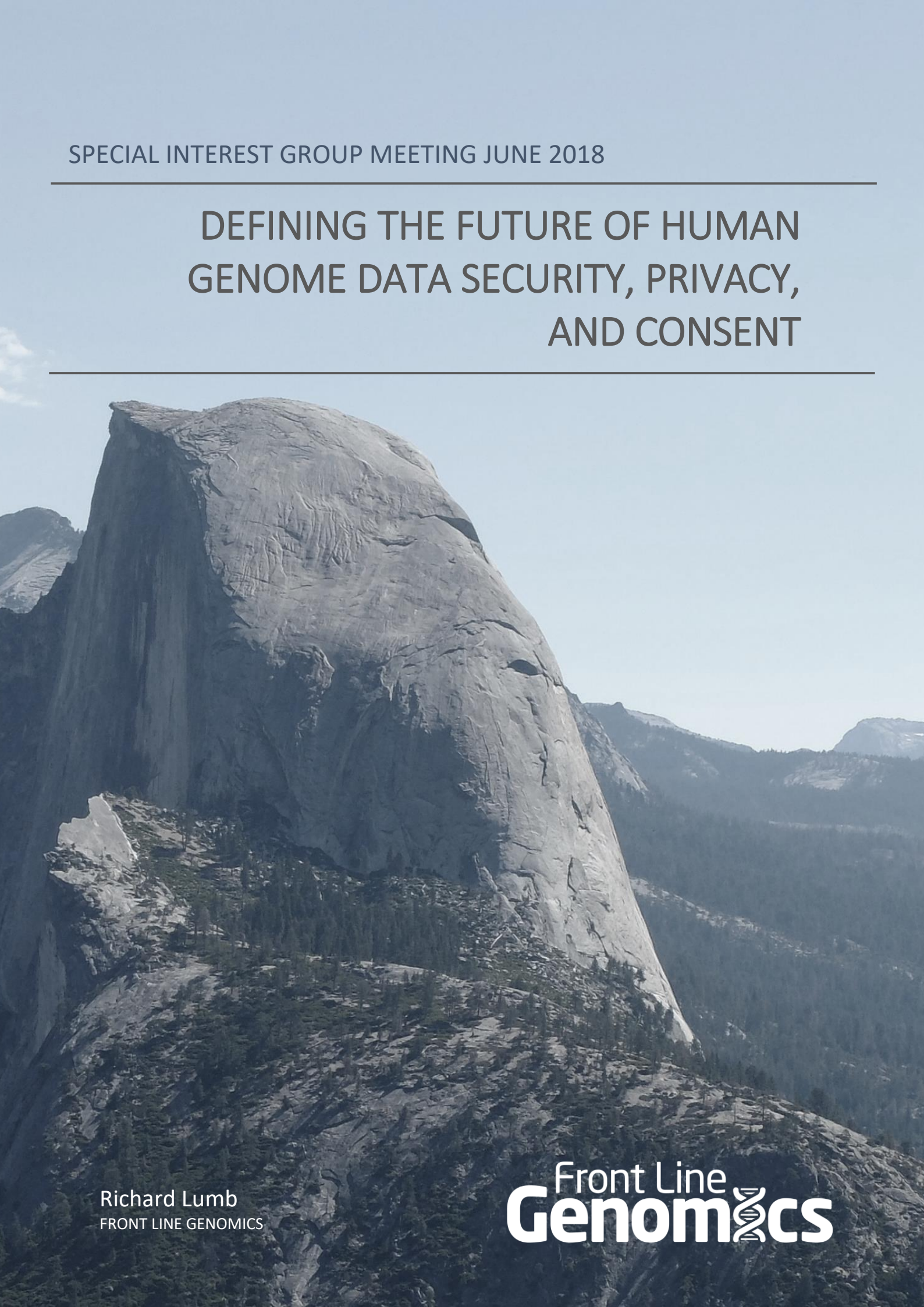
SPECIAL INTEREST GROUP MEETING JUNE 2018

DEFINING THE FUTURE OF HUMAN GENOME DATA SECURITY, PRIVACY, AND CONSENT



Richard Lumb
FRONT LINE GENOMICS

Front Line Genomics

Foreword

Robust and dynamic consent agreements with patients and research participants rely on trustworthy security and privacy measures. They are central to enabling all stakeholders access to the data needed to advance precision medicine. But how are attitudes changing towards data sharing, and how can patients and public be incentivized to participate in research?

