Computational Genomics and Bioinformatics Branch

Daoud Meerzaman, Ph.D., CGBB Chief

Computational Genomics & Bioinformatics Branch

Core Expertise of Our Computational Biology & Bioinformatics Group

Next-Generation Sequencing

Advanced genome profiling and variant discovery. Cutting-edge single-cell and spatial omics analysis.

Integrative Data Analysis

Multi-omics integration across platforms. Custom pipelines for complex biological datasets.

AI & Machine Learning

Deep learning for biomedical imaging and drug discovery. Biologically-inspired algorithms for predictive genomics.

CGBB Supports Three National Cancer Plan Objectives

Maximining Data Use
Bioinformatics analysis
User-friendly tool development

Develop Effective Treatments
AI pipelines to predict treatment response

Optimizing Workforce
Training, mentoring

How do we collaborate outside CBIIT?

NCI Center for Cancer Research

- Partner with CCR labs to develop tools and perform informatics analysis

NCI Cancer Moonshot

- Integrative proteogenomic analysis, APOLLO data

CIMAC Cancer Network

- NGS Pipelines development
- Cross trial analysis

Food and Drug Administration

- Collaborating on benchmarking algorithms for cancer research

Moffit cancer center

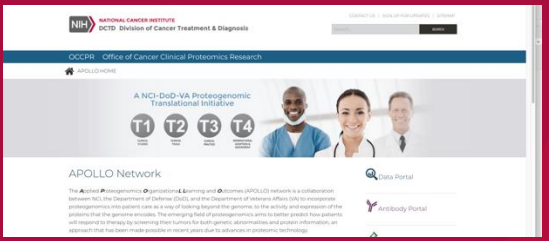
- Develop machine learning algorithm to analyze tumor subtypes for glioblastoma in a federated learning

