructure

Building **collaborative networks**

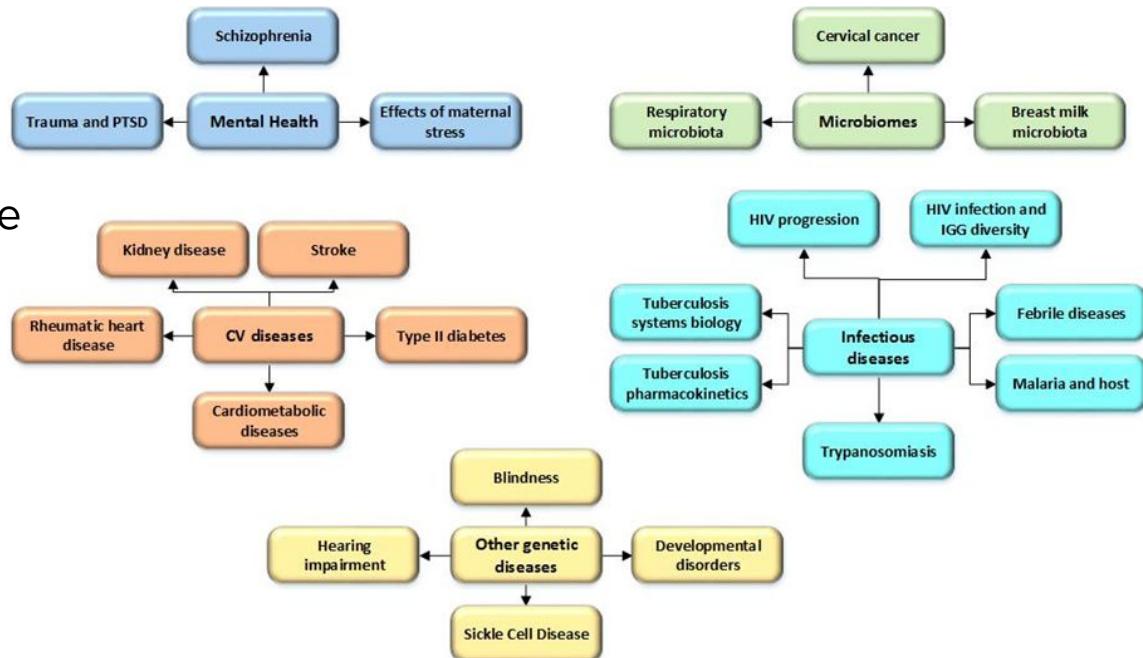
Led by Steering Committee,
cross consortium work
through working groups



Phenotyped cohort (>75k), with genomic data, focussed on genetic basis for disease

Additional WG outputs:

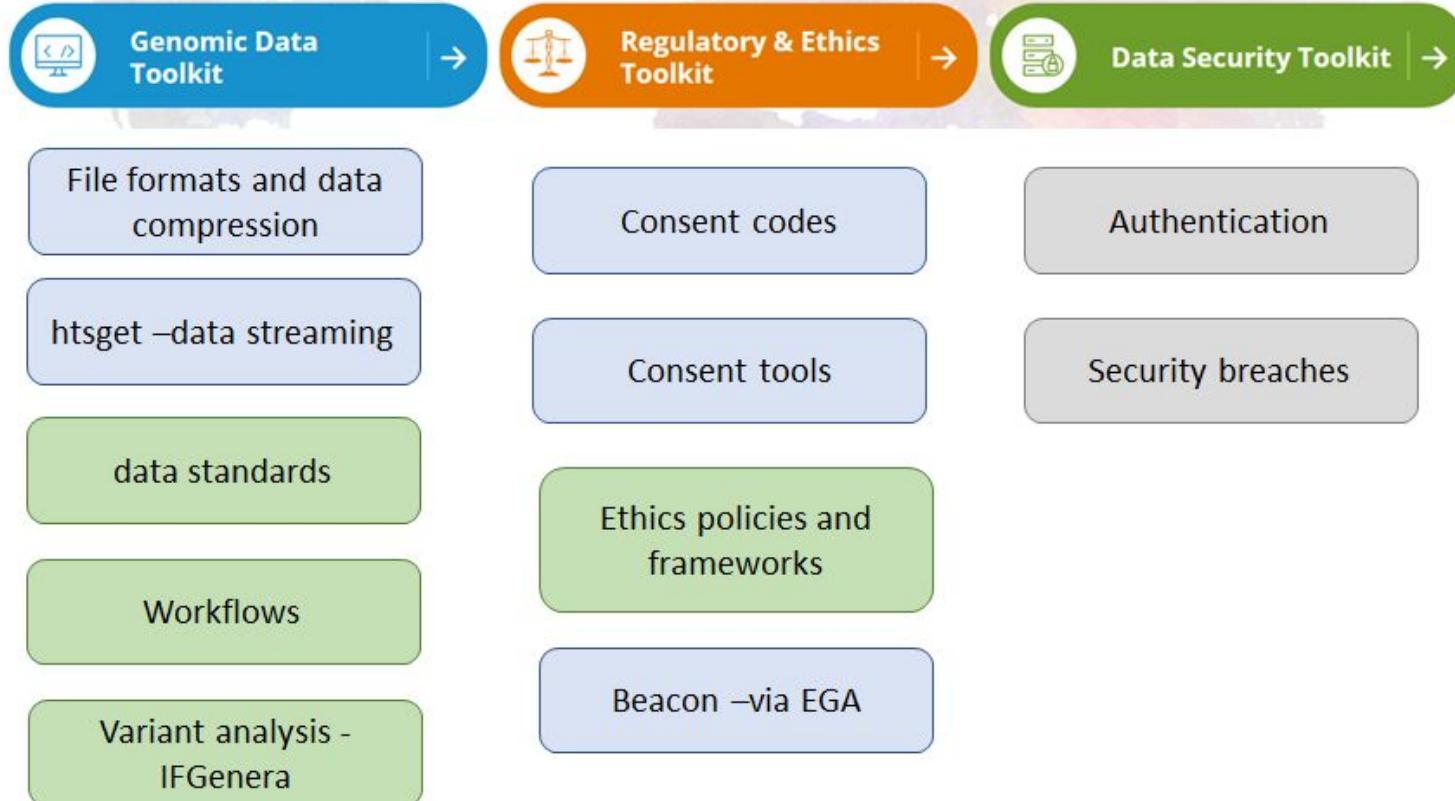
- Ethics and consent guidelines
- Phenotype harmonization
- Variant interpretation & evidence
- Data analysis tools
- Genomic medicine training



H3Africa -link to GA4GH



Global Alliance
for Genomics & Health





LOCATION: Switzerland

CHAMPIONS: Michael Baudis, Torsten Schwede, Sabine Österle, Kevin Sayers, Katrin Carmeri

New 2019 Driver Project

Vision: nationwide interoperability of biomedical information

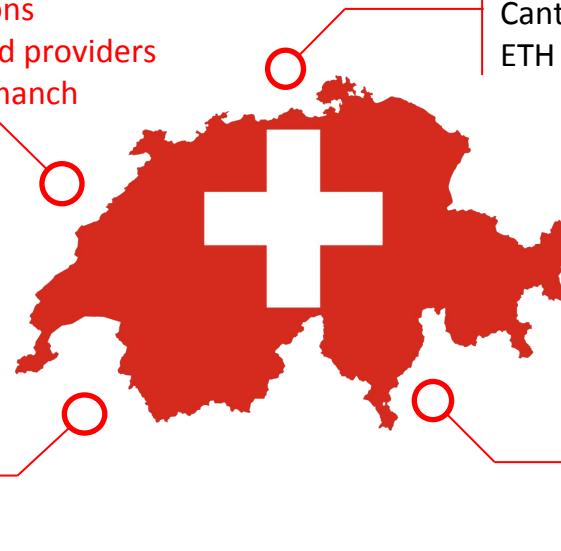
Switzerland

Population: 8.3 Million

Legal system: Confederation with 26 cantons

Healthcare: Public, private and subsidized providers

Languages: German, French, Italian, Romanch



SPHN

Government funded initiative: 68 Mio

Co-mandated to:

