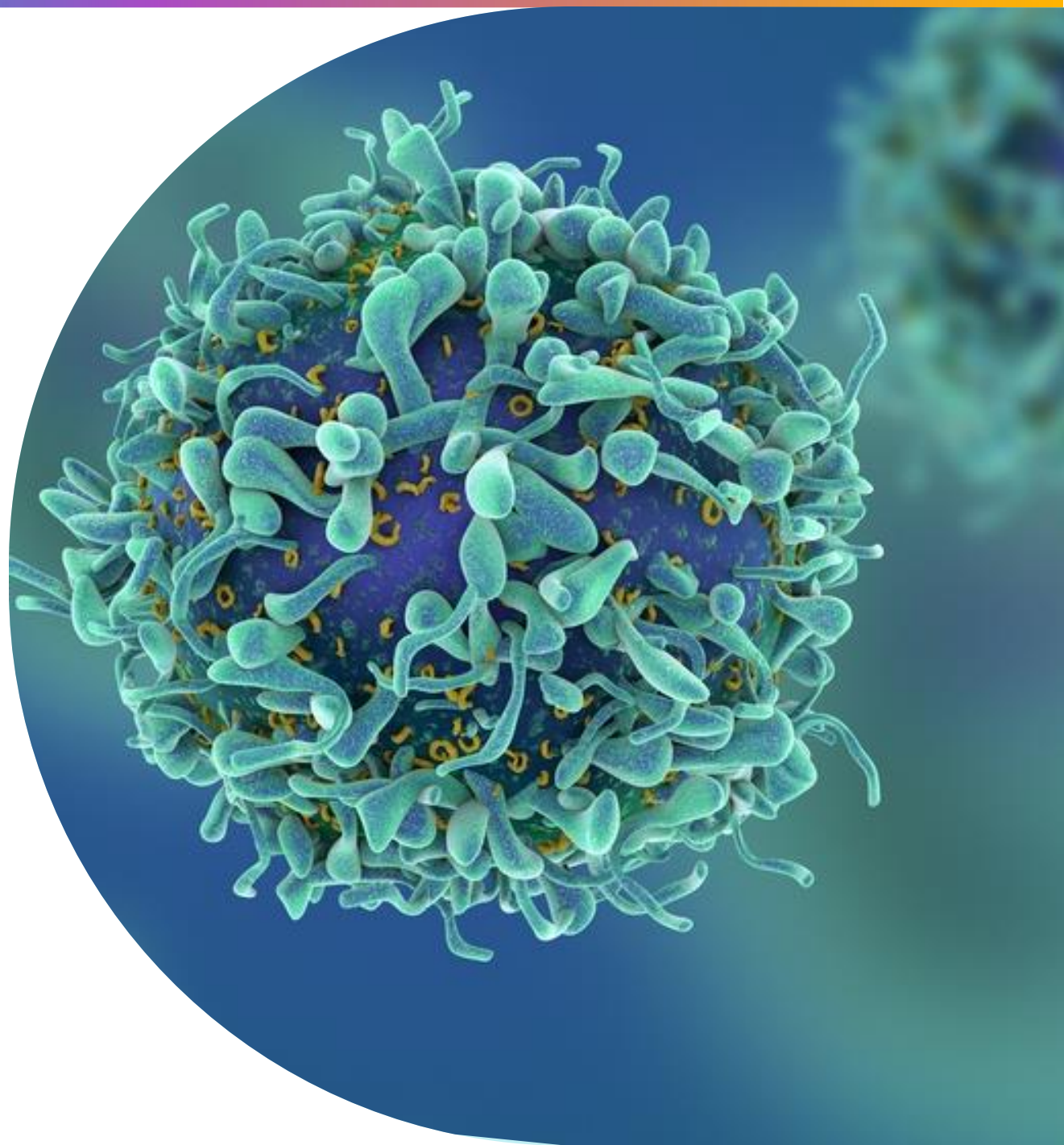
Fred Hutch Cancer Center

Stewarding Data Distributed by Core Facilities: Lessons Learned in the Cloud



Sam Minot, Ph.D.
Associate Director
Data Science Applications
Bioinformatics Core, Shared Resources

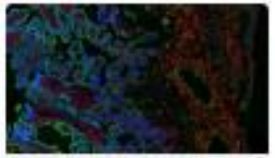
Fred Hutch: Centralized Shared Resources

Antibody Technology*

Internationally recognized team that offers novel monoclonal antibody discovery, production and characterization services.

Biospecimen Processing and Biorepository

Provides investigators with high-quality biospecimen sample p extractio services.



Cellular Imaging*

Provides light microscopy services that support both basic and preclinical projects.

Clinical Research Specimen Processing

Provides investigators with processing, storage, and

Comparative Medicine*

Enables high-quality science that proceeds with maximal efficiency, care and regulatory compliance.

Electron Microscopy*

Offers electron microscopy services to resolve structures from cellular ultrastructure to molecular

Experimental Histopathology*

Provides a range of histology, histochemistry, immunochemistry and

Flow Cytometry*

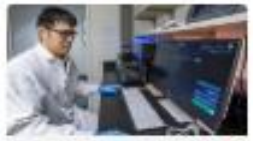
Offers top-of-the-line cell analysis and cell-sorting services, from single-laser benchtop analysis to mass cytometry and complex multi-

Genomics & Bioinformatics*

Provides integrated support to generate and analyze genomic data from DNA array, genetic analysis and high-throughput screening methods.

Proteomics & Metabolomics*

Provides mass spectrometry and high-performance liquid chromatography services for the analysis of proteins and peptides, including metabolomics studies.



Immune Monitoring*

Supports studies of cellular immunity, including assays of cell populations and responses in clinical trials.

Preclinical Imaging*

Provides state-of-the-art in vivo imaging technology and infrastructure to support basic and preclinical studies.

Therapeutic Products Program*

Offers development and manufacturing expertise for cell-based therapies for Phase 1/2 clinical testing.

