

Global Genomic Data Representativeness: Current Landscape, Gaps, and Opportunities



TOBY HOUSE, Analyst, Thought Leadership, IQVIA ; RICARDA GAENTZSCH, Principal, Head of Genomics & Precision Medicine, IQVIA ; SARAH RICKWOOD, Vice President, Thought Leadership, IQVIA; FIONA PEREIRA, Senior Principal, Integrated Research for Public Health, IQVIA; MARK BALE, Independent consultant in genomics, policy and bioethics and Chair, Ethics Advisory Board for Our Future Health UK ; AUDREY COIFFIC, Engagement Manager, Genomics & Precision Medicine, IQVIA

Contact : audrey.coiffic@IQVIA.com

Objectives

Map global genomic sample representativeness and identify opportunities to expand it, enabling less biased healthcare delivery worldwide.

This landscaping exercise informed a published report on approaches to broaden representation in human genomics.

Funding

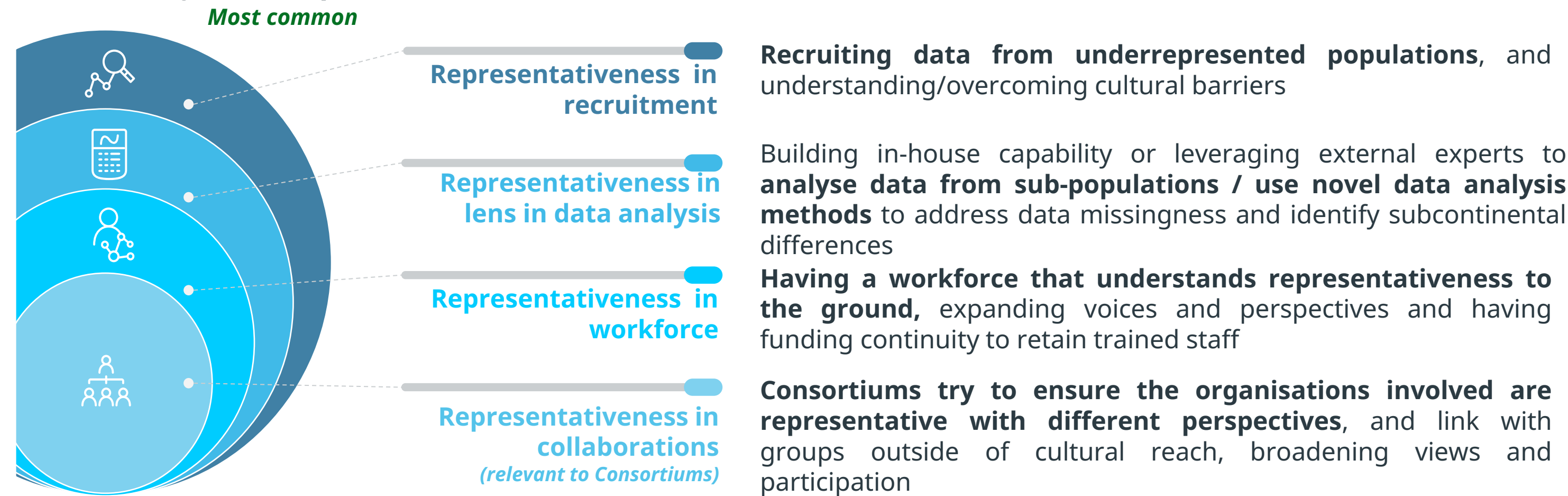
Commissioned and funded by the Wellcome Trust (WT), this landscaping exercise on genomic data representativeness is accompanied by a white paper on the evolving global landscape of genomic initiatives, funded by IQVIA.

Methodology

- Wellcome and IQVIA consolidated and identified a long list of 440 global genomics initiatives.
- IQVIA used desk research and surveys to build a representativeness profile for 198 of these initiatives.
- The representativity assessment survey received 55 unique responses.
- IQVIA carried out online qualitative interviews with 27 key initiative representatives.

Definition of representativeness in genomic research:

Different layers of Representativeness in Genomic data:



Representativeness in genomic research is defined as targeting underrepresented populations, analyzing sub-population data, cultivating a workforce with different backgrounds and cultural origins, forming international collaborative networks and reaching communities beyond cultural barriers.

