Beacon FAQs

1) What is a "Beacon" and what type of personal information does it contain?
A Beacon is an online web service that allows users to query an institution's databases to determine whether they contain a genetic variant of interest. The query is structured as a yes/no question of the form: "Do you have any genomes with an 'X' at position Y on chromosome Z?" Currently 19 organizations are sharing yes/no information about the existence of genetic variants as part of the GA4GH Beacon Network. Supporting the Beacon API on top of an existing database takes only minutes to implement.

2) What is the Beacon Project doing to mitigate the possibility and likelihood of re-identification through Beacons?
Since its inception, the Beacon Project has been actively working with experts in GA4GH's membership to mitigate any potential privacy concerns. These efforts include:

- A **data aggregation** approach, in which several private Beacons are combined into a single public Beacon. The power to re-identify is a function of the number of data points served by a Beacon and the scope of the data it contains (i.e. whole genome versus exome). Aggregate Beacons contain more data points and do not obscure the origin of the data.
- An **information budgeting** approach to track the cumulative information revealed to any one user and restrict access if information disclosure exceeds a threshold. The power to re-identify an individual varies inversely with the frequency of the alleles being queried. That is, very rare alleles reveal more about a given individual than those that are common among the population. By metering the information disclosure, Beacons can restrict access before re-identification is possible.
- **Secured access** to some of the data. Before a user is approved to access data, s/he must provide credentials and/or research plans and agree not to re-identify individuals. Some Beacons are already providing access to data in this secured way.

3) Why are Beacons important to global genomic data sharing and why is global genomic data sharing important in the first place?
Beacons present an easy-to-implement strategy for determining whether an institution has genomic data in its data set that would be useful to share. Those data are only as useful as the comparisons that can be made against them across hundreds of thousands — even millions — of individuals. Since no single institution can collect such large samples on its own, sharing is critical for genomic data to have a transformative impact on human health and disease. The initial yes/no query step is not only an interesting exercise about the willingness to share, but it also has the potential to save researchers countless hours tracking down useful data for their work.

4) What do institutions commit to when they join the Beacon Network?
When an institution "lights" a Beacon, it demonstrates its willingness and ability to share data. The organizations in the Beacon Network have created online search functions that let anyone in the world know if a certain allele exists in their datasets — but only to find a particular kind of information that was carefully chosen not to overly expose genomic data to privacy or security risks. The rapid growth of the Beacon Project shows

that many institutions across the globe are interested in working together to break down silos and share genomic information in innovative and responsible ways.

5) How is sharing data through Beacon different from traditional medical practice?
In current medical practice, doctors commonly share basic, de-identified information about patients' conditions and genetic variations in the hopes of matching a second case that could be therapeutically transformative. This practice is a long-standing tradition in the medical community but it is neither standardized nor optimized. The Beacon Project turns that informal transaction into a more efficient and effective search that can be a key initial step for learning from valuable genomic information.

Since much of that information is currently collected and held in silos, researchers lack the broad-scale ability to know whether similar data exist that could provide important learning when combined with their own data.

The Beacon Project was created both to test the willingness of organizations to share at the most basic, yes/no level, and also to allow researchers to understand if a specific allele exists in other data sets. Now that a Beacon Network has been created, researchers have the ability to discover valuable information within organizations that have lit Beacons. This is often a needed first step to aggregate and learn from large-scale genomic data.