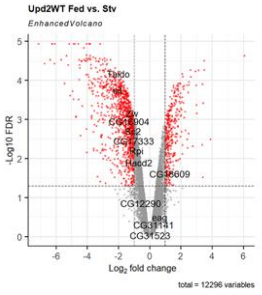
Shared Resources- Enabling Cutting-edge Research

Current Biology

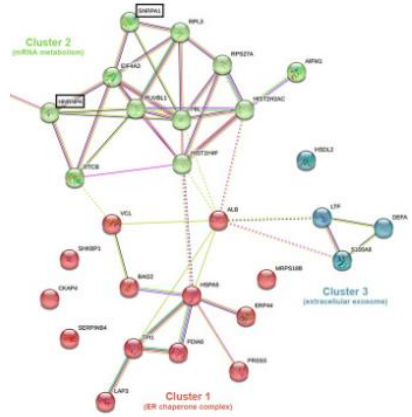
Atg8/LC3 controls systemic nutrient surplus signaling in flies and humans

Madan et al., 2024, Current Biology 34, 3327–3341
Authors Aditi Madan, Kevin P. Kelly, Patrick Bahk, …,
Mroj Alassaf, Julien Dubrulle, **Akhila Rajan**

Genomics and Bioinformatics

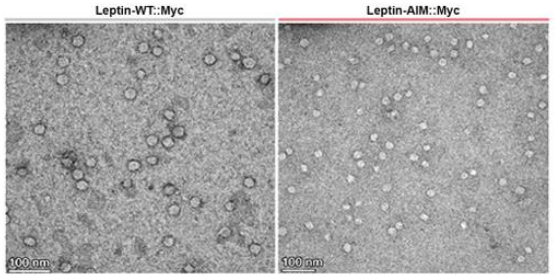


Proteomics and Metabolomics



Cellular Imaging

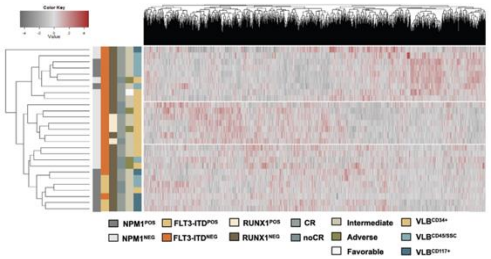
Electron Microscopy



Shared Resources- Enabling Cutting-edge Research

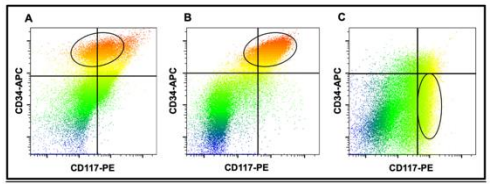
Proteogenomic characterization of highly enriched viable leukemic blasts in acute myeloid leukemia: A SWOG report

Jasmine Naru[1] | Megan Othus[1,2] | ChenWei Lin[1] | Melinda A. Biernacki[1,3] | Marie Bleakley[1,4] | Thomas R. Chauncey[1,3,5] | Harry P. Erba[6] | Min Fang[1,7] | Matthew P. Fitzgibbon[8] | Phillip R. Gafken[9] | Richard G. Ivey[1] | Jacob J. Kennedy[1] | Travis D. Lorentzen[1] | Soheil Meshinchi[1,4] | Anna Moseley[1,2] | Era L. Pogosova-Agadjanyan[1] | Vivian M. Liu[1,3] | Jerald P. Radich[1,3] | Uliana J. Voytovich[1] | Pei Wang[10] | Jeffrey R. Whiteaker[1] | Cheryl L. Willman[11] | Feinan Wu[8] | Amanda G. Paulovich[1,3] | Derek L. Stirewalt[1,3]



Proteomics and Metabolomics

Flow Cytometry*



Genomics and Bioinformatics

eJHaem
British Society for Haematology
Accepted: 23 October 2024

Synergy Across Shared Resources

Spatial Proteomics

Digital Imaging

Evident VS200: brightfield or fluorescence scanning

Akoya PhenoImager HT: multiplex IHC fluorescence imaging

Image Analysis for ROI Selection

HALO AI

VISIOPHARM®

Image analysis with HALO or Visiopharm

Laser Capture Microdissection of ROIs

Leica LMD7 laser capture microdissection microscope

Mass Spec

Integrated Research Offerings

Spatial Transcriptomics – *in situ* gene expression profiling

Biospecimen Processing and Biorepository

Provides investigators with high-quality biospecimen sample processing, DNA/RNA extraction and biorepository services.

Experimental Histopathology*

Provides a range of histology, histochemistry, immunochemistry and pathology services for human tissues and model organisms.

Genomics & Bioinformatics*

Provides integrated support to generate and analyze genomic data from DNA array, genetic analysis and high-throughput screening methods.

10X Genomics – Xenium Spatial Platform

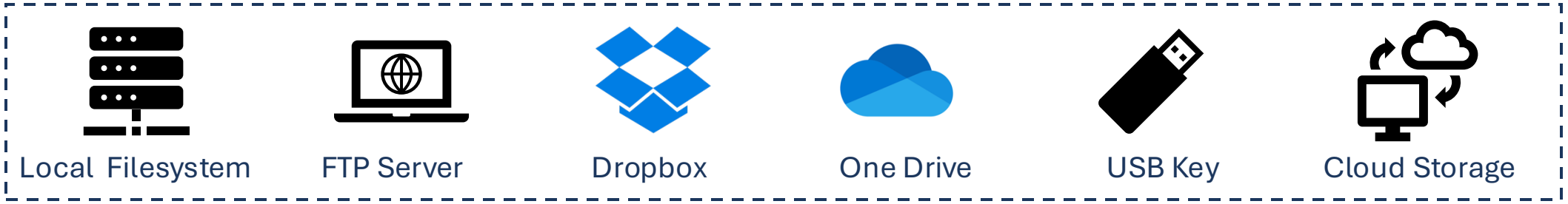
Infrastructure

Delivering Data at Scale

Fred Hutch Cancer Center

Data Delivery Mechanisms

Previously: multiple methods emerged to meet various users' needs



Core Facility / Instrument

Researcher / Customer

Data Needs: Shared Resources POV

Key Requirements from an Institutional Data Generator

1. Controlling cost (primarily on-prem usage charged to Cores).
2. Engaging with external customers.
3. Meeting rigorous data security standards.
4. Delegating data stewardship (and long-term cost) responsibilities to the customer.
5. Coordinating integrated research offerings.
6. Unlocking the unique value of advanced instrumentation.