Genomic representativeness archetypes:

Three archetypes were defined based on level of addressing representativeness:

- High maturity** : large database size, focusing on chronic disease etiology research and precision medicine
- Medium maturity** : medium database size, focusing on regional understanding of genetic factors for disease etiology
- Low maturity** : low database size, focusing on expanding databases to understand chronic disease etiology

High maturity regions
(e.g., US, Canada, UK, Estonia, Australia):

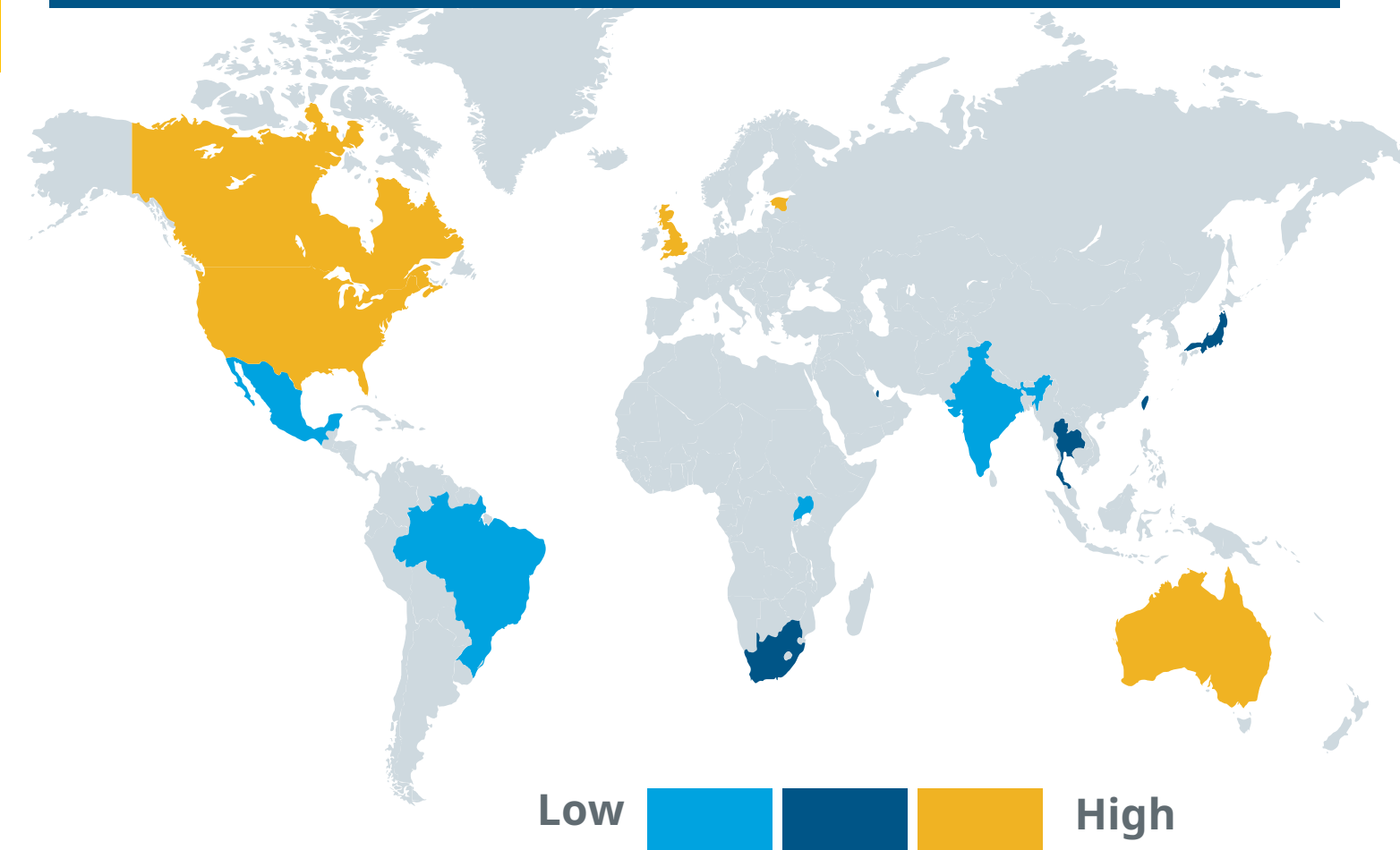
- Types of initiatives:** databases, biobanks, data aggregators, consortiums, technology
- Av. cohort size:** large (~579,000)
- Objectives:** support etiological studies of chronic diseases / ultimately contribute to precision medicine and personalized healthcare
- Representativeness efforts:** expand into underrepresented populations, make external collaborations, work on novel data analysis methods

Medium maturity regions
(e.g., Japan, Taiwan, Thailand, Qatar, Hong Kong, S. Africa):

- Types of initiatives:** mainly databases, biobanks
- Av. cohort size:** medium (~70,000)
- Objectives:** regional understanding of genetic factors associated with chronic disease etiology
- Representativeness efforts:** expand databases, overcome logistical barriers, improve data analysis, overcome legal barriers to international collaborations

