U-BRITE : a biocomputing infrastructure to enable collaborative genomic data science

Jelai Wang, Jake Y. Chen, and the UAB Informatics Institute U-BRITE team

http://ubrite.informatics.uab.edu

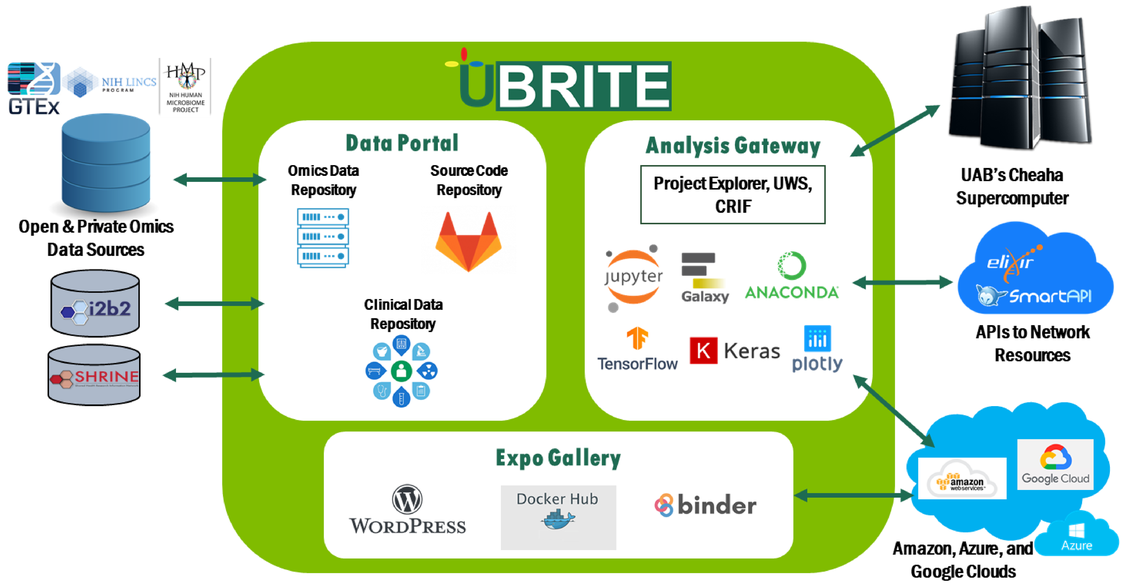
INTRODUCTION

U-BRITE (*UAB Biomedical Research Information Technology Enhancement*) assembles HIPAA-compliant, clinical informatics and bioinformatics tools layered on top of high-performance computing infrastructure. We aim to help researchers better manage and analyze genomic medicine data sets in a new "translational research commons" environment. U-BRITE will facilitate and enable interdisciplinary team science across geographic locations.

Aims

- Equip translational biomedical researchers with cutting-edge data science tools
- Facilitate computational method development for integrated genomic medicine studies
- Build an online workspace and enable data-intensive scholarly communication for interdisciplinary study teams

ARCHITECTURE OVERVIEW



CCTS

SUPPORT CONTACT

To start a project
Contact Dr. Jake Y. Chen (jakechen@uab.edu)

Technical Support
Contact Jelai Wang (jelaiw@uab.edu)

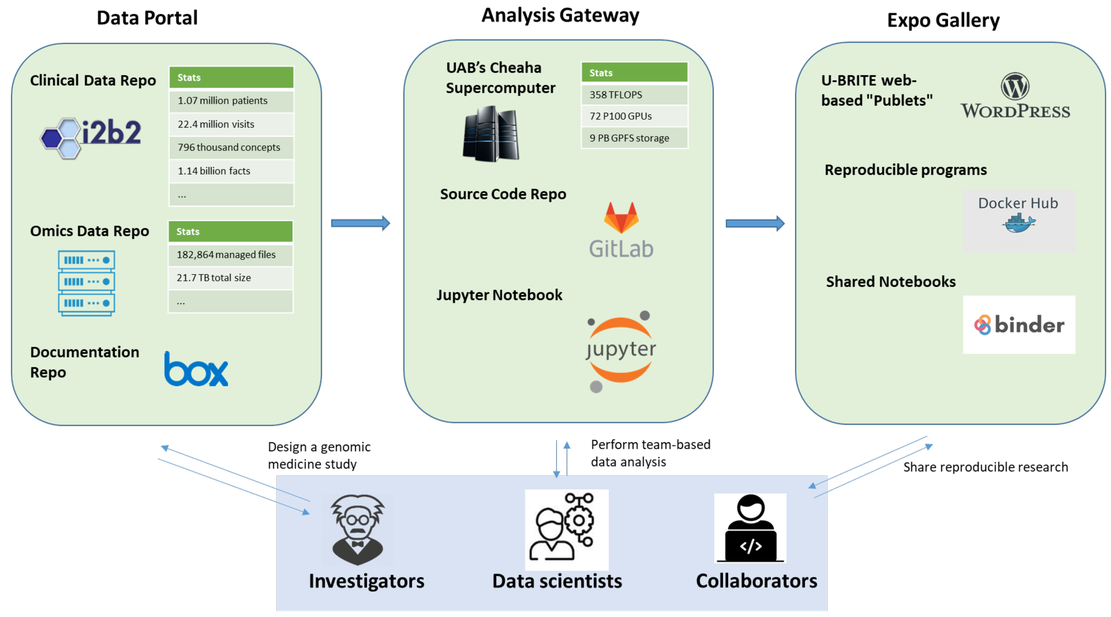
Visit the U-BRITE web site at http://ubrite.informatics.uab.edu/.

UAB INFORMATICS INSTITUTE
The University of Alabama at Birmingham

BENEFITS

- Access clinical data from **Clinical Data Repository** to build cohorts.
- Integrate multi-omics data from local databases or remote API inside the **Omics Data Repository**.
- Prototype in Jupyter Notebook and deploy to UAB's Cheaha supercomputer without leaving the **Analysis Gateway**.
- Develop team-based or open-source coding via the **Source Code Repository**.
- Achieve and ship reproducible research workflow with **Binder** supported by U-BRITE.
- Streamline NIH-compliant genomic data sharing with scholarly communication using **Expo Gallery**.
- Self-help or gain full support from data scientists in the UAB Informatics Institute.

USE CASE SCENARIOS



A genomic medicine research team can take advantage of multiple components of the U-BRITE infrastructure to perform team-based biomedical data science.

CASE STUDY

Case Study: A Multiethnic Longitudinal Study in SLE

A significant unmet need was the inability to easily integrate these existing datasets to drive further discovery research into the biological basis for disease development and progression. Accordingly, supported by U-BRITE, Dr. Jeff Edberg's team developed a master database of IDs associated with each participant (study ID for clinical data, genotyping IDs for experimental results). Using a Jupyter Notebook allowed us to merge data elements from our phenotypic and genotypic resources. We have also built an interface that allows us to derive real-time clinical data from the EMR through U-BRITE's Clinical Data Repository. This foundational infrastructure allows us to more easily integrate our WES data for analysis. Key questions to be addressed will be genotype-phenotype relationships with SLE development in minorities (African-Americans) and genotype-disease severity (such as renal disease) relationships based on up-to-date clinical data. Finally, with integration of our specimen biobank, we will be able to easily determine availability of specimens for mechanistic studies as follow-up to the genetic findings with further linkage to EHR data to determine patient availability for collection of new specimens.