To explore these questions, experts in the field were brought together for a meeting under Chatham House Rules, in Boston on June 19th. In order to maintain a balanced perspective, the participants of the meeting were selected from a range of backgrounds, with many different specialties and focuses.

The purpose of this meeting was to discuss HIPAA, GDPR, and Common Rule requirements including IT security, data access, and restrictions, and responsibilities for working with sensitive patient information. At the same time, the group looked to determine what actions are required in order to create an efficient workflow where data security, privacy, and consent is ensured without hindering research through lengthy and complicated approval processes.

This report presents an accurate reflection of the discussions that took place at this meeting and does not constitute an official statement from any of the individual participants, their organizations, or Front Line Genomics.

Attendees:

Stacey Donnelly, Senior Director, Strategic Operations, **The Broad Institute**

Michael Braxenthaler, Global Head, Strategic Initiatives, **Roche**

Carrie Blout, Senior Genetic Counselor and Project Manager, **Brigham and Women's Hospital and Harvard Medical School**

Moran Cabili, Program Manager, **The Broad Institute**

Michael Hunter, SVP, Strategy and General Manager, US Operations, **PetaGene**

Shira Rockowitz, Genomics and Bioinformatics Lead, Research Computing, **Boston Children's Hospital**

Maria Lalioti, Director, Clinical Pharmacogenomics, **Biogen**

Matthew Lebo, Director, Bioinformatics, **Brigham and Women's Hospital**

Natasha Frank, Interim Clinical Chief, Genetics, **Brigham and Women's Hospital**

Gaurav Kaushik, Associate Director, Data Products & Strategy, **Foundation Medicine**

Taunton Paine, Senior Policy Analyst, Clinical and Healthcare Policy Division, Office of Science Policy, **National Institutes of Health (NIH)**

Ingrid Holm, Associate Professor of Pediatrics, **Harvard Medical School**

Piotr Sliz, Chief Research Information Officer, **Boston Children's Hospital**

David Thomas, Director of Cachexia Research & Therapeutic Development, **The Broad Institute**

Will Chen, VP Product Management and Business Development, Precision Medicine, **Elsevier**

Johnny Kung, Director of New Initiatives, **Personal Genetics Education Project**

Brigitte Raumann, Product Manager - Life Sciences, Globus Team, **University of Chicago**

Jason Labonte, Head of Product, **DataVant**

Frances Shaw, Producer, **Front Line Genomics**

Richard Lumb, CEO and Founder, **Front Line Genomics**

Report Author: Frances Addison, Staff Writer, Front Line Genomics

Points of Discussion:

Sharing Datasets

- Accessing data is difficult for researchers and can take a long time, which limits their ability to carry out their work. Researchers are not able to rapidly assess whether a dataset contains information that will be directly relevant to their work and instead have to formulate complex queries before they begin analysis. While this is possible from a legal and protocol perspective, the majority of data storage systems are not built with this in mind. This problem may be overcome by developing automated systems for data querying or better data labelling, but this has challenges in and of itself.

- Searching data sets is hindered by the lack of standard ontologies, as well as poor or inconsistent data labelling. Equivalent data is often stored with different annotations depending on the internal ontologies of the group submitting it. If the researcher searching the dataset uses different labelling for their queries, then they may miss relevant data or have to spend more time searching for it. Large quantities of shared data have also been found to be incorrectly labelled due to errors during data processing.