FH Shared Resources – Development Timeline

Dedicated Effort to Implement Solutions for SR Needs

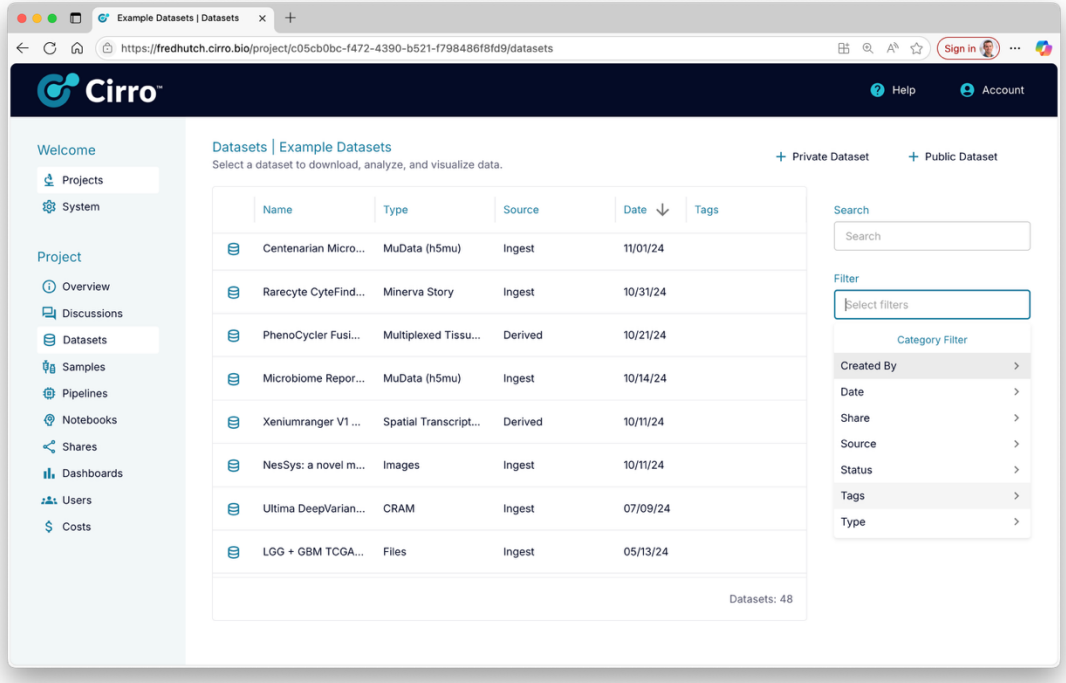
1. 2020: Formed the Data Core – a “virtual core facility” focused on immediate needs for data management.
2. 2021: Identified the **interface with cloud computing** as the primary opportunity for purpose-built solutions.
3. 2022: Deployed the **Cirro Data Platform** (originally “PubWeb”) initially to beta testing users.
4. 2023: Established **local storage gateway** to stream data directly to user-owned cloud storage accounts.
5. 2024: Continued to roll out Cirro as a data delivery option on a core-by-core basis.
6. 2025: **Cirro Bio, Inc.** spun out as independent company to provide services to peer institutions.

Cirro Data Platform

Cloud Interface Designed for Biomedical Research

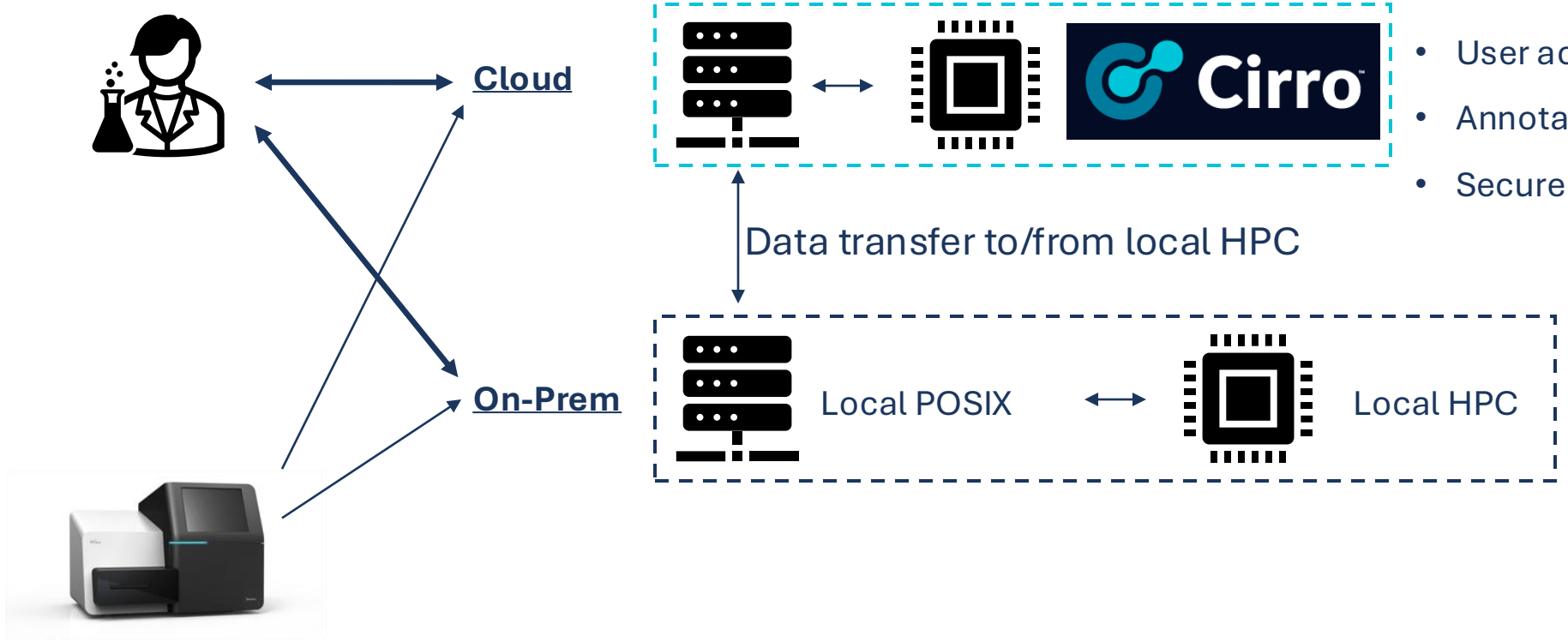
Key Features:

- Secure data storage – short-term use & long-term archival.
- Costs are isolated per lab / customer / project.
- Robust budget guardrails to prevent overages.
- Tools to enable self-service data stewardship.
- Accessible to FHCC staff and external customers.
- Compatible with the wide array of data types used in biomedical research.
- Supports interactive and HPC analysis of diverse datasets.



