Childhood Cancer Data Initiative Webinar Series

AI-Driven Multimodal Data Integration & Analysis to Improve Pediatric Cancer Diagnosis

James Amatruda, David Buckley, Jennifer Cotter, and Alexander Markowitz

NIH NATIONAL CANCER INSTITUTE

January 16, 2026

Today's Speakers from Children's Hospital Los Angeles (CHLA) and USC Norris Comprehensive Cancer Center

James Amatruda, M.D., Ph.D.

- Professor of Pediatrics and Medicine
- Chief of Hematology-Oncology
- Director, Cancer and Blood Disease Institute

Alexander Markowitz, Ph.D.

- Senior Bioinformatics Scientist

David Buckley, Ph.D.

- Clinical Bioinformatics Scientist

Jennifer Cotter, M.D.

- Associate Professor of Clinical Pathology
- Director, CHLA Pediatric Biorepository

Agenda

1. *Introduction*
2. *New CHLA datasets contributed to enhance CCDI*
   - *MethylSeq*
   - *Digital pathology*
3. *Integrated diagnostics*
   - *AI-assisted strategies*
4. *Future directions*
5. *Q&A*

Introduction

James Amatruda, M.D., Ph.D.

Pediatric Cancer Program at CHLA and University of Southern California

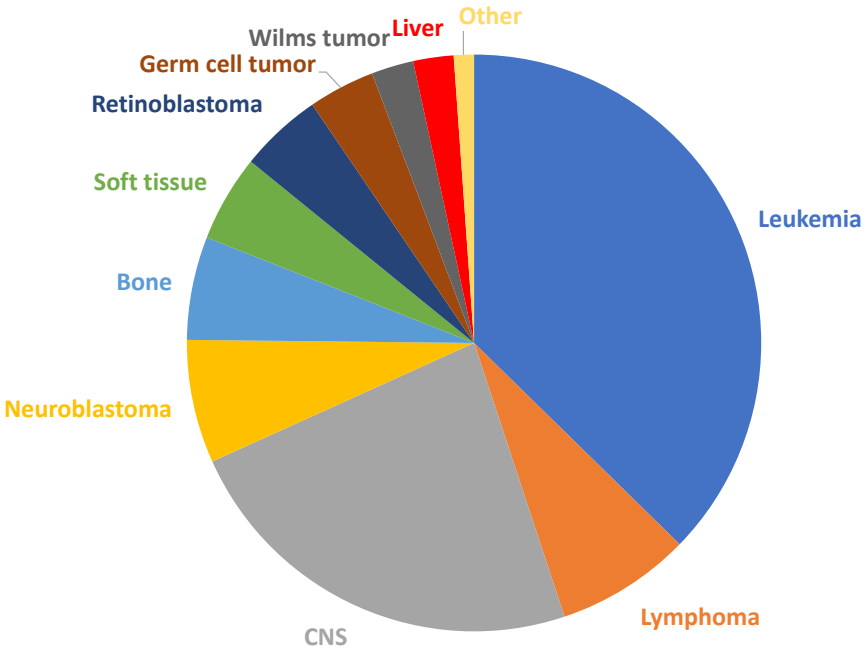
Children's Hospital LOS ANGELES
CANCER AND BLOOD DISEASE INSTITUTE

Children's Hospital LOS ANGELES
Center for Personalized Medicine

USC Norris Comprehensive Cancer Center
Keck Medicine of USC

- Largest pediatric hematology-oncology program in the Western U.S.
- 2,500 new patients and 40,000 outpatient visits in 2024
- 257 active clinical trials
- Racial and ethnic background of our patients:
  - White-Hispanic (65%)
  - Non-Hispanic White (17%)
  - Black (5%)
  - Asian (4%)
  - Pacific Islander/Native American/Other (9%)
- Two thirds of CBDI cancer patients are in the lowest 40% of socio-economic status

CHLA Cancer and Blood Disease Institute and Center for Personalized Medicine



Tumors undergo molecular characterization with CAP-CLIA certified tests:

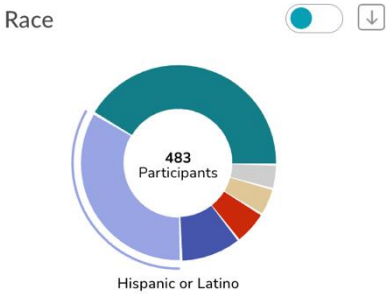
- **Chromosomal microarray**
- **OncoKids (203 cancer genes and oncogenic fusion genes)**

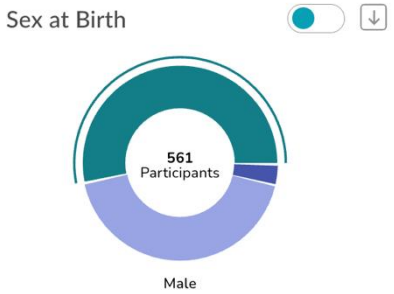
Also available:

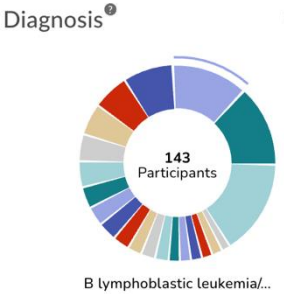
- RNASeq-Gene Fusions
- VMD4Kids (Mutations relevant to vascular malformations)
- Cancer Predisposition Panel
- LBSeq4Kids liquid biopsy copy number and targeted sequencing panel (TSP)
- Methylation Array for brain tumors

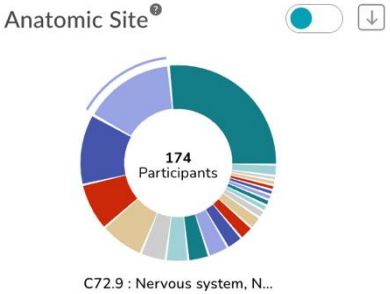
phs002518 dataset (2022)

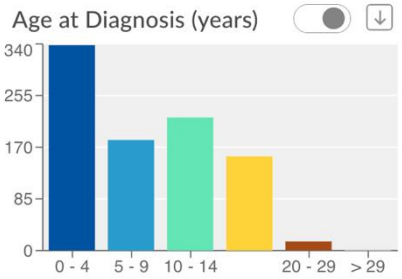
NATIONAL CANCER INSTITUTE
Childhood Cancer Data Initiative Hub

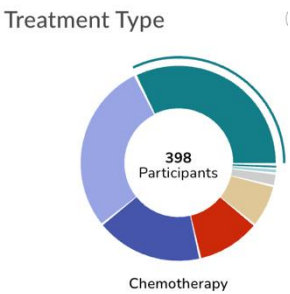












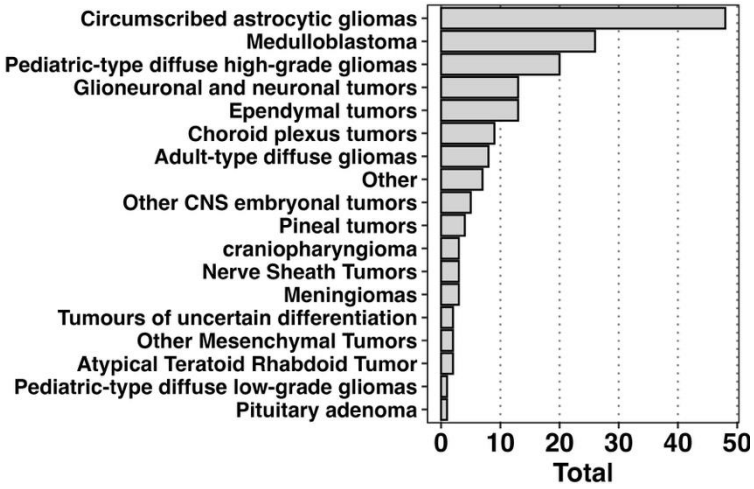
- Race/ethnicity
- Sex
- Diagnosis
- Anatomic site
- Age at diagnosis
- Treatment type
- Survival
- OncoKids results