Penn State

- Develop multimodal AI model for cancer research

The Applied Proteogenomics OrganizationaL Learning and Outcomes (APOLLO) network

Proteogenomic Integrative data analysis
APOLLO 1, 2, 4C and 5

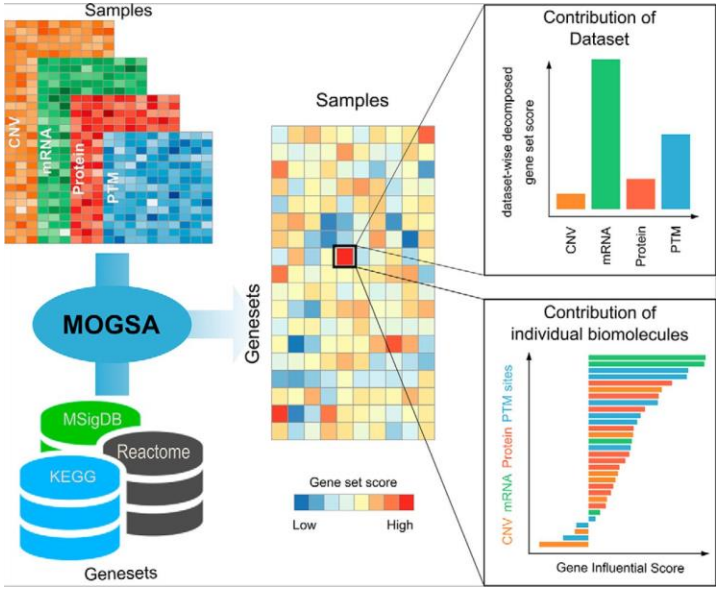


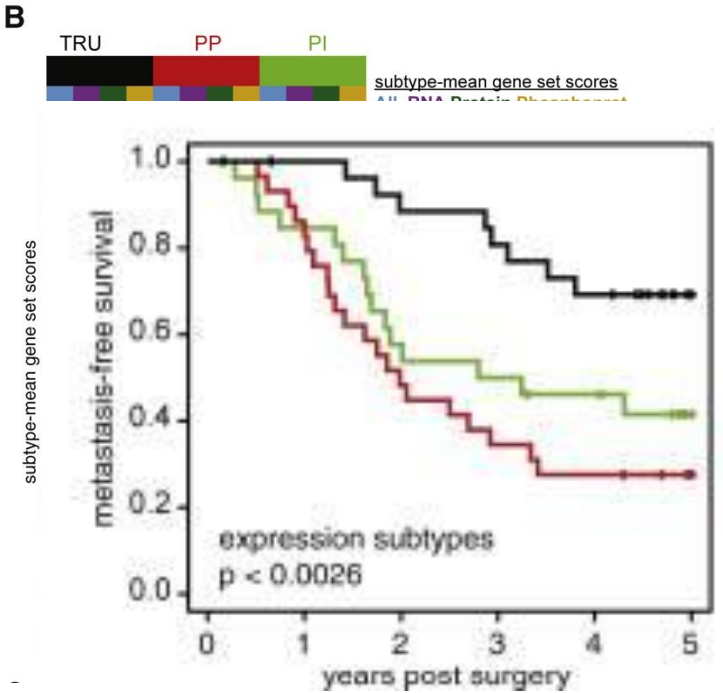
NIH NATIONAL CANCER INSTITUTE

Cell Reports Medicine

Proteogenomic analysis of lung adenocarcinoma reveals tumor heterogeneity, survival determinants, and therapeutically relevant pathways

Multi-omics gene-set analysis (MOGSA)





Chen Meng et al. Mol Cell Proteomics 2019;18:S153-S168

Soltis et al., 2022, Cell Reports Medicine 3, 100819

Bridging the gap between data scientists and biologists and clinicians

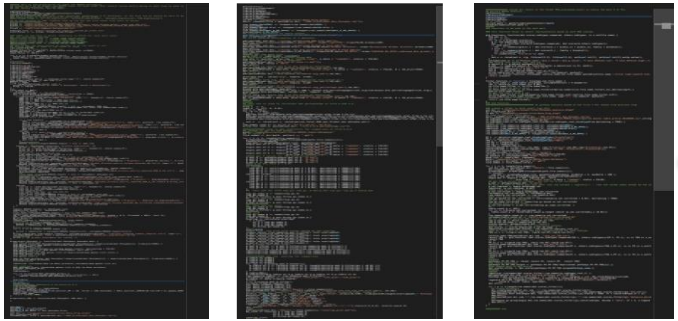
From Complex Code to Simple Click: Empowering Users with Accessible Tools using CRDC

Perform MOGSA Analysis in 3 Steps

Preparation of Inputs
A matrix of data types
At least of 2 datatypes

→ Determine the # of PCs based on MFA
To be used in Step3

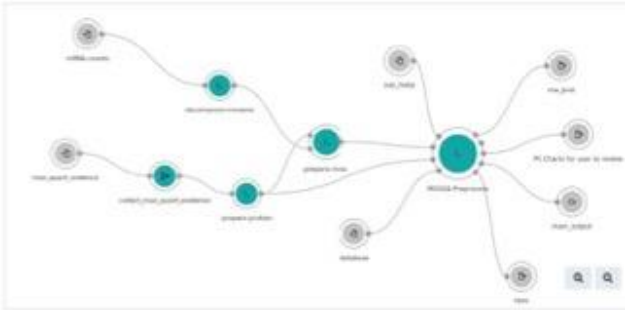
→ Perform MOGSA analysis



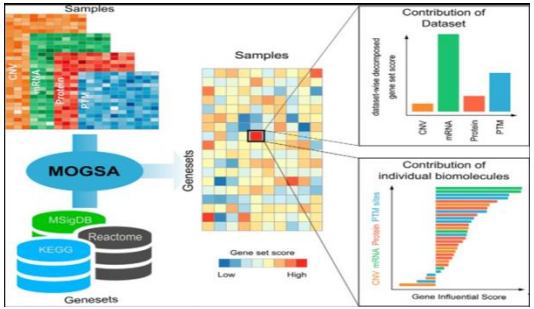
→

Make Docker image → Upload to Docker hub or SBG dock repository → Pull docker image
Create own tools → Create workflow

↓



←



How do we collaborate within CBIIT?

Data Ecosystems Branch (DEB)

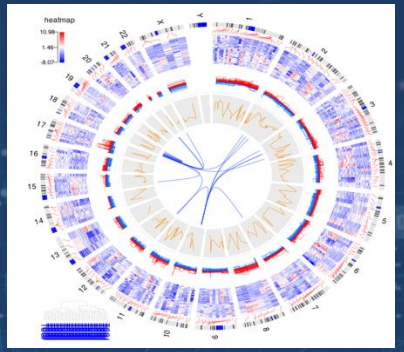
- Develop user friendly tools and pipelines for the CGC-CRDC

