International 100K Cohort Consortium

MEETING SUMMARY

Virtual Summit
Zoom

May 4-5, 2020

Hosted by the [Global Genomic Medicine Collaborative](ihccglobal.org) (G2MC)
ihccglobal.org

Vision for success:

TO CREATE A GLOBAL NETWORK FOR TRANSLATIONAL RESEARCH THAT UTILIZES LARGE COHORTS TO ENHANCE THE UNDERSTANDING OF THE BIOLOGICAL AND GENETIC BASIS OF DISEASE AND IMPROVE CLINICAL CARE AND POPULATION HEALTH

Global Genomic Medicine Collaborative

International 100K Cohort Consortium

IHCC Third International Cohorts Virtual Summit 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 participants to explore the potential of bringing them together to address scientific questions none could answer alone. This effort led to the commission of the Global Genomic Medicine Collaborative (G2MC) to bring together these cohorts through the International HundredK+ Cohort Consortium (IHCC). This group gathered for the First International Cohort Summit in the USA in 2018 followed by a Second Summit in Iceland in 2019. During planning of the Third Summit to take place in Santiago, Chile, the global outbreak of COVID-19 required a transition to virtual meeting format. Approximately 160 attendees from more than 23 countries attended the virtual Summit. The event in Santiago, Chile will be postponed to a date to be determined when attendee health and safety will not be at heightened risk due to COVID-19.

The virtual meeting objectives included:

- To galvanize the IHCC around a visionary charter and path forward (defining the IHCC organization, mission, membership, partnership opportunities, industry engagement).
- To examine how IHCC can rapidly mobilize worldwide cohorts to address the COVID-19 pandemic.
- To introduce the IHCC to a Cohort Data Atlas that can be used to stimulate and enable collaborations among cohorts.
- To engage the entirety of the IHCC membership in developing the key topics to chart a scientific agenda that can only be achieved by assembling cohorts and their data.

The Summit included keynote presentations from Francis Collins (National Institutes of Health (NIH), USA), Soumya Swaminathan (World Health Organization (WHO), Switzerland), and Jeremy Farrar (Wellcome Trust (WT), UK) along with the following sessions:

- Session 1 – IHCC Work Team Progress – Governance
- Session 2 – IHCC Work Team Progress – Science and Technology
- Session 3 – Scientific Presentations across each Work Stream
- Session 4 – External Engagement (Other Consortia and Industry Partners)

In Session 1, attendees learned details of the IHCC Charter outlining governance, membership, and expectations for participation. IHCC Policies on publications, collaborations with industry, and data sharing were also presented for review by membership. Discussion among attendees and presenters illustrated support for the Charter and Policies. These will be ratified following the Summit.

Session 2 included a live demonstration of the IHCC Cohort Data Atlas prototype, enabling cohort discovery of cohort variables, specimens, and populations of interest. User testing of the Atlas will continue in the coming months as additional cohort data is mapped and uploaded to the browser. The Scientific Strategy Team presented preliminary results of an IHCC cross-cohort pilot project on polygenic risk scores for four selected traits. A publication is in progress to highlight the proof of principle for the cross-cohort analysis as well as the Polygenic Risk Score (PRS) prediction results. Three to five additional pilot projects will be initiated on a 12-18-month timeline with funding support from the NIH and the WT.

In Session 3, presenters shared scientific progress across a variety of domains including data infrastructure, equity in polygenic risk scores, standardized phenotype measures, precision medicine initiatives, and principles and codes of conduct for collaboration.

Session 4 highlighted opportunities for collaboration though the Davos Alzheimer's Collaborative (DAC), Global Alliance for Genomics and Health (GA4GH), International Common Disease Alliance (ICDA), Global Biodata Coalition (GBC), and Global Genomic Medicine Collaborative (G2MC), and industry partners such as Regeneron, GlaxoSmithKline (GSK), and Illumina.

In recognition of the current global research priorities, IHCC is collectively launching several COVID-19 response initiatives. Presenters shared current research on COVID-19 surveillance, sex and age disparities, mental health, and host response variants. Several Scientific Working Groups (SWGs) will be established to examine the global expansion of COVID-19, the impact of mental health and environmental exposures on COVID-19, and biospecimen standards development.

Discussion during the Summit led to several key actions in accordance with the meeting objectives:

Objective One: *To galvanize the IHCC around a visionary charter and path forward (defining the IHCC organization, mission, membership, partnership opportunities, industry engagement).*	**Key Actions/Next Steps:** • Provide additional feedback on IHCC Charter as needed. A formal ratification process will follow in June 2020. • Provide responses to the data sharing survey by June 5, 2020 to inform the IHCC Core Data Sharing Principles. • Provide additional comments on policy documents as needed by May 22, 2020 to Laura Lyman Rodriguez. Final versions will be approved by the SSC and distributed to membership. • IHCC members interested in joining the working groups (Policy and Biodata Sharing, Data and Infrastructure, or Scientific Strategy) are encouraged to contact the Team leads or IHCC Secretariat.
Objective Two: *To examine how IHCC can rapidly mobilize worldwide cohorts to address the COVID-19 pandemic.*	• Provide responses to the IHCC COVID-19 cohort survey if not already complete (will be recirculated via email to cohorts). • Indicate interest in joining any of the COVID-19 SWGs by May 12, 2020. Contact Eric Plummer to join. • Cohorts measuring the impact of protective mental health interventions for COVID-19 are encouraged to contact Wellcome Trust (Jordan Smoller, Andre Brunoni, Sarah Bauermeister). • For COVID-19 research studies, register on the ICDA COVID-19 Host Genetics Initiative.
Objective Three: *To introduce the IHCC to a Cohort Data Atlas that can be used to stimulate and enable collaborations among cohorts.*	• Cohorts may continue to provide data dictionaries for IHCC Cohort Data Atlas development to ihcc-browser@googlegroups.com. IHCC cohort data will be further populated in the atlas. Members are encouraged to provide feedback to Data and Infrastructure Team on use cases for atlas queries. This Team will establish a compelling set of research and clinical showcase applications. COVID-19 phenotyping will be added to the Atlas. Cohorts interested in cross-cohort COVID-19 research may also consider providing data dictionaries for harmonization and cohort discovery. • An IHCC Resource Center with relevant data tools and other cohort resources will be created and shared on the IHCC website.

Objective Four: *To engage the entirety of the IHCC membership in developing the key topics to chart a scientific agenda that can only be achieved by assembling cohorts and their data.*	• IHCC will collectively develop a five-year roadmap. • To join the Davos Alzheimer's Collaborative, contact Drew Holzapfel and/or attend the scheduled calls on May 19/20, 2020. • Scientific Strategies team will launch three to five scientific initiatives with a 12-18-month timeframe using a federated approach with an emphasis on global diversity more than -omics. • Collaborate on the COVID-19 specific activities (see Objective Two) • Continue to develop a shared scientific agenda through implementation of the Charter with new governance structure (elected members of the Steering Committee, etc.) and virtual working group meeting in ~six months.

Abbreviations

ACTIV	Accelerating COVID-19 Therapeutic Interventions and Vaccines
AD	Alzheimer's Disease
AUC	Area under the curve
BMI	Body mass index
COVID-19	Coronavirus Disease
DAC	Davos Alzheimer's Collaborative
dbGaP	Database of Genotypes and Phenotypes
EC	Executive Committee
EHR	Electronic Health Record
ELISA	Enzyme linked Immunosorbent Assay
G2MC	Global Genomic Medicine Collaborative
GA4GH	Global Alliance for Genomics and Health
GBC	Global Biodata Coalition
GDPR	General Data Protection Regulation
GIS	Geographic Information System
GSK	GlaxoSmithKline
GWAS	Genome-Wide Association Study
HFSP	Human Frontier Science Program
HL7	Health Level Seven International
ICDA	International Common Disease Alliance
ICGC	International Cancer Genomes Consortium
ICS	International Cohorts Summit
IHCC	International HundredK+ Cohorts Consortium
ISO	International Organization for Standards
LMIC	Low- and/or Middle-Income Country
LoF	Loss-of-Function
LPS	Longitudinal Population Study
MAMA	Multi-Ancestry Meta-Analysis
MERS	Middle East Respiratory Syndrome
NIH	National Institutes of Health (USA)
NIMH	National Institute of Mental Health (USA)
PM	Precision Medicine
PRS	Polygenic Risk Score
PVI	Pandemic Vulnerability Index
RCT	Randomized Control Trial
SSC	Scientific Steering Committee
SARS	Severe Acute Respiratory Syndrome
SARS-CoV-2	Severe Acute Respiratory Syndrome Coronavirus 2
SNOMED	Systematized Nomenclatures of Medicine
SWG	Scientific Working Group
T2D	Type 2 Diabetes
WEF	World Economic Forum
WGS	Whole genome sequencing
WT	Wellcome Trust

Global Genomic Medicine Collaborative

International 100K Cohort Consortium

INTERNATIONAL COHORTS VIRTUAL SUMMIT 3

Hosted by the Global Genomic Medicine Collaborative (G2MC)

Virtual Summit, Zoom
May 4-5, 2020

SUMMIT MINUTES

Day 1: May 4, 2020, 11:00 – 17:00 UTC

SUMMIT INTRODUCTION AND BACKGROUND
CHAIR: GEOFFREY GINSBURG, MD, PHD
11:00-11:30 UTC

Welcome and Introductions – Geoffrey Ginsburg, MD, PhD *(Duke University, USA)*
The International HundredK+ Cohort Consortium (IHCC) is hosted by the Global Genomic Medicine Collaborative (G2MC) with co-chairs Geoff Ginsburg, Peter Goodhand, and Teri Manolio. The Third International Cohorts Summit (ICS3) was moved from its original location in Santiago, Chile, to a virtual format in consideration of the outbreak of COVID-19.

The vision of IHCC is to create a global network of large cohorts (with multi-dimensional data from diverse populations) for translational research that will be maximally utilized to enhance scientific understanding of the biological, environmental, and genetic basis of disease and to improve population health. By creating a consortium, the IHCC can answer globally relevant scientific questions with combined data that no individual cohort could answer alone. Cohort membership requires more than 100,000 participants, disease-agnostic selection, availability of biospecimens (current or future), and the potential for longitudinal follow up. There are currently 97 cohorts involved with the IHCC encompassing more than 50 million participants globally, with some gaps in representation across South America and Africa.

In late 2018, IHCC formed three workstreams around data infrastructure for sharing and discoverability, a scientific plan that leverages cross-cohort collaboration, and a policy agenda to support data sharing and collaboration. Additional working groups have been formed in the last 12 months for the IHCC Charter development, COVID-19 research strategy, and funding support. The Steering Committee and Executive Committee support the advancement of these work groups.

The virtual meeting objectives include:

- To galvanize the IHCC around a visionary charter and path forward (defining the IHCC organization, mission, membership, partnership opportunities, industry engagement).
- To examine how IHCC can rapidly mobilize worldwide cohorts to address the COVID-19 pandemic.
- To introduce the IHCC to a Cohort Data Atlas that can be used to stimulate and enable collaborations among cohorts.
- To engage the entirety of the IHCC membership in developing the key topics to chart a scientific agenda that can only be achieved by assembling cohorts and their data.

Keynote – Francis S. Collins, MD, PhD *(National Institutes of Health, USA)*
Since the global outbreak of COVID-19, the US National Institutes of Health (NIH) has shifted many resources toward combating this disease. The NIH is working on several vaccine programs, one of which progressed to Phase 1 clinical trials just 63 days after obtaining the SARS-CoV-2 sequence. This Phase 1 trial, in collaboration with Moderna Therapeutics, will be moving into a combination Phase 2/3 trial as early as July, with results possibly available by the fall. Research on therapeutics is also underway at the NIH. Early results of the randomized control trial (RCT) for remdesivir showed improvement in survival and length of hospital stays, but combination therapy may be needed for more significant impact on clinical outcomes. In early April, NIH organized the Accelerating COVID-19 Therapeutic Interventions and Vaccines (ACTIV) partnership across numerous pharmaceutical companies, academic institutions and government agencies in the US and Europe to develop a coordinated research strategy. The research community has been collaborating to develop large clinical trial networks in advance of RCTs and sharing promising results to advance progress globally. The USA Congress has mandated development of point-of-care diagnostic testing for COVID-19 so that rapid and widespread testing is available as communities begin to re-emerge from quarantine.