Ideal State – Unified Data Delivery

Ease-of-Use for Cloud Storage & Computing



- Cost control & budgeting
- User access control
- Annotation & organization
- Secure internal/external access

Work In Progress

Status Update – Lessons Learned

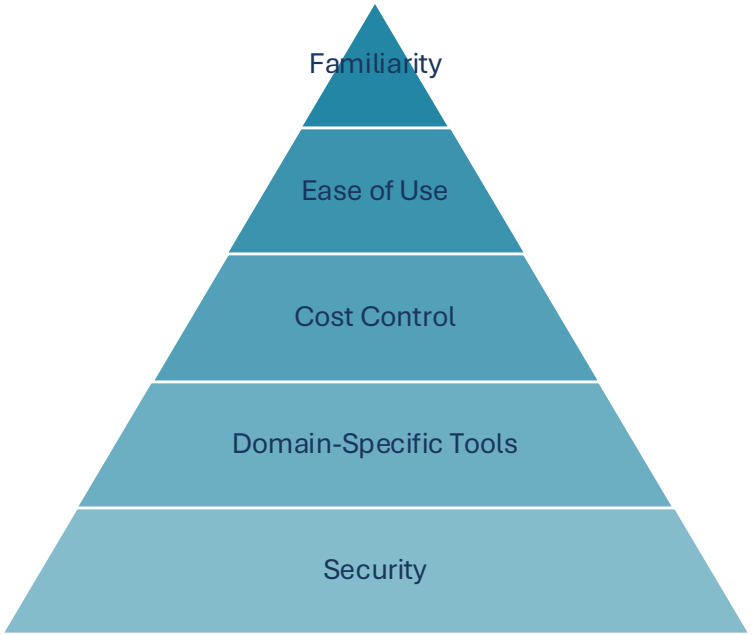
Organizational Implementation

Robust Institutions: People, Policies, Practices

- Data Retention – Formal policy delineating how data is generated, processed, delivered, and deleted.
- Billing – Costs incurred by customers need to integrate with existing billing structures.
- Data Security – Coordination with IT departments responsible for technologies and practices used for all data.
- Instrument Integration – Pathway of data transit may vary by instrument and facility.
- Visibility & Coordination with Core Staff – Datasets are often inspected, optimized, and updated as part of delivery.

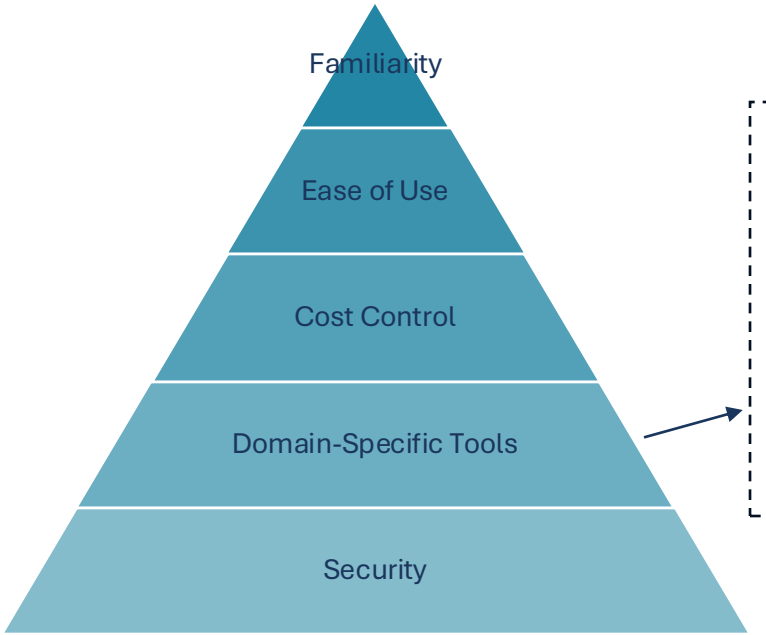
Biomedical Researcher – Hierarchy of Needs

Research Needs are Diverse and Unpredictable



Biomedical Researcher – Hierarchy of Needs

Research Needs are Diverse and Unpredictable



FLOWJO

VISIOPHARM®

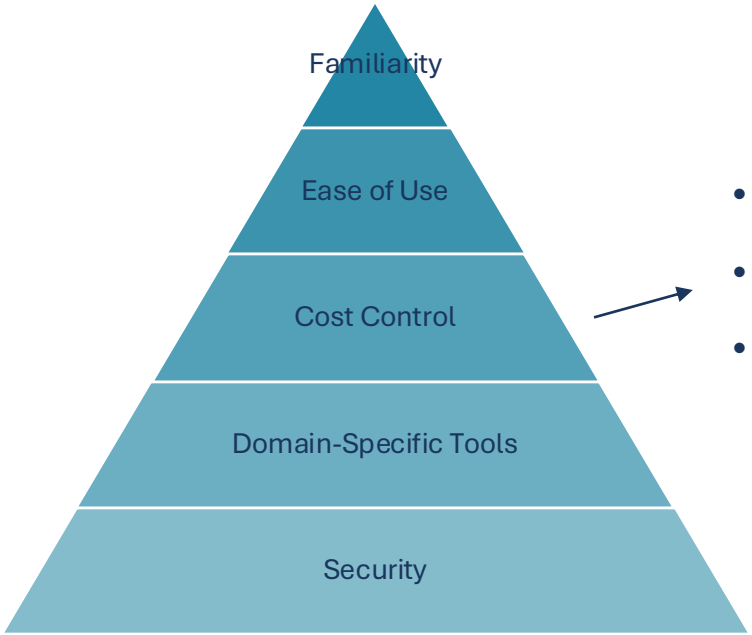
HALO®

Bioconductor
OPEN SOURCE SOFTWARE FOR BIOINFORMATICS

MATLAB®

Biomedical Researcher – Hierarchy of Needs

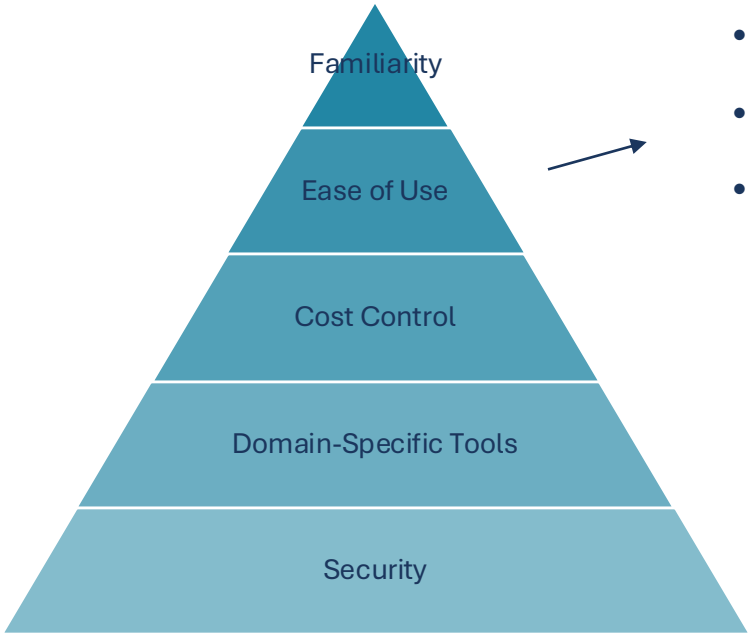
Research Needs are Diverse and Unpredictable



- What will it cost?
- How can I be sure?
- How do I allocate grant funds?