Low maturity regions
(e.g., Brazil, Mexico, Uganda, India):

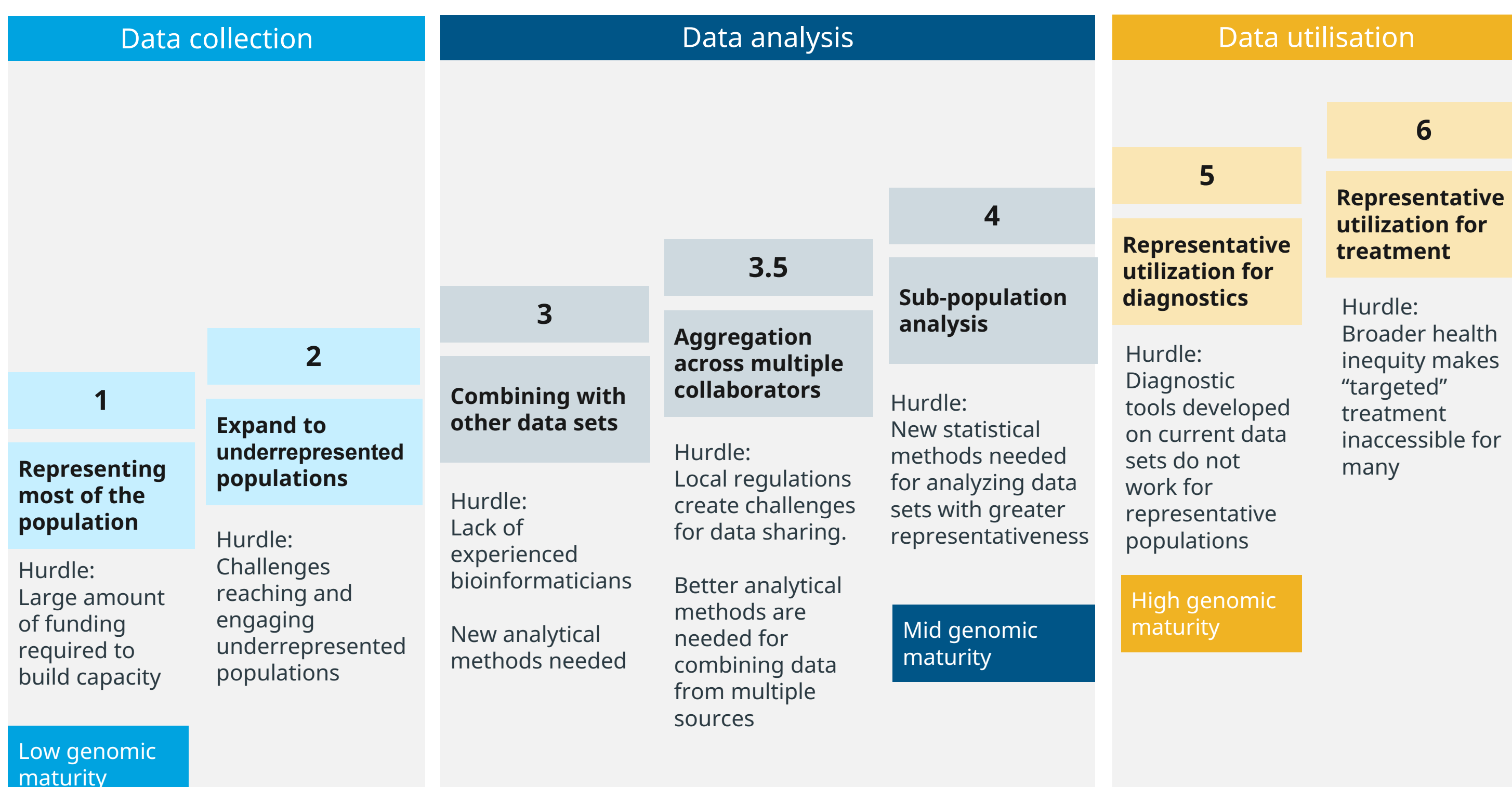
- Types of initiatives:** mainly databases, biobanks
- Av. cohort size:** low (~13,000)
- Objectives:** include less represented populations, enhance regional understanding of genetic factors associated with chronic disease etiology
- representativeness efforts:** expand databases, use alternative recruitment methods, build capacity and infrastructure



Key take-aways

- High maturity markets are focusing on expanding datasets into underrepresented populations
- Medium and low maturity markets are working on building representative databases

The path to full and unbiased utilization of genomics data



Funder opportunities to advance data representativeness in human genomics

We identified key challenges and opportunities, then asked our Scientific Advisory Board (a panel of external experts) to rank them by priority for funders and other influential organizations.

