Global Genomic Medicine Collaborative

International Cohorts Summit

Vision for success:

A global platform for translational research (cohort to bedside and cohort to bench), informing biological/genetic basis for disease and impact on clinical care and population health

1- Creating a standardized database and registry—pros, cons, how best to do it

- Start with cohort-level metadata, work toward unified database of individual-level data
- Leverage existing metadata collections, automate as possible
  - NCI Cohort Consortium, UK Dementia Platform
- Tiered structure: each contributes at level of comfort, queryable
  - 0: Cohort website
  - 1: Cohort description (like program book), mechanisms for access
  - 2: Data description (demographics, data collection instruments, omic data
  - 3: Counts of phenotypes, sample types, update
  - 4: Individual-level data

1- Creating a standardized database and registry—pros, cons, how best to do it

- Funding and infrastructure needed for fielding queries, depositing data, companies may have solutions but should not hold the data
- Projects that could be facilitated: ***build reproducibility network***
- Address data security with cloud solutions
- Need to understand country-specific regulations on use of servers and data produced from them

- Create scalable, transportable systems for extracting follow-up information, outcomes
- Develop repository of SOPs for sample collection and storage, assays and protocols, assessments of validity/quality, analytical methods

2-IT considerations for enabling coordination, communication, centralization

- Need to define cohorts, what data stored and used
- Federation vs centralization
  - Centralizing reduces duplication, more sustainable funding
  - Federated addresses jurisdictional restrictions, sheer size, sustainability of smaller portals in question
    - ***Did we agree on this?***
  - Pharma works this way, most cohorts do not
- Standard interfaces needed, some extant examples
- Interoperable analysis a goal of GA4GH workstream
- Authorization and access: once fully authorized get expedited access
- Operationalizing consent– funding to harmonize and standardize consent, ensure access c/w consent

2-IT considerations for enabling coordination, communication, centralization

- Discoverability: searchable databases, GA4GH
- Develop ***Data Use Ontology*** to tag datasets with use restrictions
- Specifying hypotheses not always possible, enable exploration and pattern identification through hypothesis-free machine learning
- Challenge of legacy data
- Measuring impact
- Opportunity to negotiate with cloud venders

Highest priority:

- Work with Group 1 to get accurate registry
- Select data standard

3-Scientific agenda with short- and long-term goals

- Enhancements to existing cohorts:
  - Collect new samples for repeated measures or novel assays
  - Population-specific genotyping arrays
  - Standardize novel assay types: ***prioritize/choose***
  - Support cohort-wide genotyping and other –omic assays

    Did we agree on this?
  - Standardized phenotyping approaches esp with EHRs
  - Novel environmental measures and mobile technologies not possible through EHR linkage: ***prioritize/choose***
  - Generic data visualization methods

3-Scientific agenda with short- and long-term goals

- Scientific questions
  - Good examples in past, almost always require close collaboration and independent support
  - Determinants of health
  - Rare conditions or subgroups
  - Harmonize/standardize only what really matters– novel -omic assays
  - Develop systems for long-term health outcomes in LMIC
  - Facilitate research access to health outcome data, data protections

Will require engagement of funds with governments to convince relevance of research to health care

4-Policy agenda to facilitate and optimize impact of assembling these cohorts

Global Genomic Medicine Collaborative

International Cohorts Summit

- Challenges:
  - Defining “what’s in it” for… investigators, cohorts, countries
  - Differing institutional interpretations of regulations
  - Sharing samples more difficult vs. data/metadata
  - Differing approved uses by cohorts
  - Including native/aboriginal communities
  - Including for-profit entities
- Needs/Benefits: combine scientific and policy discussions
  - Define what collaboration trying to do, high-level principles
  - Develop/adopt international principles, build on GA4GH
  - Develop protocol for project review– selection and participation
  - Resource/platform to share lessons learned

4-Policy agenda to facilitate and optimize impact of assembling these cohorts

- Pilot efforts
  - Governance description
  - Cohort policy “traits” added to metadata
  - Identify current collaborations, what’s worked, lessons learned
- Understand implications of GDPR

  Engage primary funders for shared benefit and power
- Potential for common consent, at least for given project
- Develop pre-competitive spaces for industry to interact with cohorts
- Use policy frameworks for responsible sharing, obtaining consent, ensuring privacy– All of Us and GA4GH
- Develop international strategic agenda for CoC coordination