Biomedical Researcher – Hierarchy of Needs

Research Needs are Diverse and Unpredictable



- Browsing files in folders.
- Managing who has access.
- Finding the right data quickly.

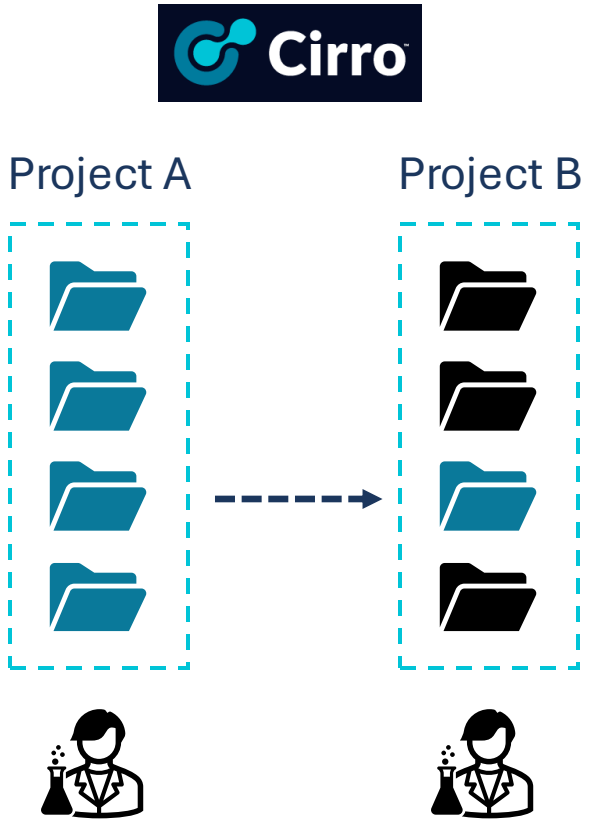
Planning Ahead

Biomedical Research: New Abilities

Easily Sharing Large-Scale Datasets

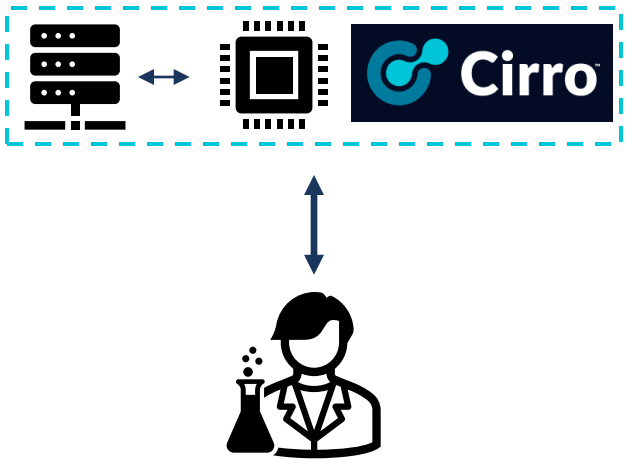
Collaborative projects often challenge data resources

- Data stewards can elect to share datasets between account in Cirro.
- No copies are made of the underlying files in cloud storage.
- Seamlessly enables multi-lab or cross-institutional collaborations.
- Allows core facility staff to maintain visibility on delivered datasets.