With so much variation in the manifestations of COVID-19 (asymptomatic cases, mild to severe symptoms, mortality), risk factors such as age and comorbidities do not seem to fully explain the progression of the pandemic. Research is needed to explore these modifiers; like why some high risk individuals recover from COVID-19 while others do not, why therapeutics like remdesivir are effective for only certain individuals, and why some cases of the disease have mild or no symptoms. Highly powered studies like those possible within the IHCC are uniquely positioned to examine these global variations, maximizing the utility of cohorts with consent for recontact. Reconsented participants with existing genomic data may aid understanding of the host genome response to SARS-CoV-2. Vaccine trials will require many thousands of participants in areas where the virus is still circulating; involving participants with genomic data would be extremely valuable to deepen the understanding of the disease. IHCC has progressed substantially since its inception just over two years ago, now able to work on real projects and supporting population diversity amongst cohorts. The NIH will continue to be invested in the IHCC's progress and looks forward to competitive grant submissions from the Consortium.

SESSION 1 – IHCC WORK TEAM PROGRESS - GOVERNANCE
CHAIR: GEOFFREY GINSBURG, MD, PHD
11:30-14:15 UTC

Charter Work Team Progress Update – Mary De Silva, MSc, PhD *(Wellcome Trust, UK)*
During the Second International Cohorts Summit (ICS2) in April of 2019, attendees expressed the need to draft a charter to guide engagement with the IHCC. After a call to members to join a Charter working group, a small group of individuals representing a diversity of interests and geographies was formed:

- Chair: Mary De Silva, Wellcome Trust, UK
- Secretariat: John Connolly, Children's Hospital of Philadelphia, USA
- Brazil: Alexandre Pereira, Laboratory of Genetics and Molecular Cardiology (Sao Paulo Brazil Cohort)
- Iran: Arash Etemadi, NCI (Golestan Cohort Study & PERSIAN Cohort)
- Japan: Norihiro Kato, National Center for Global Health and Medicine
- South Africa: Kobus Herbst, African Health Research Institute (South African Population Research Infrastructure Network - SAPRIN)
- Spain: Josep Maria Haro, Saint John of God Health Park Barcelona (SYNCHROS Cohort)
- UK: Anthony Swerdlow, Institute of Cancer Research (Generations Study, UK)
- USA: Elizabeth Jensen, Wake Forest School of Medicine (Environmental influences on Child Health Outcomes - ECHO)

The group first met in October of 2019 and an initial draft was available by November of 2019. After incorporating comments from the Steering Committee and Policy Team, the Charter was revised and circulated to IHCC membership for comments in April of 2020. No major comments or concerns were received from membership prior to the Summit.

The IHCC Charter outlines membership requirements and expectations and will link to the policies that will guide IHCC member activities and collaborations. It is not a legally binding document but rather a Membership Agreement guided by principles of transparency, fairness, inclusivity, and equity by which the members agree to participate. The main points of the Charter are summarized at a high level below, but the full text of the Charter (ratification pending) can be accessed here.

Governance
The Charter formalizes the existing IHCC Governance structure and proposes minor changes. The seven-member Executive Committee (EC) will include three co-chairs with five-year renewable terms and four additional members nominated by IHCC membership for three-year terms. All seven members will be voted on the EC by the Scientific Steering Committee (SSC). The EC is responsible for operationalizing the IHCC vision and strategy by acquiring and managing funding, directing administrative activities, assigning priorities to working groups, and resolving disputes.

The SSC will include 15 voting members comprised of cohort members that have been nominated and elected by IHCC membership for three-year terms. Working group chairs will also be among the 15 voting members of the SSC, elected by the other members of the SSC for three-year terms. The EC may observe and participate in the SSC, but do not have voting rights. The SSC will be responsible for approval of new IHCC members, election of new EC members, guiding the EC on long term strategy, approval of program goals set out by the EC, as well as commission, review, and formal approval of IHCC policies and procedures. The SSC will include a Scientific Projects Sub-Committee, chaired by a member

of the SSC. This committee will review and approve IHCC scientific projects, including those involving partnerships with external organization or Industry Members.

There will be various work groups to be altered as the needs of IHCC evolve (e.g. according to the 5-year strategic plan). There are currently three work groups: Data Standards and Interoperability, Scientific Strategy and Cohort Enhancements, and Policy/Bio-data Sharing. These groups are open to membership for collaboration and led by SSC-elected IHCC members with relevant expertise. Sub-groups may be formed (e.g. project teams), but these sub-groups will not be individually represented on the SSC.

Membership
IHCC will recognize three types of members.

- *Cohort* members will include cohorts from any country with more than 100,000 enrolled participants, without selection for a specific disease/condition, longitudinal follow up, and biological samples possible. Cohorts with less than 100,000 participants can apply to be a Cohort Member if they include low- and middle-income country/countries, and/or cohorts of disadvantaged populations in high income countries, and/or collect data from exceptional, unique, or difficult to accrue groups. Cohort members have voting rights for IHCC-wide decisions (one vote per cohort).
- *Affiliate* members may not meet the requirement of full cohort membership and voting rights but bring specific valuable expertise and perspective to the IHCC. This category may also include cohorts planning to enroll more than 100,000 participants in the future, or cohorts with more than 100,000 participants who do not intend to collect biological samples.
- *Industry* members may participate in IHCC in accordance with the *Collaboration with Industry Partners* Policy. Industry members may also qualify separately as Cohort members if they meet the *Cohort* member criteria. While they will also not have voting rights, participating industry members will strengthen collaborations with academia and commercial researchers.

Next Steps
The Charter will be a living document with a process for both major revision (IHCC membership ratification) and minor revision (SSC vote) approvals. The IHCC Core Data Sharing Principles, Publication Policy, and Guidance for Collaboration with Industry are referenced in the Charter, allowing for evolution of the Charter along with IHCC needs. The Charter will be ratified by vote among IHCC membership in the coming weeks. Additional comments are welcomed for incorporation prior to ratification. Attendees of the virtual Summit voted with preliminary support of the current Charter content.

Policy and Bio-Data Sharing – Laura Lyman Rodriguez, PhD *(USA)*
The IHCC work group on Policy and Bio-Data Sharing has developed three policy/guidance documents guided by the IHCC vision and values. Each document was identified as a need based on prior Summits and meetings of the IHCC Steering Committee. They are open for discussion with membership prior to approval. Final approved versions of the policies below will be available on the IHCC website.

(1) Publication Policy
This document will support timely and open dissemination of IHCC research findings as well as fair and equitable opportunity for members to participate in authoring IHCC publications. Publication types covered by this policy include *Consortium Papers*, *IHCC Project Papers,* and *Individual (or non-IHCC) Collaboration Papers. Consortium* papers are developed and managed by the IHCC SSC and open to all

IHCC members for contribution. These publications might illustrate IHCC policies, views, or methods. *Project* papers will stem from IHCC cross-cohort project proposals that utilize IHCC resources and infrastructure. Project-related papers describing broad collaborative efforts across IHCC members may be considered to be *Consortium Papers* while *Individual Collaboration* papers may involve work of a small number of IHCC cohorts but do requiring development and review as an official IHCC project through the Scientific Projects Sub-Committee.

This policy also provides guidance for authorship and acknowledgements, timeline for dissemination of results and the use of open-access formats. *IHCC Project* proposals will require a publication and authorship plan addressing elements included within an example template included in the policy. The plan will include tentative authors and expected contributions, target journal, and expected milestones toward publications. While one project may result in multiple publications, only one publication/authorship plan is expected to be submitted with the project proposal.

Several discussion points were identified by IHCC members for further consideration with the Policy and Bio-Data Sharing Team:

- Defining a *Consortium* paper: minimum criteria, number of participating cohorts?
- *Consortium* papers: authorship/acknowledgement of individuals vs. cohorts given Journal policies. Ideally individuals may be named in authorship so they may be cited. SSC may draft a standard acknowledgement statement and/or an appendix of the individuals representing each cohort.
- Defining an *IHCC Project* vs. *Individual* paper: minimum criteria, number of cohorts involved for an *IHCC Project* publication?
- Review timeline: Current proposal is three weeks for review/contribution from authors. This would also be the timeline for institutions with an internal review process.

(2) Guidance for Collaboration with Industry

This document will support the establishment of IHCC collaborations with industry partners through agreements that promote the IHCC vision and values of transparency, equity, and inclusivity. Three types of collaboration are covered in the guidance:

- Collaboration between industry-led and non-industry-led IHCC cohort members
- Collaboration between industry and IHCC cohort members for data generation or analysis
- Resource provision from an industry partner to support an IHCC project in exchange for access to data or analysis

The guidance provides principles upon which these collaborations should be built including fairness and equity, transparency in aims and interest, and the timely and open dissemination of research data and findings. Each collaboration should provide clear value with unique scientific opportunity or resources and/or building capacity when appropriate. An Industry Collaboration Checklist is included in the guidance document appendix. The checklist includes points of discussion with partners such as regulatory issues, indemnity, fairness and equity, responsibilities of involved parties, governance, transparency, intellectual property, and dissemination timelines.

Partnerships with tobacco or nicotine industry is strongly discouraged for collaborative IHCC projects; however individual cohorts reserve the right to make autonomous decisions on their activities, partnerships, and funding sources outside of the IHCC.

Several discussion points were identified by IHCC members for further consideration with the Policy and Bio-Data Sharing Team:

- Types of industry collaboration: comprehensive list?
- Standards for declarations/disclosures currently reference the International Committee of Medical Journal Editors: other appropriate standards?
- Defining timely dissemination of findings: 12 months post-data cleaning adequate?
- Restrictions on collaboration with specific industries: Perhaps add alcohol to this list. Consider upgrading the recommendation to a strict prohibition against collaboration for IHCC projects.

(3) Core Data Sharing Principles

This document establishes fundamental data sharing expectations for IHCC activities and provides a foundation for future development of data sharing policies. Principles were drafted based on survey responses from IHCC cohort members. The data sharing survey remains open for additional responses to further guide IHCC strategies for data sharing.

The current draft outlines foundational principles of what is immediately required to collaborate as well as aspirational principles to facilitate efficient long-term data sharing. The principles are designed to work in conjunction with local cohort needs and legal requirements and assume the appropriate participant consent and oversight is in place.

Foundational principles include:

- Data access will occur through robust data management practices consistent with research participant/data donor consent.
- All institutional, local, national, etc. laws and norms will be followed.
- IHCC generated findings will be accessible to the scientific community within 12 months of data cleaning.
- No individually identifying information will be shared without appropriate approvals.
- IHCC members participating in approved IHCC projects will promote maximally appropriate sharing of cohort data.

Aspirational principles include:

- Streamlining data access procedures consistent with ethics oversight for long term collaboration.
- Supporting maximally appropriate secondary research access for the scientific community to relevant cohort metadata and summary results.
- Optimization of access procedures to relevant cohort data for the scientific community consistent with data donor consent within two years.

Several discussion points were identified by IHCC members for further consideration with the Policy and Bio-Data Sharing Team:

- Ethics principles based on GA4GH Framework for the Responsible Sharing of Genomic and Health Related Data standards; others to incorporate?
- Cohort membership requires willingness to share data or metadata. Is there a minimum or core set of data that should be required for sharing? Consider maximally appropriate data sharing as

a requirement with details on what is deemed appropriate. Will sharing of limited metadata alone be sufficient for membership?
- Federated data access should be discussed as a method of sharing for cohorts with restrictions on data movement.
- For data sharing outside of IHCC, how broad of sharing is required?
- Capacity building for data management in low-resource settings may be appropriate as an aspirational principle. Responding to the data sharing survey will aid in identifying needs.

IHCC members are encouraged to share feedback on any of the points of discussion from the Summit or other content in the presented policy documents by contacting Laura Lyman Rodriguez by May 22nd, 2020.

Mobilizing IHCC Cohorts to Address COVID-19 Pandemic – Keri Atlhoff, PhD, MPH *(Johns Hopkins Bloomberg School of Public Health, USA),* Kelly Gebo, MD, MPH *(All of Us Research Program, USA)*
In light of the COVID-19 pandemic with 3.2 million cases and more than 233,000 deaths globally, the IHCC is uniquely positioned to contribute to research questions of global importance. The five speakers below shared early findings of COVID-19 specific research and/or opportunities to provide a basis and inspire discussion for IHCC COVID-19 research.