Global Genomic Medicine Collaborative

International Cohorts Summit

5-Developing a collaborative genomic sequencing (and other -omics?) strategy

- Key questions only through large-scale collaborations
  - Genomic and exposure diversity, ***migrant studies***
  - Rare diseases: human knock-out and homozygous deletions
  - Drug repurposing opportunities
  - Global problems: obesity, toxic exposures, alcohol-related diseases
  - Microbiome across ethnicities and exposures
- Very large projects will drive down costs
- Centralized coordinating function enabling queries to identify most informative cohorts for specific question

5-Developing a collaborative genomic sequencing (and other -omics?) strategy

- Types of data of most value
  - WGS, to be shared, WES not enthusiastic
  - Leveraging existing GWAS data and SNP-array genotyping ($50-100M)
  - Would need cohort-specific sequencing (few hundred?)
  - Phenotyping data (no sharing issues?) and metabolomic data
- Design collaborative sequencing strategy
  - Leverage existing GA4GH efforts at standardization, reduce artifacts
  - LMIC need funding, level playing field
- Methods/tools
  - Imputed SNP arrays rather than exomes (too much variability)
  - WGS data file harmonization (TOPMed > 100K)
  - May need charter or principles

5-Developing a collaborative genomic sequencing (and other -omics?) strategy

- Conceptualize as agnostic platform vs. science-driven questions
  - Choose handful of grand challenges: genomic variation (knock-outs) and exposure variation (alcohol)
- Genomics is easy: same in all cell types, stable
- Many complexities of proteomic data: population validation, interpretation, biologic variation
- Need close partnerships with developers of assays, can work iteratively to improve them
  - ***How to decide when assay ready for millions?***

- Large numbers reference samples in key subgroups (elderly?)

6-Translation and clinical impact

- Opportunities:
  - Advance practice– dx/px/rx; Mendelian, PGx, genetic risk scores
  - Drug development
  - Generic: health literacy, exemplars for teaching, evidence generation using large simple trials, learning healthcare systems
  - Population health and policy: new knowledge moves to policy
- Barriers:
  - Variable healthcare systems, disparities, diversity, evidence
  - Hand-off from evidence to implementation
  - Regulatory, reimbursement, ethics, “academic territorialism”
  - Engaging industry (real and perceived)

6-Translation and clinical impact

- Exemplar projects
  - Standardize implementation of RoR: FH or cancer
  - Country/cohort-specific risk prediction with standardized methods
    - GRS and non-genetic risk factors across ancestries

- Provide continuously updated estimates of individual risk and health behaviors of neighborhoods and populations
  - Mexico City Cohort: Poor glycemic control identified as likely cause of high mortality from renal disease and other causes, leading to a rapid public health intervention in Mexico
- Integration of personal, clinical, biological information

Compelling scientific questions addressable with millions of individuals

- Rare conditions (CKB venomous snakebite), subgroups, exposures
- Rare genotypes: human knock-outs project, extremes of risk
- Consanguinity and founder population studies: for collaborations?
- Critical bottlenecks: drive technology development
- Pilot studies:
  - Utilize repository of e-phenotyping algorithms (PheKB) and test transportability across different countries' EHRs
  - Apply NLP to cohort studies' data collection instruments (or consents?) to extract data just as doing with EHRs
  - Identify high-risk individuals for early disease detection, recognize when undetected disease biasing early outcomes

Funding needs

- Register and deposit data
- Review country-specific data access policies and ensure compliance
- Harmonize consents, re-consent
- Scalable phenotyping of outcomes: ascertainment (suspected cases), confirmation (case-ness), classification (subtypes, details)
- Collaborative analyses
- Adding value to existing cohorts– cohort wide assays, novel methods
- Patience: invest for long term, avoid pushing for quick publications
- Sequencing/genotyping support in LMIC to level playing field
- Support for open-source data platforms, analysis environments, data deposition

Possible outcomes from this meeting

- Creation of a searchable registry to facilitate collaboration across the cohorts– initially “members” vs broader global scientific community?
- Foundational principles for creating consortium of cohorts (CofC) and agreement to further explore creating it
- Identification of potential key work streams to create a foundation for a possible CofC
- Organizational entity to support exploratory activities– likely G2MC and GA4GH partnership
- Outreach to cohorts not in attendance
- White paper of opportunities and challenges
- Follow-up working groups, second summit

2-IT considerations for enabling coordination, communication, centralization

- Rory: bring researchers to data, convert images and other tech to data, provide creative and informative visualization, make possible to use and understand data
- Josh: cloud solutions, data biosphere
- Cathy: major potential output could be pilot of comparing and validating phenotype algorithm performance internationally
  - Not only disease outcomes but risk exposures, treatment
  - Some common data models have been used internationally

Improve prospects for interoperability and compatibility of instruments, data formats, phenotypic and clinical measures, etc.

