Breakout Session 3: Track B

Enhancing Kids First Digital Pathology Datasets Via Scalable, Cloud-based Data Management, Processing, and Analytics

Dr. Ariana Familiar
Senior Data Scientist, Children's Hospital of Philadelphia

Children's Hospital of Philadelphia®

D3b Center for Data-Driven Discovery in Biomedicine

Enhancing Kids First digital pathology datasets via scalable, cloud-based data management, processing, and analytics

Ariana Familiar, PhD

Translational Imaging Research Unit

Center for Data-Driven Discovery in Biomedicine,

Division of Neurosurgery, Children's Hospital of Philadelphia

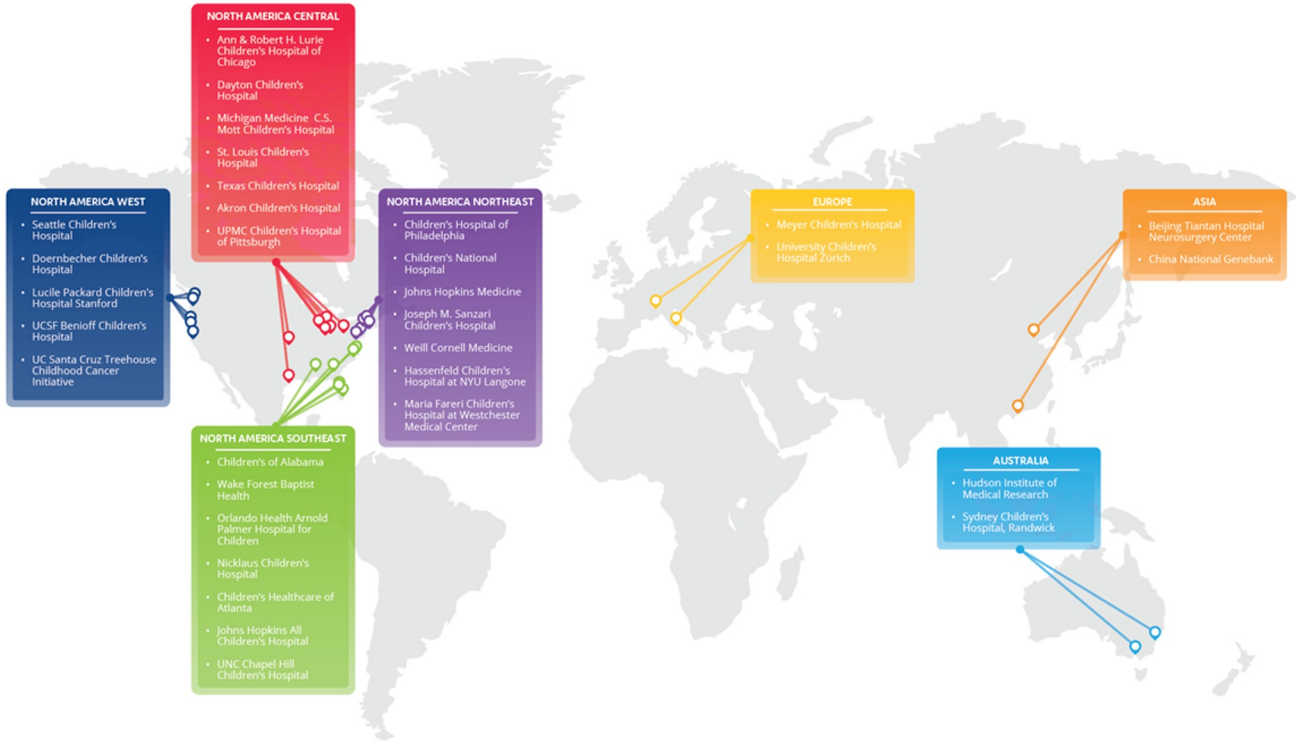
Gabriella Miller Kids First pediatric research program

- The Gabriella Miller Kids First Data Resource Center (KFDRC) focuses on empowering research in the rare disease space via centralized collection and public release of data repositories, and empowering data exploration to analysis.
- The KFDRC has integrated large amounts of genomic and clinical data from different disease types, while in parallel developing data-driven platforms and resources
- Through this program, some of the largest, multi-modal pediatric datasets have been generated, harmonized, and released for use by the research community.