- Providing authentication to access shared data is complicated and takes a long time. Many institutions that maintain their own databases utilize a two-step approach to approving the use of their data in research: a staff member who is able to analyze the training of the researcher involved to evaluate their suitability, and a data access committee who evaluate the specific research proposal. This process ensures that only people who are equipped to handle the data responsibly are able to use it, but it can also substantially delay research. An automated approvals process may minimize this problem, but it would not remove the need for human interaction to determine feasibility and morality of proposed research.

- The introduction of GDPR is prohibiting access to European datasets. A lot of people are unclear as to who the new regulations affect and how they will impact research. Without that clarity, it is easier for many groups to stop using European data instead of adapting their workflows, which limits the scope of research being carried out.

Consent

- There are different levels of consent possible that need to be considered when approaching patients. When asked about sharing their data with researchers, they may provide no consent, specific consent (relating to a single, determined use), tiered

consent, broad consent, or dynamic consent. Broad and dynamic consent models are likely to be the most applicable for genomic data moving forwards, but consent forms are not standardized. It needs to be clearly communicated to researchers when a patient has placed limits on how their data can be used.

- Consent forms are complicated and unclear for patients. Patients are unlikely to be willing to give consent for their data to be shared if they cannot be certain what they are agreeing to. The complicated information being provided by the consent forms currently in use are exacerbating this problem, particularly among patients from poorer socio-economic backgrounds with lower levels of education. A standardized, easy-to-understand language that is shared across all consent forms may help patients to become more accepting of sharing their data.

- Patients may not want to provide their consent forever, and there needs to be mechanisms in place that allow them to retract their consent once given. However, this requires the data stewards from tracking how and when each patient's data is used, which can become impossible if it was entered into a publicly accessible database. Withdrawing consent while research is being carried out can also set studies back dramatically. One solution being employed is to ensure that if consent is retracted, the data is prevented from being used in any further work, but remains a part of any ongoing studies.

- Keeping in contact with participants over time is challenging. Often, genomic studies can take place over many years, during which time the participants will move to a new house, change their phone numbers, etc. Keeping in contact with all participants can be a very complicated process and, in some cases, could be expensive. For a patient to be able to retain the right to withdraw their consent, they need to be contactable, but there isn't currently an effective system to manage this.

- Genomic data isn't officially considered protected health information, and so clinicians are not legally required to tell their patients that their information may become identifiable in future. Nonetheless, many labs are treating genomic information in the same way they handle data that is legally considered PHI.

Security

- Data is best protected using standard safeguards. For instance, a common approach for genomic datasets is to require researchers to apply for use and then ask them to travel to the data instead of sending it out. In this way, data never leaves the control of the stewards. Similarly, it is possible to develop and deploy smart data access exchanges that only enable authorized access and limit the data available to only what is needed.

- There are still many misconceptions about the security of the cloud. Some research groups prefer to store their data on their own hard drives that they transport between labs. However, in many cases, the data is at greater risk of damage or theft during this transfer than it would be if it was kept on an isolated cloud network.

Culture

- Genomic information is difficult to understand for patients and clinicians. Many of the primary care physicians being asked to use genomic testing will not have received extensive formal training in the field because of its novelty. Similarly, members of the general public are unlikely to have been educated about next generation sequencing or genetic tests. This could be minimized with better, more accessible education streams being offered at various levels of understanding.

- Often the benefits of sharing data for research purposes are misunderstood. Many patients who consent to share their data do not realize that doing so is unlikely to bring an immediate benefit to them personally. To prevent these misunderstandings from occurring, there needs to be a push to enable better public education and awareness, in a similar fashion to early efforts highlighting the dangers of smoking.

- Only patients with health insurance are likely to benefit from genomic research. Very few patients that do not have their own health insurance are likely to pay for genetic testing or treatments, and so there are less likely to be willing to share their data. This presents a barrier for people of poorer socio-economic backgrounds and excludes their data from research.