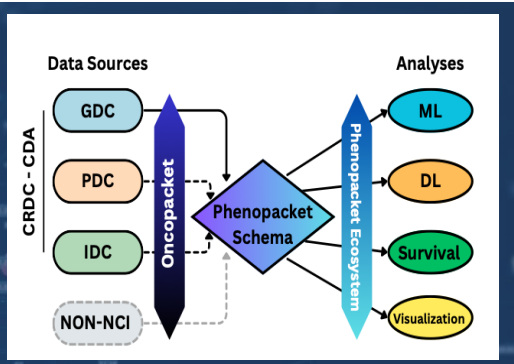
Office of Chief Information Officer (OCIO)

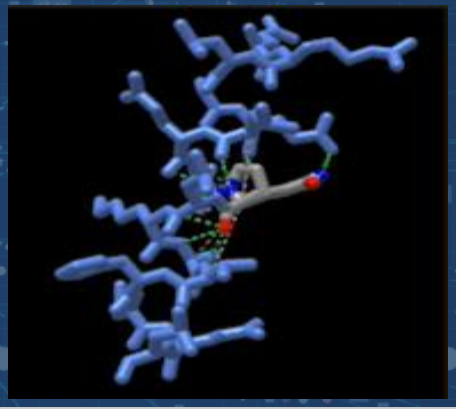
- Convert CGBB tools into bioinformatics platforms
- Scientific Software tools training

Clinical and Translational Informatics Research (CTRIB)

- Collect and sequence data samples for the CIDC
- Developed analysis tools for Federated Learning project
- Sequence and analyze data from MATCH-COMBO
- Develop pipelines that work with EVS

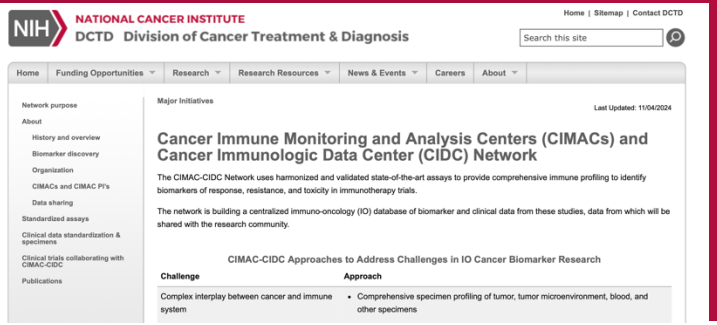






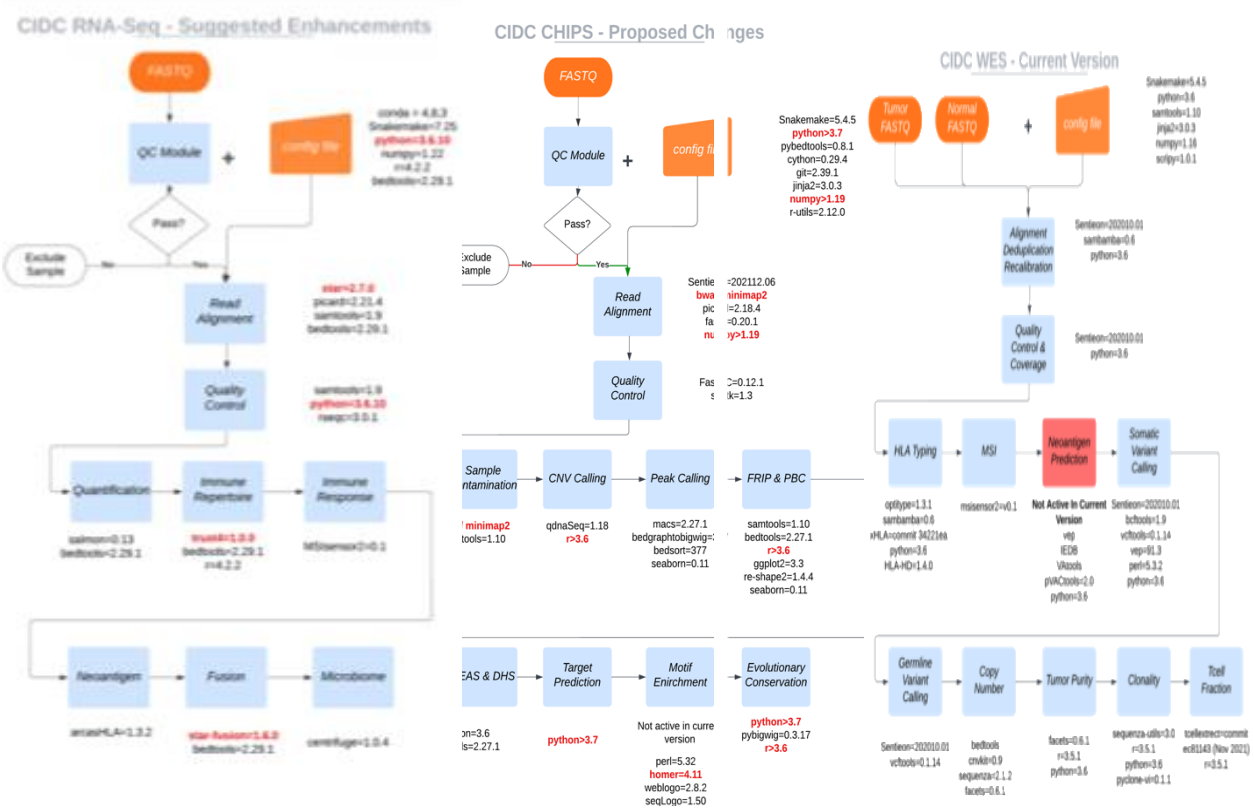
Cancer Immune Monitoring and Analysis Centers (CIMACs) and Cancer Immunologic Data Center (CIDC) Network