Children's Brain Tumor Network (CBTN)

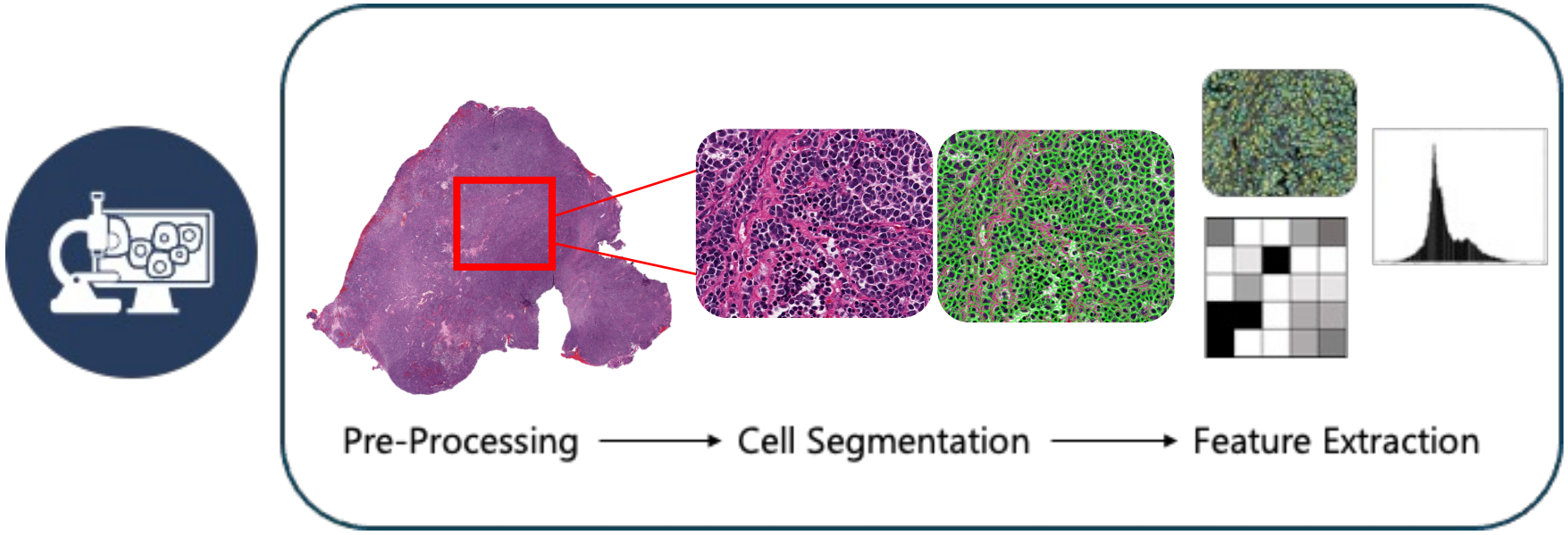
- An international consortia of pediatric healthcare institutions focused on advancing brain tumor research and scientific discoveries

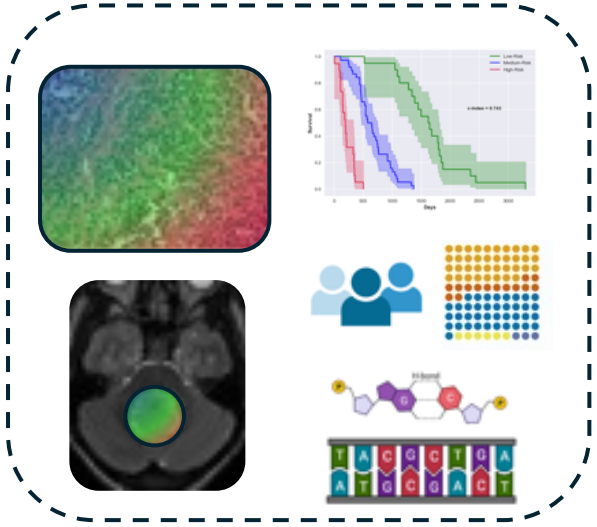
	October 2023
CBTN Sites (including onboarding)	34
Total Consented Patients	5,300
Approved Research Projects Using CBTN Data/Specimens	342
Biospecimen Collected	69,368
Molecularly characterized parents & participant surgical events	+6,000
Preclinical Models	+180
Imaging Sessions	+23,000