Note: Only the highest-priority opportunities appear below. The full report covers every opportunity and our recommendations for global collaboration.

High Maturity Archetype

For high maturity regions, efforts should focus on underrepresented and community groups engagement, alternative recruitment methods and respondent return on investment

HIGHER RANKED OPPORTUNITIES

Community efforts

Support efforts to collaborate with underrepresented populations (indigenous advisory boards, local community outreach plans)

Support community engagement, to develop perspectives on use and storage of genetic data

Recruitment

Support with alternative recruitment methods to reach representative populations (e.g., mobile blood collection vehicles to reach distant, rural areas)

Return of interest

Support in the provision of individual outputs to provide patients with a "return of investment" - responsibility for sharing genomic insight back with original country + health data

Mid Maturity Archetype

In those regions, the most important opportunity is to support regional teams with education and training

HIGHER RANKED OPPORTUNITIES

Education / training

Support regional teams with educating and training researchers - bioinformatics, database development and data interpretation to drive use of data

Low Maturity Archetype

In those regions, efforts should focus on growth and retention of local talent and local infrastructure development

HIGHER RANKED OPPORTUNITIES

Local infrastructure

Support in infrastructure building (sample storage, sequencing capabilities, required materials and technologies)

Support with growth and development of local genomic research teams to promote long-term research capacity

Talent growth and retention

Support with the retention of talent and expertise to prevent local "brain drain"

Specific support for educating and training in bioinformatics, database development and data interpretation to drive local use of data

SOLUTIONS / RECOMMENDATIONS

- Engage with Patient Advocacy Groups so they can see value of this type of research. Fund conference attendance and specific research that drive community engagement.
- Support research that connect with local leaders (e.g. youth leaders) or include within winning criteria.
- Create funding programs, that seek to assess community views on consent to data usage (to take pressure off individuals) and supports community empowerment. Ideally lead by PIs from these communities.
- WT to request PIs to identify country-specific challenges and support research groups who understand and target these representative groups
- Fund a network or work with existing network to encourage bulk blood collection
- Funding social programs of work for academics to develop individual understanding on research purposes.
- Behavioral change initiatives addressed to doctors and pharmacists and understand their existing challenges and motivators to change engagement with representative groups.

SOLUTIONS / RECOMMENDATIONS

- Capacity building for human genomics research for both wet-lab and data analysis; grant calls for locally-led genomic data generation/analyses, pairing them with expertise on databases and interpretation, with train-the-trainer models to build regional expertise for further training and scale
- Funding for 'low-hanging fruit' initiatives, e.g., country-led initiatives to build local genomic data repositories, well-characterized biobanks
- Acknowledgement and credit to data contributors and data owners to be built into such collaborative projects, potentially explore benefit-sharing mechanisms

The evolving global landscape of Genomics initiatives

The global landscape of genomic initiatives is rapidly evolving, driven by falling sequencing costs and scientific advances. IQVIA published a white paper reviewing developments since 2020, covering applications in drug development, personalized medicine, pharmacogenomics, and global health, alongside key challenges such as funding and ethics.

The analysis was done in parallel of the Wellcome Trust landscaping analysis and approached the analysis of genomic initiatives with a different angle. It identifies major trends, opportunities, and disparities, promoting collaboration and unbiased progress in genomic medicine worldwide.

Key findings of the global landscape of genomic initiatives :

- Global Landscape:** Genomic initiatives are poised to play an increasingly vital role in human health, with significant progress and developments since 2020.
- Regional Developments:** Whilst North America leads by total number of genomic initiatives, Asia and Europe have experienced the greatest growth since 2020.
- Genomics at an Inflection Point:** There is a pressing need for public-private partnerships (PPPs) to capitalize on genomic data and overcome key challenges.



Conclusion :

Regions with low or medium genomic maturity face major structural constraints, including limited infrastructure, lack of specialized expertise, insufficient funding, regulatory barriers and public trust. Despite recent progress, socio-economic data collection remains limited, hampering the translation of genomic knowledge into health policies. Other challenges include the rise of data nationalism, ethical issues related to the sensitivity and sustainability of genomic data, as well as the long-term sustainability of databases, where public-private partnerships could play a key role.

The representativeness of global genomic data can be improved through early community engagement, integration of socio-economic data, strengthening training in bioinformatics and genetic consulting, and targeted investments in sequencing capacity. These elements are essential for globally relevant precision medicine.