Childhood Cancer Data Initiative Virtual Symposium Series

Catherine Cottrell and Michael Rusch

NIH NATIONAL CANCER INSTITUTE

April 2, 2026

Today's Speakers

Catherine Cottrell, Ph.D., FACMG

- Section Chief of the Institute for Genomic Medicine Clinical Laboratory at Nationwide Children's Hospital
- Professor-Clinical in the Departments of Pathology and Pediatrics at the Ohio State University College of Medicine

Michael Rusch

- Director of Bioinformatics Software Development at St. Jude Children's Research Hospital

Agenda

1. *Developing a Clinical Methylation Classifier for Pediatric Brain Tumors and Sarcomas*
   - *Q&A with Dr. Catherine Cottrell*
2. *CCDI Pediatric Cancer Diagnosis Ontology*
   - *Q&A with Michael Rusch*

Childhood Cancer Data Initiative Virtual Symposium Series

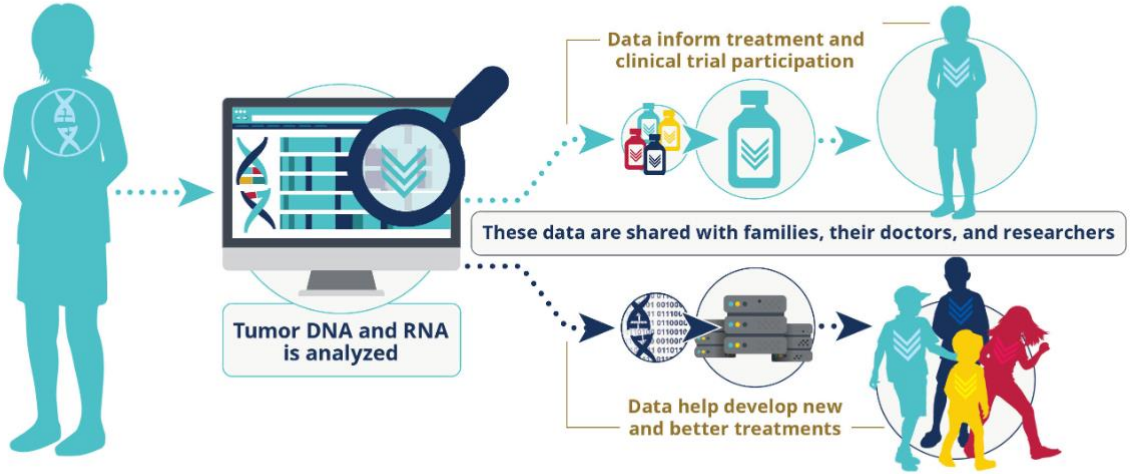
Developing a Clinical Methylation Classifier for Pediatric Brain Tumors and Sarcomas

Catherine Cottrell, Ph.D., FACMG

NIH NATIONAL CANCER INSTITUTE

NATIONAL CANCER INSTITUTE

WHAT IS THE CCDI Molecular Characterization Initiative?



cancer.gov/CCDI-molecular

Somatic Disease/ Germline Comparator Exome

Solid Tumor Fusion Analysis

Methylation Array

Assays Performed within MCI

Clinical Molecular Characterization Initiative Assays and Analytics		
Sample	**Assay**	**Result Type**
Tumor + normal DNA	Enhanced ES	Germline + somatic SNVs, INDELs, CNV & LOH
Tumor RNA	Targeted RNA Solid Tumor Fusion assay	Fusion/ITD detection
Tumor DNA	DNA based methylation array for CNS tumor classification	Disease classification