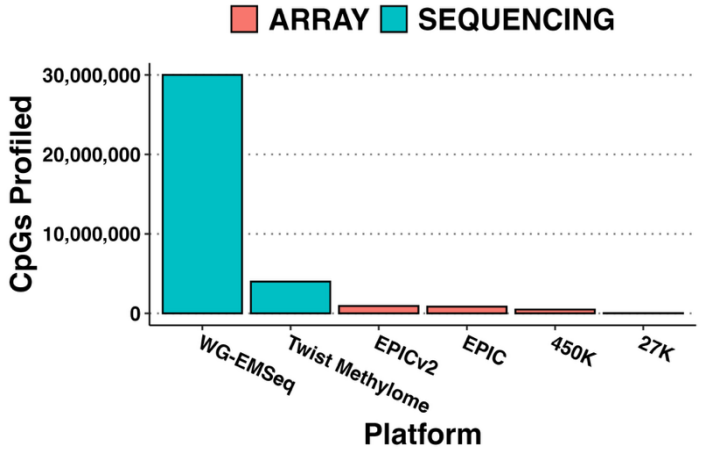
Methylation Sequencing

David Buckley, Ph.D.

MethylSeq Sample Summary

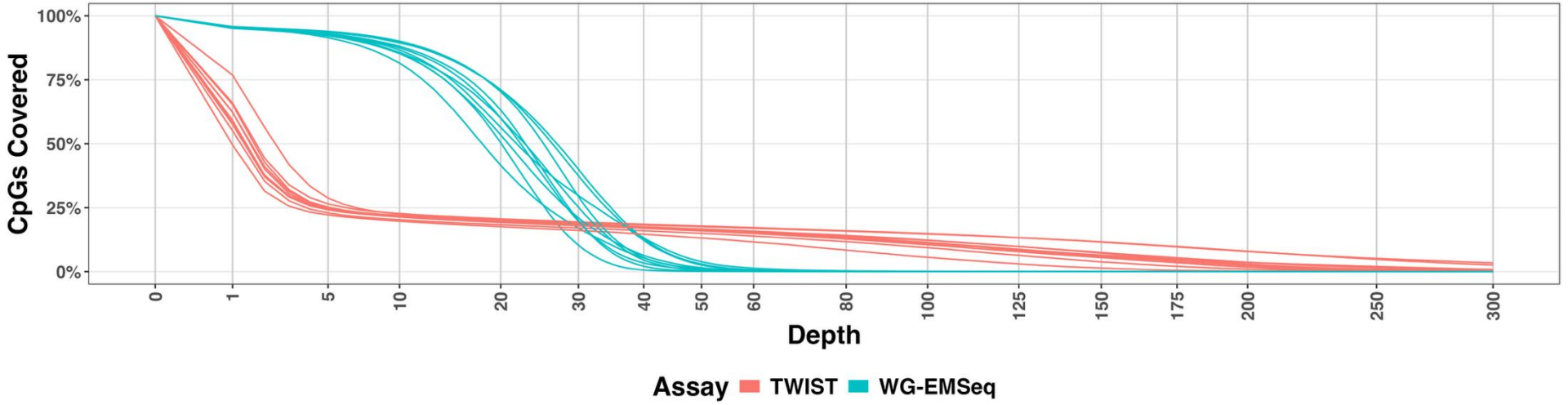
- Total patient samples: **170** (138 fresh frozen, 32 FFPE), which have matching OncoKids data deposited on CCDI – all samples sequenced by methylation profiling
- Dataset well-suited for benchmarking methylation-based CNS classification models
- Sequencing-based methylation profiling captures substantially more CpG loci than array platforms; EPIC v2 captures only ~3% of CpGs, ignoring the vast majority of the methylome
- Sequencing-based methylation profiling is now cost-competitive (and falling fast)
- No batching constraints – each sample can run independently
- Whole-genome (WG)-EMSeq enables high-resolution copy-number analysis from the same data





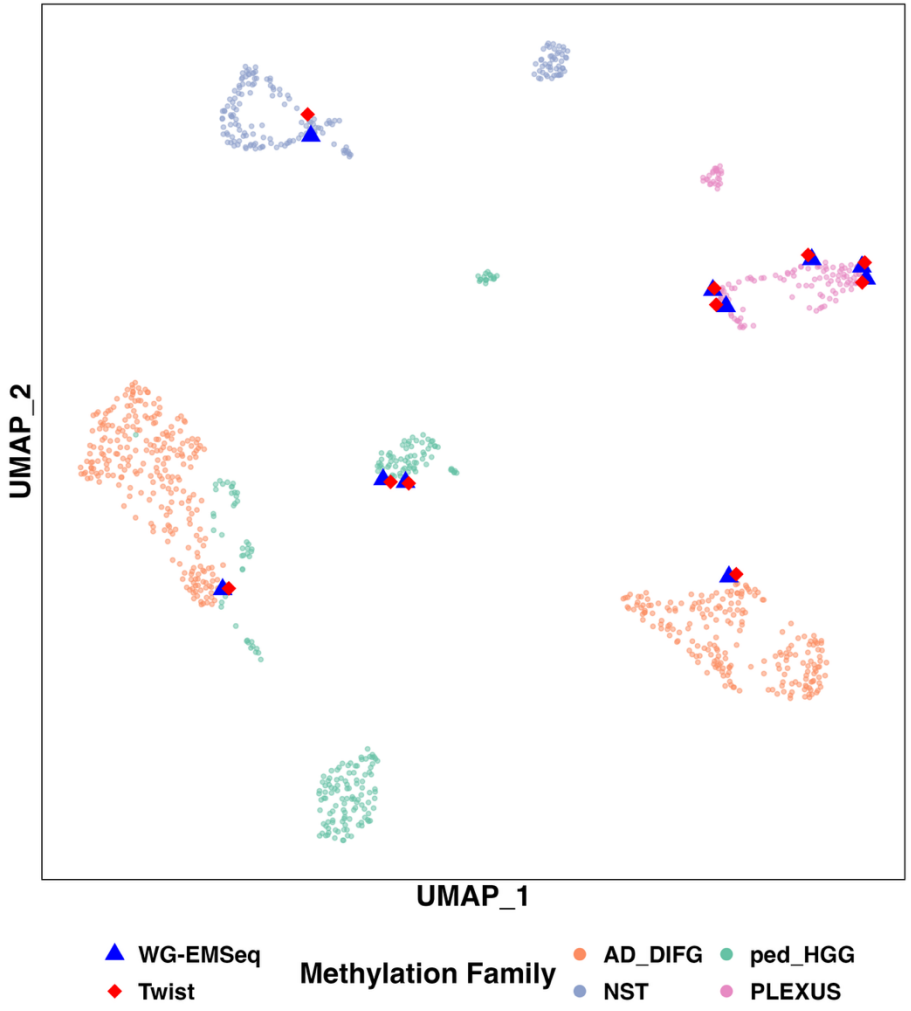
Twist vs. WG-EMSeq Sequencing Depth

- Pilot study of 10 paired (Twist & WG-EMSeq) samples processed and sequenced per manufacturer-specified protocol
- Total CpGs covered higher in WG-EMSeq compared to Twist; coverage in Twist target regions higher than WG-EMSeq; these results confirm that the two approaches trade breadth for depth, which is important context when interpreting downstream classification performance