(1) Liz Cirulli, PhD (*Helix, USA*)
Helix offers an Exome+ assay that provides both clinical grade exome information plus the GWAS backbone. They have Exome+ data from 200,000 individuals for research and development purposes with consent for recontact. With these resources, Helix set out to identify some of the genetic variants influencing the human response to SARS-CoV-2, starting with an examination of the SARS-CoV-2 receptor (ACE2 protein) coding variants. Helix researchers found 332 coding variants including many missense variants, several loss-of-Function (LoF) variants, and several copy number variants.

Among the missense variants were 40 coding variants that affect the binding regions for SARS-CoV-2 (present in ~1% of the sample population). These variants may block binding of ACE2 to SARS-CoV-2 and thus provide partial or full immunity while variants that enhance binding would provide increased susceptibility. LoF variants in ACE2 are known to be poorly tolerated based on publicly available data. This gene is on the X-chromosome and thus fewer LoF alleles were found in males vs. females. The effective difference in sample size for the X-chromosome between males and females was accounted for by using allele frequency rather than the proportion of individuals with the variant. It is not yet definitively known if the LoF variants stop SARS-CoV-2 from entering cells but may be demonstrable with more data. These results are available on the Helix blog.

Helix recontacted the genotyped individuals to obtain COVID-19 phenotypic data. Approximately 25,000 respondents answered a COVID-19 survey, with 17% reporting exposure to COVID-19. Twelve percent of respondents experienced a fever during the pandemic, 6% with additional COVID-19 symptoms and 3% were tested for COVID-19. Of those 3% tested, 10% tested positive. Helix has begun to analyze the positive cases' phenotypic data including exposures, risk behavior, and clinical symptoms with their genotypic data. There are not yet clear signals from genetic variation, however, Helix is working to increase sample size by joining international efforts such as the International Common Disease Alliance (ICDA)-hosted COVID-19 Host Genetics Initiative.

(2) Eileen Scully, MD, PhD (*Johns Hopkins, USA*)

Sex is a known modifier in the immune response to viral infections such as hepatitis C, influenza, and HIV. These effects vary with age: females show more severe disease during reproductive years while males show more severe disease during early/late stages of life. For this discussion, sex will be defined by chromosomes while gender is defined within the social-behavioral context. Previous coronaviruses such as Severe Acute Respiratory Syndrome (SARS) in 2003 and Middle East Respiratory Syndrome (MERS) in 2013-2014 had higher mortality rates in males, but some increased risk was associated with gender-related exposures. Early SARS-CoV-2 data from China demonstrated a higher mortality rate, higher risk of death among the severely ill, and increased risk association from comorbidities in males. Although there is not a significant disparity in the number of cases for males vs. females or likelihood of infection after exposure, global mortality rates support the sex disparities observed in the early China data.

Differences in testing methods have also suggested that women are more commonly asymptomatic. In South Korea, screening was widespread rather than symptom-based and 60% of identified cases were among women. This may be compared to symptom-based screening in China identifying 55% of cases in males. When China conducted screening based on contact-tracing instead of symptoms, only 28% of the cases were among males. These early data are not yet adjusted for comorbidities or other risk factors such as smoking. Given the likelihood of many asymptomatic cases among females going undetected, this suggests that the case-fatality ratios between males and females are even larger than currently documented. These disparities in mortality persist beyond reproductive years but are not yet observable in children. The significant presence of asymptomatic infection among females support the need for testing beyond those who are symptomatic or high risk due to comorbidities.

These observations suggest that research would be valuable on the sex differences at each stage of disease – from acquisition and early responses (receptor expression), to progression to the lower respiratory tract (adaptive responses, comorbidities, risk factors), to progression to severe disease (alveolar inflammation, clearance of infected cells, etc.). Mortality differences may be impacted by genetics, hormones, disparities in comorbidities, etc. Further study could define the populations at highest risk and likely outcomes at each stage of the disease.

(3) Karestan Koenen, PhD (*Harvard T.H. Chan School of Public Health, USA*)

The COVID-19 pandemic is an extreme stressor with multiple characteristics known to be toxic to mental health, such as novelty, unpredictability, lack of control, social isolation, bereavement, stigma, etc. Previous epidemics illustrated outcomes of these stressors; SARS-related social isolation increased the number of suicides in older adults in Hong Kong and longer lasting psychiatric morbidities were persistent at four-year follow up. Emerging evidence from earlier stages of the pandemic in China show mental health impacts such as depression and anxiety among healthcare workers caring for patients, but also healthy individuals exposed to COVID-19 saturated social media.

IHCC has unique access to pre-exposure data across many populations globally. With a concerted public health effort to address mental health in the COVID-19 pandemic, the mental impact of COVID-19 on both individuals with and without the disease may be lessened. Individuals and communities can support mental health during the pandemic using the **REACH** strategies below:

1. **R**ecognize the problem: 47% of US adults report sheltering-in-place is adversely affecting their mental health.

2. **E**xpand the safety net: keeping individuals in their homes and jobs is protective for mental health.
3. **A**ssist those most at risk: consider those with existing mental health challenges, addiction, high-conflict environments, health care workers, etc. Participate in community research or improvement efforts.
4. **C**ultivate resilience.
5. **H**ave empathy (for ourselves and others): doing positive things for others may have protective effect on mental health.

There are numerous questionnaires in use for COVID-19, many of which include information about symptoms, exposures, and mental health. Many of these questionnaires can be found on the COVID-19 Host Genetics Initiative or the NIH PhenX Toolkit. Additional questionnaires are already in use by numerous IHCC cohorts; these will be discussed further on Day 2 of the Summit. Use of standardized metrics for mental health such as PHQ9, GAD7, RCADS25 is encouraged by the Wellcome Trust (WT) and US National Institutes of Mental Health (NIMH). Cohorts actively measuring the impact of protective mental health interventions are encouraged to contact Jordan Smoller, Andre Brunoni, or Sarah Bauermeister with this information.

(4) Kári Stefánsson, MD (*deCODE, Iceland*)

Iceland began screening for COVID-19 cases on January 31, 2020, restricted to persons travelling from regions of high risk. The first case was found on February 28 and in early March, deCODE began screening the general population. As of May 3, 2020, deCODE has screened more than 32,000 individuals with an infection rate of 0.58%. Another 4,000 randomly selected individuals resulted in a similar infection rate of 0.54%. In addition to deCODE, Iceland's National Hospital screened 18,000 high risk individuals with a resulting infection rate of 9.6%. All rates have been corrected for exposure. These efforts encompass testing of 14% of the total population in Iceland.

Using nasal swab specimens, deCODE sequenced the virus in positive cases and found 1,106 mutations in the virus along with numerous mutations in the receptor binding domain of the spike protein. Mutations create common viral haplotypes in certain geographic regions. Identification of the viral haplotypes therefore indicate its point of origin (concordant with contact tracing). As the pandemic progressed in Iceland, changes in haplotype frequencies illustrated initial sources from Austria and Italy, later from the UK, and eventually from the US. The timing of the UK origin haplotype in Iceland revealed that the virus was circulating in the UK earlier than previously thought. The observed rate of change of the virus is not particularly high, but the large number of infected individuals enable relatively rapid mutation in the population. Given this information, is difficult to conclude how long a vaccine may be effective or if the antibodies produced from one haplotype provide immunity for another viral haplotype.

With 1799 confirmed COVID-19 infections, 72 currently isolated, 4 currently hospitalized and no patients in intensive care, the measures taken in Iceland seem to be making an impact. These measures include wide screening of the general population, agile contact tracing, use of isolation of infected individuals (including one-week post-symptoms) and quarantine for exposed individuals. From mid-March to early April, the transmission rate declined 2% per day.

Aside from the haplotypes, the Icelandic data has yielded various other preliminary findings. Among the confirmed cases, there are no observed associations of ACE2 in the HLA region, however, the sample size is quite small. This limitation has also made association of viral mutation with clinical outcomes

difficult at this stage. Preliminary data shows that females do not become as severely sick as males and that children have a lower tendency to become infected. If children become infected, they are unlikely to transmit this infection to adults; there has only been one observed instance of this in the Icelandic data (using viral haplotypes for tracing sources). Of recovered individuals, 50% of the 900 screened still test positive for the virus. deCODE has also begun antibody testing using enzyme-linked immunosorbent assays (ELISA). Thus far they have screened 6000 individuals and are hoping to validate ELISA antibody results for asymptomatic individuals. Thus far the titer is lower on asymptomatic individuals, but a second ELISA will be used to confirm these results.

(5) WT COVID-19 LMIC-LPS Response – Geoffrey Ginsburg, MD, PhD *(Duke University, USA)*

WT has been actively working on a standardized COVID-19 questionnaire for low- and middle-income countries (LMICs) with longitudinal population studies (LPS) including cohorts, survey panel, and biobanks. To date, approximately 20 LMIC cohorts have joined the effort to create the questionnaire, with questions across several domains including physical health, behaviors, mental health, economics, social environment, etc. The standardized questionnaire will enable harmonized measurement of COVID-19 impact across WT funded LMIC cohorts. The group plans to agree upon a core set of standard questions by May 14th, 2020. These will be made available with a REDCap survey and available to IHCC and WT cohorts for use in the weeks following the Summit. Use of this questionnaire will facilitate cross-cohort analysis to inform COVID-19 national policies and research.

Next Steps – Keri Atlhoff, PhD, MPH *(Johns Hopkins Bloomberg School of Public Health, USA),* Kelly Gebo, MD, MPH *(All of Us Research Program, USA)*

With the background of the COVID-19 data presented, there will be an in-depth discussion during Day 2 of the Summit on how the IHCC may be positioned to advance COVID-19 scientific discovery during the pandemic. IHCC cohort COVID-19 survey data (collected in the weeks prior to the Summit) supported several initial research questions, to be examined on Day 2:

1. The global expansion and timing of COVID-19
    a. Serology in biospecimens collected November 2019 - Present
2. The impact of the host genome and environmental exposures on COVID-19
    a. Sex differences
    b. Tobacco/environmental pollution, population density/urbanization, public health infrastructure
3. The impact of social distancing on mental health/anxiety

Some initial suggestions from Summit attendees included:

- Impact of different country-level approaches to/timing of lockdowns and social distancing; examine the impact on transmission rates.
- Cohorts may want to coordinate on biospecimen collection standards (types, collection procedures, processing, storage, etc.). Additional considerations for biospecimens:
    - Planning for data sharing
    - Use of older serum samples (prior to 2019) to validate serotesting results
    - Blood spots, blood samples for serology
    - Funding resources for serology-related biospecimen collections
    - Availability of specimen collection supplies such as nasal swabs, viral transport media for -80°C freezer storage
    - Utility of saliva and/or gargle in place of nasopharyngeal or nasal swab specimens

    - Cytokine activity, inflammation variability
- Isolated populations
- Immune response in populations with high HIV-burden

Day 1 discussions on COVID-19 highlighted the utility of creating a resource toolkit for IHCC cohorts engaging in research on the pandemic. Several of the current resources are listed below. Note that this list is not comprehensive but is inclusive of resources discussed during Summit presentations and shared by attendees during discussion. A more complete resource toolkit is in development.