https://ccdi.cancer.gov/MCI

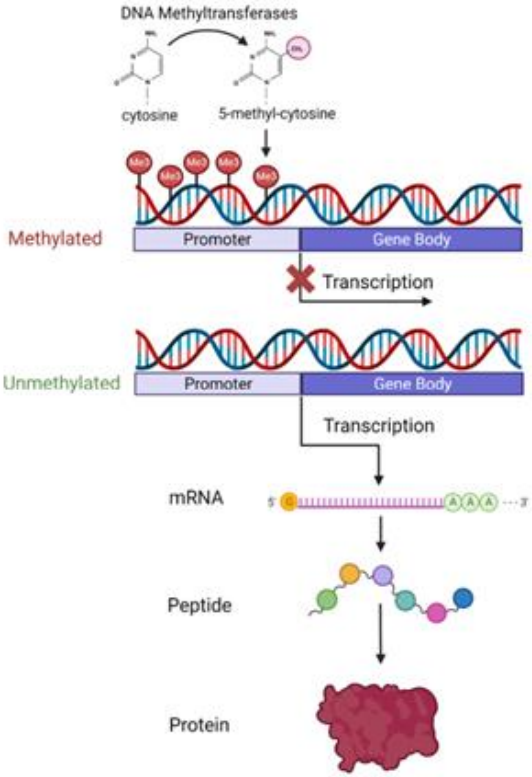
Clinical & Research Components

Disease Group	Clinical Return of Results	Research Testing*	Not performed
CNS	1. Somatic Disease/Germline Comparator Exome 2. Solid Tumor Fusion Analysis 3. Methylation Array Tumor Classification		
STS	1. Somatic Disease/Germline Comparator Exome 2. Solid Tumor Fusion Analysis	Methylation Array Tumor Classification	
RAR	1. Somatic Disease/Germline Comparator Exome 2. Solid Tumor Fusion Analysis	Methylation Array Tumor Classification	
NBL	1. Somatic Disease/Germline Comparator Exome	Methylation Array Tumor Classification	Solid Tumor Fusion Analysis
EWS	1. Somatic Disease/Germline Comparator Exome 2. Solid Tumor Fusion Analysis	Methylation Array Tumor Classification	

*No return of clinical results, however de-identified data are deposited back to the CCDI

Fundamentals of DNA Methylation

- DNA methylation = epigenetic modification in which a methyl group is added to a cytosine nucleotide of DNA
- Increasing evidence supports that the DNA methylation pattern of a tumor reflects its cellular origin in addition to its somatic composition
- Differences in methylation can be measured at:
  - CpG islands
  - Enhancers
  - Promoters



www.cap.org

DNA Based Methylation Array Tumor Classification

- The epigenetic landscape of a tumor can be assayed through a variety of methods, including DNA-based methylation array
- Bisulfite conversion distinguishes methylated vs unmethylated cytosine
- Whole genome amplification
- Fragmentation
- Purification
- Hybridization
- Washing
- Single base extension with a labeled nucleotide
- Staining and Imaging
  - IDAT Files – store intensity data from the probeset on the array

Central Nervous System Methylation Classifier Version in Use for MCI

3/2022- 4/2023

- Heidelberg DKFZ v11b6
- Illumina v1 (850k)
- Random Forest
- Capper et al., 2018
  - PMID: 29539639
- Family, Class

4/2023-5/2024

- Heidelberg DKFZ v12.5
- Illumina v1 (850k)
- Random Forest
- Super Family, Family, Class, Subclass

5/2024-Present

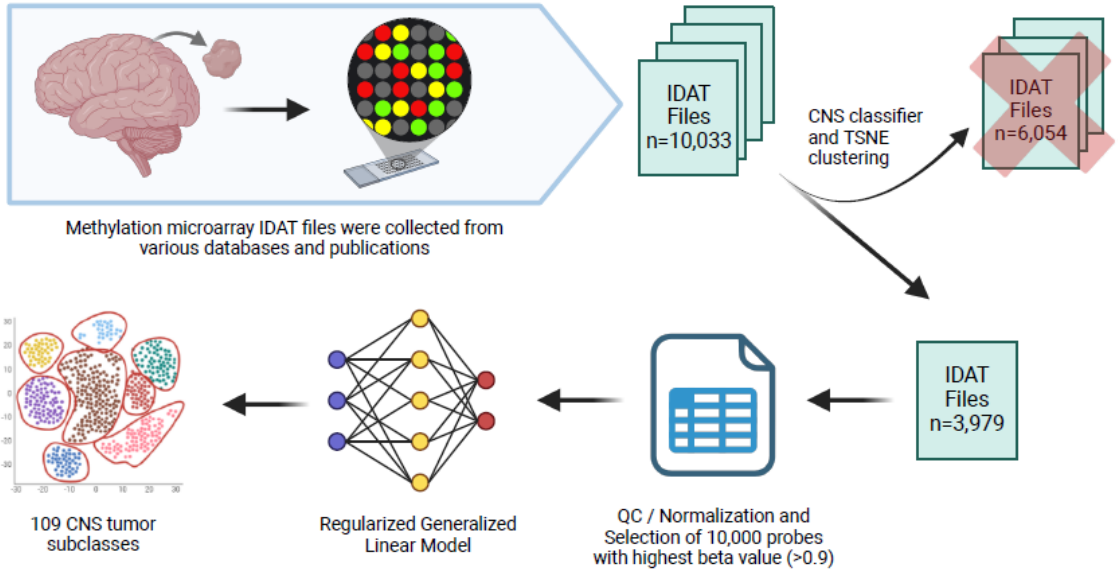
- IGM v1
- Illumina v2 (935k)
- Regularized Generalized Linear Model
- Super Family, Family, Class, Subclass