Extending the Impact of Bioinformaticians

Deploying automated bioinformatics workflows



nf-core

gatk

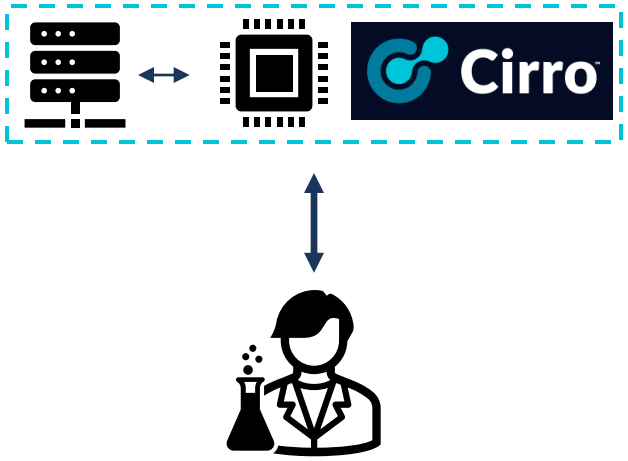
AWS HealthOmics

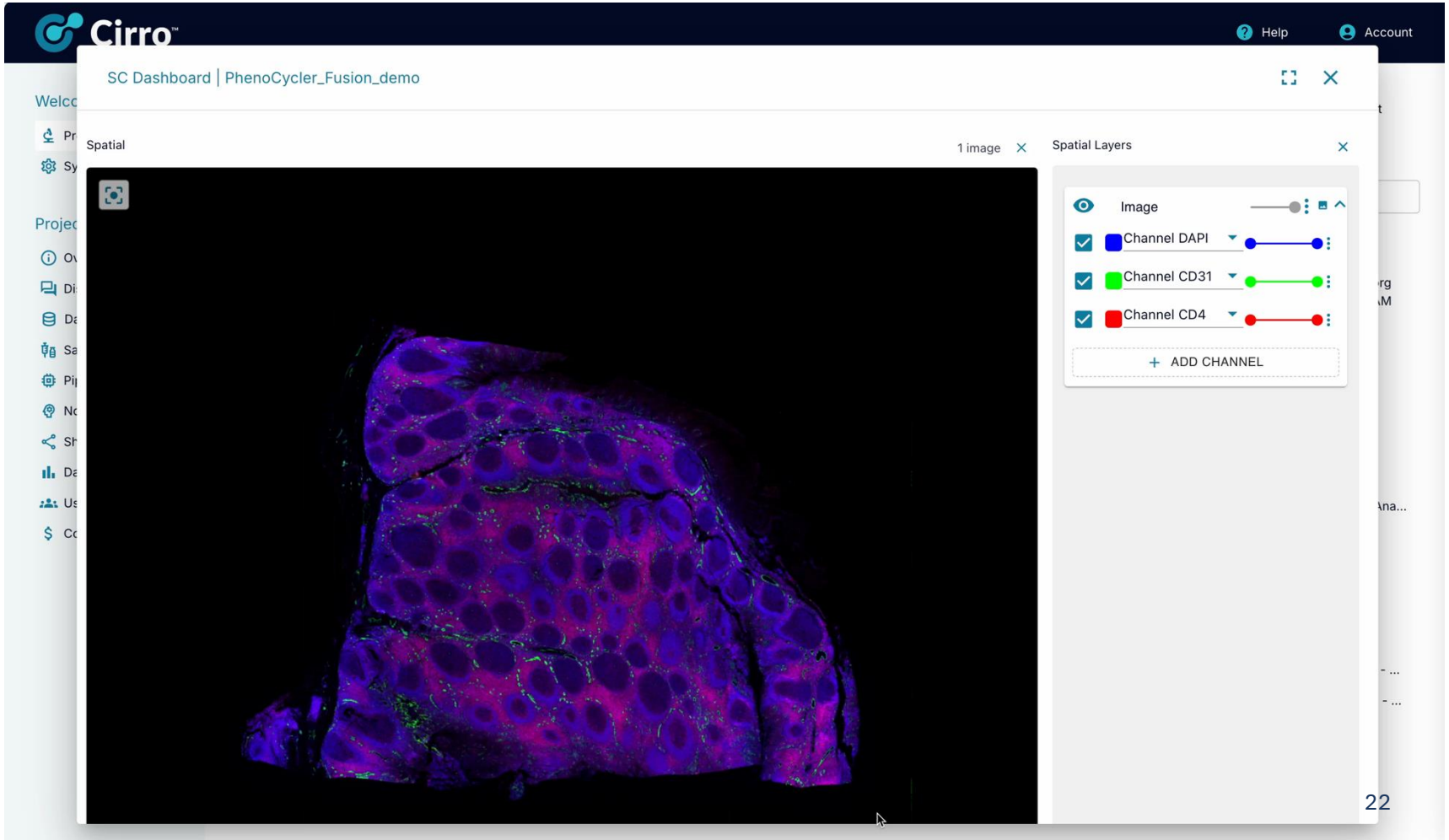
Resource-Intensive & Standardized Tooling