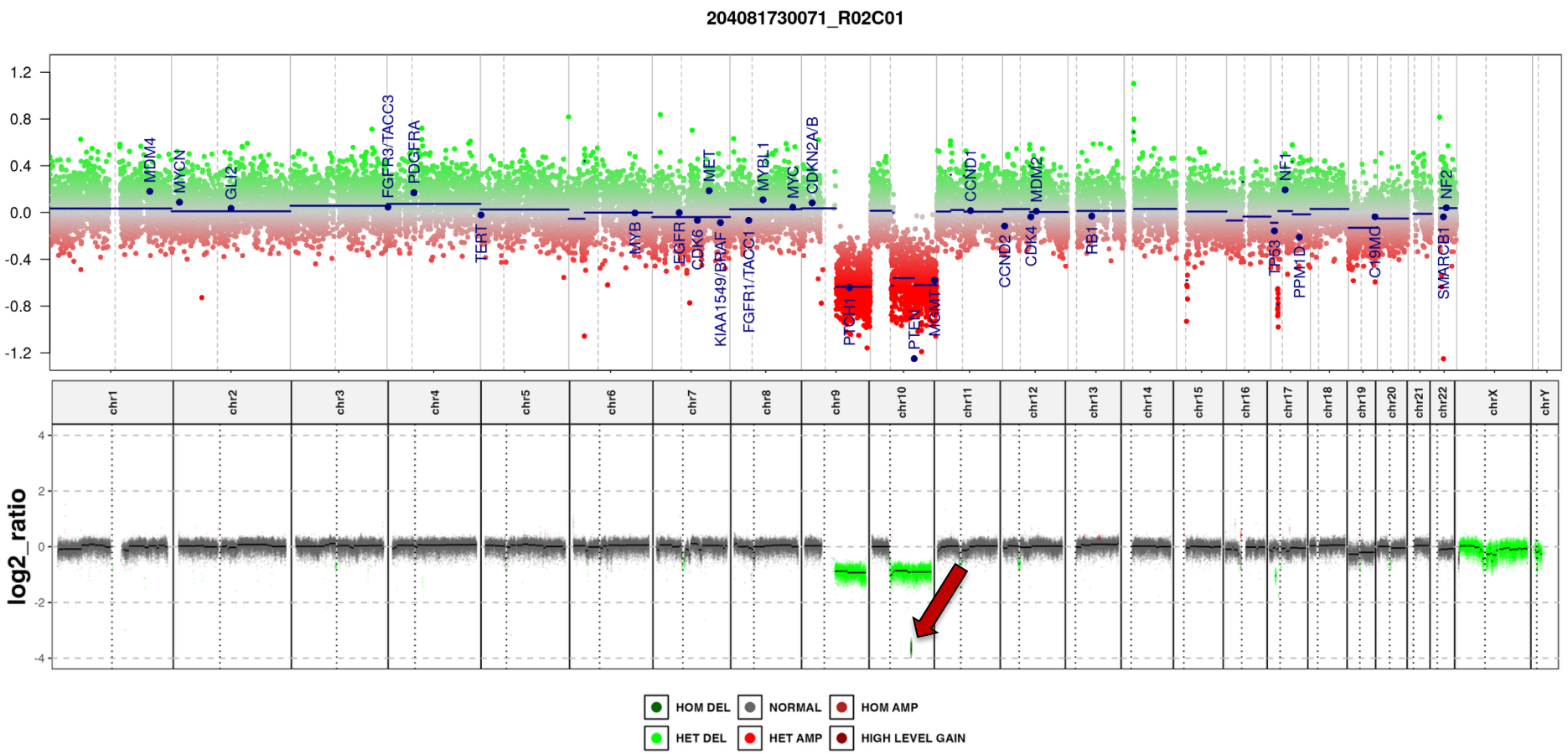


Methylation Profile Clustering (WG-EMSeq & Twist)

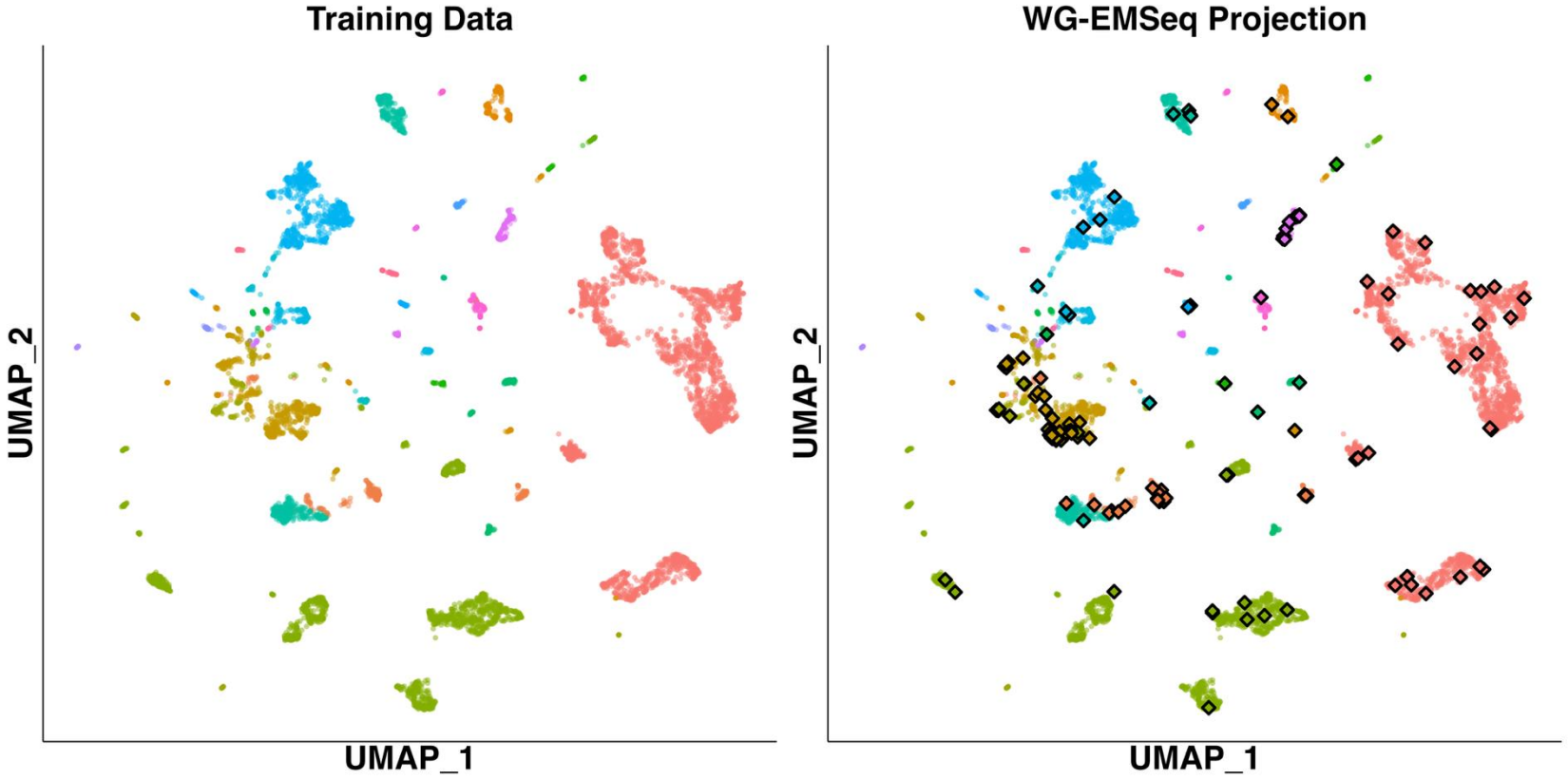
- Beta values used to project samples on UMAP generated from publicly available array methylation profiles of corresponding tumor types (high-grade gliomas, choroid plexus tumors, and nerve sheath tumors)
- Twist and WG-EMSeq samples clustered according to **expected tumor type**, regardless of profiling platform, indicating that both sequencing approaches preserve **diagnostically relevant methylation signal**.