- Swiss Academy of Medical Sciences
- Swiss Institute of Bioinformatics

Phase 1: 2017-2020 patient data

Phase 2: 2021-2024 Consolidation and expansion of the network, healthy citizen data and citizen controlled data

Main collaboration partners

5 University Hospitals

Cantonal Hospitals

ETH Domain and Universities

SPHN funded projects (examples)

Oncology

Variant Interpretation Platform

Pathology

Frailty

Immune repertoire

Immunotherapy

Sepsis

Citizen data

Radiomics

Ophthalmics

Heart failure

Data governance

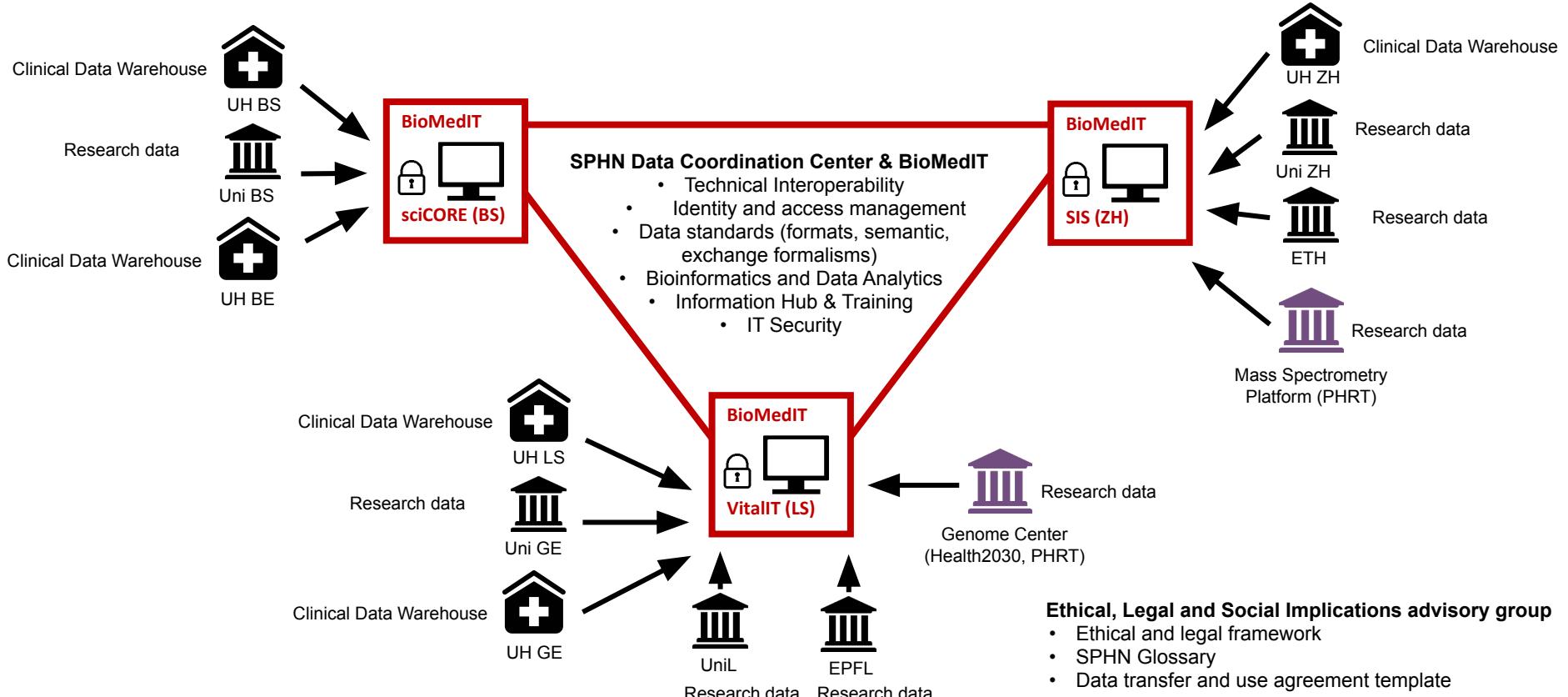
E/dynamic consent

De-identification

Data standards

Data security / privacy

SPHN – A decentralized approach



GA4GH Workstreams

- Cloud
 - Federated multi-site workflow execution
 - ELIXIR-TESK implementation study
 - BioMedIT interoperability WG (chair Kevin Sayers)
- Discovery
 - Co-chair Michael Baudis
 - Beacon & SchemaBlocks
- Data Security
 - Co-chair Jean-Pierre Hubaux
- Potential interaction
 - Data registry service (DRS), Tool registry service (TRS)
 - to be discussed

Swiss involvement or collaboration with GA4GH Driver projects





CLOUD AND AAI

LOCATION: Europe

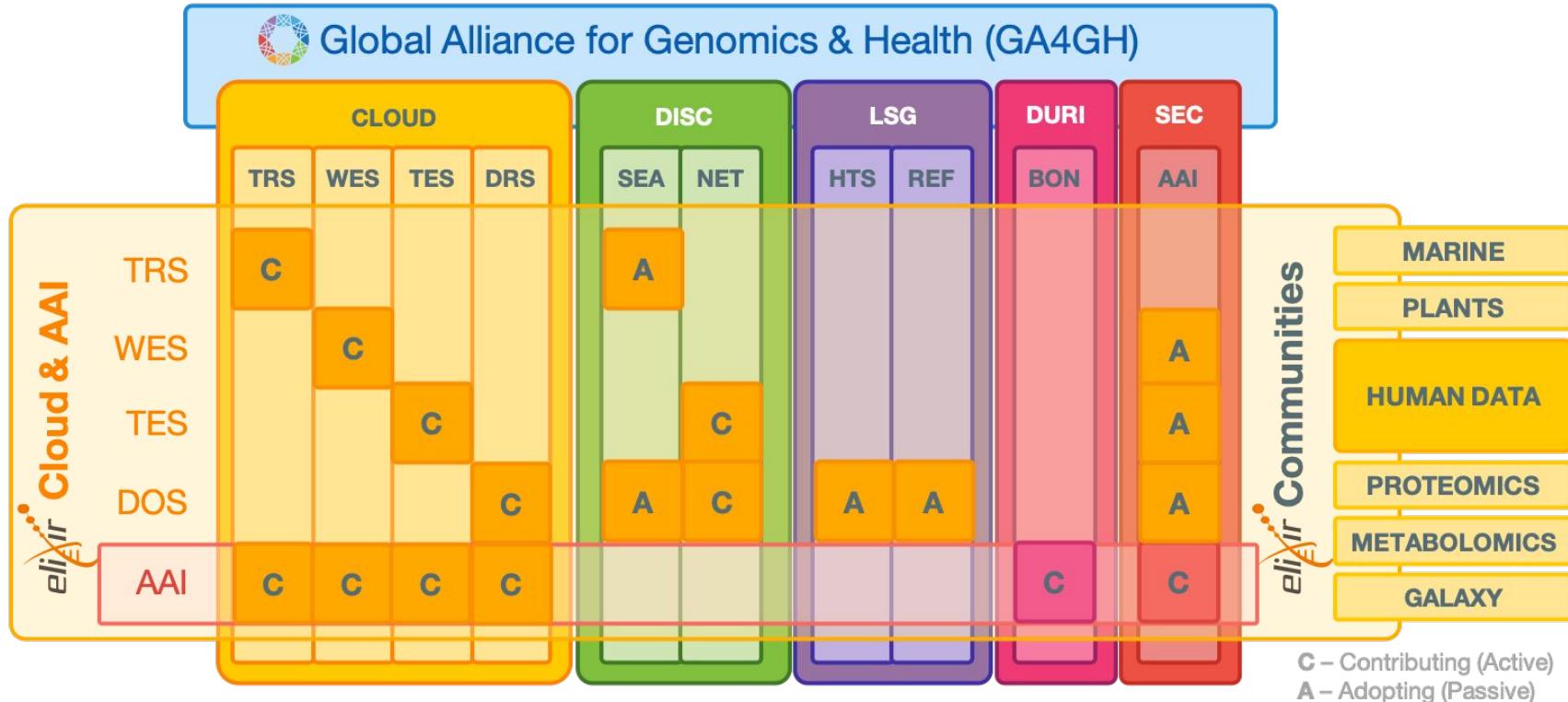
CHAMPIONS: Susheel Varma, Mikael Linden, Jonathan Tedds (coordinator), Steven Newhouse, Tommi Nyrönen, Salvador Capella