Benchmarking, Development and Enhancing Bioinformatic Pipelines
Single cell and CyToF analysis



NIH NATIONAL CANCER INSTITUTE

Development and Enhancing Bioinformatic Pipelines for CIMAC project



Development and Enhancing Bioinformatic Pipelines for CIMAC project

CyTOF (cyttools pipeline)

Single cell pipeline

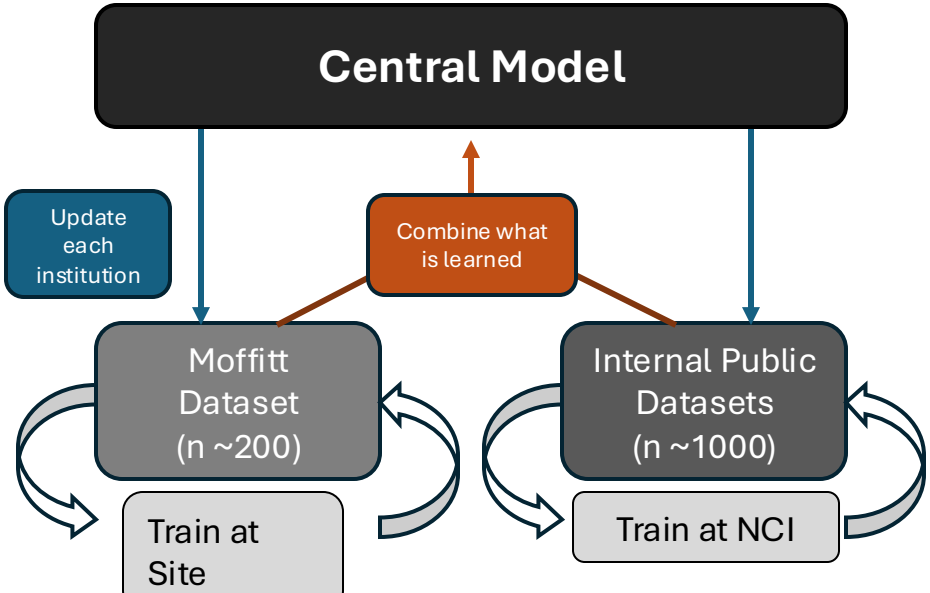
$600K

Saved through CGBB developed tool

Utilizing Federated Learning to Improve Performance and Generalizability of Predictions of Overall Survival in GBM

Purpose

- Utilize the ***Privacy Preserving Federated Learning Network Consortium*** to expand our model
- Incorporate private data from Moffitt Cancer Center



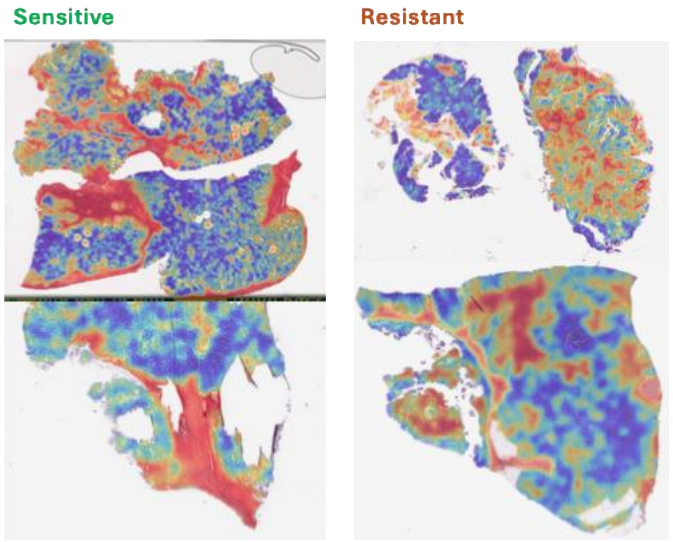
Methods

- Training is done at each institution. While training, each institution shares what it learns, but not the data itself
- ***Avoids data centralization, a challenging and time-consuming task***