Array & WG-EMSeq Genome-wide Copy Number Profiles

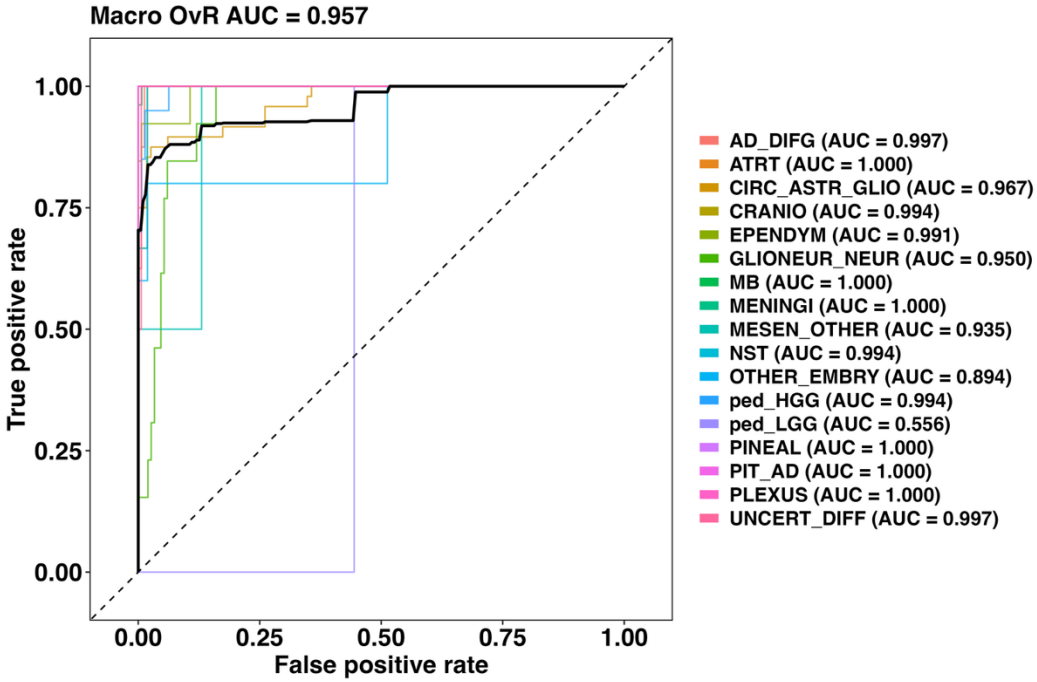


Methylation Profile Clustering WG-EMSeq



WG-EMSeq NNet Classification

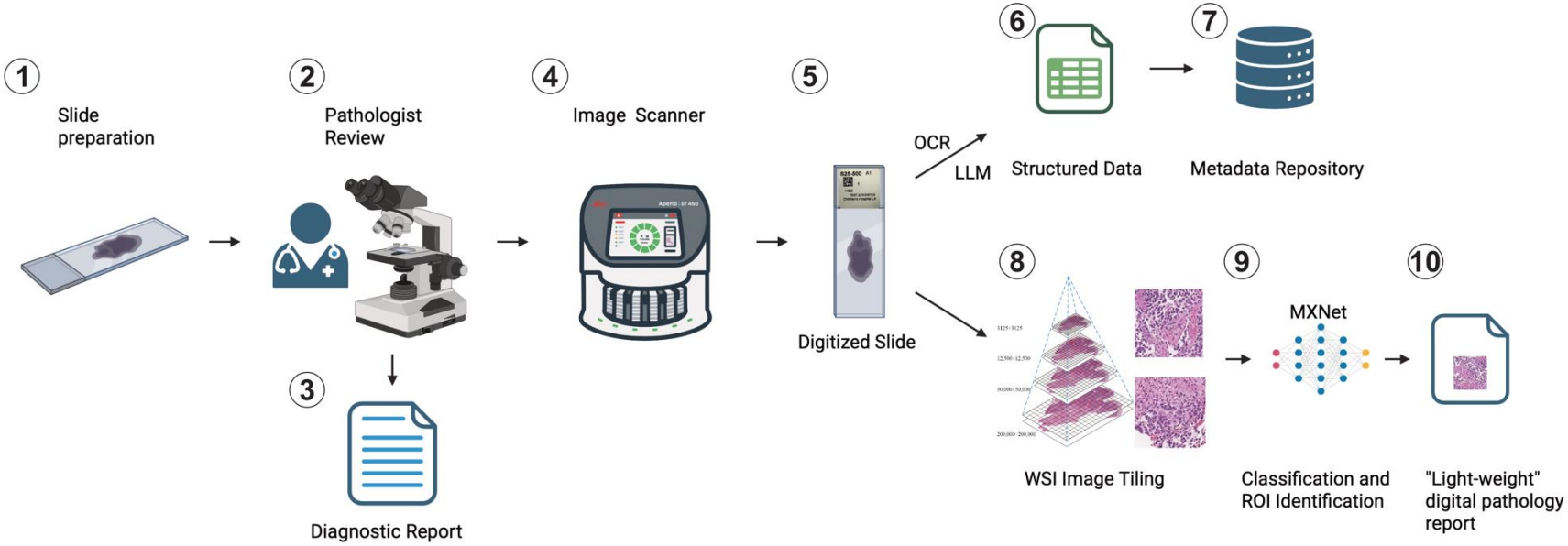
- 163 WG-EMSeq samples used for concordance analysis
- One-versus-rest (OvR) AUC = 0.957
- 144/163 (88%) ‘matched’ to a methylation family (score above threshold = 0.75)
- Sensitivity = 0.90, specificity = 0.99 in matched samples at the family level
- 71% of samples achieved a class score above threshold; sensitivity = 0.922, specificity = 0.985



Digital Pathology and Integrated Diagnostics

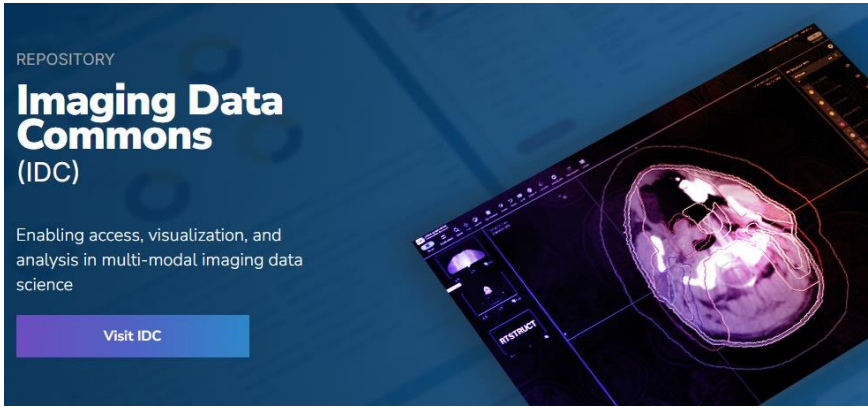
Jennifer Cotter, M.D.