New 2019 Driver Project

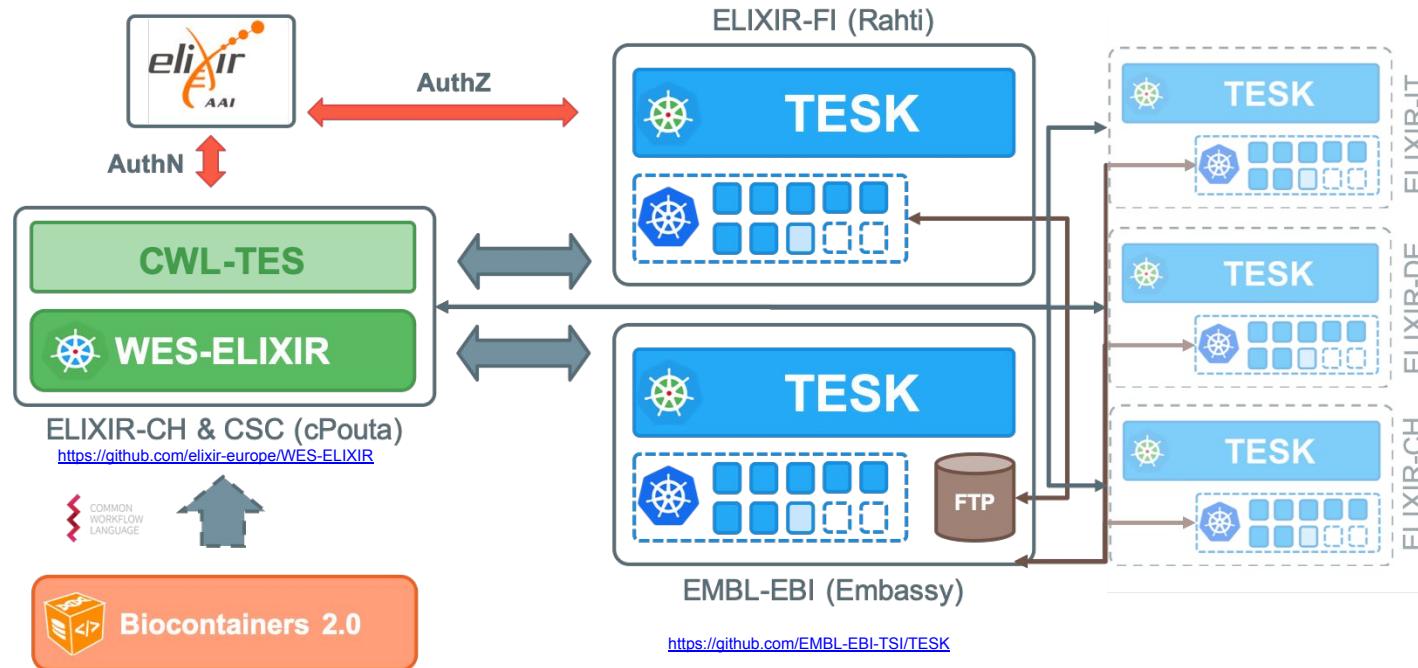
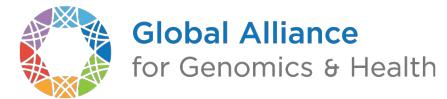
ELIXIR Cloud & AAI

- Integrate user federation into local compute and data deployments - **ELIXIR AAI**
- Rationalise a ELIXIR-wide Data Distribution Network – starting with Reference datasets - **RDSDS**
- Collect & curate community-driven containers and workflows for life-sciences - **Biocontainers & ELIXIR Tools**
- Develop Task Distribution Network using Task orchestration engines – e.g. Kubernetes - **TES**
- Support national or regional Workflow Choreography Engines – e.g. **CWL-TES**, Cromwell, Nextflow, Galaxy, etc. - **WES-ELIXIR**
- Drive **ELIXIR Compute Platform** support to Nodes to develop hybrid cloud/HPC deployments - **EOSC-Life RAP**





ELIXIR Cloud & AAI Ecosystem



<https://github.com/BioContainers/biocontainers-backend>

ELIXIR Cloud & AAI Ecosystem - Status



Biocontainers

BioContainers : A set of resources proposed as a common standard for tool representation.	
<code>get /metadata</code>	Show/Hide
<code>get /stats</code>	Return some metrics
<code>get /ToolClasses</code>	List all tool types
<code>get /Tools</code>	List all tools
<code>get /Tools/{id}</code>	List one specific tool, acts as an anchor for self-references
<code>get /Tools/{id}/similar</code>	List similar tools
<code>get /Tools/{id}/versions</code>	List versions of a tool
<code>get /Tools/{id}/versions/{version_id}</code>	List one specific tool version, acts as an anchor for self-references
<code>get /Tools/{id}/versions/{version_id}/containerfile</code>	Get the container specification(s) for the specified image
<code>get /Tools/{id}/versions/{version_id}/{type}/descriptor</code>	Get the tool descriptor for the specified tool
<code>get /Tools/{id}/versions/{version_id}/{type}/descriptor/relative_path</code>	Get additional tool descriptor files relative to the main file
<code>get /Tools/{id}/versions/{version_id}/{type}/files</code>	Get a list of objects that contain the relative path and file type
<code>get /Tools/{id}/versions/{version_id}/{type}/tests</code>	Get a list of test JSONs
BioContainerWorkflows	
<code>get /Workflows</code>	Show/Hide List Operations Export Operations
<code>post /Workflows</code>	List all BioContainer workflows
[BASE URL: https://api.biocontainers.pro/api/qa4gh/v2/ , API VERSION: 2.0.0]	

<https://api.biocontainers.pro/api/qa4gh/v2/ui/>

Biocontainers 2.0

- 7500+ Tools
- 600+ Workflows
- TRS v2 API



WorkflowExecutionService

<code>GET /runs</code>	Show/Hide List Operations
<code>POST /runs</code>	Run a workflow.
<code>GET /runs/{run_id}</code>	Get detailed info about a workflow run.
<code>POST /runs/{run_id}/cancel</code>	Cancel a running workflow.
<code>GET /runs/{run_id}/status</code>	Get quick status info about a workflow run.
<code>GET /service-info</code>	Get information about Workflow Execution Service.

TaskService

<code>GET /tasks</code>	List tasks. TaskView is requested as such: "v1/tasks?view=BASIC"
<code>POST /tasks</code>	Create a new task.
<code>GET /tasks/service-info</code>	GetServiceInfo provides information about the service, such as storage details, resource availability, and other documentation.
<code>GET /tasks/{id}</code>	Get a task. TaskView is requested as such: "v1/tasks/{id}?view=FULL"
<code>POST /tasks/{id}:cancel</code>	Cancel a task.