CNS Tumor Classifier Transitioned Over Time

- Key points in the development of an institutional CNS methylation classifier
  1) Illumina announced plans to replace the existing 850K Methylation EPIC (v1.0) arrays with a 935K Methylation EPIC (v 2.0) array
  2) DKFZ v12.5 was not compatible with the 935k (v2) array initially
  3) DKFZ initiated plans to offer methylation-based classification as a commercial service
  4) Establishment of an in-house classifier enabled greater control over software versioning, and timelines for updates, consistent with lab regulatory requirements
  5) Ability to refine the classifier over time as new disease entities are identified and defined

Content courtesy of Dr. Elaine Mardis

Overview of Classifier Training



Training IDAT files collected from 22 databases (primarily GEO)
www.ncbi.nlm.nih.gov/geo

- Includes n=3979 cases from 168 different CNS subtypes as reference datasets for model training
  - *Superfamily (n=32)*
  - *Family (n=73)*
  - *Class (n=131)*
  - *Subclass (n=168)*

Methylation **A**rray **C**lassification **D**iagnostic **A**nd **D**ata **I**ntegration (MACDADI; IGMv1)

Visualization courtesy of Ke Qin, Ben Kelly and Dr. Marco Leung

IGMv1 Classifier Training Verification

- RGLM Machine Learning model training is *fast*: under four hours to train the classifier models for raw and calibrated scores
- Our significant experience with the DKFZ classifiers allowed for use as a benchmark

Classification results at subclass level	N = 143
Concordant classification with both subclass score > 0.8	101
Neither classifiers reached subclass score of 0.8	17
DKFZ classifier had subclass score > 0.8, MACDADI < 0.8	6
DKFZ classifier had subclass score < 0.8, MACDADI > 0.8	19

- For the 17 cases that did not receive a clinically reportable subclass score, 10 had **family/class** scores greater than 0.8

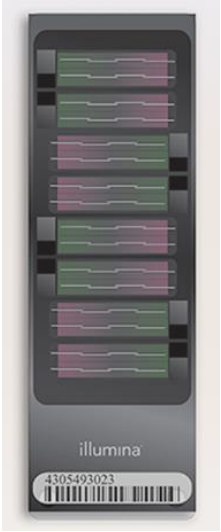
Validation Performance Characteristics

Samples	Run 1 - [redacted]	Run 2 - [redacted]	Run 3 - [redacted]
1	x	x	x
	x		
	x		
2	x	x	x
		x	
		x	
3	x	x	x
			x
			x

- Precision and Reproducibility =100%
  - 3 Samples, 5 replicates in total
  - Based on a high confidence score >0.9 in all categories (superfamily through subclass) and *MGMT* methylation status (unmethylated score = <0.3582; methylated score =>0.3582).
- 30 samples were included to compare the clinical indications of CNS samples (based on histological diagnosis) to the IGM v1 CNS Classifier results using Illumina MethylationEPIC V2 chips.
- 92.5% (27/30) concordance for CNS classification based on a high confidence score ≥0.9 in the categories of superfamily through class.

Validation Performance Characteristics

- Further concordance studies
  - 50 tumor samples were studied to compare output from DKFZ v12.5 850k v1 to IGMv1 935k v2 arrays
  - *MGMT* methylation status was concordant in all cases
- 43/50 (86%) of tumor samples were concordant



Reference and Reportable Range IGMv1

Super Family (n=32)
Family (n=73)
Class (n=108)
Subclass (n=109)

Three confidence score levels
Reporting based on the highest score of the most granular tier

Classifier Score	Description
≥0.9	Consistent with …
0.7-.89	No match, but potentially suggestive of …
<0.7	Tumor classification not determined.