Digital Pathology: Whole Slide Imaging



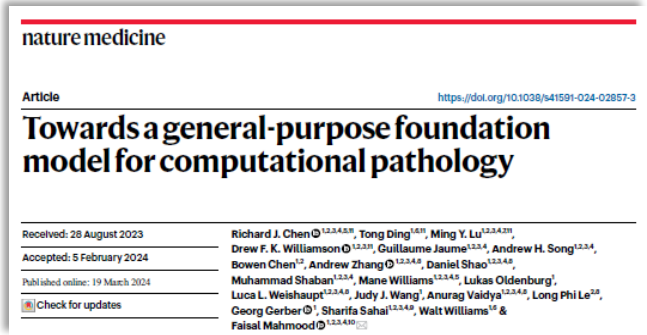
WSI Dataset partnered to OncoKids Dataset (phs002518)

- >700 cancer cases from CHLA screened and digitized
- Key slide for each case was selected by pathologist review to be contributed to the NCI Imaging Data Commons
- H&E whole slide image for any available case linked to OncoKids and WG-EMSeq data via common identifiers

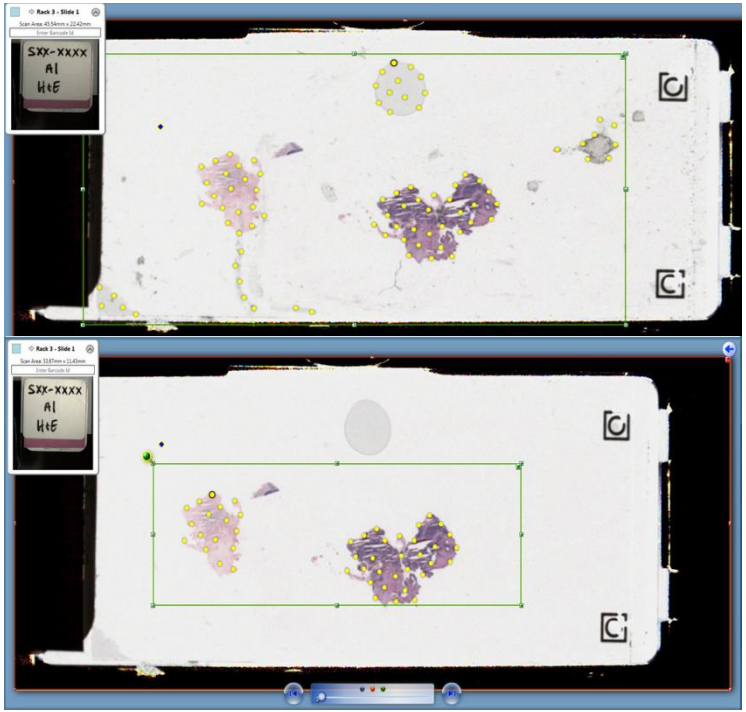


Adoption of Digital Pathology Has Been Slow

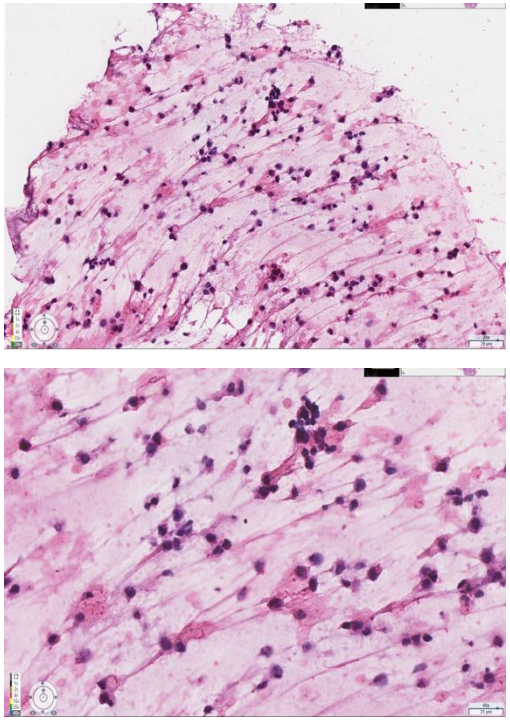
- Many institutions have had the capacity to scan slides into WSI for years, but fewer have fully transitioned to digital workflows for operational reasons
    - Staffing
    - Accurate data entry/labeling
    - Data storage costs (one WSI can be >1 GB)
- Automation of scanning, quality checking, and data organization will drive more labs to digital format in the next few years
- Delay to digital transition in pathology has limited progress in development of machine learning/AI tools for WSI, but early work is promising



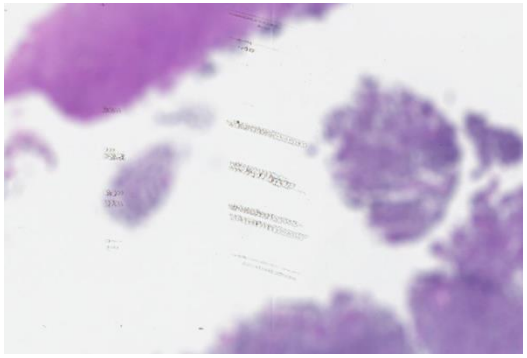
Slide Quality Importance



(top): Poorly prepared H&E scan
(bottom): Properly prepared H&E scan



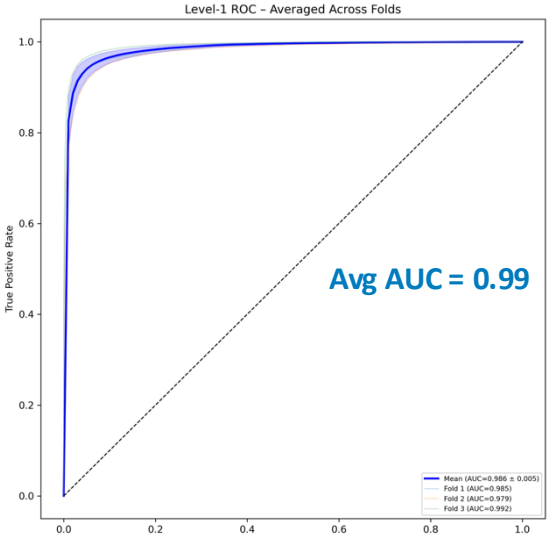
In focus images; 20x vs. 40x magnification



Out of focus (scratched plastic coverslip)