<http://193.167.189.73:7777/qa4gh/wes/v1/ui/>
<https://tes.ts1.ebi.ac.uk/qa4gh/wes/v1/ui/>

WES-ELIXIR & TESK [PoC]

- ELIXIR AAI Integration, CWL-TES
- PoC Oct 2018



WES-ELIXIR



TESK



Global Alliance
for Genomics & Health



RSDS

DRS

DataRepositoryService

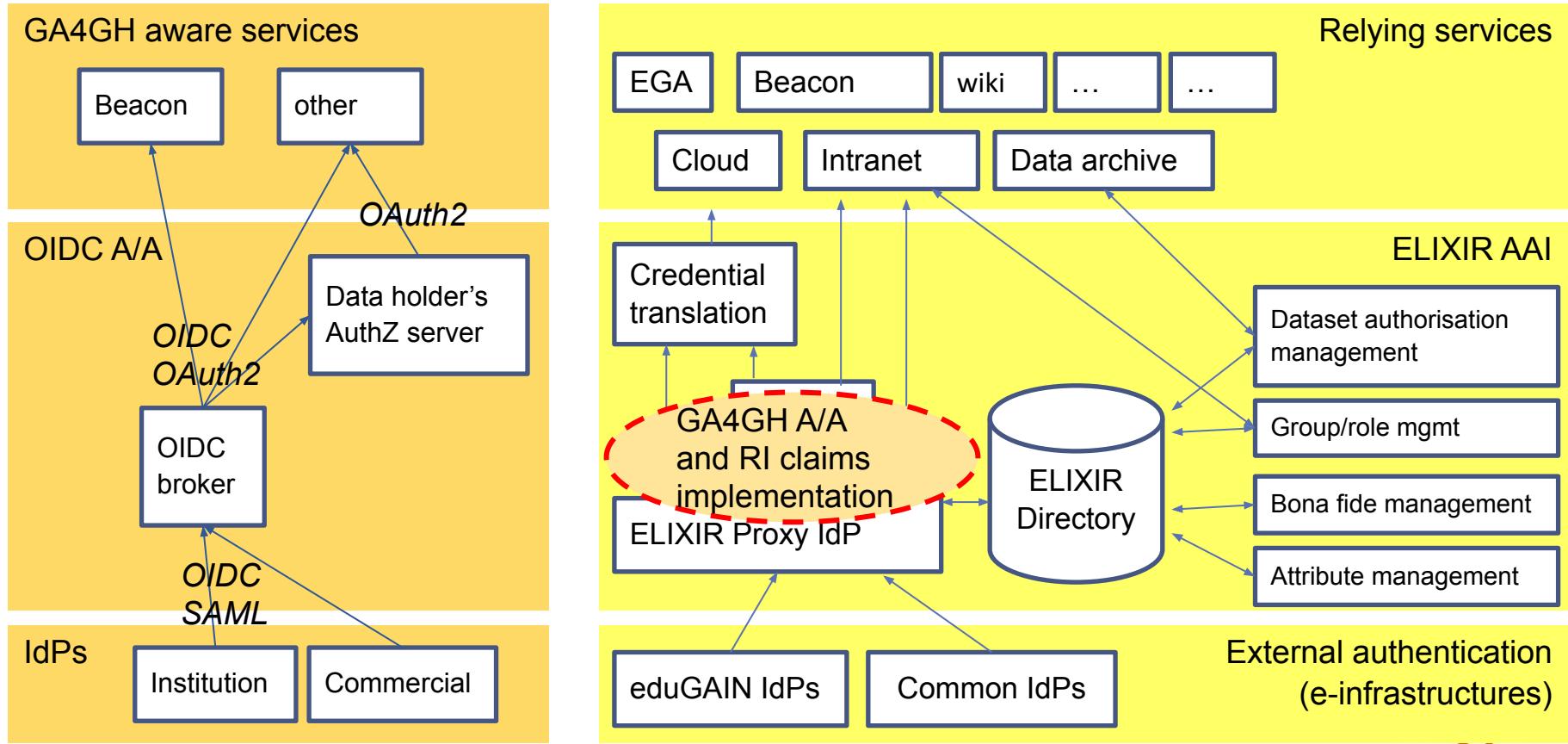
<code>GET /service-info</code>	Get information about this implementation.
<code>GET /bundles/{bundle_id}</code>	Get info about a Data Bundle.
<code>GET /objects/{object_id}</code>	Get info about a Data Object.
<code>GET /objects/{object_id}/access/{access_id}</code>	Get a URL for fetching bytes.

<https://dsds-service.ebi.ac.uk/>

RSDS v2 [PoC]

- DRS 0.1 API
- EMBL-EBI Reference Dataset Index:
 - Metabolights
 - PRIDE
 - ArrayExpress
 - ENA
- Interested in Data Transfer & Subscription Services
- PoC Q2 2019

GA4GH A/A and RI claims in ELIXIR AAI





LOCATION: Japan

CHAMPIONS: Makoto Suematsu,
Hidewaki Nakagawa, Kenjiro Kosaki

New 2019 Driver Project





GEnome Medical alliance Japan

- Variant Aggregation and Allele Frequencies
- Collection and integration of Pathogenic Variants
- Contribution to Standardization and Data Sharing
- Biobank Cross-search System

