Workshop on Establishing a Central Resource of Data from Genome Sequencing Projects

Certified Researchers and a Research Commons Data Access Model

Ewan Birney, Michael Boehnke, Lisa Brooks, Wylie Burke,
Adam Felsenfeld, Laura Rodriguez, Stephen Sherry

Human subject information is nearly always obtained under specific consents signed by the individuals. Generally, the overall goal of research participants and researchers is to maximise the utility of these data for the understanding and ultimate treatment of disease. This drive for utility of the information demands that we continually seek to improve access to the data consistent with the consents signed by individuals. A particular opportunity has arisen with the advent of far broader consents, written from the perspective of general human health investigation, without specific restrictions on the diseases that are studied or the researchers involved beyond certain "obvious" behaviours, e.g., keeping the information anonymous. This conceptually allows an inversion of the current practice where a researcher must be authorised to see each dataset, to a scenario where the researcher is "pre-authorised" to see all datasets that fit a set of criteria. We will describe this researcher as a "Certified" researcher for the rest of this document (although note formally this is not changing any aspects of the relationship between participant consent and researcher).

This process would allow a Certified researcher to analyse a far greater range of scientific projects than can be routinely analysed currently, and may also allow the development of novel analyses. For example a Certified researcher might routinely run any new result, say in the molecular understanding of disease, against a full panel of disease studies, without having to specifically enumerate a hypothesis each time and obtain authorisation for each one. The concept of a "Certified researcher" is independent of where the data from the many studies would be stored.

Central databases might pool the storage and processing of disease studies for Certified researchers in a cost-effective manner, both saving resources and (probably more importantly) lowering barriers for running analyses. Having appropriate data sets with consent for broad use in a central database would allow a "Research Commons", where Certified researchers could do analyses across all the data sets in one database. There could be a few such Research Commons, at central databases such as at NCBI and EBI.

It must be stressed that this process has to be consistent with the consents. New consents allowing for broad use by Certified researchers would allow studies to be placed in a Research Commons, with IRB approval. IRBs would need to look carefully at the consent forms and processes for existing studies, to determine whether they allow for broad use by Certified researchers and could be placed in a Research Commons.

Another possible section of a Research Commons could include data sets with disease restrictions on use. Researchers accessing this section would have to agree to abide by the disease restrictions, just as they do now for each study they access through systems such as dbGaP. Even so, it could be valuable to have a central database with many data sets on T2D, cancer, or neurodegenerative diseases. If the consents allowed broad use but limited who could access the data, then some users (such as company researchers) may not be allowed access to certain data sets, but they could do any analyses on the ones they could access. For future studies, consent for broad use and broad range of users would be encouraged.

From a database perspective, placing studies in a Research Commons would be similar to placing them in dbGaP, allowing access by Certified researchers to the entire set of studies.

Studies with very sensitive information, such as drug addiction, or ones working with disadvantaged groups where limiting the information release is critical for obtaining participation, may not be appropriate for inclusion in a Research Commons or for automatic access by Certified Researchers.

Certification

Current authorisation for access to a particular dataset requires a request from an individual researcher, backed by an institution, with a specific analysis task. The data access request is then considered by the relevant data access committee, and the researcher is given access to the dataset or not. Certification would not change this general process, but rather would utilise the commonality of modern, broad consents to change the practice, specifically:

1. The analysis task would be considered to be "performing analyses with the broad goal of understanding human health and disease biology" and there would be standard restrictions on keeping the data secure and not attempting to deanonymise any information.
2. The researcher would have some regular certification to ensure he or she understands the responsibilities undertaken in 1 and the broader issue of the ethics involved in analysing human subject information.
3. The institution would have the same requirements as currently, that it provides appropriate data security measures and oversight systems.

Compared to the current situation, the provision of a formal certification procedure is likely to increase the compliance of researchers and also prevent the common practice of postdocs or students using data authorised to a senior researcher, as all the individual researchers would be certified. Certification of researchers might be done by a funding agency (such as the NIH).

International cooperation

The shift towards certification of individual researchers could greatly foster international coordination as well, with the agreement to recognise appropriate certification procedures in different countries as being equivalent, and thus allowing certification to cross between borders. In theory, there could be more than one certification process inside a country, but one-per-country seems the most logical and efficient scheme. It should be noted that different countries have different national legislation about investigator access to patient data; nevertheless, broad consortia in genetics have already been formed consistent with laws in multiple countries. As the certification process is not a fundamental shift in how the consent/researcher interaction works, this process should be achievable.