Output and Visualization Tools

- tSNE plots with R package Rtsne
- Copy number variation (CNV) with R package conumee (850k)/ conumee2.0(935k)
- *MGMT* prediction with R package mgmtstp27

Comprehensive Molecular Data Support Classifier Calls

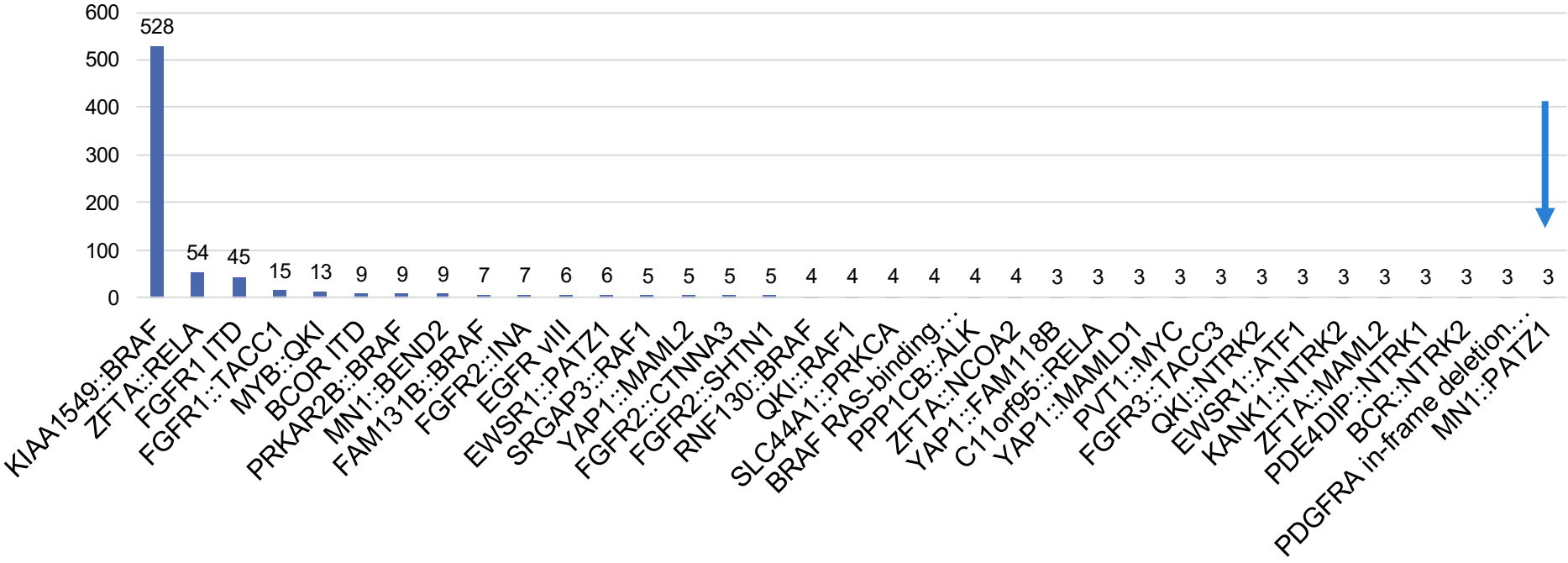
	Predicted Classification	Classifier Score
Super Family	Neuroepithelial tumor with PATZ1 fusion	0.9603
Family	Neuroepithelial tumor with PATZ1 fusion	0.9603
Class	Neuroepithelial tumor with PATZ1 fusion	0.9603
Subclass	Neuroepithelial tumor with PATZ1 fusion	0.9603

1. AMP/ASCO/CAP Tier I or Tier II Structural Variants (Potentially Actionable)

Gene Fusion	5' Fusion Partner	3' Fusion Partner	Classification (AMP/ASCO/CAP)
MN1::PATZ1	*MN1*:NM_002430.2 |exon: 1| (GRCh37) chr22:28192751	*PATZ1*:NM_014323.2 |exon: 1| (GRCh37) chr22:31740754	Tier I (Level A)