- COVID-19 Questionnaires
    - WT LMIC-LPS standard questionnaire (in development, will be shared by IHCC)
    - COVID-19 Host Genetics Initiative has collated questionnaires used by many of the international studies registered on the site.
    - Harvard University (contact: Karestan Koenen)
    - German National Cohort (contact: Klaus Berger)
    - Million Veteran Program (contact: Michael Gaziano)
    - All of Us (contact: Kelly Gebo)
    - CanPath (contact: Philip Awadalla)
    - UK-Wellcome Trust cohorts (contact: John Connolly)
- Biosampling/Biospecimen standards from WT for LMICs (contact: Laura Lyman Rodriguez or John Chambers)
- NIH Resource List (includes *All of Us* tools)
- CovIDentify Duke University; use of wearable data (heart rate, heart rate variability) to detect pre-symptomatic viral infection (contact: Geoff Ginsburg)

Session 2 – IHCC Work Team Progress – Science and Technology
Chair: Laura Lyman Rodriguez, PhD
15:00 – 17:00 UTC

Data and Infrastructure – Philip Awadalla, PhD *(Canadian Partnership for Tomorrow Project, Ontario Institute for Cancer Research, Canada),* Thomas Keane, PhD *(European Bioinformatics Institute, Global Alliance for Genomics and Health, UK)*

The Data and Infrastructure Team's mission is to deliver interoperable cross-cohort infrastructure to enable population scale biomolecular data accessible across international borders accelerating research and improving the health of individuals resident across continents. This mission will be operationalized through enabling cohort data discovery (e.g. finding relevant variables for research questions), cohort metadata standardization and harmonization, cohort data access (e.g. federated model, security, etc.), consideration of ethical/legal issues, enriching cohort diversity and building capacity. The Team's work to date has been focused on data discovery and metadata standardization/harmonization through the *Cohort Data Atlas* project. To leverage existing resources and support interoperability, the Data and Infrastructure Team has been utilizing existing global standards from GA4GH and conferring on lessons learned with other cross-cohort projects such as Human Heredity and Health in Africa (H3Africa) and Common Infrastructure for National Cohorts in Europe, Canada, and Africa (CINECA).

Cohort Data Atlas Methodology - Melanie Courtot, PhD (European Bioinformatics Institute, UK)

The IHCC Cohort Data Atlas will enable cross-cohort queries for discovery, supporting highly powered research questions with global reach. The process of building the Atlas began with a call to IHCC cohorts

to collect their data dictionaries. Of the 97 IHCC cohorts, 66 have public metadata which has been incorporated into the Atlas browser. Twenty-five cohorts from diverse geographies (South Korea, UK, Iran, South Africa, etc.) shared their data dictionaries with the Data Team in a variety of formats (PDF, Microsoft Word, Excel, etc.). Use of some data dictionaries required licensing clarification; this could be added as a standard consideration for IHCC cohorts in the future. Due to the short timeframe to operationalize the Atlas for pilot testing, nine cohorts were prioritized for recontact, and seven of those cohorts responded with additional information.

Data received from the cohorts was incorporated into a metadata model to maximize interoperability (e.g. CINECA, Maelstrom) and anticipate queries (use-case driven). Several existing ontologies were recruited in the model design including Web ontology language (OWL), experimental factor, biomedical investigations, and clinical measurement ontologies. Six cohort data dictionaries have been loaded into the Atlas browser thus far. Harmonization was performed with Genomic Cohort Knowledge Ontology (GECKO) methodology and individual mapping. This time-consuming step has resulted in two initial cohorts mapped and loaded into the Atlas browser and a third in process.

Efforts are underway to automate the process of harmonizing and mapping the data dictionaries. The Team also aims to develop a registry of publicly available data dictionaries and implement versioning/change detection for revisions to data dictionaries. Eventually cohorts would be able to generate updates to their data dictionaries and upload to the repository with change detection. The Atlas will use the registry APIs to pull the up-to-date harmonized data. Minor resource allocations will be required at the cohort level to keep the Atlas up to date. As mapping is still a manual process, the Data and Infrastructure Team may recontact mapped cohorts for a preview of their data before upload to the browser.

Cohort Data Atlas Demonstration - Christina Yung, PhD (Ontario Institute for Cancer Research, Canada)
The Cohort Data Atlas will enable cohort and variable discovery on a searchable browser hosted privately by the IHCC. There are three components on the current browser: a list of harmonized search attributes, a main panel of search results, and a panel of results visualizations. As data harmonization is in progress, the current Atlas contents are a mix of real and mock-up data for user/platform testing.

The search panel includes 14 facets of search criteria with sub-categories and real-time counts of cohorts meeting each sub-category criteria. The sub-categories may be updated with ontology changes. These search panel facets include:

- Socio-demographic and economic characteristics
- Physiological measurements
- Lifestyle and behaviors
- PI lead
- Microbiology
- Current enrollment (number of participants)
- Sample types
- Biospecimens
- Countries
- Cohort Name
- Data types available: basic cohort attributes
- Data types available: phenotypic clinical data
- Data type available: Genomic data
- Data type available: Environmental data

Sortable results are listed with cohort names, countries, and enrollment data. The results also include the type of data available, PI name, and a link to the cohort's external site (currently variable

formats/information on the external links). A future iteration of the Atlas may standardize the links to cohort information. A standard cohort page may be created for IHCC cohorts that include information on study design, data availability, marker publication, population, and the data dictionary in original form (not harmonized). The selected search criteria are listed individually at the top of the browser and can be deselected as needed.

The panel with visualizations of the search results currently includes two charts, but more may be added. The first is bar chart of the cohort results by country. The second shows the types of biospecimens available (saliva, blood, urine, or stool) among search results; this chart can be filtered by the specimen or cohort of interest.

The Atlas browser is not yet open to the public as it contains mock-up data which should not be used for scientific questions. However, feedback on user interface and functionality will be valuable as additional data is mapped and loaded into the browser. Data and Infrastructure Team members and/or another small group will be recruited for user testing in the coming weeks.

Next Steps
Over the next year, the Data and Infrastructure Team will continue to develop the IHCC Cohort Data Atlas, populating it with additional data dictionaries and collecting feedback from membership for iteration and public launch. The medium term of the project (1-3 years) will build out the Team structure for additional bandwidth and sub-groups and task leaders from the IHCC community and establish a compelling set of research and clinical showcase applications. In the long term (5+ years), the Team aims to support a fully operational federated IHCC cohort network. IHCC members interested in participating in the Data and Infrastructure Team activities or providing their data dictionaries should contact ihcc-browser@googlegroups.com.

Scientific Strategy – Adam Butterworth, PhD *(University of Cambridge, UK),* Hákon Hákonarson, MD, PhD *(Children's Hospital of Philadelphia, USA)*, Gad Rennert, MD, PhD *(Carmel Medical Center and Technion, Israel)*
The Scientific Strategy Team mission is to grow a highly collaborative and vibrant scientific initiative encompassing dozens of large scale programs with research deliverables that are integral to clinical care and population health. Phase 1 of the program has focused on building infrastructure to support pilot projects concurrent with efforts on the Cohort Data Atlas and IHCC Charter.

Phase 2 of the Scientific Strategy program will support the IHCC-wide research programs with high impact, low investment projects. During ICS2, many project ideas were submitted from IHCC members across diverse scientific domains. From these ideas, a short list of projects was identified for pilots including genetic risk assessment in underserved populations, pharmacogenomics, LoF variants, and PRS. The risk assessment project would genetically characterize well-phenotyped samples in low-income countries with low-depth whole genome sequencing. This optimal scientific approach is also the most expensive and funding will be required for next steps. The LoF project would use existing exome and genome data to develop a public database of LoF variants with associated phenotypes, biospecimen and health record data. This project would also develop a framework for recontact of individuals who are heterozygous or homozygous for the LoF variants. These projects were prioritized based on use of existing resources, broad scope across cohorts, results on a finite timeline, potential attractiveness to funders, expertise among researchers and development opportunities for junior researchers. Additional pilots and/or subsequent projects will emphasize scientific strategy priorities of cohort enhancement in

low resource settings and existing data sets as well as improved understanding of causes of disease and variability in response to treatment.

Preliminary Results of the PRS Pilot - Patrick Sleiman, PhD (Children's Hospital of Philadelphia (CHOP), USA)

With NIH funding support, the Scientific Strategy Team launched the first IHCC cross-cohort pilot project on an accelerated timeline in late 2019. The project selected common diseases/phenotypes (asthma, body mass index (BMI), type-2 diabetes (T2D) and blood pressure) for study. To accelerate the timeline for results, traits/diseases were selected with large-scale meta-analysis and publicly available genome-wide summary statistics. Participating cohorts included in the data are: ELSA-Brazil, Norwegian Mother and Child (MoBa) Study, CHOP repository, Nurses' Health Study (I and II), Shanghai Men's and Women's Health Study and UK Blood Donor cohorts. Additional cohorts have expressed interest and will be added to the analysis after the Summit (e.g. 23andMe, East London Genes and Health, Estonian Genome Project).

LDpred methodology was used to account for the effects of linkage disequilibrium and improve prediction accuracy in multi-ethnic populations. Summary statistics from individual studies were coordinated and Gibbs sampling was used to generate weights centrally while score generation was performed at the individual sites. Based on the participating cohorts, LDpred weights were generated for African (AFR), Hispanic (AMR), East Asian (EA), Northern European (EUR), and all ethnicities/trans-ethnic (TE). Using trait summary statistics, the research team imputed 2500 samples from the CHOP repository to create eight unique weight files with different fractions (from 1 to 0.001) of causal variants. These eight files were distributed to each cohort to generate scores on their data and report area under the curve (AUC) results for the fraction with the highest coefficient of determination.

During the timeframe of analysis, shifting priorities due to COVID-19 resulted in a focus on the BMI and T2D. The asthma and blood pressure summary statistics were not sufficient to generate predictive PRS on the accelerated timeline. In the Shanghai Men's and Women's Health Study, the TE score outperformed the EA score for BMI (no T2D data available). The MoBa study showed best performance in the EUR score for both BMI and T2D; however very few T2D cases were included in this cohort. The ELSA-Brazil cohort showed best performance in the TE score for BMI and T2D. The Nurses' Health Study I and II cohorts demonstrated best performance in the EUR score for both BMI and T2D, however the only very slightly outperformed the TE scores. The collective data shows TE scores to outperform all population specific scores in non-European cohorts with similar predictive values. European cohorts had best performance with EA scores.

Participating cohorts had varied sample sizes which may have affected score performance; additional cohorts will be added to the analysis to provide more comprehensive results. While the modular approach of generating scores at each site had challenges, this process worked well for efficient analysis. However, data sharing agreements with participating cohorts would improve future analysis. Individual level genotypes among the combined cohort would enable a common weight across participant sites, reducing effort at individual sites (generate scores only once). Population specific LD files and superior methodological tools (e.g. LDpred2) would improve results; these may be used for next iterations.

Next Steps

As the PRS project continues, more cohorts will be added across other ethnic groups. The Scientific Strategy Team will also engage with the COVID-19 research opportunities discussed during the Summit.

Aside from these topics, the Team would like to agree on methods for prioritizing scientific ideas from membership. At ICS2, there was interest in broadening the initial focus of IHCC to include observational epidemiology, epigenetics, and other non-genetic/omic topics. Prior to ICS3, research themes were solicited from membership to develop priority research questions during break-out meetings during the planned Summit in Santiago. As the face-to-face meetings have been postponed, the Scientific Strategy Team will reconsider how to solicit and agree upon these research priorities virtually.

With multiple IHCC proof of concept successes (PRS, Atlas), the Scientific Strategy Team will leverage existing infrastructure for collaborative proposals to various funding bodies and industry partners. NIH and WT have agreed to support three to five IHCC $100,000 pilots over a 12-18-month timeline. Work on these projects will support additional collaborative funding applications amongst membership. Preliminary results on the cross-cohort PRS project provide an opportunity for the first IHCC scientific publication.

Day 2: May 5, 2020, 11:00 – 17:30 UTC

SCIENTIFIC PRESENTATIONS AND EXTERNAL ENGAGEMENT

CHAIR: TERI MANOLIO, MD, PHD

11:00-11:50 UTC

Keynote: Soumya Swaminathan, MD, PhD *(World Health Organization, Switzerland)*

Dr. Swaminathan is currently the Chief Scientist at the WHO. In 2002, the WHO published *Genomics and World Health*, outlining the importance of genomics for world health, raising issues of justice, resource allocation, and ethics in a post-genomic era. While there has been significant healthcare capacity advances in LMICs since the publication, there are still large disparities in the application of genomic technology in many developing countries. Genetic technologies in healthcare can achieve their true potential only if the evidence base is truly representative of the world population.