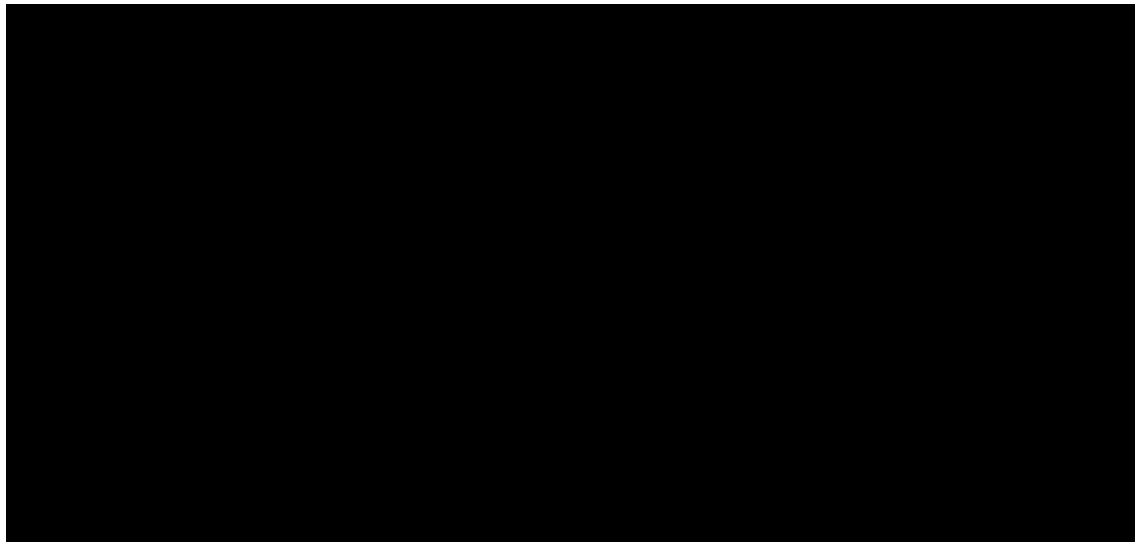
- Almost all

Promote data sharing and open access policies

Global Genomic Medicine Collaborative

International Cohorts Summit

- Rory: convert samples to data, whole cohort assays, industry support

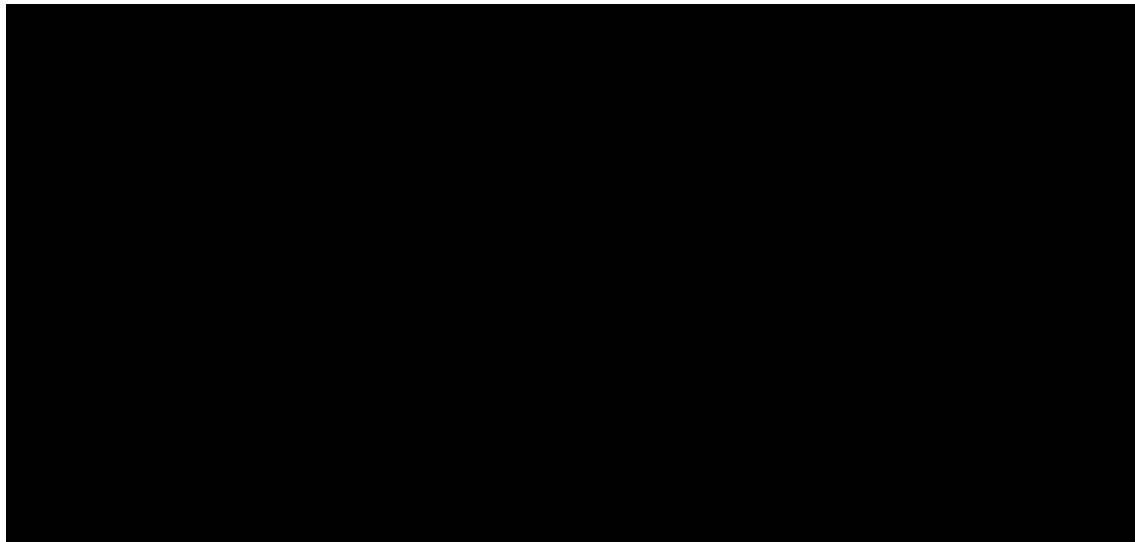


Broaden international cooperation through existing tools and resources

Global Genomic Medicine Collaborative

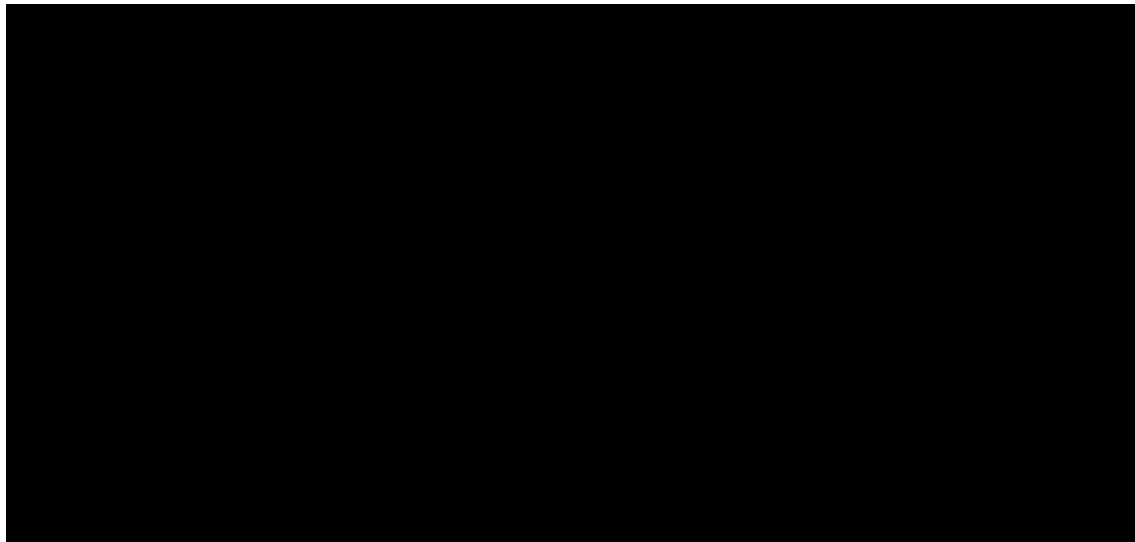
International Cohorts Summit

- Rory: create scalable, transportable systems for extracting follow-up information, outcomes