Nation-wide systems for genomic medicine



IRUD

Initiative on Rare
and Undiagnosed
Diseases



Medical Genomics
Japan Variation
Database

Biobanks in large scale



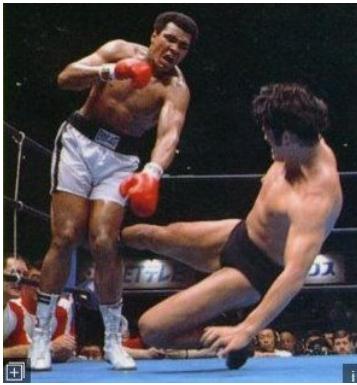
157,000
healthy
individuals



270,000 patients
with ~50 adult
common
diseases



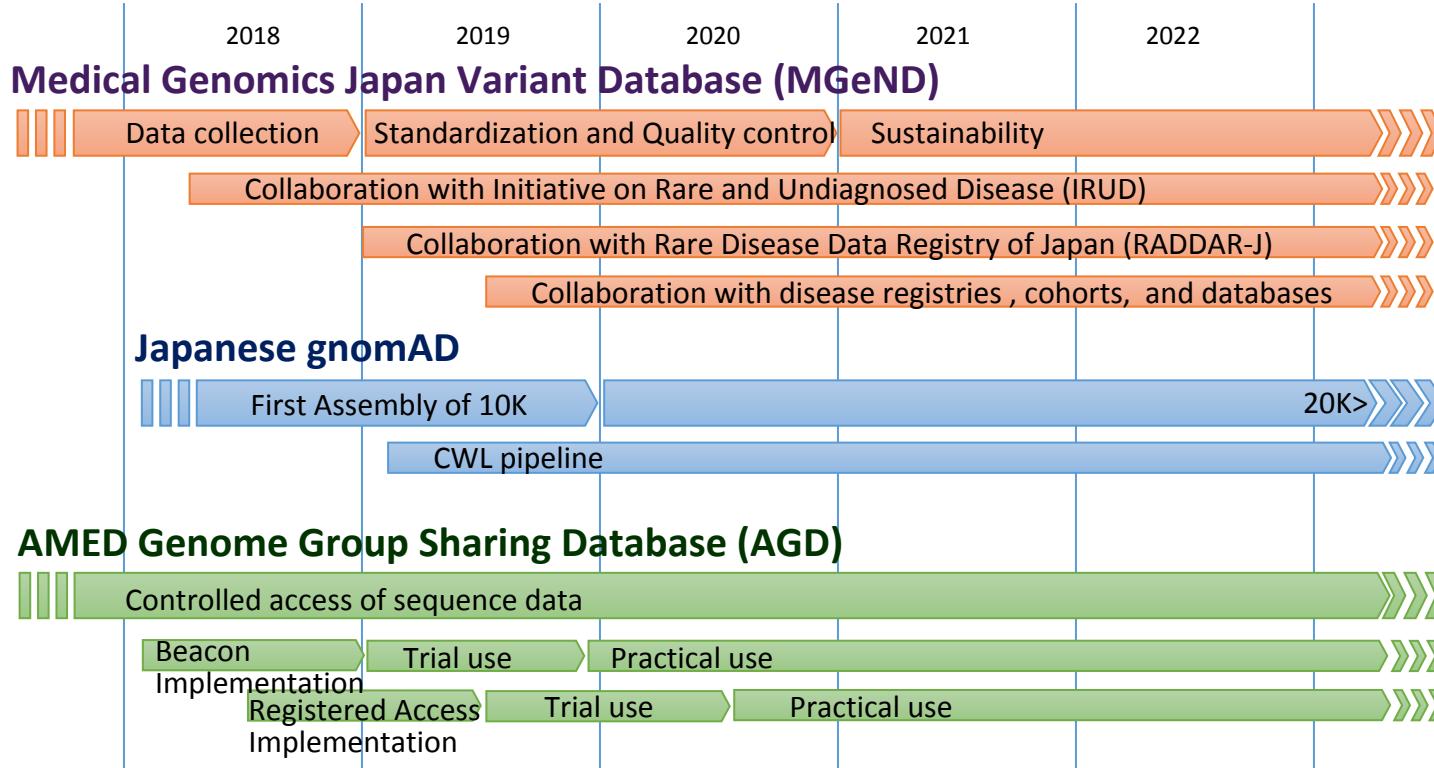
74,000 patients with
major common and
various rare
diseases



Vision / Mission / Goals

- To harmonize our existing projects with the international standards for the genome based health system
- To promote sharing our Japanese genomic and phenotypic information with global community
- To develop the linkage of genome and clinical information between East Asian populations which are thought to be genetically close to Japanese population





From “beyond domestic borders” to “beyond global borders”

GEM Japan



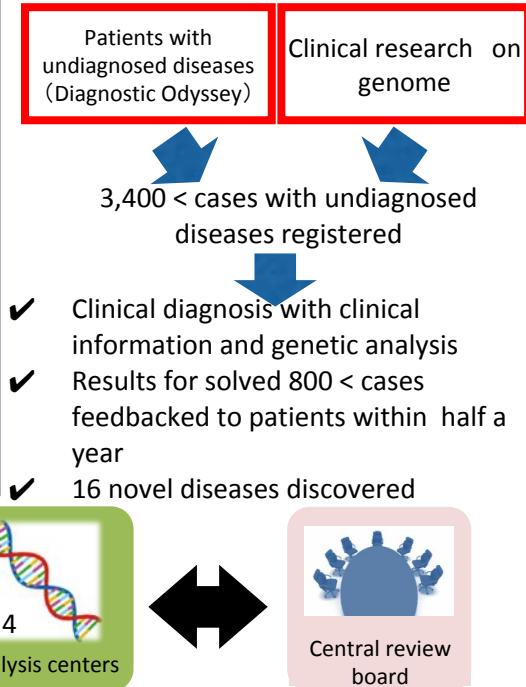
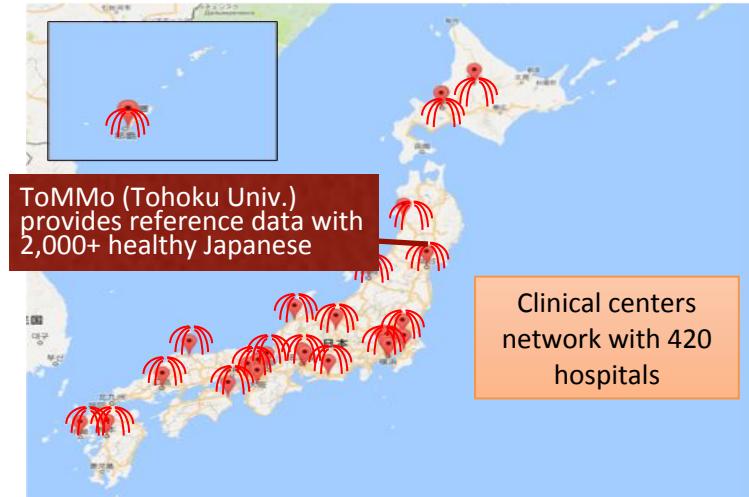
Global Alliance
for Genomics & Health

		Technical Work Stream								
Category	Database / Project	VCF Aggregator	Large Scale Genomics	HLA	Clinical & Phenotypic Data capturing	Family Health History	Genomics Knowledge Standards	Discovery	Data Use & Researcher IDs	Cloud
Data Platform	Patient Archive (IRUD Exchange)			☆	☆	○	○			
	HPO Japanese Version			☆	○			Beacon API	Researchers ID	
	Biobank Cross-search				○	○			Data Use Ontology	
	AGD						○	○	○	○
Integrated DB (Open DB)	MGeND	☆	☆				○	○	○	○
	Japanese gnomAD	☆					○	○	○	○

IRUD

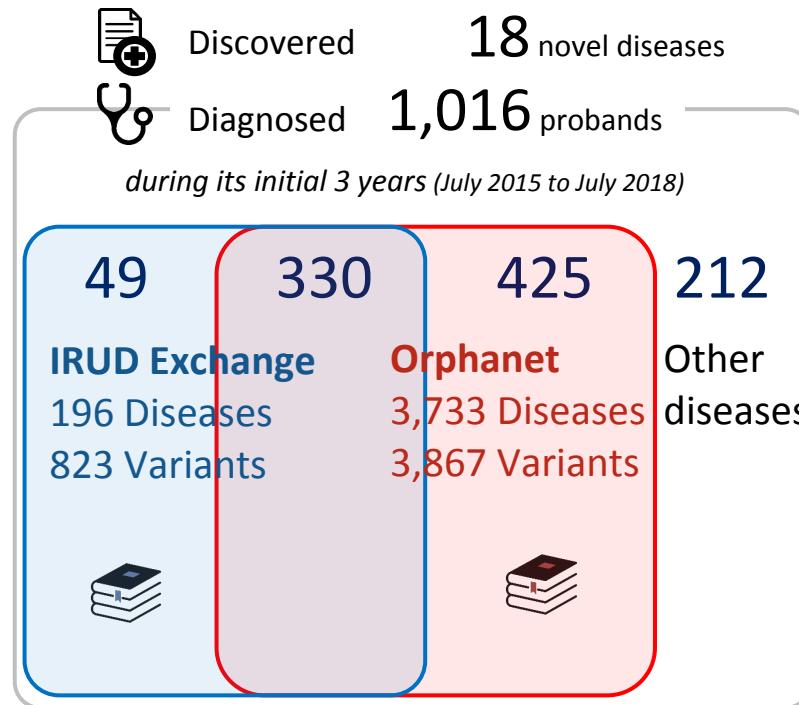
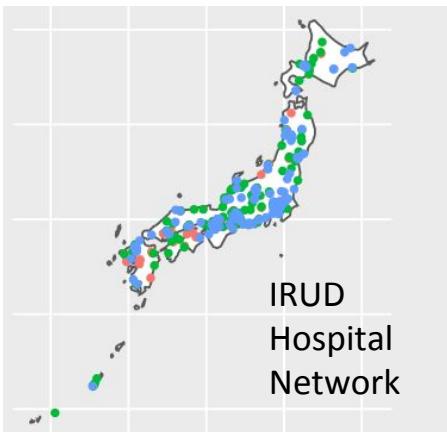
Initiative on Rare and Undiagnosed Diseases

