“Cavatica: empowering research with a pediatric genomic cloud”

Pichai Raman, Alex S Felmeister, Phillip B Storm, Rishi R.Lulla, Angela J Waanders, and Adam C Resnick, on behalf of the complete team membership of the Children's Brain Tumor Tissue Consortium (CBTTC).

The pediatric cancer genome is severely under-represented in genomic databases as existing data portals have primarily focused on adult cancers. Furthermore, large-scale pediatric datasets like TARGET lack pediatric central nervous system (CNS) data. To address this unmet need, we have developed a new cancer genomic platform named Cavatica. The rationale behind Cavatica is to provide a sustainable application cloud based eco-system that supports many of the aspects associated with basic & translational research. Cavatica is the first of its kind pediatric genomic portal for disease research, and its goal is to serve as a central hub to promote collaborative research between investigators. Cavatica supports the sharing and creation of pipelines, data, algorithms, visualizations, and hypotheses. Currently, one of the biggest barriers and challenges to collaborative research is the transfer and processing of ‘big data’ such as cancer genomes. By placing data, pipelines, computation, and visualizations on the Cavatica cloud we provide a centralized area for researchers to collaborate on projects and bring their data, algorithms, and expertise. Currently, Cavatica features the following applications designed to work together: a biorepository and specimen query tool (Harvest, harvest.research.chop.edu), a data visualization application (PedcBioPortal, pedcbioportal.org), data storage in S3 buckets, and data processing via Seven Bridges Genomics. Users can move seamlessly between these applications, and thus can go from points on a graph to physical samples. Cavatica also protects data or pipelines on an individual and group basis so various team members can share a common working space with controlled or a single individual can store experiments in a private space. All current solutions will be constantly evaluated and replaced as technology evolves. Cavatica is set to house data from a number of sources including the Childhood Brain Tumor Tissue Consortium (CBTTC), Pacific Neurooncology Consortium (PNOC), Stand Up to Cancer (SU2C), NCI Therapeutically Applicable Research To Generate Effective Treatments (TARGET), and The Cancer Genome Atlas (TCGA). Cavatica’s framework will also allow unique opportunities for data scientists, statisticians, data engineers, programmers, application developers, bioinformaticians, and pre-clinical and clinical researchers to contribute and expand the reach and impact of this application.