



# International Cohorts Summit

## EXECUTIVE SUMMARY

Duke University  
JB Duke Hotel, Ballroom ABC, 230 Science Dr, Durham, North Carolina, USA

March 26-27, 2018

*Hosted by the Global Genomic Medicine Collaborative (G2MC)*

### **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

## EXECUTIVE SUMMARY

The International Cohorts Summit, hosted by the Global Genomic Medicine Collaborative (G2MC, <https://g2mc.org/>) at Duke University in Durham, NC on March 26-27, 2018, and sponsored by All of Us, National Institutes of Health, Medical Research Council, and Wellcome Trust was conceived in 2015 by NIH and endorsed in June 2016 by the Heads of International Research Organizations (HIROs). The vision for success from this meeting and future collaborations is the creation of ***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.***

Cohorts represented at the Summit were selected based on 4 criteria: having 100K participants or more, not being disease-specific, having available biospecimens, and having at least the potential for longitudinal follow-up of participants.

Approximately 100 investigators from 24 countries representing 60 cohort studies attended and represented greater than 25 million people at their current sizes and greater than 36 million based on future recruitment targets, some with available data from the 1960s. From a pre-meeting survey, the majority of the represented cohorts have samples available, including DNA and genotyping, and almost half have whole genomic/exome sequencing data on at least some samples. Most indicated willingness to share data with appropriate patient consent, and agreed that the benefits of data sharing (enable increased cohort size, statistical power, associations, effect augmentation, and the advancement of scientific/medical knowledge/research and foster collaborations and new approaches/ideas) are understandably offset by the challenges of costs, regulations, and data harmonization.

Drs. Francis Collins and Jeremy Farrar spoke to the importance of combining large-scale cohort programs to encourage data sharing and pooling to improve scientific discovery, improve efficiencies, and maximize investments on a global scale. They emphasized that the scientific community has a responsibility to partner and share, to lead by working across borders and at a global level.

The five objectives for this Summit were:

- Improve prospects for harmonization of data, data formats, phenotype measures, consent, etc.
- Promote data and specimen sharing, and open access policies
- Examine the potential for a collaborative (global) sequencing project
- Explore the feasibility of a searchable on-line global registry of large-scale cohorts
- Create a vision: *Where do we want to be in ten years?*

The Summit was organized into 8 sessions to allow for the staging of the issues to be discussed during the breakout working sessions, and subsequent reports and identification of next steps. **Session 1** provided the goals for the meeting and set the stage for the value and challenges of combining large cohorts and the opportunities for translational impact for health. **Sessions 2 and 3** discussed the opportunities for collaboration in the broad areas of phenotype and outcomes data, biospecimen collection, genomic and other -omic information, environmental and nutritional information, and multi-ethnic data. **Session 4** addressed data standards and privacy and **Session 6** provided an overview of the EU experience in assembling “cohorts of cohorts.” **Sessions 5 and 7** were devoted to break-out groups in the following areas:

- Group 1: Creating a standardized database and registry
- Group 2: IT considerations for enabling coordination, communication, centralization
- Group 3: Scientific agenda with short- and long-term goals
- Group 4: Policy agenda to facilitate and optimize impact of assembling cohorts
- Group 5: Developing a collaborative genomic sequencing (and other -omics?) strategy
- Group 6: Translation/clinical impact.

**Session 8** concluded the Summit with a Summary of the Break Out group report-outs (See session 8, page 15 of meeting summary) and outline of possible outcomes as articulated by Drs. Collins and Farrar.

Drs. Collins and Farrar emphasized the enormous potential for great benefit to the general population and desire to assist with realizing the possible outcomes. They encouraged the exploration of opportunities and synergies with funding outside of NIH, Wellcome Trust, and other existing funders, and suggested a model of providing support for cohorts from within each country, similar to the genome project model and national infrastructure. Emphasis was placed on investment for the long term (decades) with periodic looks (every few years) to ensure appropriate productivity, timeliness, leadership and governance. They encouraged sharing best practices across cohorts to enable more effective approaches world-wide, rather than imposing a single unifying structure, and highlighted that joining a consortium could have significant benefits to individual cohorts, which will have much stronger voices with the power of this community behind them than if speaking alone.

Summit attendees articulated an initial set of compelling scientific questions (Breakout Report Summary Slide 15; also see page 19 of Meeting Summary) that would be addressable through the access to millions of individuals, such as investigating rare conditions and genotypes, enabling consanguinity and founder population studies, addressing bottlenecks with new technology development, and initiating novel pilot studies. The summit attendees recognized that this desired global platform however would require funding and other resources to address several initial desirable goals: registration and data deposition, review and compliance of country-specific data access policies, ensuring consent or re-consent processes, sequencing/genotyping support, and support for open-source data platforms and analysis platforms.

The next steps and possible outcomes from this Summit include:

- Creation of a searchable registry to facilitate collaboration across the cohorts– initially “members” vs broader global scientific community
- Establishment of 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
- Creation of an 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 to be planned in China as offered