Digital pathology data for clinical decision-making

- Digital pathology data has the potential to provide novel, rich characterizations of pediatric diseases





Predictive Analysis

CBTN: 2,006 subjects with 8,219 pathology files to-date

Data management and processing challenges

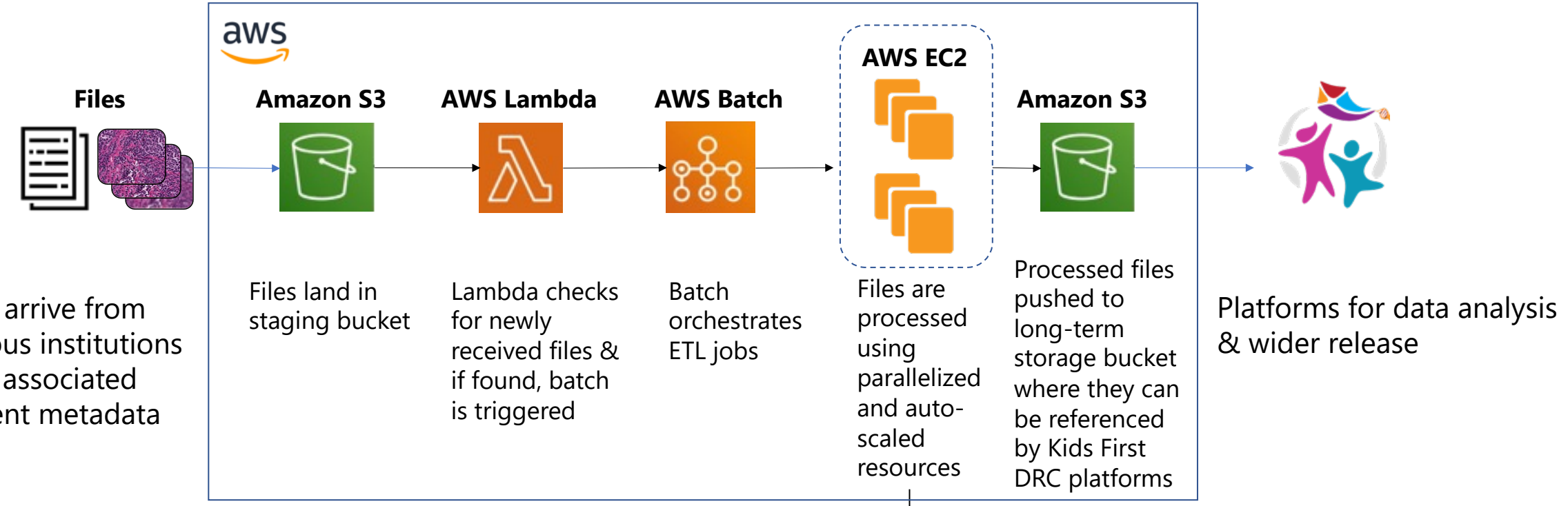
- Collecting and utilizing digital slide images in research contexts is met with significant operational and technical challenges, such as:
  - existing open-source software not readily integrated with cloud services
  - high demands of data file storage & computational resources
  - non-uniform file formats & PHI/PII de-identification requirements

- In this project we implement and test cloud platforms and services for scalable intake and preparation of digital pathology slides under the KFDRC.
- Our goal is to establish an end-to-end infrastructure and workflow in a high-performance cloud environment to facilitate rapid, uniform collection and preparation of digital pathology slides for scientific research

Enhance the ingest of digital pathology data for Kids First projects through AWS cloud-based services with automation, parallel batch processing, and auto-scaled computing resources.

