About Synapse

The past two decades have seen an amazing exponential growth in the technical ability to generate genetic and biomolecular data fueled by incredible advances in measurement technologies. These breakthroughs have resulted in an increasing amount of resources directed at genomic research by both industry and academia. However, with a few exceptions, these investments have failed to improve prevention or treatment of common human disease. For example, the numbers of new drugs approved by the FDA has actually declined over this period.

A fundamental reason for this discrepancy between data generation and clinical improvement is the immature development of analytical techniques to meaningfully interpret these new data types. As with any new field, analytical methodologies need to be iteratively developed and refined. The difficulty of accessing, understanding, and reusing data, analysis methods, or models of disease across multiple labs with complimentary fields of expertise is a major barrier to the effective interpretation of genomic data today. Additionally, much of the relevant data to answer a particular research question is spread among multiple public and private repositories. Because each research group protects their own data, the end result is enormous duplication of effort and missed opportunities across both industry and academia.

Sage Bionetworks' mission is to catalyze a cultural transition from the traditional single lab, single-company, and single-therapy research paradigm to a model founded on broad precompetitive collaboration on analysis of large-scale biological data. This model would benefit future patients by accelerating development of disease treatments, and society as a whole by reducing costs of health care and biological research. It could even benefit the pharmaceutical industry by seeding increased innovation at a time when the sector is struggling to replace revenue lost as old medicines go off patent production. Sage Bionetworks is actively engaged with academic, industrial, government, and philanthropic collaborators in developing this research model.

The technology component of Sage Bionetworks' solution strategy is Synapse, an informatics platform for open data-driven collaborative research. Synapse will serve as a public resource for the broad scientific community.

Catalyzing a transformation to collaborative research requires a platform that helps scientists solve a series of problems:

1. **Finding and using relevant data –** Currently, scientists have difficulty tracking down and gaining access to data and resources generated by others, even within the confines of the same organization. Even when data is available, it is often not useable. Indeed, an estimated ~75% effort within each analytical project is devoted to interpreting data structures and appropriately preprocessing data. Synapse provides a mechanism to access data in a uniform manner that leverages common formats, controlled vocabularies, and annotation standards.
2. **Understanding analysis workflows** - Synapse is built with the understanding that most analytical research is experimental and ad hoc in nature, with hardened analysis methods only emerging over time. Tracking who has run what version of code on what version of the data immediately helps projects run more smoothly, and ultimately enables reproducible workflows that allow others to build off of prior work.

3. **Supporting genome-scale analysis** - Analyzing datasets with information on whole genomes is currently limited to those with access to large computational resources and significant IT support. Synapse makes cloud computing technologies accessible to scientists.
4. **Forming and maintaining productive collaborations** – Scientists tend to start from scratch on a project rather than take work in an unknown state. The platform must help scientists track what work has already been done in a particular area and help create and sustain collaborations.

To ultimately support this vision, multi-disciplinary users will need to interact with Synapse. Initially the Synapse system will focus on supporting the statistical or computational biologists that are directly involved in manipulating clinical and genomics data sets; over time support will grow for more biological or clinically focused researchers.

Synapse leverages a web service-based architecture in which a common set of services is accessed via different sets of client applications to support a growing set of use cases over time. One client application is the Synapse web portal: an environment for scientists to discover and share data, models, and analysis methods. The portal is organized around projects, which any scientist can create and invite collaborators to join. These online workspaces then serve as the glue to help teams of researchers collaborate to solve complex scientific analysis problems. Additional capabilities to visualize data and interact with disease models through the web portal will be developed over time.

We expect our initial target users to already be proficient in using data analysis tools and to want to continue using those tools to work with data. Synapse's first analysis client will be an integration with the R / Bioconductor statistical package allowing users to track analyses performed with published packages or custom scripts to complete their work. Additional analysis platforms (e.g. Gene Pattern or Matlab) will be supported over time. All analysis platforms would interact with the Synapse system by calling the Synapse web service layer. This is a set of REST-based services providing support for annotating, querying, and updating data, analysis code, and models, and controlling access to these resources. These services will also allow tracking of the provenance of a multi-step analysis procedure, and executing analysis steps on cloud-based resources.

For more information, please see the Synapse Vision Document or contact synapseInfo@sagebase.org.