Clinical research program for supporting the diagnosis of patients with undiagnosed diseases in Japan



IRUD research group supports the diagnosis of patients with undiagnosed disease via data sharing with a globally compatible and active data sharing system

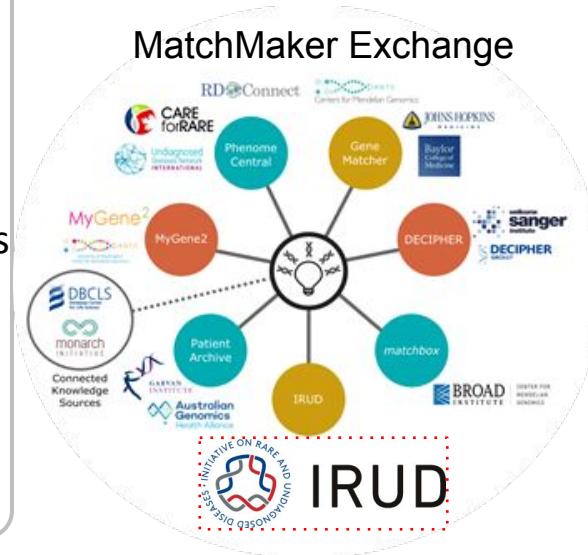
Domestic Data Sharing

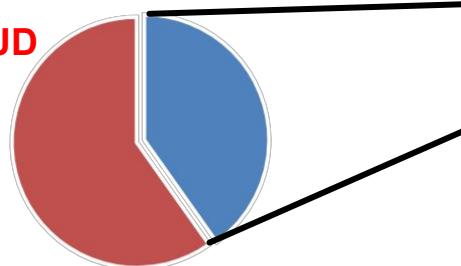


Global Data Sharing



MatchMaker Exchange





ClinVar
Clinically relevant variation

GEM Japan



Global Alliance
for Genomics & Health

■ Registered ■ Unregistered



All Japan Alliance
Universities - Institutes - Hospitals

As of September 2017

438
17,384
13,513

111
14,557
25,073

82
28,429
157,082

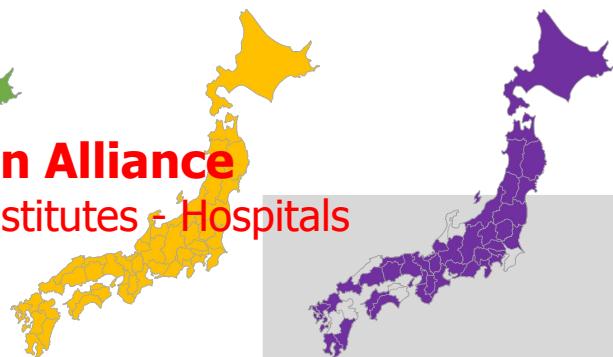
Dementia /Sensory

99 sites
23,464 patients
19,339 variants

Rare Variants

Common Variants

MGeND : Medical Genomics Japan Variant Database



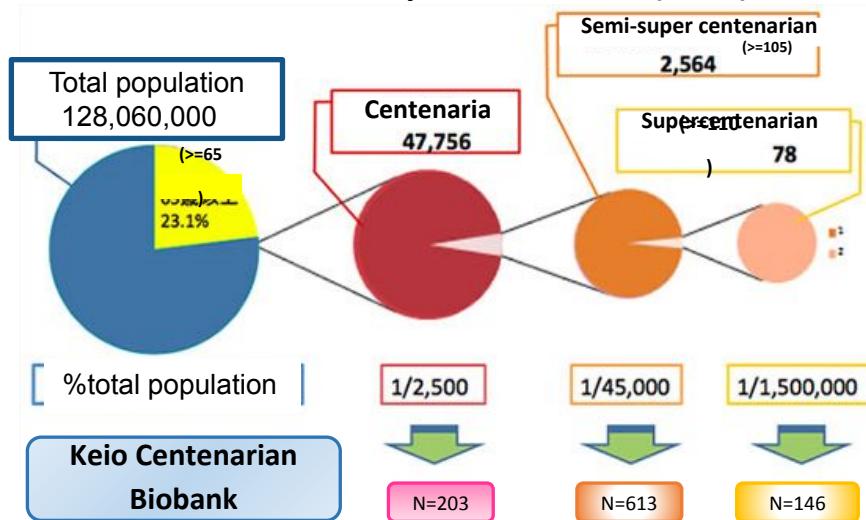
Clinical
Annotation
Committee



As of April 2019

**Super centenarian >110 y/o
Semi-super centenarian 105~110 y/o
YOUNG centenarian 100~105 y/o**

% of Centenarian and Supercentenarian (2010)



Distribution of Supercentenarians



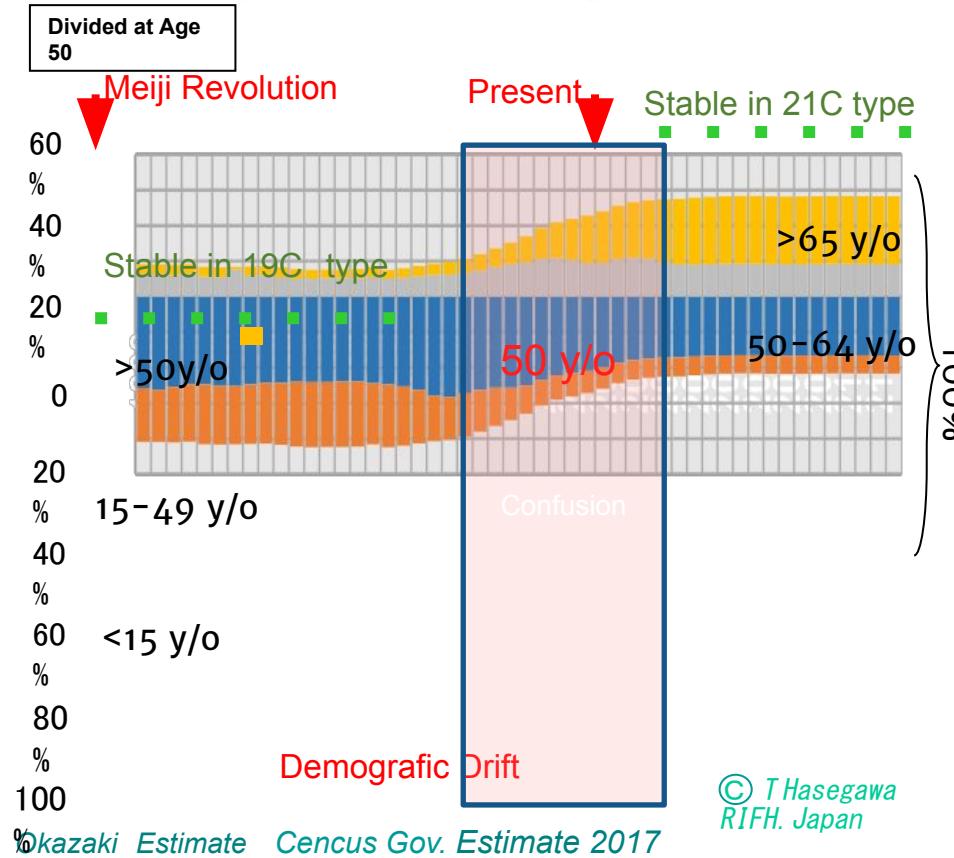
International Collaborative Researches
- 5 Country Oldest Old Project (SCOOP)