DEVELOPING EFFECIVE TREATMENTS

Multimodal AI Model to predict chemotherapy response to ovarian cancer

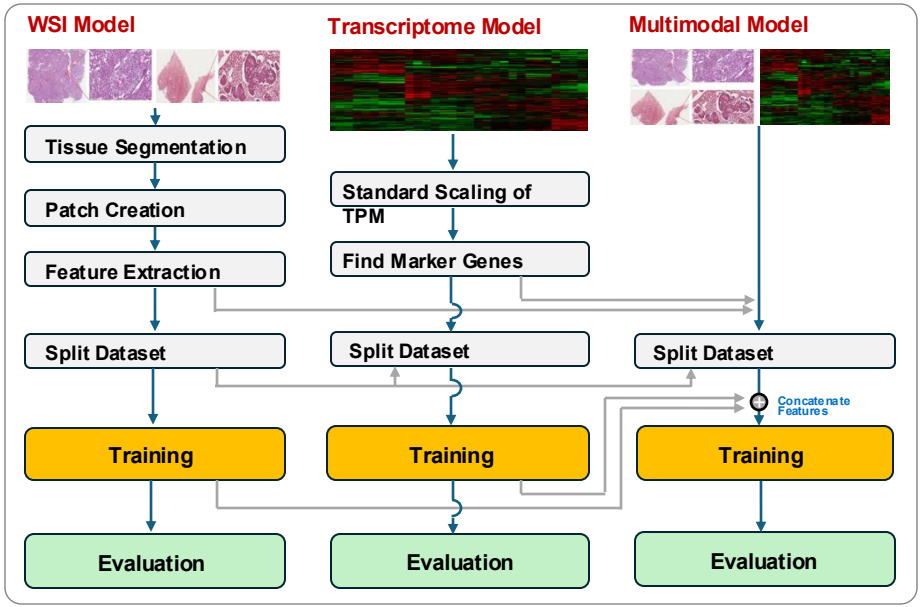
- Tissue Segmentation
- Patch Creation
- Feature Extraction
- Attention Scoring
- Model Development



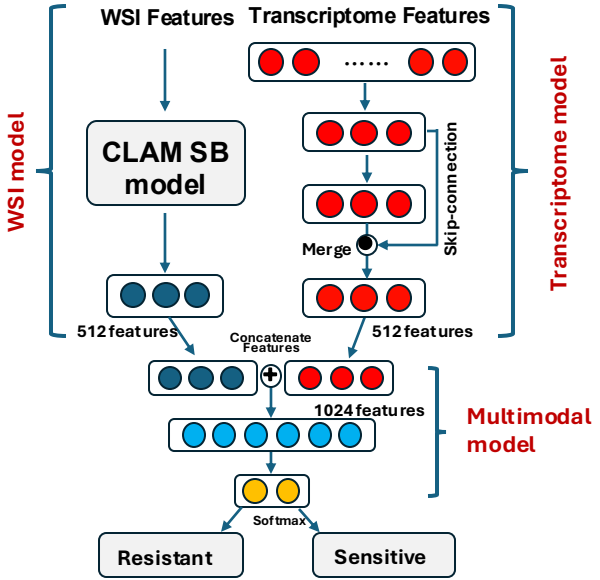
=> Higher attention area is used more importantly for training 23/2

