



# International 100K Cohort Consortium

## EXECUTIVE SUMMARY

Grand Hotel  
Gullteigur Conference Room, Reykjavik, Iceland

April 23-24, 2019

*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

In 2015, the National Institutes of Health (NIH) launched an effort to identify all large-scale prospective cohort studies involving at least 100,000 (100K) participants to explore the potential of bringing them together to address scientific questions none could answer alone. The goal was, and remains, to create a global network for translational research that will utilize large cohorts to enhance understanding of the biological, environmental, and genetic basis of disease and to improve clinical care and population health. This effort was discussed at the June 2016 Heads of International Research Organizations (HIROs) meeting and reached an agreement to commission the Global Genomic Medicine Collaborative (G2MC), working with the Global Alliance for Genomics and Health (GA4GH), to bring the cohorts together for an initial summit, which was held in March 2018 in Durham, North Carolina. It involved 61 cohorts from 33 countries, representing current and expected cohort sizes of 30 million and 35 million individuals, respectively.

Cohorts were invited based on four criteria: having 100K participants or more, were not selected for disease, having available biospecimens, and having at least the potential for longitudinal follow-up of participants. It soon became clear, however, that regions such as Africa, the Middle East, South America and South Asia were under-represented in such a compilation, so these criteria were relaxed somewhat to facilitate their participation. The outcome of the summit and initial plans for an International Hundred Thousand Plus Cohorts Consortium (IHCC) were discussed at the June 2018 HIROs meeting. They reached an agreement to support a broader effort to develop IHCC and hold a second summit in spring 2019. Cohort representatives were invited to volunteer for three teams focused on data standards and interoperability, scientific strategy and cohort enhancements, and policy and data sharing.

On April 23-24, 2019, at the Grand Hotel in Reykjavik, Iceland, G2MC hosted the second International Cohorts Summit. The objectives of this meeting were to identify scientifically meritorious cross-cohort research projects and collaborators willing to participate in them and to develop a scientific agenda for IHCC to bring to funding bodies. Prior to the Summit, IHCC solicited a Request for Ideas (RFI) to review and discuss during the second summit.

Approximately 117 representatives from 59 cohorts and 29 countries attended. Twenty-three proposals from the RFI solicitation were included for discussion at the meeting. The proposals encompassed several key themes, including data platforms, -omics studies, rare variant detection and human knock-outs, diversity and ethnicity gaps, and polygenic risk scores.

The Summit was organized into nine sessions so Teams could discuss issues during their breakout working sessions for subsequent reports and to identify next steps.

Five sessions were held on the first day of the conference: **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. **Session 2** discussed opportunities for collaboration across broad geographies. **Session 3** addressed data standards and infrastructure in accordance with the Data Standards and Infrastructure Team focus. **Session 4** provided an overview of IHCC's scientific strategy and enhancements of our cohorts per the focus of the Scientific Strategy and Cohorts Enhancement Team. **Session 5** discussed current policy and data sharing challenges per the focus of the Policy and Bio-Data Sharing Team.

Four sessions were held on day two, which started with a keynote address about big data and opportunities in healthcare, followed by a charge to the breakout groups. **Session 6** hosted the breakout sessions for each of the working group teams. The team leads then gathered during **Session 7** for the Team Lead synthesis in preparation for **Session 8**, which provided reports from the breakout sessions and discussion of synergies. **Session 9** concluded the Summit with a summary of progress and action plans from the meeting.

Each of the teams progressed toward their goals during the summit. The Data Standards and Infrastructure Team identified two project proposals to establish data infrastructure and interoperability standards, starting broadly with a Cohort Discovery Atlas, then narrowing focus with a driver project involving a few large cohorts from discovery to federated analysis. The Scientific Strategy and Cohorts Enhancement Team focused the scientific agenda on projects that could demonstrate the cohesiveness of IHCC and the value of the collective group. They acknowledged a variety of possible projects that might use existing data, some of which could be low-resource, quick deliverables for IHCC, as well as possible projects that would require generating new data. The Policy and Bio-Data Sharing Team refocused their primary goal from purely data sharing to generating *collaboration among cohorts*, which might enable data sharing. They will develop a “first principles” document that would define the value of participation along with guidance for fair collaboration.

In addition to the team progress, the meeting revealed considerable interest in becoming a more formal consortium with a charter, guiding principles, membership guidelines, etc. Meeting attendees emphasized the need to engage under-represented populations and cohorts in low and middle income countries (LMICs) to advance the global scientific agenda.

The next steps and outcomes from this Summit include

- Develop and distribute a summary of the meeting
- Develop exemplar proposals for presentation to HIROs from each IHCC Team (Data Standards, Scientific Strategy, and Policy/Bio-Data Sharing)
- Bring priorities to the HIROs
- Develop a charter and guiding principles for the consortium, potentially including a more formalized sign-up process and consideration of the expectation that Team co-leads cover the breadth of time zones to promote inclusivity
- Conduct a second solicitation of RFIs with sufficient advance notice, time to respond and proposal reviews by a broader section of the consortium
- Promote opportunities for trainees and junior investigators to participate in IHCC activities
- Circulate the list of cohorts in attendance and those that IHCC has knowledge of to ensure that no cohorts have been overlooked
- Based on the teams’ and proposed projects’ progress, host a third summit in 2020 that includes as many cohorts represented as possible and industry participation. This summit will present progress on the ongoing projects.