- Geoff: repository of SOPs for sample collection and storage, assays and protocols, assessments of validity/quality, analytical methods



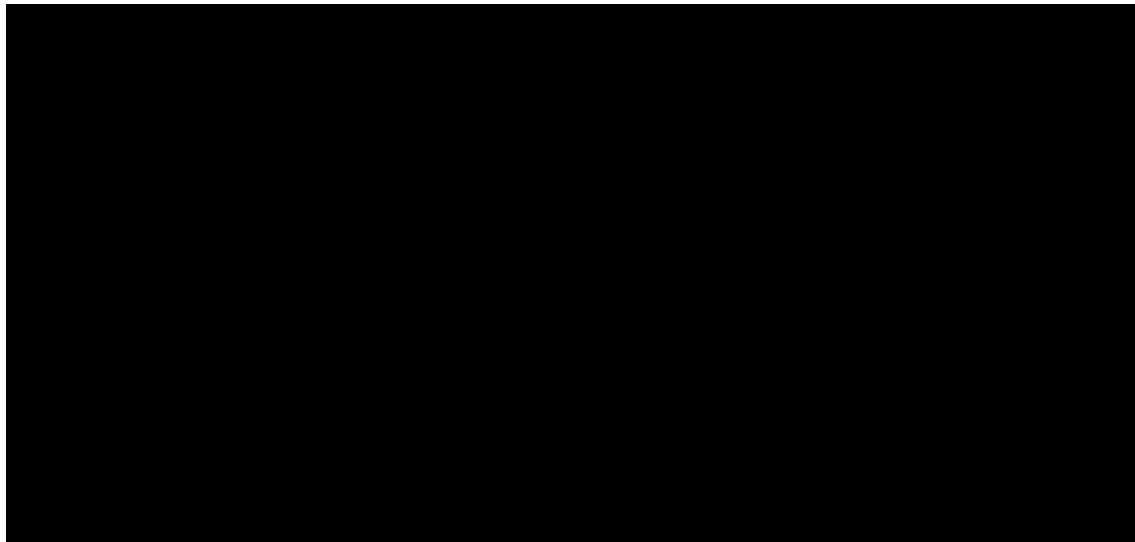
Explore the feasibility of a "digital" platform, or web-based, evolving registry of large-scale cohorts, in searchable format

- Almost all



Examine the potential for a collaborative sequencing (and other -omics?) strategy

- Rory: find ways to get samples from those who don’t have them?



Consider strategies for translating findings for health impact

- Almost all

Advance a collective vision: where do we want to be in ten years?

- Almost all