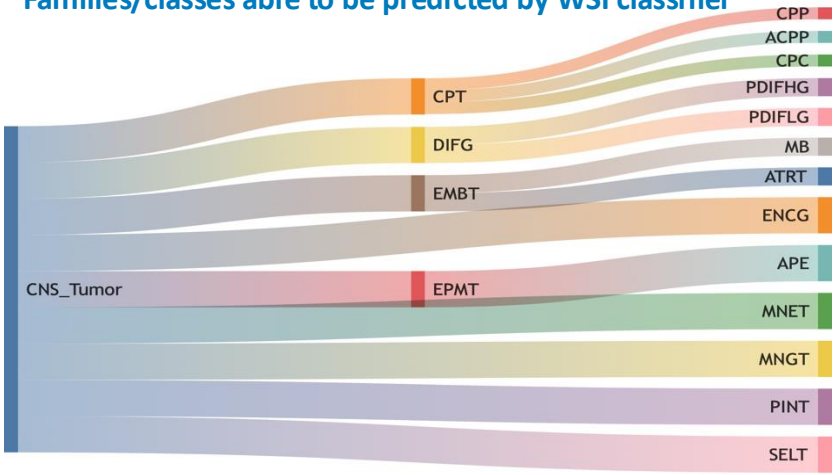
Whole Slide Image Classification

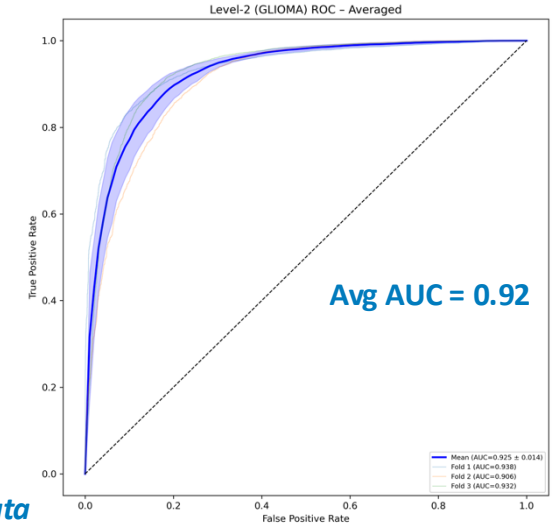
- Tile-based computer vision approach
- Levels of predictions:
  - OncoTree family (e.g., CPT, EMBRY, DIFG)
  - Tumor subclass (e.g., PDIFHG, PDIFLG)



Prediction of OncoTree family

Families/classes able to be predicted by WSI classifier





PDIFHG vs. PDIFLG (diffuse high grade vs. diffuse low grade)

Unpublished data

Integrated Diagnostic Approach

- Morphology alone may not be sufficient to classify a tumor, but WSI contain a large and complex set of visual data
- Molecular profile drives
  - Classification
  - Prognostication
  - Treatment options
- Assembling study cohorts benefits from more comprehensive case characterization
- Adaptability to future classification changes is key

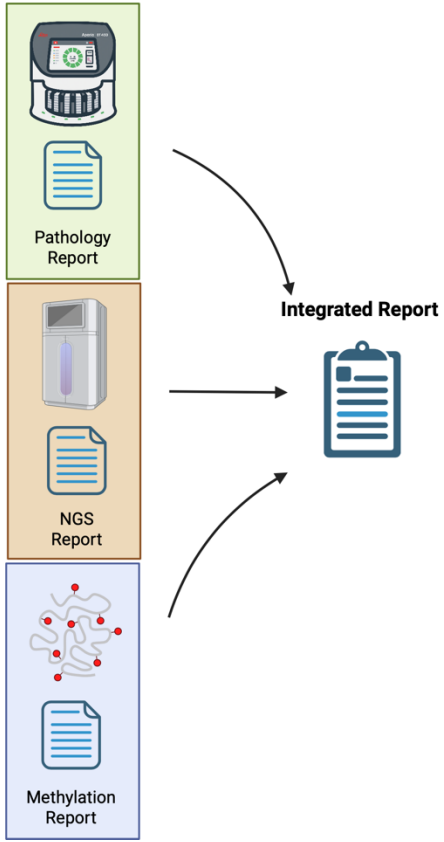
WHO 2007 (4th ed.)	WHO 2016 (4th ed., revised)	WHO 2021 (5th ed.)
Glioblastoma	Glioblastoma, IDH-mutant	Astrocytoma, IDH-mutant
Glioblastoma	Diffuse midline glioma, H3 K27M-mutant	Diffuse midline glioma, H3 K27M-altered
Glioblastoma	Glioblastoma, IDH-wildtype	Diffuse hemispheric glioma, H3 G34-mutant
Glioblastoma	Glioblastoma, IDH-wildtype	Infant-type hemispheric glioma

AI-driven Integrated Diagnostics

Alexander Markowitz, Ph.D.

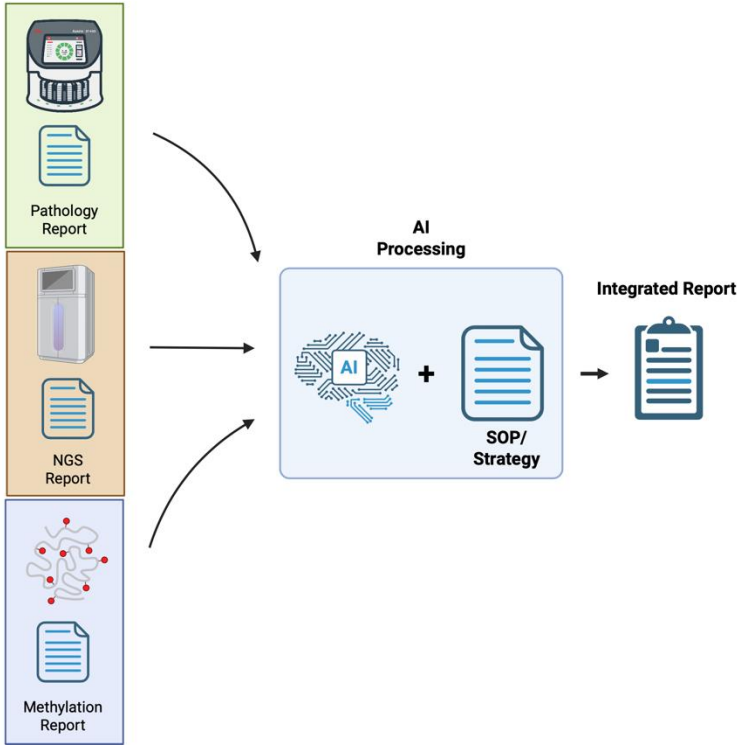
Current Challenges of Integrated Reporting

- Labor intensive process requiring coordination of clinical and bioinformatics teams
- Classifications can change over time as new insights are generated and new subtypes are discovered
- Retrospective classifications is performed manual ad-hoc basis
- Data silos inhibit automation of integrating new results when reviewing past cases



Vision and Motivation for AI-driven Integrated Diagnostics

- Develop AI solutions that provide trustworthy outputs when performing integrated diagnostics and classification tasks
- Motivated by:
  1. Promising ML/AI tools:
     - Digital Pathology
     - Variant Classification
     - Methylation Classification
  2. Clinical multi-modal datasets (CCDI)