Central Nervous System Fusions - MCI

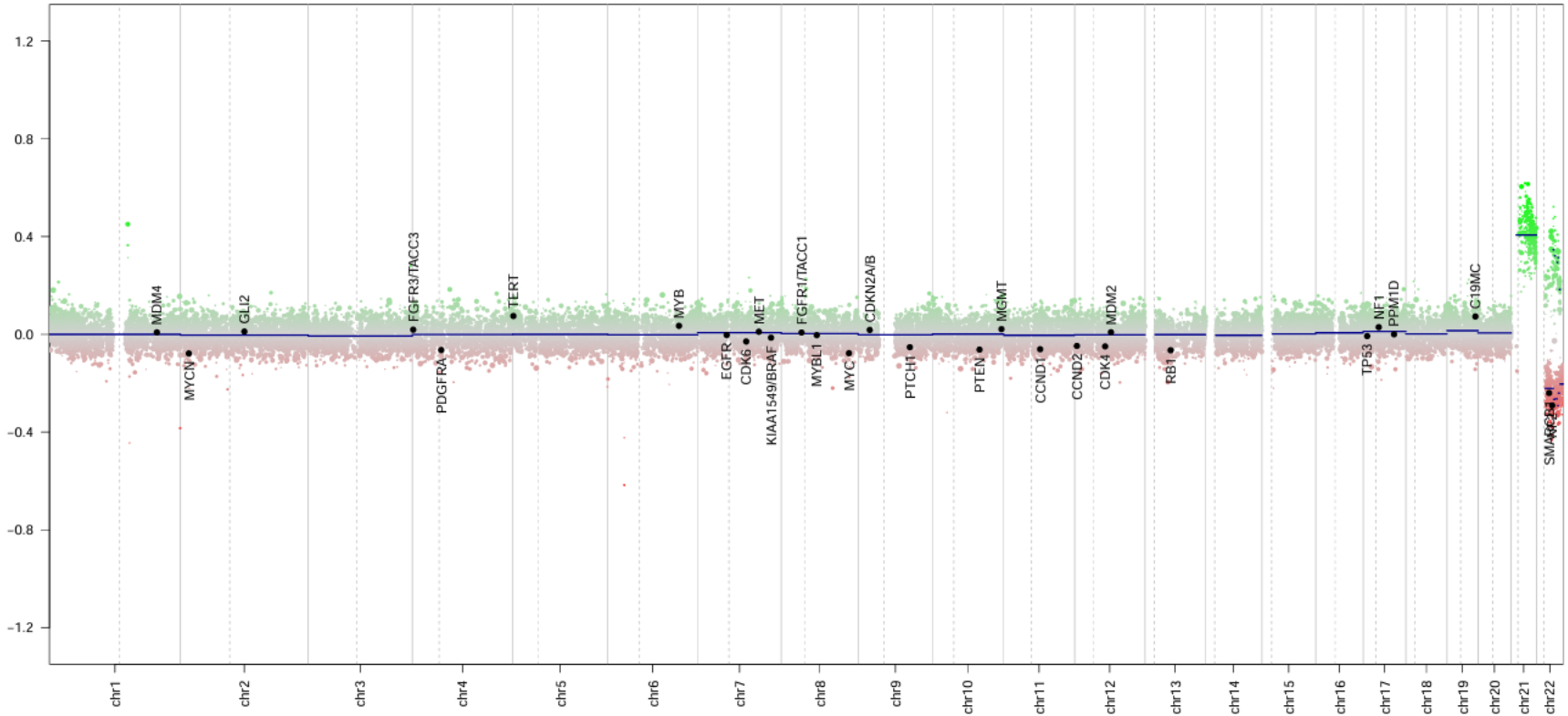
Total Cases	Positive Results
3312	940 (28.4%)