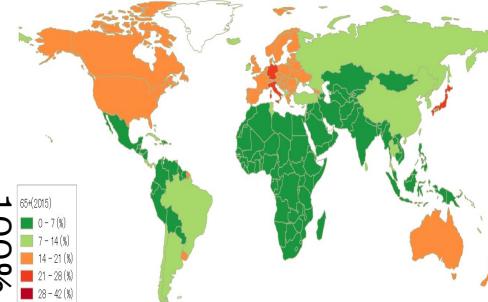
- WGS: 496 centenarians including 126 Supercentenarians.
- Allele frequency of variants for 57 genetic disease genes
- Allele frequency of variations in 40 dementia-associated genes

Data Sharing

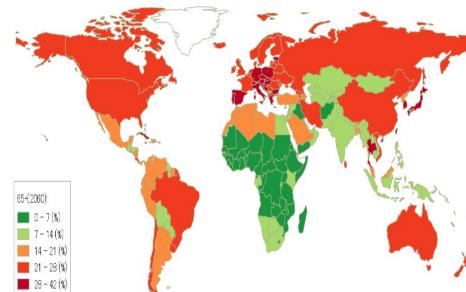
Great Demographic Transition: Japan first, and your countries will share



Population(%) age over 65 years in 2015



in 2060



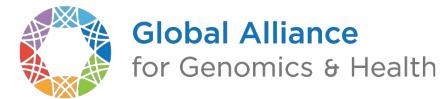


LOCATION: Pan-Europe

CHAMPIONS: Serena Scollen, Gary
Saunders

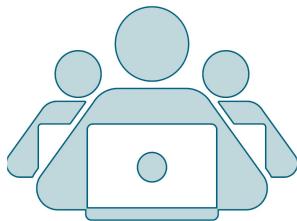
New 2019 Strategic Partnership

ELIXIR::GA4GH Strategic Partnership



Global Alliance
for Genomics & Health
Collaborate. Innovate. Accelerate.

Simplify the way people search for and request access to potentially identifiable data in international and national genomic data resources



Working towards GA4GH standards, APIs and toolkits to be used throughout ELIXIR Nodes for human data discovery and access – **GA4GH into Europe**

GA4GH Work Stream

Clinical and Phenotypic Data Capture
(e.g. PhenoPackets)

Cloud
(e.g. TES, WES, TRS, DRS)

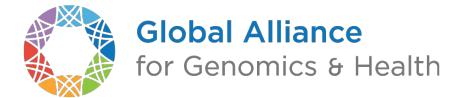
Data Use & Researcher Identities
(e.g. DUO, AAI)

Discovery
(e.g. Beacon, SearchAPI, MME)

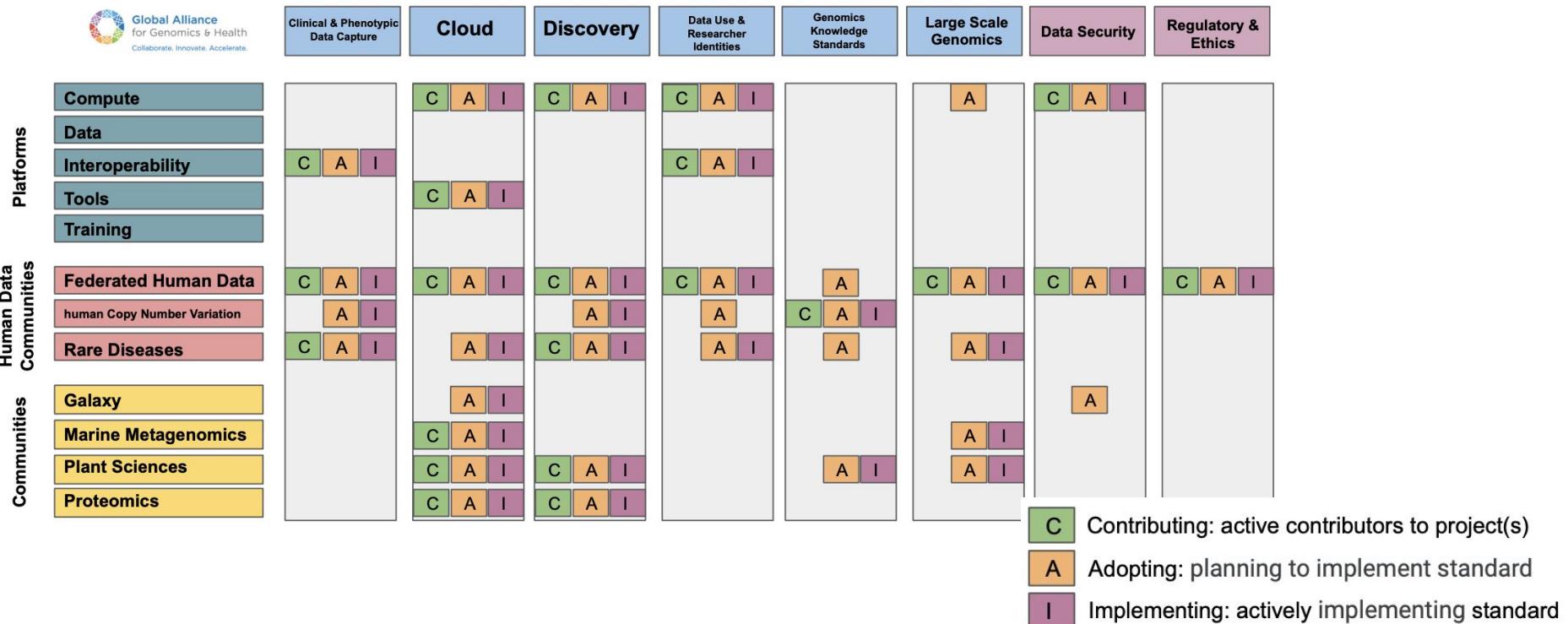
Genomics Knowledge Standards
(e.g. VMC)

Large Scale Genomics
(e.g. HTSGET, REFGET)

ELIXIR::GA4GH Strategic Partnership



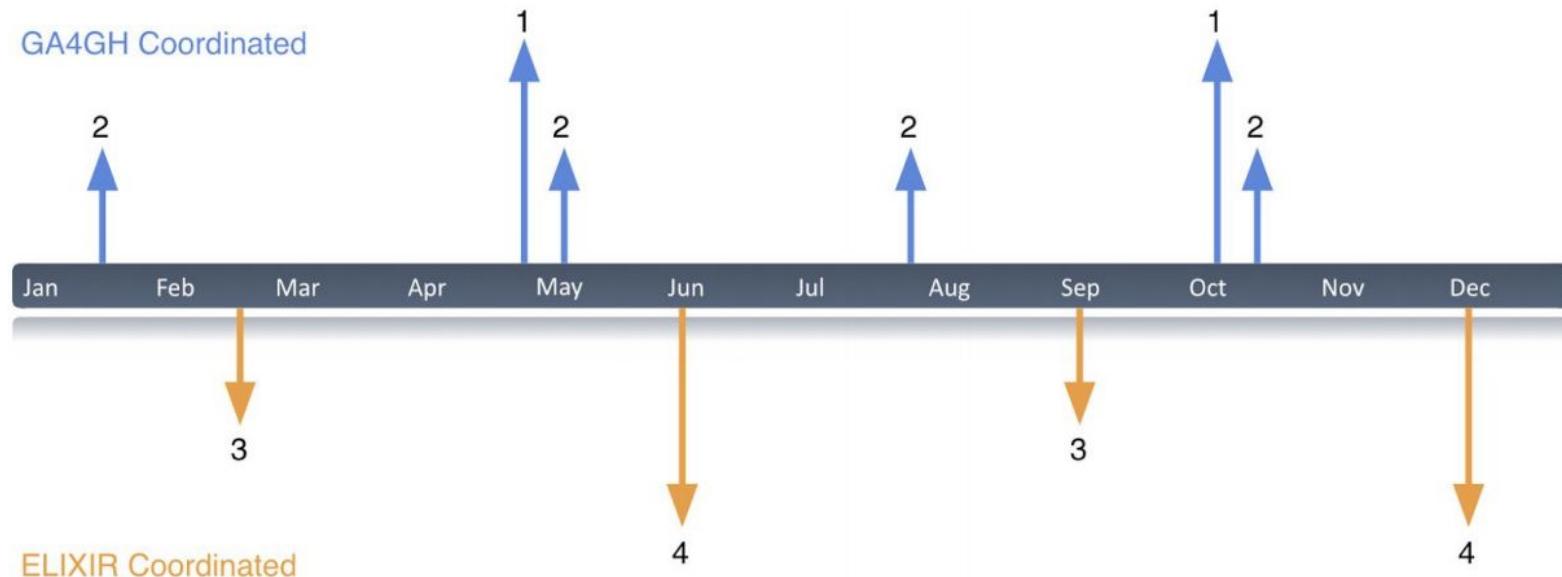
ELIXIR Platform and Community activities (contributing, adopting, implementing) across the GA4GH Work Streams.