Demo Case: Using AI-integrated Tools to Update a Classification

Original

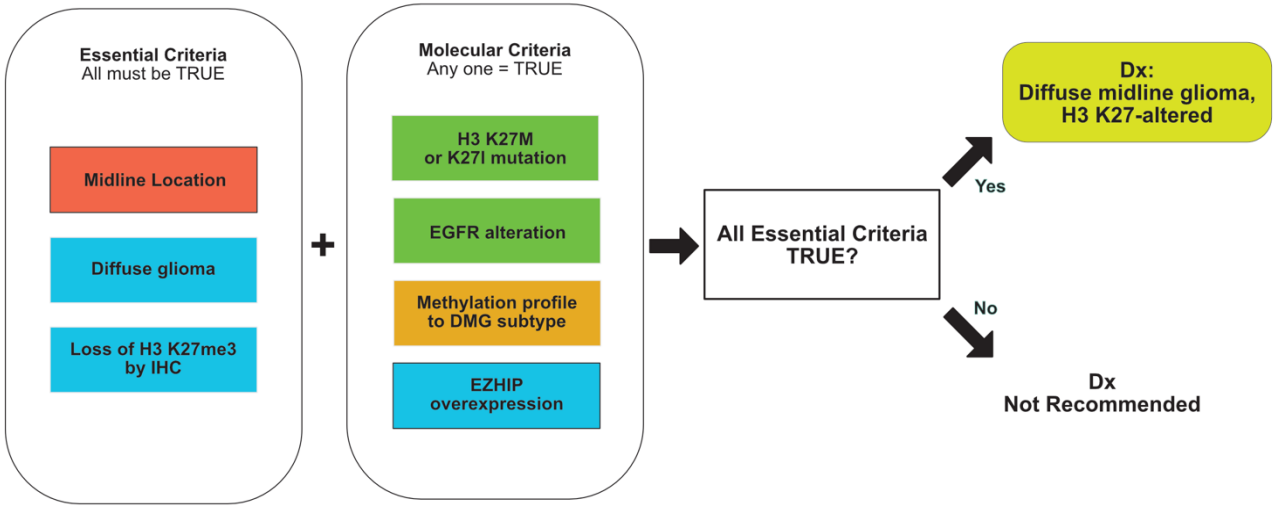
- CCDI Participant ID: R96341124 (phs002518)
- Descriptive Classification:
  - ICD; Glioma, malignant
- Data Sources:
  - Pathology report

Update

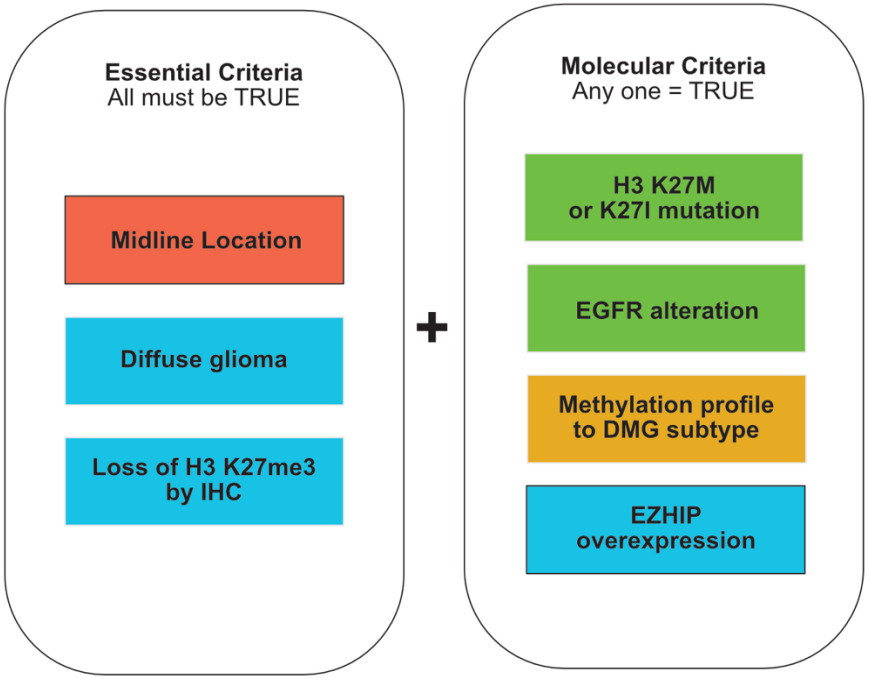
- CCDI Participant ID: R96341124 (phs002518)
- AI-driven Classification:
  - **WHO CNS5; Diffuse Midline Glioma, H3 K27-altered**
- Data Sources:
  - OncoKids Cancer Panel (BAM)
  - **Digital Pathology WSI (new*)**

WHO Classification Logic Schema

Precise tumor classification involves the accumulation of diagnostic test results. Classification systems, such as the WHO CNS5, provide logical schemas to help deduce a classification.



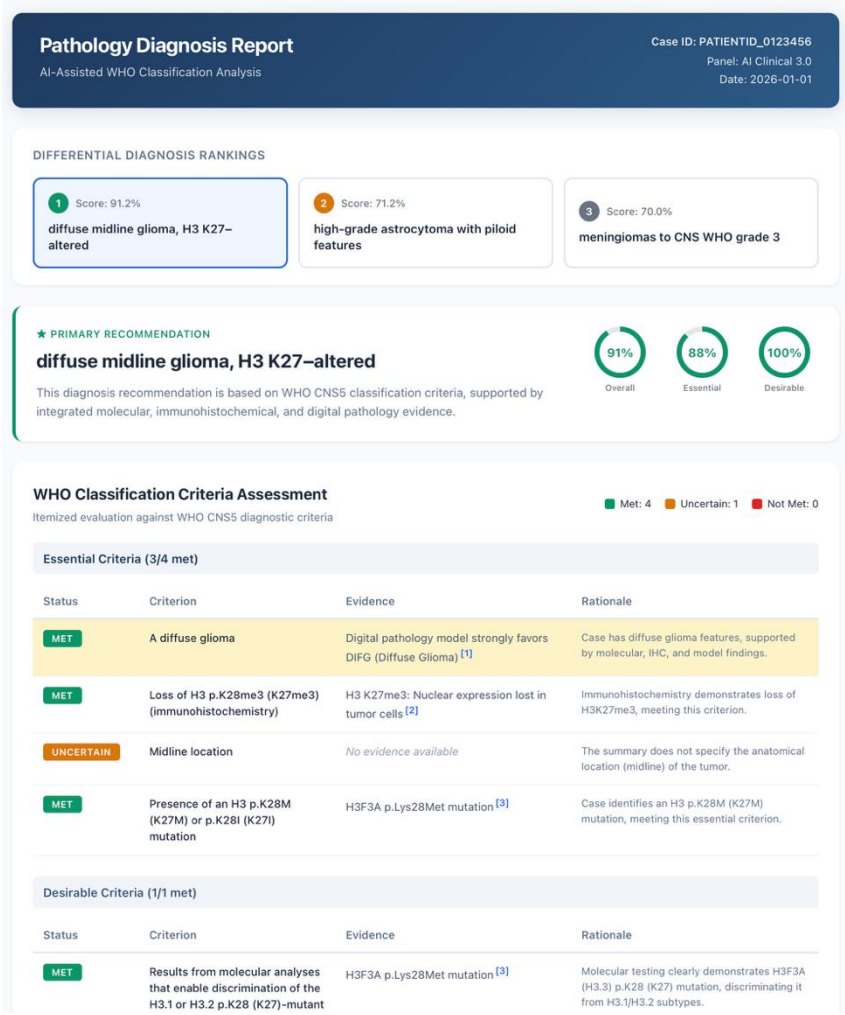
Combination of Custom and Commercial Classification Tools Can Be Used to Evaluate Criteria