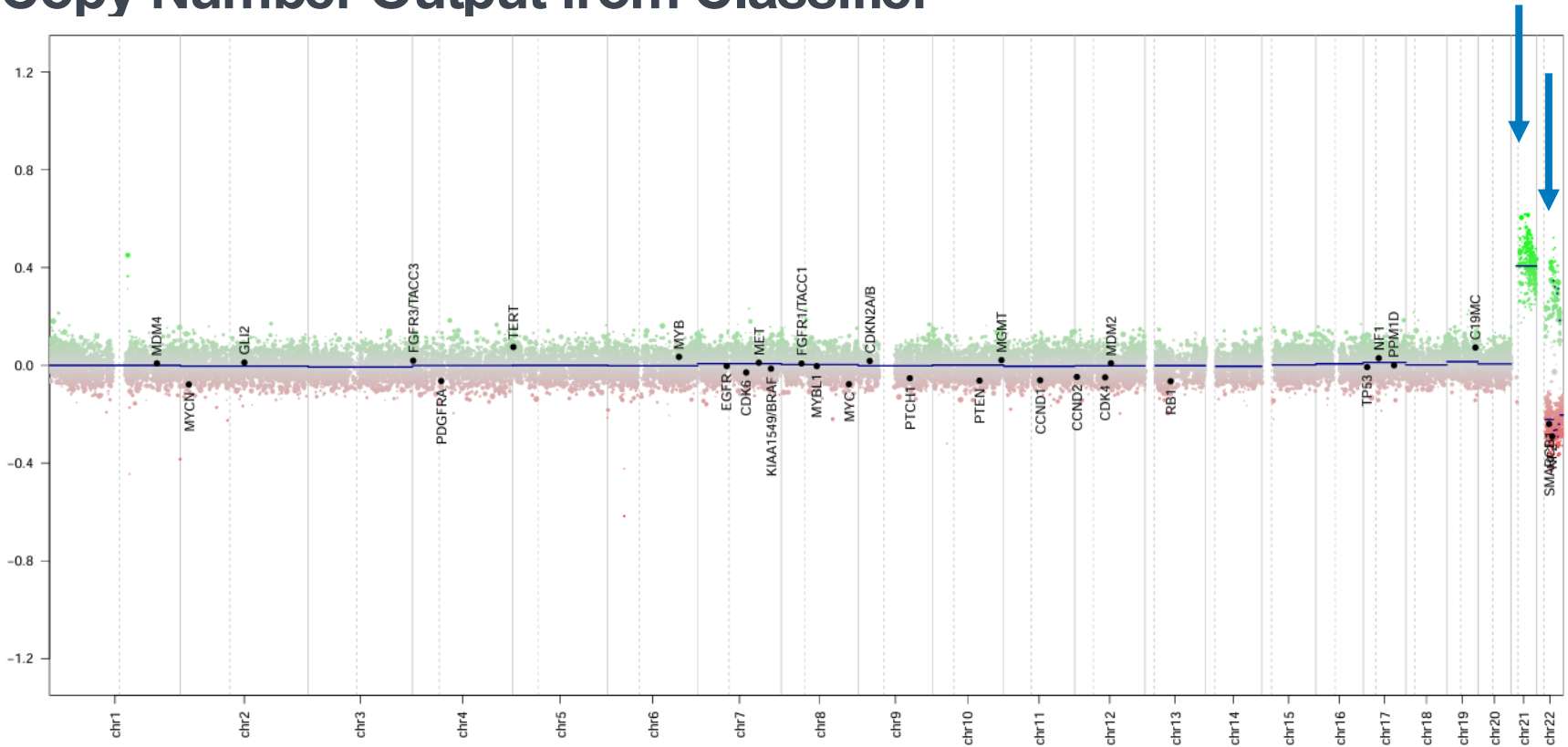
Fusions found in at least 3 tumors

3/31/22-7/31/25

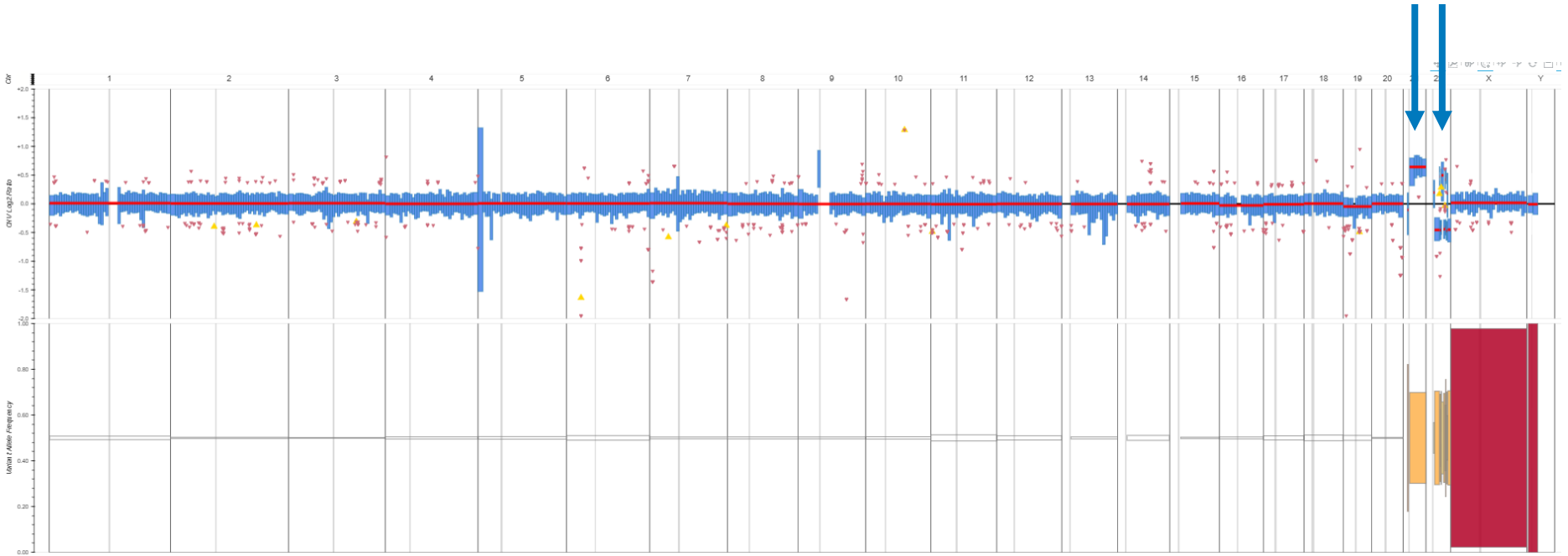