The WHO has emphasized three main aims for global implementation of genomic technologies:

1. Global must truly mean global, with research representation from all sub regions of the world. For example, research of populations in South Africa should not have to be generalized to all of Africa.
2. Governance and Regulatory frameworks in LMICs need to facilitate genomic technology implementation. The WHO has been working on ethical frameworks for some of these technologies from the LMIC perspective, while others (e.g. NIH, Gates Foundation) have been working on low cost genetic tools for the low resource setting. The LMIC perspective needs to be considered from the beginning of the research and technology design.
3. Data must be shared in both directions. Researchers from the South are often reluctant to share data as they may lose control of it. They often become coauthors on the collaborative research publications, but do not receive adequate credit for their contribution. Standards are needed for federated models for data hosting, sharing, reuse, etc. such that global data sets and subsequent research questions can be equitably driven by researchers from the North and South.

Innovation in genomics cannot be transformative without access. Local scientists are best placed to drive innovation to meet the needs of their local populations; engagement must begin during research and development to accelerate the access timeline.

In recent months, WHO has also been focused on the COVID-19 pandemic. They convened researchers in early February for a Research and Innovation Forum to create a research roadmap. This forum led to the creation of nine workstreams focused on the pandemic with team leaders from around the world. These groups meet virtually each week to discuss emerging issues, best practices, etc. Work products of these groups form WHO technical guidance and drive international research collaborations on therapeutics and vaccines.

1. Animal-Human Interface
2. Clinical Disease
3. Infection Control and Prevention
4. Diagnostics
5. Therapeutics
6. Vaccines
7. Social and Behavioral Science
8. Ethics
9. Capacity Building

Creating truly global collaborations is challenging, but the current global quarantine has created a good opportunity to bring everyone together virtually. Outreach through global research networks and engagement with young investigators may help facilitate collaboration. These global networks are numerous, including some specifically for COVID-19:

- A global coalition to accelerate COVID-19 clinical research in resource-limited settings, announced in the Lancet in early April, 2020
- Funding call for COVID-19 from the African Academy of Sciences with deadline of June 30, 2020
- Alliance for Accelerating Excellence in Science in Africa (AESA)
- European and Developing Countries Clinical Trial Partnership (EDCTP); funding call specifically for COVID-19 research has passed
- Global Outbreak Alert and Response Network (GOARN)

Plans for IHCC COVID-19 Research – Keri Althoff, PhD, MPH *(Johns Hopkins Bloomberg School of Public Health, USA),* Kelly Gebo, MD, MPH *(All of Us Research Program, USA)*

IHCC is uniquely positioned to contribute to research on the COVID-19 pandemic by combining global longitudinal cohorts for highly powered research questions that will inform scientific, clinical, programmatic, and policy decision making communities globally. After reviewing suggestions from meeting attendees, IHCC narrowed the collaborative COVID-19 research focus on the following themes:

1. Global expansion and timing of COVID-19
    a. Serology testing in biospecimens collected from November 2019 – Present
    b. Impact of social distancing interventions (timing, approach, country level differences), including isolated populations
2. Impact of host genome and environmental exposures on COVID-19
    a. Sex, age, ethnicity, comorbidity differences for infection rates and disease severity

   b. Exposures such as tobacco, environmental pollution, population density/urbanization, public health infrastructure
   c. Biomarker associations with clinical outcomes
3. Impact of social distancing on mental health/anxiety
   a. Are some subgroups more vulnerable than others?
   b. Attempt to harmonize mental health measures and collect social distancing metadata of cohorts across different geographies. Utilize existing tools (e.g. Mental Health Futures Project – WT/United Nations)
   c. Examine coping methods, resilience, economic and occupational impacts.

In the weeks prior to the summit, IHCC conducted a survey among members with 60 cohort respondents. Half of the cohort respondents reported access to one or more types of COVID-19 data, including disease status, mental health/well-being survey results, access to close contacts of confirmed cases, health records, molecular data, and biospecimens. These data types can be grouped into the core constructs below:

- SARS-CoV-2 test results
- COVID-19 disease outcomes (symptoms, diagnosis, treatments, resilience, etc.)
- COVID-19 exposures

While the cohorts have varied access to these data types, more than one-third of the respondents have access to data to support research theme #1, two-thirds have data to support research theme #2, and over half have data to support research theme #3. The IHCC COVID-19 cohort survey will be sent to cohort members again after the Summit to solicit additional responses.

The IHCC will form SWGs around each of the research themes to draft research aims, hypotheses, study design and measurement tools, statistical approaches, and deliverables. A fourth SWG will be formed to consider biospecimens and standard data measurements to support each research question. This group would also explore harmonization of different assays to define cases and use of metadata to adjust credibility intervals. The SWGs can be supported administratively by the IHCC Secretariat; funding support is needed for in-depth research. Cohort members are requested to reply with their interest in participation by May 12, 2020.

Summit attendees shared ideas about additional aspects of COVID-19 research including:

- Examining risk factors like obesity and alcohol use with existing cohort infrastructure and prospective data collection.
- Social distancing indices from technology providers and validation efforts.
- Return of antibody test results to participants without understanding of immunity.
- Running discovery queries on the IHCC Cohort Data Atlas; cohorts interested in participating in IHCC COVID-19 collaborative research are encouraged to share their data dictionaries with the Data Standards and Infrastructure Team.
- Feasibility of dried blood spot testing.
- Use of longitudinal data on adverse health effects from previous socioeconomic disruptions (e.g. 2008-2010 economic recession). Use this historical data to project future increase in mortality and health effects from the pandemic unrelated to COVID-19 disease status.
- Social distancing/shelter-in-place adherence – are there cohorts collecting this data?
- Impact of closing the healthcare system to elective healthcare on longer term health, grouped by geography.

SESSION 3 – SCIENTIFIC PRESENTATIONS
CHAIR: TERI MANOLIO, MD, PHD
11:50 – 14:15 UTC

Data and Infrastructure Presentations

The Data Standards and Infrastructure Team aims to support the mission statement: "deliver interoperable cross cohort infrastructure to enable population scale biomolecular data accessible across international borders accelerating research and improving the health of individuals resident across continents."

David Glazer *(Verily, USA)*

Data sharing has traditionally involved moving data to individual researchers, downloading from a centralized source. In an effort to enable the next generation of collaborative biomedical research, Verily, along with Vanderbilt University, the Broad Institute, the University of California Santa Cruz and the University of Chicago, created the cloud-based Data Biosphere using the Terra platform (optimized for Google cloud but compatible with other cloud providers). The modular, community driven, open, and standards-based environment serves three groups of researchers: data generators, tool developers, and biomedical researchers. The Biosphere hosts a workflow execution service, data repository service, and researcher identities. The centralized platform increases security and accessibility, supports a shared research tool ecosystem, and decreases the cost of storage.

At the NIH sponsored *All of Us* Research Program (USA), researchers have agreed to share data widely and wisely. Thus their infrastructure must support the storage of data in a reliable way, safe from inappropriate use, and yet widely available. The Data Biosphere environment has been applied to this project to enable collaborative researcher and participant engagement. *All of Us* has numerous participating institutions collecting data for a centralized research center with a collaborative researcher workbench and shared tools. Researchers use the workspace to build sub-cohorts/data sets for cloud-based analysis with R or Python. Individual sites need to have modest computational and human resources to analyze and store their data/results. Research data providers may create their own workspaces and grant researchers' permission to work in that space.

Alison Motsinger-Reif, PhD *(National Institute of Environmental and Health Sciences, USA)*

The USA National Institute of Environmental and Health Sciences (NIEHS) has created an Environmental Polymorphisms Registry with more than 17,000 participants to interrogate gene-environment interactions. Survey data details general health questions, family health, lifestyle, etc. along with Exposome A (external exposures) and Exposome B (internal exposures – e.g. vitamins, medications). This is combined with electronic health record (EHR) and Geographic Information System (GIS) data to consider additional exposures. The participants will have whole genome sequencing (WGS) to investigate gene-environmental interaction mapping. The WGS results will also enable interrogation of pleiotropy, immune mediated diseases, and support deep phenotyping with clinical callback.

NIEHS is also hosting a new initiative during COVID-19 with the Pandemic Vulnerability Index (PVI). The PVI has an online data dashboard with a set of analysis tools to which researchers can bring their data. The index aims to guide decision making; using their best model to hypothesize which communities are most vulnerable to becoming the next COVID-19 hot spot, where interventions should be focused or relaxed, etc. Data informing PVI are updated daily at the county level, including measures of infection

rate, population concentration, interventions, and healthcare infrastructure/demographics. The index is rank based, yielding a pie chart visualization of risks for each county selected. As PVI has been rapidly built during the outbreak, additional data streams are continuously added (e.g. all-cause mortality, symptom tracker app data, etc.). With the data currently integrated in PVI, prediction of disease trajectory up to 28 days ahead has been highly accurate. While PVI does incorporate information about implementation of social distancing interventions by local governments, the variation in implementation and adherence make impact difficult to measure thus far. The index is publicly available and could be applied in non-USA locations but would require that similar data sources are available. The dashboard allows downloads of the PVI aggregated data, which can be connected with geocoding data or other data sets as desired. NIEHS can assist researchers with integration of their PVI data as needed.

Goncalo Abecasis, PhD *(Regeneron, USA)*
Like many other companies in biotech currently, Regeneron has been shifting resources toward COVID-19. They are engaged in an ongoing clinical trial evaluating the role for sarilumab (Kevzara) for investigational use in treatment of hospitalized COVID-19 patients. In January, they used diseases similar to COVID-19 such as MERS and Ebola to develop new antibodies that bind to the SARS-CoV-2 spike protein, blocking interactions with host cells. They have narrowed hundreds of antibodies down to just a few top candidates in April, now moving into production for human testing over the summer months. As a speedy response to COVID-19 is urgent, Regeneron is using human genetic data to guide their hypotheses and project trial outcomes to guide decision making. For example, a cohort of patients with COVID-19 outcomes could be genotyped for the specific interleukin-6 polymorphism associated with better clinical outcomes in similar diseases.

For their COVID-19 and other initiatives, Regeneron has optimized their technology with individual labs able to sequence more than a million exomes annually, perform analysis on the cloud, and use the DNAnexus platform. They have improved computation efficiency, reducing analysis time from 1000 hours with traditional approaches like SAIGE and BOLT to less than 10 hours for queried traits with equivalent results. In partnership with the UK Biobank, Regeneron will sequence all 500,000+ participants in 2020 and make all data available to their researchers. The first 150,000 participants sequenced showed many LoF positive controls. This number is expected to increase as the remaining samples are sequenced.

<u>Scientific Strategy and Cohorts Enhancement Presentations</u>

Alicia Martin, PhD *(Broad Institute, USA)*
Although the field of genomics has progressed significantly in the last decade, scientific advances have not been equitable across the globe. Most genetic studies are still among those of European decent, despite this population encompassing only 16% of the global population. PRS, for example, use genetic indicators to predict an individual's phenotype by using large scale genetic studies and multiplying effect sizes by the genotypes present in a target population cohort. However, application of European discovery cohorts to non-European ancestry cohorts has been inaccurate, limiting the benefit of these discoveries to a small percentage of the global population.

Given this disparity, the Broad Institute is using novel methods such as Multi-Ancestry Meta-Analysis (MAMA), with promising results on initial simulations and test phenotypes in global biobanks. Using GWAS summary statistics from two or more populations, cross-population LD analysis is used to

recalibrate effect sizes. The LD correlation matrices will eventually be updated with user-guide documentation and shared with fellow biomedical researchers.

Historically, European-based PRS perform the most poorly when applied to African ancestry populations. A current initiative from the Broad, neuroGAP, aims to study 35,000 participants across Africa (Uganda, South Africa, Ethiopia, Kenya) to learn about the genetic risk of psychosis. Recognizing that capacity building is needed for genetic studies in low resource settings, in the last 2.5 years, Broad has implemented a research capacity program (GINGER) to implement workshops, virtual classrooms, onsite trainings, etc. Prospective sample collection and data generation, while likely to lead to novel discoveries, will take time for translation. However, large global biobanks of existing samples and data contain useful diversity. The UK Biobank contains substantial sample sizes of Central and South Asian (~9,000), African (~6700), East Asian (~2700), Middle Eastern (~1600), and Admixed American populations (~1000) populations. These cohorts provide immediate opportunities for discovery.