- Radiology/Imaging:
  - Location identification
- Digital Pathology:
  - Tumor Classification
  - IHC Scoring
- DNA Variant Classification:
  - Golden Helix Varseq
- Methylation Classification:
  - EMSeq, Methyl Array

AI-driven integrated reporting

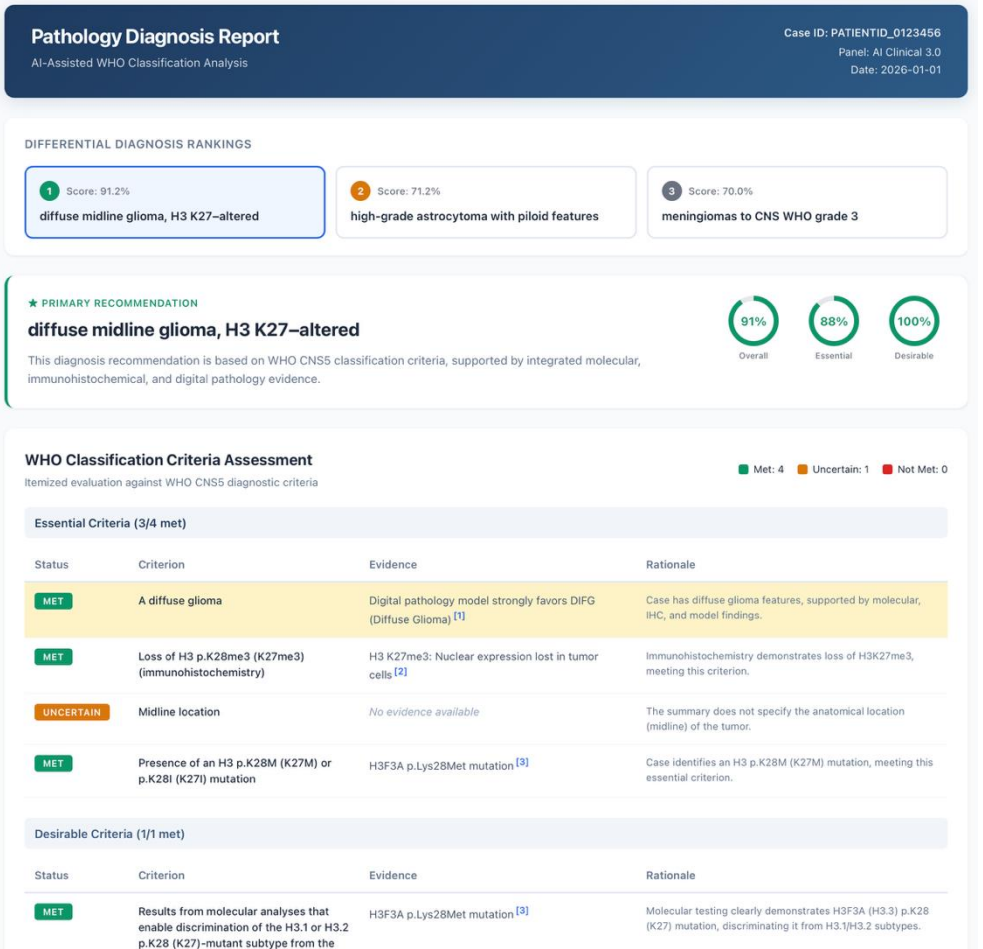
- Report provides a recommended diagnostic classification:
  - Input WSI Images
  - OncoKids Cancer Panel
- Itemized list of whether the sample met each criterion of the diagnosis.
  - When data is not provided, the status is listed as uncertain
- Scoring of diagnosis is weighted among the essential and desirable criteria



AI-driven Integrated Report Presents Evidence and Rationale For Its Classification

Integrated report provides transparent access to:

1. Narrative diagnostic summary
2. ML predictions and scores
3. Underlying raw image and genomic data (via footnoted citations)



Pathology Diagnosis Report
AI-Assisted WHO Classification Analysis

Case ID: PATIENTID_0123456
Panel: AI Clinical 3.0
Date: 2026-01-01

DIFFERENTIAL DIAGNOSIS RANKINGS

1 Score: 91.2%
diffuse midline glioma, H3 K27–altered

2 Score: 71.2%
high-grade astrocytoma with piloid features

3 Score: 70.0%
meningiomas to CNS WHO grade 3

★ PRIMARY RECOMMENDATION

diffuse midline glioma, H3 K27–altered

This diagnosis recommendation is based on WHO CNS5 classification criteria, supported by integrated molecular, immunohistochemical, and digital pathology evidence.

91% Overall
88% Essential
100% Desirable

WHO Classification Criteria Assessment

Itemized evaluation against WHO CNS5 diagnostic criteria

Met: 4 Uncertain: 1 Not Met: 0

Essential Criteria (3/4 met)

Status	Criterion	Evidence	Rationale
MET	A diffuse glioma	Digital pathology model strongly favors DIFG (Diffuse Glioma) [1]	Case has diffuse glioma features, supported by molecular, IHC, and model findings.
MET	Loss of H3 p.K28me3 (K27me3) (immunohistochemistry)	H3 K27me3: Nuclear expression lost in tumor cells [2]	Immunohistochemistry demonstrates loss of H3K27me3, meeting this criterion.
UNCERTAIN	Midline location	*No evidence available*	The summary does not specify the anatomical location (midline) of the tumor.
MET	Presence of an H3 p.K28M (K27M) or p.K28I (K27I) mutation	H3F3A p.Lys28Met mutation [3]	Case identifies an H3 p.K28M (K27M) mutation, meeting this essential criterion.

Desirable Criteria (1/1 met)

Status	Criterion	Evidence	Rationale
MET	Results from molecular analyses that enable discrimination of the H3.1 or H3.2 p.K28 (K27)-mutant subtype from the	H3F3A p.Lys28Met mutation [3]	Molecular testing clearly demonstrates H3F3A (H3.3) p.K28 (K27) mutation, discriminating it from H3.1/H3.2 subtypes.

Benefits of AI-assisted Integrated Classification

- Improved accuracy, consistency, and transparency in case classifications
- This approach:
  - Detects misclassified or edge-cases
  - Provides quantitative confidence to support decisions
  - Enables transparent review and future reinterpretation

Acknowledgements

USC Norris Comprehensive Cancer Center
Caryn Lerman, Ph.D., Director

CHLA CCDI Team
Fariba Navid, M.D.
Bruce Pawel, M.D.
Shengmei Zhou, M.D.
William Mango
Anya Zdanowicz

Grant Support
P30CA014089 Supplement

Q&A

How You Can Engage with CCDI

Learn about CCDI and subscribe to our monthly newsletter:
cancer.gov/CCDI

Access CCDI data and resources:
ccdi.cancer.gov

Questions? Email us at:
NCIChildhoodCancerDataInitiative@mail.nih.gov

Thank you for attending!

cancer.gov

cancer.gov/espanol