WSIomics: An End-to-End Pipeline for Training Multimodal AI Models

Pipeline architecture



Model structure



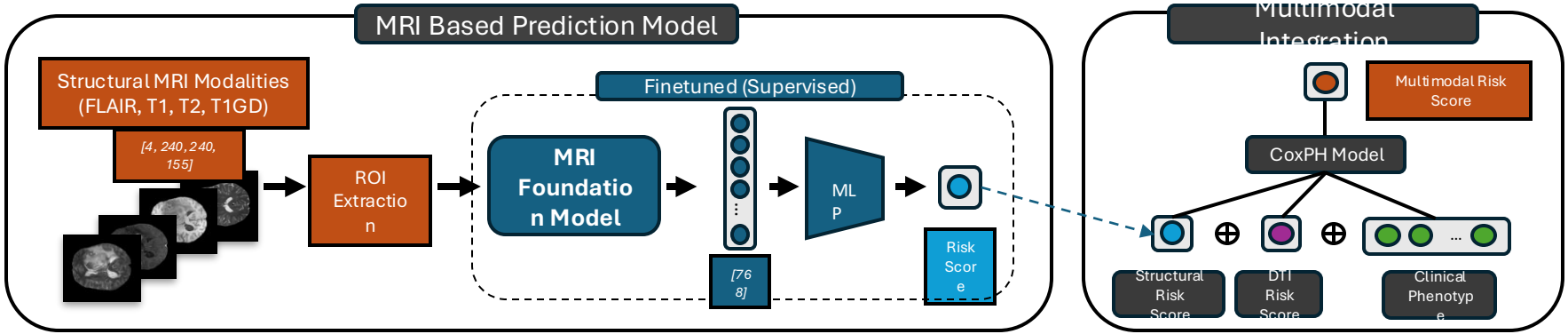
- Integrates Whole Slide Images (WSIs) and transcriptome data for multimodal AI models

- Automatically identify marker genes for transcriptome modality

Leveraging an MRI Foundation Model to Enhance Survival Predictions in Glioblastoma: A Multimodal Deep Learning Approach

Purpose

- Finetune an CNS-specific MRI foundation model for the purpose of predicting OS
- Combine imaging predictions with prognostically important clinical variables

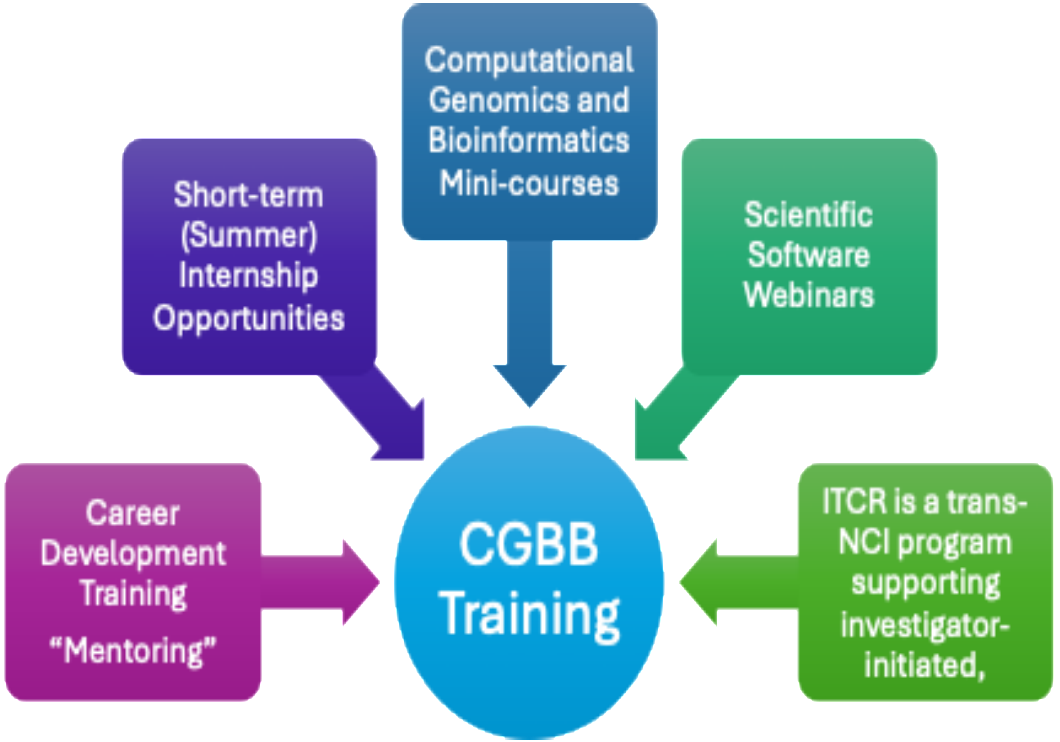


Key Result:

An MRI foundation model can improve survival prediction in GBM patients and highlights the synergy of imaging derived risk scores and clinical phenotype.

OPTIMIZING THE WORKFORCE

Empowering and Equipping NCI with Bioinformatics Skills



The CGBB Training Approaches

Registration numbers: 3,183

Attendance numbers: 2,318