Digital pathology data & workflows

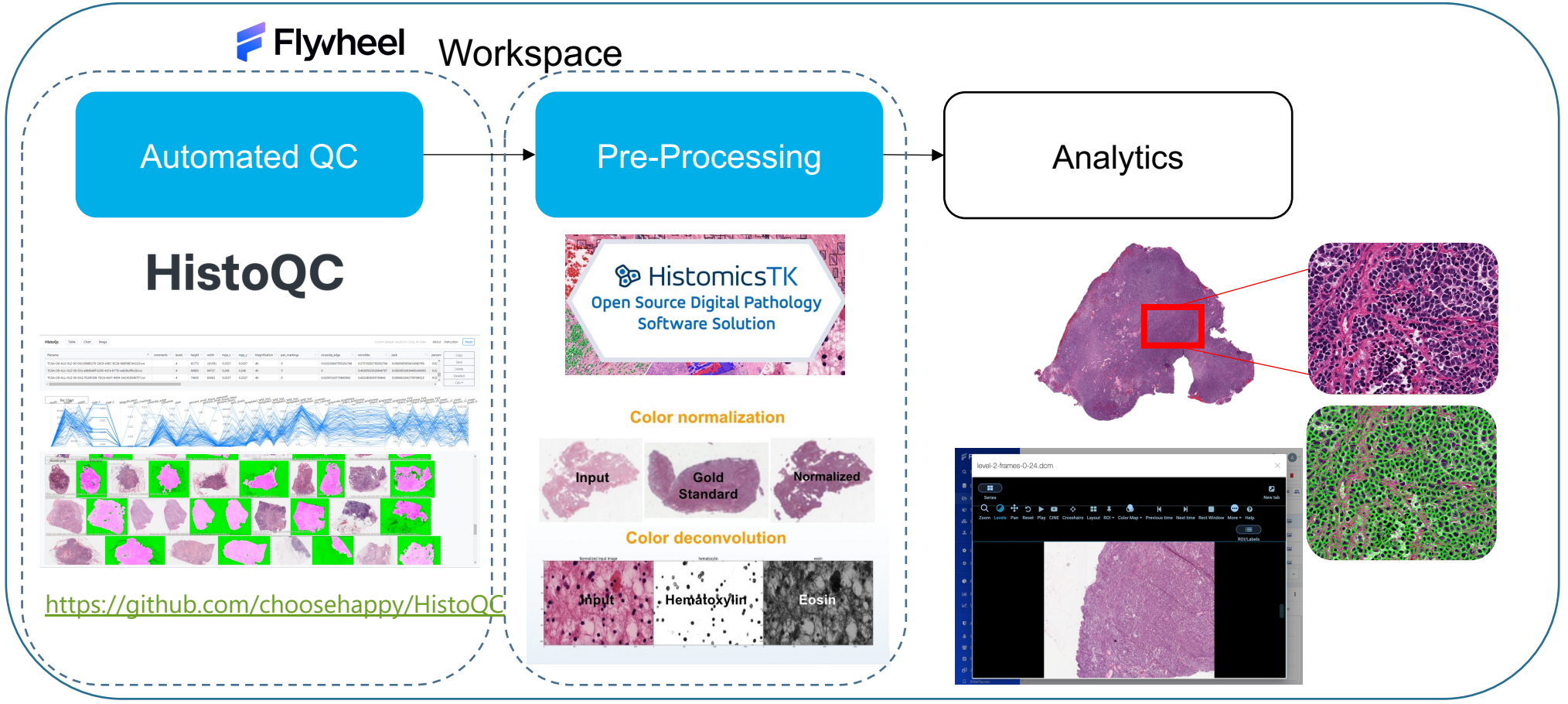
- Phase 1: Scaling pathology infrastructure & processing pipelines using AWS cloud services



Enable use of pathology data in AI/ML analytics with automated image QC and preparation tools in a cloud-hosted analytics platform.

Digital pathology data & workflows

- Phase 2: Integrating automated QC and data preparation tools into Flywheel to make digital pathology slides useable in downstream analysis



Thank you!

familiara@chop.edu