1. Genome Sequencing
2. Single-Cell Analysis
3. Microbiome Analysis
4. Image Processing

Unlocking the Value of Advanced Instrumentation

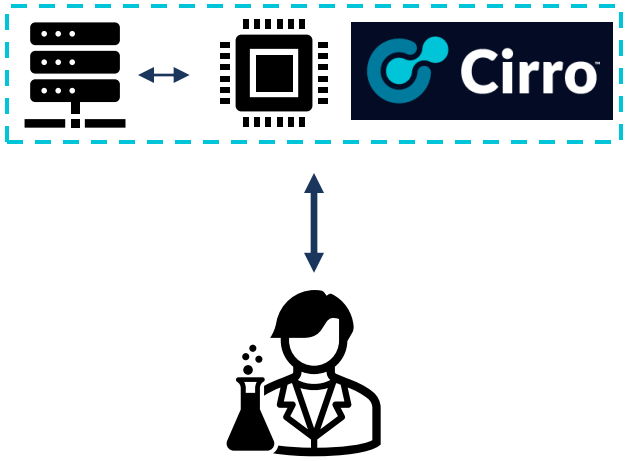
High-resolution image exploration

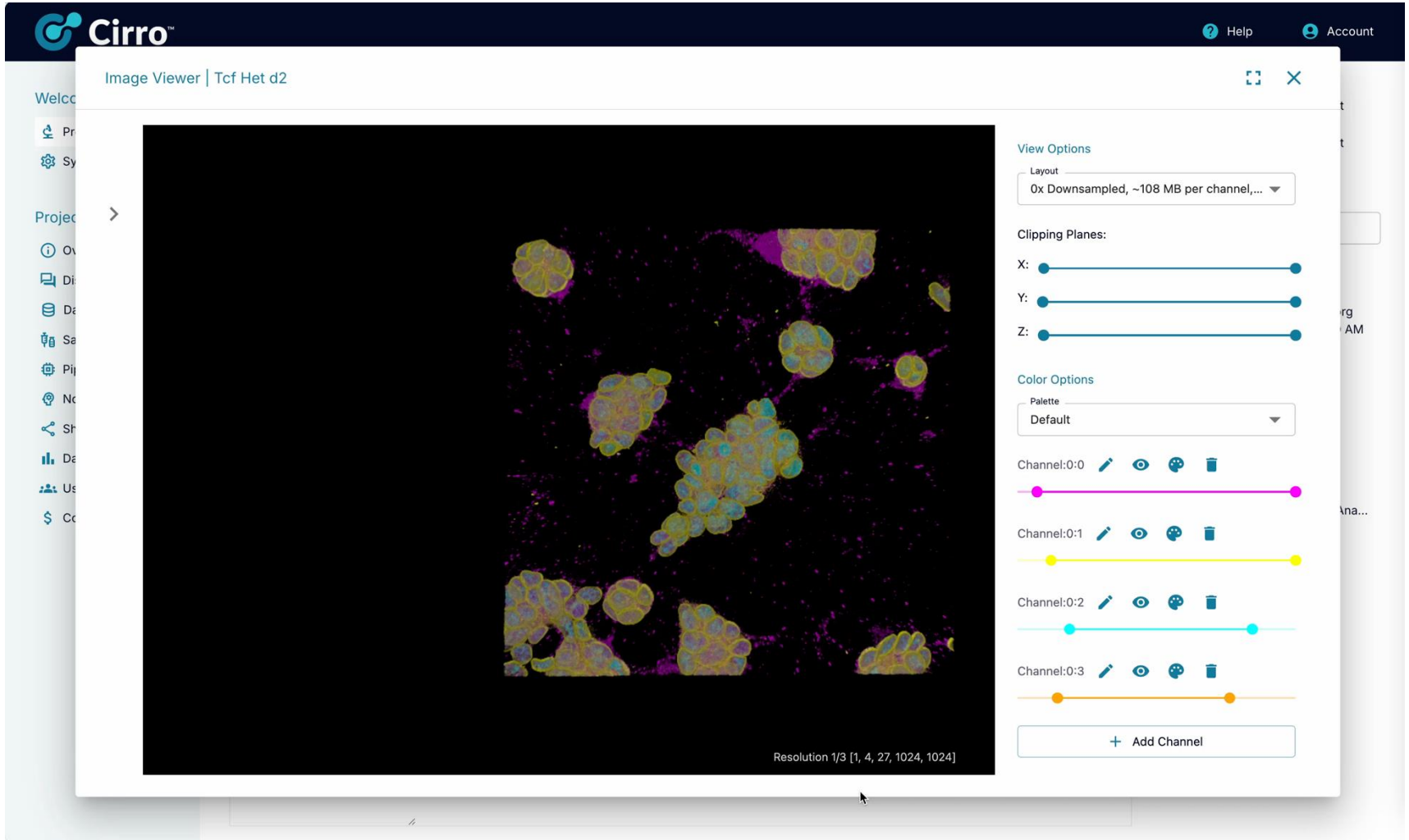




Unlocking the Value of Advanced Instrumentation

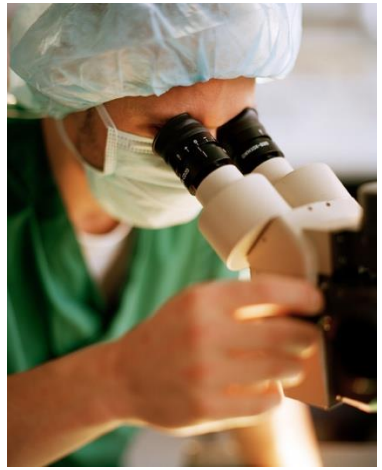
High-resolution image exploration





Integrated Research Offerings

Spatial Transcriptomics – *in situ* gene expression profiling



Biospecimen Processing and Biorepository

Provides investigators with high-quality biospecimen sample processing, DNA/RNA extraction and biorepository services.

Experimental Histopathology*

Provides a range of histology, histochemistry, immunochemistry and pathology services for human tissues and model organisms.

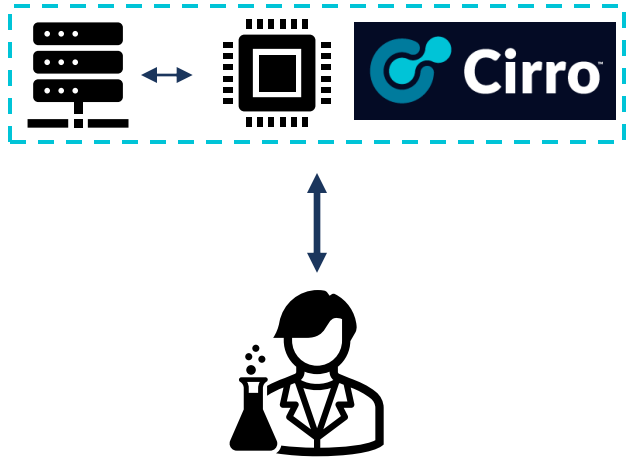
Genomics & Bioinformatics*

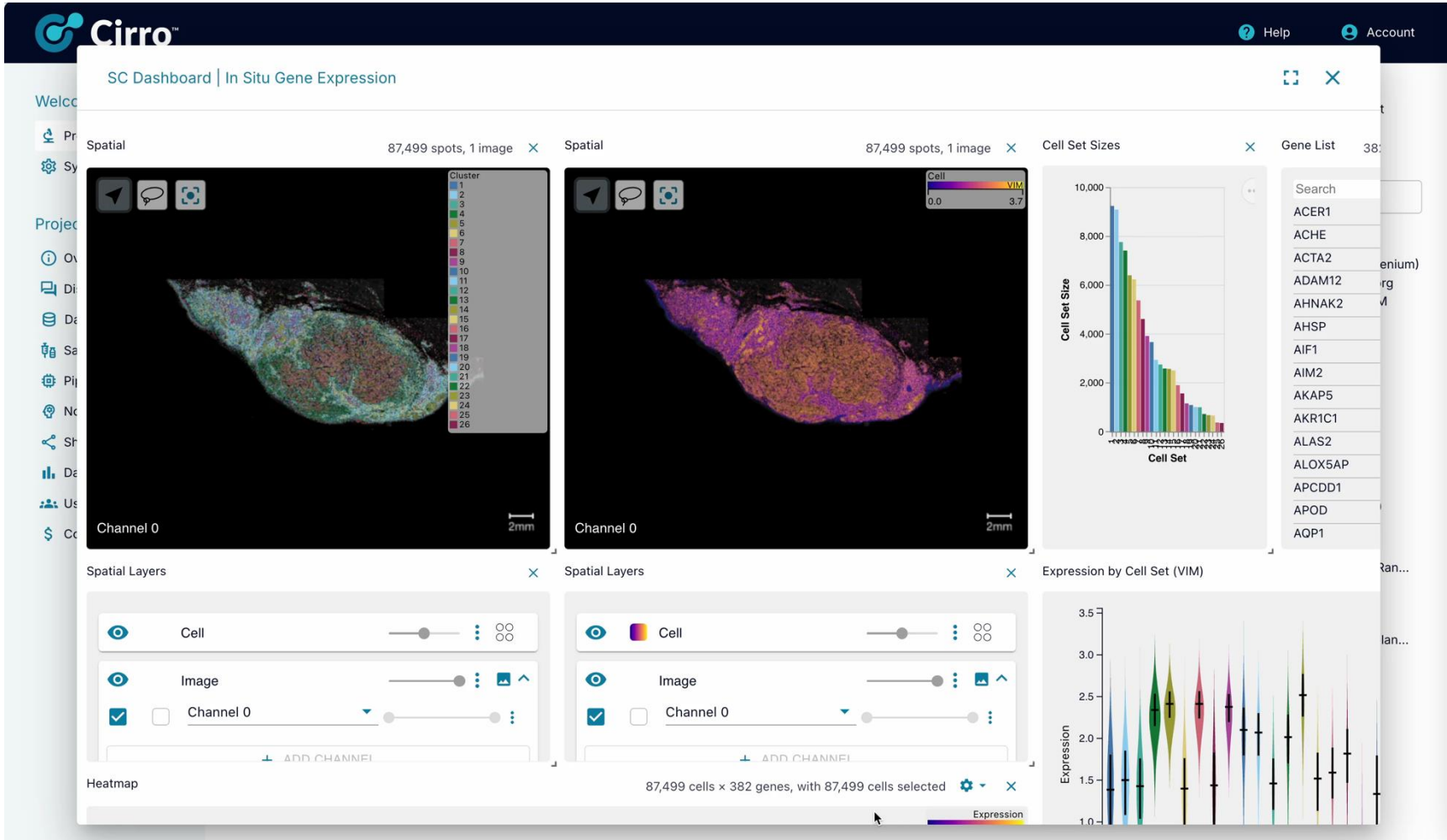
Provides integrated support to generate and analyze genomic data from DNA array, genetic analysis and high-throughput screening methods.