ELIXIR::GA4GH Strategic Partnership



Next steps



1: GA4GH Plenary and Strategic Meetings (2x yearly)

2: GA4GH Steering Committee (4x yearly)

3: ELIXIR::GA4GH Strategic Meetings (2x yearly)

4: ELIXIR meeting to discuss Strategic Partnership (2x yearly)



Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Before Lunch...

U24 Grant Application



TITLE: Setting the Standard for Genomics and Health-Related Data Sharing: The Global Alliance for Genomics and Health

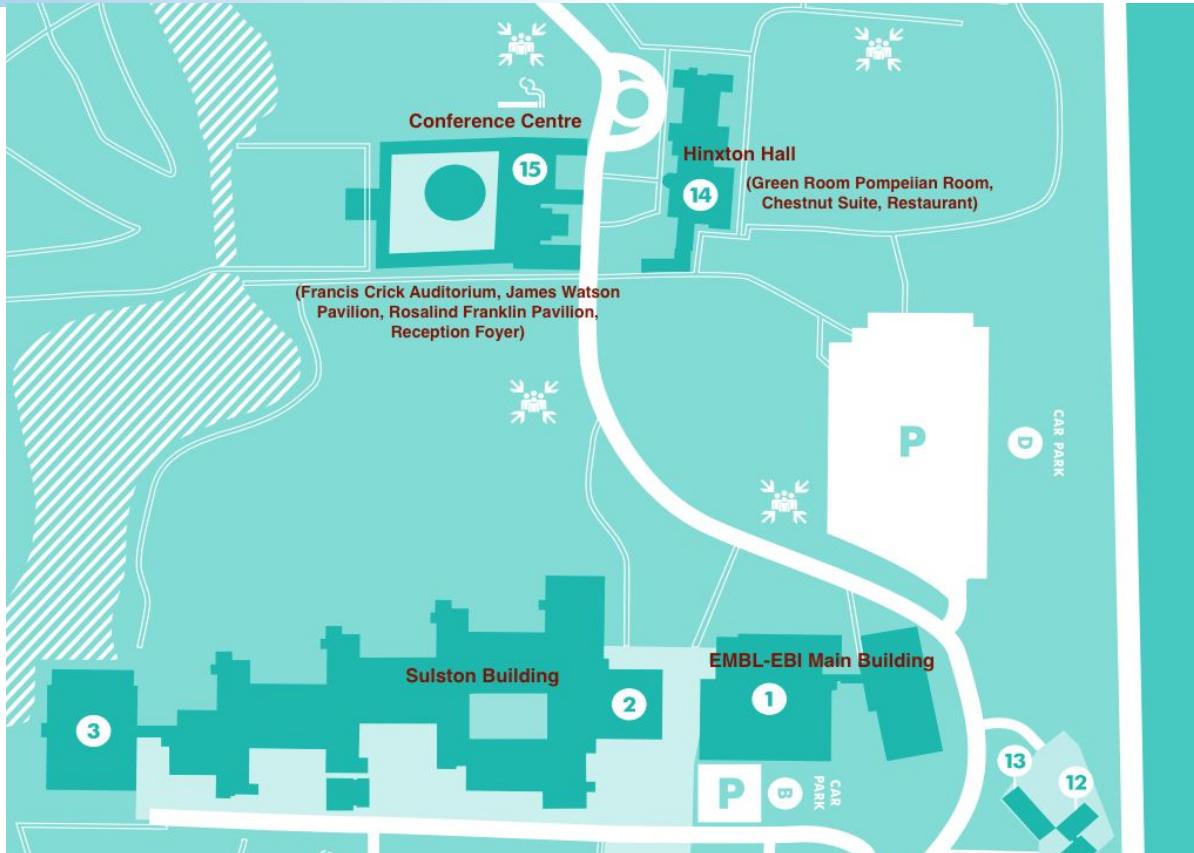
APPLICATION DEADLINE: May 25

DRIVER PROJECTS - SEEKING LETTERS OF SUPPORT: Any person or group providing specific resources to GA4GH standards development

LOS DEADLINE: May 20

WORK STREAM LEADS - EXPECT TO HEAR FROM US: Input on requested technical resources

Campus Buildings



For Meetings in:

EMBL-EBI Main Building
Garden Room
Courtyard Room

Sulston Building
Sulston C3-02

meet at reception and walk over together.



Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

LUNCH



Breakout Session 1



SESSION	LOCATION
Security Review Process	Courtyard Room (EBI Building)
Family History / Pedigree Standard	Francis Crick Auditorium
Researcher IDs	Garden Room (EBI Building)
GKS Strategy Meeting	Green Room
Cloud API Review & DRS	James Watson Pavilion
RNASeq	Loft Room 2
SchemaBlocks	Rosalind Franklin Pavilion
REWS Meeting:	Sulston C3-02 (Sulston Building)



Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

BREAK



Breakout Session 2



SESSION	LOCATION
Acute Paediatric Care Information Models	Francis Crick Auditorium
DUO	Garden Room (EBI Building)
Variant Annotation / Variant Representation	Green Room
TES & TRS	James Watson Pavilion
Crypt4GH	Loft Room 2
JSON Schema Workshop	Pompeiiian Room
Work Stream Lead & Driver Project Champion Speed Meetings	Rosalind Franklin Pavilion



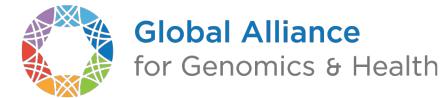
Global Alliance
for Genomics & Health

Collaborate. Innovate. Accelerate.

Dinner



Afternoon Plans - OLD



Work Stream & Project Breakouts

Security Review Process: Courtyard Room (EBI Building)

FHIR / Phenopackets: Francis Crick Auditorium

Researcher IDs: Garden Room (EBI Building)

Genomic Knowledge Standards Strategy Meeting: Green Room

Cloud API Review & DRS: James Watson Pavilion

RNASeq: Loft Room 2

SchemaBlocks: Rosalind Franklin Pavilion

Regulatory and Ethics Work Stream Meeting: Sulston C3-02 (Sulston Building)

Work Stream Lead & Driver Project Champion Speed Meetings

Acute Paediatric Care Information Models: Francis Crick Auditorium

DUO: Garden Room (EBI Building)

Variant Annotation / Variant Representation: Green Room

TES & TRS: James Watson Pavilion

Crypt4GH: Loft Room 2 JSON

Schema Workshop: Pompeian Room

Work Stream Lead & Driver Project Champion Speed Meetings: Rosalind Franklin Pavilion