George Vradenburg, *(UsAgainstAlzheimers, USA),* Elias Zerhouni *(Johns Hopkins University, USA)*
The Davos Alzheimer's Collaborative (DAC), in partnership with the WEF and the Global CEOinitiative is developing a global cohort for Alzheimer's Disease (AD) study. The CEOinitiative is a patient-led collaborative of pharmaceutical and biotech companies committed to combating Alzheimer's Disease. DAC is actively seeking IHCC cohorts interested in collaborating on this initiative, in hopes that AD, affecting more than 50 million families globally, may be a use case for global technology in genomics. There is currently an incomplete understanding of the heterogeneity of AD outside of European ancestry populations, including genetic risk factors and LoF mutations. Rather than combining individual AD participants from many cohorts, the collaborative intends to collaborate with cohorts, working around new ideas, platforms, and technologies. Participating cohorts would coordinate on a shared scientific plan, solicit additional funding, and agree on standardized methods for phenotyping. DACs goals include creating a global cohort for biomarker development, creating a support system for globally coordinated clinical trials, and supporting healthcare system preparedness for AD. Existing blood and plasma samples may be used to detect underlying pathophysiology and additional biospecimens may be collected in the future to evaluate specific proteins.

Interested IHCC cohorts are encouraged to contact Drew Holzapfel and/or attend one of the scheduled meetings below with the DAC to consider the scientific plan, 5-year budget, and business plan.

- May 19, 2020 at 9:30-10:30am EDT
- May 20, 2020 at 2-3pm EDT

Summit attendees suggested collaboration with several initiatives such as Dementias Platform UK (PI John Gallacher), the HUNT study (Norway), and the Canadian Partnership for Tomorrow's Health (CanPath) Project.

Erin M. Ramos, PhD, MPH *(National Human Genome Research Institute, USA)*
NHGRI has created the PhenX toolkit, a catalogue of recommended measurement protocols for researchers to facilitate cross-study analysis. The toolkit currently contains 746 protocols across 26 research domains (e.g. demographics, reproductive health, social environments, etc.) and some special collections on topics like mental health, substance abuse, sickle cell disease, etc. Each consensus-based protocol is added to the toolkit through discussions with the Steering Committee, NIH liaisons, expert working groups and community outreach. Toolkit domains are reviewed regularly for protocol relevance and updates by expert review panels. Included protocols must be broadly applicable, incur low burden

to participants and investigators, be validated and reproducible, open-source and non-proprietary, and have a standard measurement protocol available. Protocols in the toolkit are searchable with variable IDs and mapping to the database of Genotypes and Phenotypes (dbGaP) and include instructions for use. The toolkit memorizes protocol redundancy and highlights one protocol for use.

In recent months, 45 protocols related to COVID-19 have been added on symptoms, knowledge and attitudes, adherence to mitigation behaviors, social and economic impacts, etc. The PhenX Team is working on adding multi-language and multi-culturally relevant protocols but currently is most easily applied to USA-based studies. The PhenX toolkit is free online with over 1.5 million visitors.

Policy and Bio-Data Sharing Presentations

Benjamin Neale, PhD *(Broad Institute, USA)*
The ICDA is currently working on a Maps to Mechanisms to Medicines Initiative: translating variants (Maps) to the Mechanisms of underlying disease and leveraging for therapeutic development (Medicines). The more than 70,000 associations across hundreds of human diseases and traits are starting points for investigating their biological pathogenic mechanisms. More are being identified continuously through GWAS and other association studies. However, there is a bottleneck in translating these associations to mechanistic insights (cell types, pathways, etc.) and eventual novel medicines and therapies. The associations can answer questions beyond *who* may get sick; they may also indicate *when* someone may get sick.

As IHCC's large data sets with genetic and phenotypic data are becoming available, harmonization is key to enabling progress in this initiative. ICDA has an organizing committee and has set up three working groups (Maps, Mechanisms, and Medicines), a funder's council, and a pharma council. All three working groups interact with issues across data platforms, data sharing, governance, and ethics. The initiative has developed a set of recommendations to drive foundational resources and suggest flagship diseases to serve as use cases.

ICDA is also supporting a COVID-19 Host Genetics Initiative by creating a forum for communities working on COVID-19 to collaborate and drive scientific progress. The forum invites cohorts around the globe to register their study with details on sample size, data types, planned analysis, etc. There are currently 170 registered studies, all of whom have agreed to the basic principles of collaboration set out by ICDA. There is a COVID-19 Slack channel for engagement with 954 current members. Individual cohorts will decide how to collaborate and how much data sharing is feasible within the principles of the Initiative. IHCC cohorts and global biobanks are invited to register their studies through the website or email idca-office@icda.bio for more information.

Fruzsina Molnár-Gábor, PhD *(Heidelberg Academy of Sciences and Humanities, Germany)*
From 2013-2019, the International Cancer Genomes Consortium (ICGC) supported the Pan-Cancer Analysis of Whole Genomes (PCAWG) project, collecting over 800 terabytes of genomic data across 30 countries and 460 institutions. PCAWG aimed to identify genetic traits across various types of cancer while navigating challenges with data sharing, participant rights, shared tools, and cloud providers. During this process, PCAWG found the need for an International Code of Conduct for Data Sharing that would account for General Data Protection Regulation (GDPR).

PCAWG prioritized the following points for an International Code of Conduct:

- Identifiability of Data
- Broad Consent
- Return of individual findings, data portability and access
- Withdrawal (of consent)
- Compelled Disclosure

The Codes of Conduct offer an independent norm regime that can be adapted to local needs. Other benefits include a bottom-up approach, the ability to make the code specific to a project or study, serving the best interest of the group, decision-making roots, timely implementation, less regulation, ethics integration, expertise and compliance, and legal relevance. These codes can become benchmarks for specific types of research. After creation of the Codes of Conduct (timeline ~1-1.5 years), individual national bodies will be able to consider whether it is GDPR compliant.

Lynsey Chediak, MSc *(World Economic Forum, USA)*
The World Economic Forum (WEF) is an international organization with equal funding support from both public and private corporations. The WHO has partnered with the WEF on their healthcare platform, guiding industry participation in healthcare delivery systems.

Within this platform, the Precision Medicine (PM) portfolio was launched in 2017 with a policy lab in San Francisco (USA). PM evolved as one of nine emerging technology project areas included in the launch of the Center for the Fourth Industrial Revolution, also in San Francisco. This portfolio will shape the trajectory of the PM field for the acceleration of societal benefits with minimal risk. Barriers to implementation include data sharing/infrastructure, pricing and reimbursement, regulatory environments, integration with clinical practice, patient/public engagement, and evidence generation on effectiveness.

The first project in this portfolio, launched in July 2018, set out to unlock global data silos for a patient-led rare disease use case. This project involved partners in several countries with a federated data system. For this use case, WEF wanted to define the technical scope of a federated data system, define funding allocation and economic incentives for data sharing, and outline how to efficiently operate a federated data system with proper security and patient privacy. The project is set to complete in June of 2020 with subsequent scale up planned with additional country partners. WEF will apply this framework to other future projects such as wearable data for COVID-19 surveillance.

IHCC members are invited to join any of the WEF project working groups, or to participate in pilots of their deliverables. Multi-cohort projects within IHCC may also benefit from piloting deliverables of this WEF portfolio.

SESSION 4 – EXTERNAL ENGAGEMENT (OTHER CONSORTIA AND INDUSTRY PARTNERS)
CHAIR: PETER GOODHAND
15:00 – 17:30 UTC

Partner Presentations

Ewan Birney, PhD *(Global Alliance for Genomics and Health and European Bioinformatics, UK)*
GA4GH launched in 2013 with the mission of accelerating progress in genomic science and human health by developing standards and framing policy for responsible genomic and health related data sharing. In addition to Ewan Birney, GA4GH leadership includes Peter Goodhand (Canada), Heidi Rehm (USA), and Kathryn North (Australia). Membership includes leaders from University and research institutes, academic medical centers and health systems, disease advocacy and patient organizations, consortia and professional societies, funders/agencies, and life science and information technology companies.

More than 500 members across 90 countries have collaborated on several initiatives since the 2013 launch, including plenary meetings, establishing roadmaps, etc. In 2016, GA4GH began to focus on a small number of driver projects with collaborating partners (e.g. All of Us, ClinGen, H3Africa, etc.) to identify concrete technical data sharing issues. The organization remains standards based; members support the driver projects with the development and iteration of standards to support the projects, but scientific work on the projects is external to the organization. GA4GH created technical work streams, each with two co-leads, to develop standards for:

- Discovery
- Large-scale genomics
- Data use & researcher IDs,
- Cloud data
- Genomic knowledge standards
- Clinical and phenotypic data capture
- Regulatory/ethics
- Data Security

GA4GH aligns work products with existing standards such as the International Organization for Standards (ISO) technical committee 215, Systematized Nomenclatures of Medicine (SNOMED), Health Level Seven International (HL7), etc. In addition to a publicly available framework for responsible sharing of genomics and health related data, GA4GH has a number of approved standards from each work stream in their genomic data toolkit. IHCC members are welcome to join the GA4GH technical workstreams at ga4gh.org.

Eric Green, MD, PhD *(Global BioData Coalition and National Human Genome Research Institute, USA)*
The Global BioData Coalition (GBC) brings together international funders to address issues with data resources. A foundational publication in Nature (2015) from three NIH leaders highlighted the need for funders to support a global big data ecosystem. While most data is nationally funded, international collaborations on research is highly valuable. Funding mechanisms need to adapt beyond agency and national borders to enable international research.

More than 3000 biomedical data resources exist from academia, government, healthcare, biotechnology, and pharmaceutical industries, with more added each year. A subset (~100) of these interconnected core data resources are used millions of times each day. These vital sources are

supported with an annual budget of approximately $500 million USD. Exponential growth in data generation, increased demand from open-access policies, fragmented funding, and new technologies that require data resources contribute the vulnerability of these core data. Curtailment of the funding risks data resources retreating behind firewalls, making them inaccessible to many researchers in low-resource settings. In response to this concern, the Human Frontier Science Program (HFSP) convened an international workshop. This gathering yielded two publications in Nature (2017) and BioRxiv (2017) outlining core data sustainability considerations and launched the GBC.

Hosted by HFSP, over the last four years, the GBC steering committee has been working toward a public launch in 2020. GBC aims to provide meaningful coordination among international funders of the core biomedical data resources, ensure sustainability of resources and open access through global stewardship, expand the number of international funders involved, and enhance data resource-related expertise worldwide. The GBC has closely coordinated with the Heads of International Research Organizations (HIROs) during organization planning. Outreach with community stakeholders is ongoing while the group establishes a process for selecting the Global Core Data Resources and recruits Secretariat support staff.

Rachel Liao, PhD *(International Common Disease Alliance and Broad Institute, USA)*
The International Common Disease Alliance (ICDA) officially launched in September of 2019 with a White Paper (v0.1). They are an international forum of diverse stakeholders invested in progressing scientific research on common disease through the Maps to Mechanisms to Medicines (M2M2M) Challenge. ICDA released their first set of recommendations for removing bottlenecks and accelerating progress in this challenge in February of 2020 with v0.9 of the White Paper. ICDA planned to release version 1.0 of these recommendations at their in-person meeting in Copenhagen in March of 2020, but this event was postponed due to COVID-19. The recommendations use exemplar projects to highlight the M2M2M framework for any common disease. Recommendations include 23 proposed efforts including foundational genetic/genomic resources, critical computation tools, activities to propel therapeutics, priority areas for research and training, and strategies to promote global equity.

ICDA's main strategic directions include:

- Flagship Diseases: these will test the M2M2M challenge. They aim to launch projects for at least 10 diseases and will consider selection from global burden of disease data.
- Maps: how to promote diversity, utility, and analysis across biobanks/cohorts and develop genetic resources to catalogue genetic variation and GWAS summary statistics.
- Mechanisms: map the cis-regulatory elements and trans-effects to identify activated cellular programs and systematically characterize protein coding variants.
- Medicines: understand the genetic basis of common disease for development of therapies and provision of clinical care with development of best practices, tools, models, methods for biomarker identification.
- Data and Policy: consider diverse data from various locations with federated ecosystem of software resources and ethical principles/policies for collaboration.
- Innovation: accelerate progress on M2M2M challenge with new methods and technology while promoting equitable global collaborations.

IHCC cohorts are invited to participate with ICDA on their strategic directions or their COVID-19 Host Genetics Initiative (also discussed in prior presentations at this Summit). IHCC and ICDA project discoveries may be able to synergistically support progress in both organizations.