10X Genomics – Xenium Spatial Platform

Delivering Integrated Research Offerings

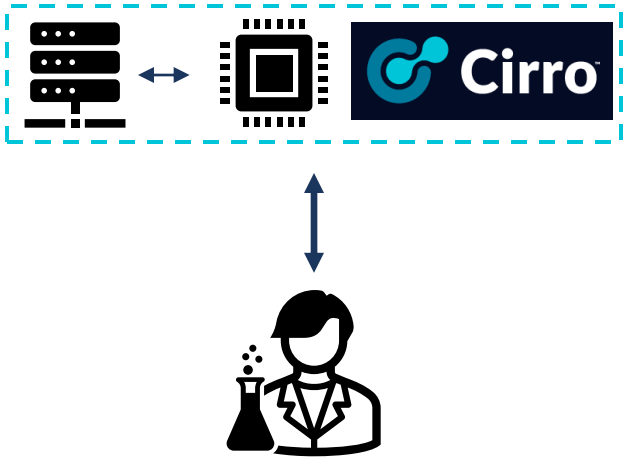
Spatial transcriptomics

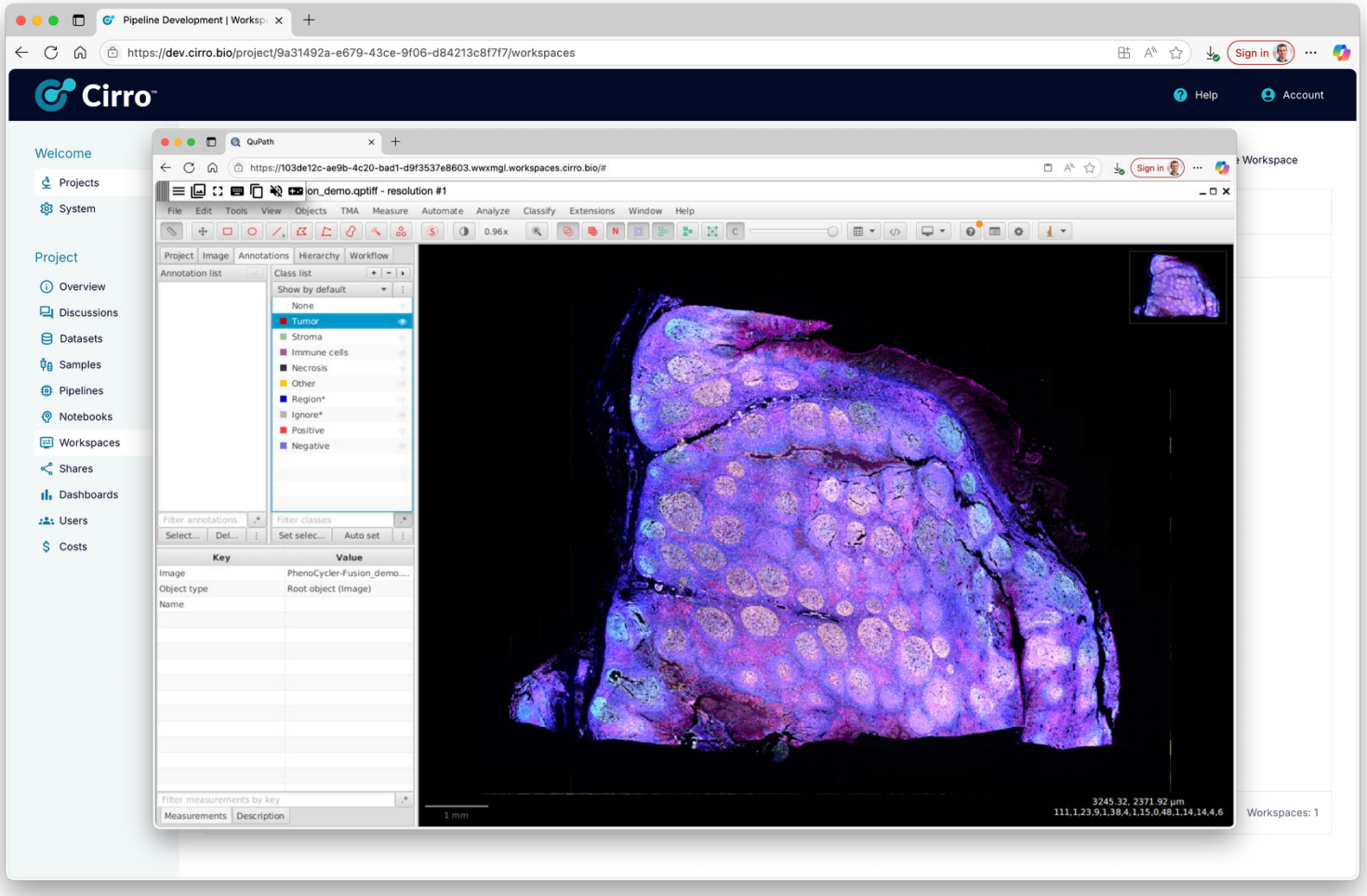




Co-locating Data and Compute

Desktop Applications in the Cloud





Lessons Learned

Philosophical

- Needs of the institution and customer exist in dialogue.
- Change must be worth its inherent cost.

Practical

- Carefully consider the input of every institutional stakeholder.
- Data systems must accommodate instrument design and operations, but not vice versa.

Scientific

- Researchers are diverse in every way imaginable, and their complete data needs can not be fully anticipated.
- Unblocking scientists yields unforeseen rewards.

Fred Hutch
Cancer Center

Thank You

Sam Minot, Ph.D.
Associate Director
Bioinformatics Core, Shared Resources
sminot@fredhutch.org