Keynote – Jeremy Farrar *(Wellcome Trust, UK)*
The emergence of the COVID-19 pandemic will alter the way scientific collaborations occur moving forward. The necessity for global collaboration to combat a critical shared burden has increased willingness to share and inspired changes to infrastructure to support rapid progress. These changes are positioned to optimize future research in domains beyond infectious disease.

WT is in the process of a strategic review. However, given the context of COVID-19, large scale longitudinal population studies will continue to be critical to the WT mission. IHCC cohorts may aid in global understanding of real-time progression of COVID-19 (or rebound effects, vaccinations, immunity) and future pandemics. Historically, epidemics occur every 18-24 months. With changes in ecology, animal-human interfaces, travel frequency, these epidemics will only become more frequent and complex. With deep phenotyping and known background data of individuals, IHCC cohorts may be able to act as sentinels for emerging epidemics, enabling detection before large number of severe syndromes present in clinical facilities.

There is likely to be more political support for public health and preparing for future epidemics both of infectious and non-infectious disease (e.g. mental, metabolic, cardiovascular health). Sustaining large cohorts for many years is often viewed as a drain on significant resources, but COVID-19 may demonstrate how this investment could aid in identification of global disease. National agencies struggle to provide long-term funding commitments, however cohort sustainability beyond 3-5-years is vital. IHCC may be able to consider approaches for sustainability so that researchers can explore long term high-impact research questions (like efforts of GBC on sustainable funding for core data).

Industry Partner Panel Discussion
Alan Shuldiner, MD *(Regeneron, USA)*
Regeneron Pharmaceuticals' Genetics Center (RGC) applies large scale, fully integrated human genetics approaches to advance science, guide the development of therapeutics, and improve patient outcomes. The RGC was established five years ago with ultra-high throughput sequencing, biobanking with automated sequencing and genotyping sample prep workflows, cloud data storage, and automated analysis pipelines. RGC plans to sequence more than five million exomes in the next few years.

Large cohorts provide opportunity for pharmaceutical partners to identify new drug targets/pathways and new indications for those targets. Cohorts with genomic information enable de-risking of drug studies by confirming lack of on-target adverse effects in drug target LoF carriers. They may also help develop pharmacogenomic markers to predict drug response. RGC has engaged in more than 90 academic collaborations around the world and is actively working with four IHCC cohorts, with another 10 IHCC cohorts in progress. IHCC's guidance on industry partnerships (see Day 1 presentation from Policy and Biodata Sharing Team) is in sync with Regeneron's principles, including mutual respect for capabilities, fairness and equity, transparency, disclosures, timely dissemination of research, and freedom to operate. While each partnership will be different based on cohort needs, RGC's scientists will participate as equal research collaborators to advance understanding of human health and disease.

Meg Ehm, PhD *(GlaxoSmithKline, USA)*
In 2015, GSK published findings that drugs with human genetic evidence are twice as likely to be successful. GSK subsequently pivoted to gain better understanding of disease by using genetic research. Using a hypothesis-free approach, GSK uses genetic and patient data from sources with a broad number of traits such as 23andMe, UK Biobank, FinnGen, and public sources. This data is combined with

phenotypic information to identify genes for evaluation as drug targets, alternative drug indications, and/or on-target safety hazards.

IHCC is uniquely positioned to contribute to biomedical research with industry partners like GSK by supporting study of underrepresented populations and cross-cohort/biobank research. This element is critical to achieving the large sample sizes needed for drug discovery in less common conditions. GSK supports collaboration with large cohorts with high disease burdens that may enable identification of progression to severe outcomes.

Phil Febbo, MD *(Illumina, USA)*
For genomic information to be incorporated into healthcare systems and clinical decision making, there needs to be sufficient evidence of clinical utility. While the cost of genomic information has decreased dramatically since the first human genome was sequenced at the cost of $3 billion USD in 2003, evidence is still needed to justify the added cost of a $1000 USD genome or eventual $100 USD genome. Genomic data may reveal the hereditary risk of cardiovascular disease, cancer, rare disease, or provide a polygenic risk score for a trait. However, until clinicians see the value in requesting this genetic information, genomic medicine will not become a standard part of healthcare infrastructure. Even for high-risk children, it is not common to perform whole genome or whole exome sequencing. Of individuals with advanced cancer, only 15% receive comprehensive genomic testing to inform clinical decision making.

The aggregation of large cohorts facilitates deep analysis of genomic data with quality phenotypic data. IHCC's collection of cohort resources across diverse populations can catalyze movement to clinical utility, and thus Illumina is highly motivated to partner with IHCC to achieve this shared goal. COVID-19 is serving as a use-case for the importance of genomics in health. In response, Illumina has shifted resources to build infrastructure/capacity to identify host-genomics associated with differential transmission in China, South East Asia, and Africa.

Panel Discussion with Industry Partners
Summit attendees shared questions about opportunities and challenges for IHCC-Industry collaboration. Several topics were discussed amongst the panelists.

LMIC Engagement: LMICs represent a strategic gap in genomic databases and development of novel therapeutics and diagnostics. Low-resource settings often lack infrastructure for sampling, data generation, and data management. Partnerships between industry and LMIC cohorts will need to prioritize trust and transparency. Regeneron has a *Founder and Special Populations* Group dedicated to understanding the allelic architecture of founder populations (phenotype agnostic). They are actively seeking global collaborators with access to patient populations with consent and legal allowances to share samples and deidentified data for sequencing. Sequence data is available to collaborators in real-time. Illumina is considering how to build infrastructure in LMICs for sustainability and independence beyond initial grant funding. In addition to sequencing, the capacity for processing and sharing the clinical information will be a priority. COVID-19 may heighten the importance of in-country genomics infrastructure to enable surveillance. Organizations like GSK will have a better understanding in the future of what is available and what scientific questions can be explored in global and LMIC populations.

Challenges of Industry Partnerships: Industry partners have struggled to find data, or know what data already exists. The IHCC Cohort Data Atlas is a promising solution to this issue. A partnership with one or

multiple cohorts may include a wide breadth of phenotype data; agreements for sharing of this data must also be broad. Standardization of these phenotypes and genotypes is also a challenge for larger partnerships. Aside from specific data, developing trust has also been a challenge in industry partnerships. While a collaboration with many cohorts or a consortium such as IHCC is relatively comfortable for most, individual labs/centers are more wary of safeguarding data and discoveries. Setting standards for industry engagement as the IHCC is doing is a helpful step toward building trusting partnerships.

Data Sharing with Industry and Exclusivity: IHCC's values prioritize timely and open-source dissemination of scientific discoveries. However, this can be difficult to merge with stipulations around periods of exclusivity and intellectual property ownership that often accompany industry partnerships. Regeneron has experience with more than 90 collaborators with varied agreements, but the guiding principles remain the same between all partners. There is some period of exclusivity (sometimes requested by either side), but academic collaborators usually have immediate access to data. Study consent forms also pose limitations for data sharing and exclusivity; new partnerships will likely require re-consent. Collaborations that involve multiple industry representatives or funders often enable more relaxed exclusivity requirements.

At present, cohorts are collecting information on scientific questions believed to be important, but it would be helpful to know up front what pharmaceutical and biotechnology companies are investing in to address unmet medical needs. IHCC and individual cohorts may be able to align strategic priorities to collect additional valuable data in the future.

Summary, Consensus and Next Steps – Geoffrey Ginsburg, MD, PhD, Peter Goodhand, and Teri Manolio, MD, PhD

The emergence of COVID-19 as a global pandemic resulted in the ICS3 taking place in virtual format rather than in-person in Santiago, Chile. At this virtual gathering, approximately 160 attendees from more than 23 countries learned details of the IHCC Charter outlining governance, membership, and expectations for participation. IHCC policies on publications, collaborations with industry, and data sharing were also presented for review by membership. Virtual discussion between attendees and presenters illustrated support for the Charter and policies. These will be ratified/approved in the weeks following the Summit.

IHCC's Data and Infrastructure Team presented a live demonstration of the IHCC Cohort Data Atlas prototype, enabling cohort discovery of variables, specimens, and populations of interest. User testing of the Atlas will continue in the coming months as additional cohort data is mapped and uploaded to the browser. Once the Atlas is fully operational, the Data and Infrastructure Team will help to disseminate awareness with a publication.

IHCC's Scientific Strategy Team presented preliminary results of an IHCC cross-cohort pilot project on polygenic risk scores for four selected traits. A publication is in progress to highlight the proof of principle for the cross-cohort analysis as well as the PRS prediction results. Three to five additional pilot projects will be initiated on a 12-18-month timeline with funding support from the WT.

IHCC is launching several COVID-19 response research initiatives based on discussion during the Summit. Presenters showed current research on COVID-19 surveillance, sex and age disparities, mental health,

and host response variants. Several SWGs will be established to examine the global expansion of COVID-19, the impact of mental health and environmental exposures on COVID-19, and biospecimen standards.

Presenters shared scientific progress across a variety of domains including data infrastructure, equity in polygenic risk scores, standardized phenotype measures, precision medicine initiatives, and principles and codes of conduct for collaboration. Several speakers highlighted opportunities for collaboration though the DAC, GA4GH, ICDA, GBC, and G2MC and industry partners such as Regeneron, GSK, and Illumina.

Summary Actions for Cohorts

IHCC Internal Initiatives

- Provide additional feedback on IHCC Charter as needed. A formal ratification process will follow in June 2020. Implementation of the charter will enable new governance with elected members to guide shared scientific agenda.
- Provide responses to the data sharing survey by June 5, 2020 to inform the IHCC Core Data Sharing Principles.
- Provide additional comments on policy documents as needed by May 22, 2020 to Laura Lyman Rodriguez. Final versions will be approved by the SSC and distributed to membership.
- Scientific Strategies team will launch three to five scientific initiatives with a 12-18-month timeframe using a federated approach with an emphasis on global diversity more than -omics.
- IHCC members interested in joining the working groups (Policy and Biodata Sharing, Data and Infrastructure, or Scientific Strategy) are encouraged to contact the Team leads or IHCC Secretariat.
- Cohorts may still provide data dictionaries for IHCC Cohort Data Atlas development to ihcc-browser@googlegroups.com. IHCC cohort data will be further populated in the atlas. Members are encouraged to provide feedback to Data and Infrastructure Team on use cases for atlas queries. This Team will establish a compelling set of research and clinical showcase applications. COVID-19 phenotyping will be added to the Atlas. Cohorts interested in cross-cohort COVID-19 research may also consider providing data dictionaries for harmonization and cohort discovery.
- An IHCC Resource Center with relevant data tools and other cohort resources will be created and shared on the IHCC website.
- Host virtual working meeting in ~six months.
- IHCC will collectively develop a five-year roadmap of planned progress

IHCC COVID-19 Research

- Provide responses to the IHCC COVID-19 cohort survey if not already complete (will be recirculated via email to cohorts).
- Indicate interest in joining any of the COVID-19 SWGs by May 12, 2020. Contact Eric Plummer to join.
- Cohorts measuring the impact of protective mental health interventions for COVID-19 are encouraged to contact Wellcome Trust (Jordan Smoller, Andre Brunoni, or Sarah Bauermeister) with details.

Opportunities for Collaboration

- For COVID-19 research studies, register on the ICDA COVID-19 Host Genetics Initiative
- To join the Davos Alzheimer's Collaborative, contact Drew Holzapfel and/or attend the scheduled calls on May 19/20, 2020.

ATTENDEES

NAME	AFFILIATION	COUNTRY
Goncalo Abecasis	Regeneron	USA
Malak Abedalthagafi	KIng Abdulaziz City for Science and Technology	Saudi Arabia
Jahad Alghamdi	King Abdullah International Medical Research Centre	Saudi Arabia
Keri Althoff	Johns Hopkins Bloomberg School of Public Health	USA
Narges Amini	University College London	UK
Sophia Apostolidou	University College London	UK
Ricardo Armisen	Universidad de Desarrollo	Chile
Philip Awadalla	Ontario Institute of Cancer Research	Canada
Radja Badji	Qatar Foundation	Qatar
Rosi Bajari	Ontario Institute of Cancer Research	Canada
Kate Balaconis	Broad Institute	USA
Cori Bargmann	Chan-Zuckerburg Initiative	USA
Lisa Bastarache	Vanderbilt University Medical Center	USA
Sarah Bauermeister	Dementias Platform UK	UK
Klaus Berger	University of Muenster	Germany
Ewan Birney	The European Bioinformatics Institute	UK
Dan Brake	Sequence Bio	Canada
Adam Butterworth	University of Cambridge	UK
Lynsey Chediak	World Economic Forum	USA
Rajiv Chowdhury	University of Cambridge	UK
Liz Cirulli	Helix	USA
Francis Collins	National Institutes of Health	USA
John Conollly	Children's Hospital of Philadelphia	USA
Melanie Courtot	The European Bioinformatics Institute	UK
Mary De Silva	Wellcome Trust	UK
Josh Denny	All of Research Program	USA
Rajesh Dikshit	Tata Memorial Centre	India
Meg Ehm	GlaxoSmithKline	USA
Robert Eiss	National Institutes of Health	USA
Joanne Elena	National Cancer Institute	USA
Heather Eliassen	Brigham and Women's Hospital and Harvard University	USA
Jonathan Emberson	University of Oxford	UK
Arash Etemadi	National Cancer Institute	USA
Jeremy Farrar	Wellcome Trust	UK
Phil Febbo	Illumina	USA
Catterina Ferreccio	Pontificia Universidad Católica de Chile	Chile

NAME	AFFILIATION	COUNTRY
Laurie Findley	National Human Genome Research Institute	USA
Andy Forbes	Otsuka Pharmaceuticals; Davos Alzheimer's Collaborative	USA
Neal Freedman	National Cancer Institute	USA
Daniel Freitag	Bayer	Germany
John Gallacher	Dementias Platform UK	UK
Bruna Galobardes	Wellcome Trust	UK
Gergana Gateva	Nightingale Health	Finland
Mia Gaudet	National Cancer Institute	USA
Kelly Gebo	All of Us Research Program	USA
Alex Gentry-Maharaj	University College London	UK
Geoff Ginsburg	Duke University	USA
David Glazer	Verily	USA
Marcel Goldberg	INSERM	France
Peter Goodhand	Global Alliance for Genomics and Health	Canada
Eric Green	National Human Genome Research Institute	USA
Joseph Grzymski	Renown Health and Desert Research Institute	USA
Marc Gunter	International Agency for Research on Cancer	France
Sue Haadsma-Svensson	CEO Initiative on Alzheimer's Disease	USA
Carolina Haefliger	AstraZeneca	Germany
Hakon Hakonarson	The Children's Hospital of Philadelphia	USA
Jennifer Harrow	ELIXIR	UK
Nonye Harvey	National Cancer Institute	USA
Kathy Helzlsouer	National Cancer Institute	USA
Kobus Herbst	South African Population Research Infrastructure Network	South Africa
Lucia Hindorff	National Human Genome Research Institute	USA
Drew Holzapfel	Davos Alzheimers Collaborative, High Lantern Group	USA
Rebecca Jackson	Knocean, Inc	USA
Jae-Pil Jeon	Korea National Institute of Health	South Korea
Farin Kamangar	Morgan State University	USA
Norihiro Kato	National Center for Global Health and Medicine	Japan
Thomas Keane	European Bioinformatics Institute	UK
Hannah Kennel	Global Genomic Medicine Collaborative	USA
Giselle Kerry	The European Bioinformatics Institute	UK
Zomoroda Knana	Carmel Hospital Haifa	Israel
Karestan Koenen	Harvard T.H. Chan School of Public Health	USA
Ida Konstantopoulos	Palladian Partners	USA
Steinar Krokstad	HUNT Research Centre	Norway
Alissa Kurzman	High Lantern Group	USA
Jim Lacey	City of Hope	USA
Debra Lappin	Us Against Alzheimers	USA
Richard Leach	Sequence Bio	USA

NAME	AFFILIATION	COUNTRY
David Ledbetter	Geisinger	USA
Dong Li	Children's Hospital of Philadelphia	USA
Rongling Li	National Human Genome Research Institute	USA
Rachel Liao	Broad Institute	USA
Paulo Lotufo	University of Sao Paulo	Brazil
James Lu	Helix	USA
Beatrice Lucaroni	European Commission	Belgium
Erin Luetkemeier	National Institutes of Health	USA
Chris Lunt	All of Us Research Program	USA
Teri Manolio	National Human Genome Research Institute	USA
Katherine Marcelain	Universidad de Chile	Chile
Josep Maria Haro	Parc Sanitari Sant Joan de Deu	Spain
Alicia Martin	Massachusetts General Hospital & Broad Institute	USA
Elena Martinez	University of California, San Diego	USA
Nico Matentzoglu	The European Bioinformatics Institute	UK
Matthew McIntyre	23andMe	USA
Martin Mcnamara	Sax Institute	Australia
Usha Menon	University College London	UK
Arshiya Merchant	ELIXIR	UK
J. Michael Gaziano	Brigham and Women's Hospital	USA
Sadra Mirhendi	Sequence Bio	Canada
Fruzsina Molnar-Gabor	Heidelberg Academy of Sciences and Humanities	Germany
Takayuki Morisaki	BioBank Japan, University of Tokyo	Japan
Alison Motsinger-Reif	National Institute of Environmental Health Sciences	USA
Nicky Mulder	University of Cape Town	South Africa
Yoshinori Murakami	Biobank Japan, The University of Tokyo	Japan
Sarah Murphy	Helix	USA
Øyvind Næss	Norwegian Institute of Public Health, University of Oslo	Norway
Benjamin Neale	Broad Institute	USA
Soichi Ogishima	Tohoku Medical Megabank, Tohoku University	Japan
James Overton	Knocean, Inc.	Canada
Christina Park	National Institutes of Health	USA
Hyun-Young Park	Korea National Institute of Health	South Korea
Alexandre Pereira	University of Sao Paulo	USA
Annette Peters	Helmholtz Zentrum Muenchen	Germany
Slavé Petrovski	AstraZeneca	UK
Michael Phillips	Sequence Bio	Canada
Paul Pinsky	National Institutes of Health	USA
Eric Plummer	Global Genomic Medicine Collaborative	USA
Cecilia Poli	Universidad del Desarrollo	Chile
Tejinder Rakhra-Burris	Global Genomic Medicine Collaborative	USA

NAME	AFFILIATION	COUNTRY
Erin Ramos	National Human Genome Research Institute	USA
Gadi Rennert	Carmel Medical Center and Technion	Israel
Gabriela Repetto	Clinica Alemana-Universidad del Desarrollo	Chile
Jessica Reusch	All of Us Research Program	USA
Stephen Riffle	Helix	USA
Laura Lyman Rodriguez	Global Genomic Medicine Collaborative	USA
Albert Sanchez-Niubo	Health Park Sant Joan de Déu	Spain
Norie Sawada	National Cancer Center	Japan
Minouk Schoemaker	Generations Study	UK
Serena Scollen	ELIXIR	UK
Eileen Scully	Johns Hopkins University	USA
Svati Shah	Duke University	USA
Suyash Shringarpure	23andMe	USA
Alan Shuldiner	Regeneron	USA
Patrick Sleiman	Children's Hospital of Philadelphia	USA
Lindsay Smith	Global Alliance for Genomics and Health	Canada
SUNG SOO	Korea National Institute of Health	South Korea
Kari Stefansson	deCODE	Iceland
J. Clay Stephens	Genomics GPS	USA
Soumya Swaminathan	World Health Organization	Switzerland
Anthony Swerdlow	Institute of Cancer Research, London	France
Jonathan Tedds	ELIXIR	UK
Nikolaos Tertipis	Nightingale Health	Sweden
Elin Thysell	Umeå University	Sweden
Jason Torres	Nuffield Department of Population Health	UK
Meredith Towery	Global Genomic Medicine Collaborative	USA
Shoichiro Tsugane	National Cancer Center	Japan
Ricardo Verdugo	University of Chile	Chile
Daniela Villarroel	Masso Eventos	Chile
George Vradenburg	CEO Initiative on Alzheimer's Disease	USA
Nicole Washington	Helix	USA
Su Ming Wei	Taiwan Biobank	Taiwan
Quinn Wells	Vanderbilt University	USA
Lynne Wilkens	University of Hawaii Cancer Center	USA
Christine Williams	Ontario Institute for Cancer Research	Canada
Peter Wurtz	Nightingale Health Ltd	Finland
Christina Yung	Ontario Institute for Cancer Research	Canada
Elias Zerhouni	Johns Hopkins University	USA
Wei Zheng	Vanderbilt University Medical Center	USA
Brittany Zick	Global Genomic Medicine Collaborative	USA
Heather Zurel	Sequence Bio	Canada

PARTICIPATING COHORTS

23andMe
45 and Up Study
Africa Health Research Institute (AHRI) Population Cohort
Apolipoprotein MORtality RISk study (AMORIS)
AstraZeneca Integrated Genomics Initiative
Barshi Cohort
Southeast Asian Cohorts Network
Biobank Japan
BioVu Vanderbilt
California Teachers Study (CTS)
Canadian Partnership for Tomorrow Project
Cancer Council Victoria Cohort Studies
Cancer Prevention Study II (CPS-II)
Cancer Prevention Study II Nutrition Cohort
Children's Hospital of Philadelphia (CHOP) Biorepository
China Kadoorie Biobank
China PEACE (Patient-centered Evaluative Assessment of Cardiac Events) Million Persons Project
Constances Project
Danish National Birth Cohort
Dementias Platform UK (data repository)
East London Genes and Health
ELSA-Brazil Project
Environmental influences on Child Health Outcomes (ECHO) Cohort
Environmental Polymorphisms Registry
EPIC (European Prospective Investigation into Cancer, Chronic Diseases, Nutrition and Lifestyle)
EpiHealth
Estonian Genome Project
Finnish Maternity Cohort Serum Bank
Generations Study (GS)
Genomics England / 100,000 Genomes Project
German National Cohort (NAKO)
Golestan Cohort Study
H3Africa
Healthy Nevada
Israel Genome Project
Japan Public Health Center-based Prospective Study (JPHC)
Japan Public Health Center-based Prospective Study for the Next Generation (JPHC-NEXT)
Kaiser Permanente Research Program on Genes, Environment, and Health
Korea Biobank Project

Korean Cancer Prevention Study (KCPS-II Biobank)
Korean Genome and Epidemiological Study (KoGES)
LifeGene
LIFEPATH (Lifecourse biological pathways underlying social differences in healthy aging)
Malaysian Cohort
Maule Cohort (MAUCO Study)
Mexico City Prospective Study
Million Veteran Program
Million Women Study
Multiethnic Cohort Study (MEC, NCI)
MyCode Community Health Initiative
NCI Cohort Consortium
National Israeli Cancer Control Center
Netherlands Twin Registry
Newfoundland 100K Genome Project / Sequence Bio
Newfoundland and Labrador Genome Project
NHS (Nurses' Health Study, NCI)
NHSII (Nurses' Health Study II, NCI)
Nord-Trøndelag Health Study (HUNT)
Northern Sweden Health and Disease Study
Norwegian Family Based Life Course Study
Norwegian Mother and Child Cohort Study (MoBa)
Pakistan Genomic Resource (PGR)
Persian Cohort Study
PLCO (Prostate, Lung, Colorectal and Ovarian Cancer Screening Trial, NCI)
Qatar Genome Project
SAPRIN (South African Population Research Infrastructure Network)
Sardinia Project
Saudi Human Genome Program
Saudi National Biobank
Shanghai Men's Health Study
Shanghai Women's Health Study
Singapore National Precision Medicine Program
SYNCHROS
Taiwan Biobank
Tohoku Medical Megabank Project
U.S. Precision Medicine Initiative / All of Us
UK Biobank
UK Blood Donor Cohorts
UKLWC (UK Collaborative Trial of Ovarian Cancer Screening Longitudinal Women’s Cohort)
Vorarlberg Health Monitoring and Promotion Programme (VHM&PP)
WHI (Women's Health Initiative)

SPONSORS

Thank you to the sponsors of the International Cohorts Summit for their generous contribution and continued support!

NIH National Institutes of Health
Turning Discovery Into Health

wellcome