Copy Number Output from Classifier



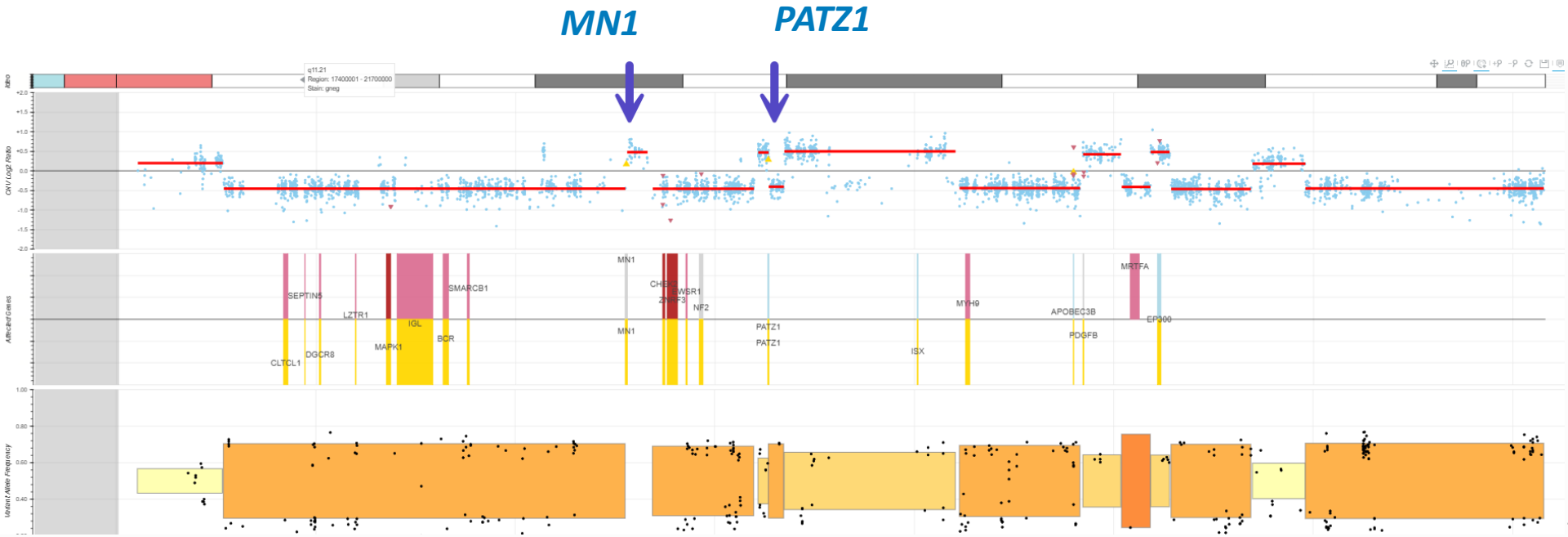
Copy Number Output from Classifier



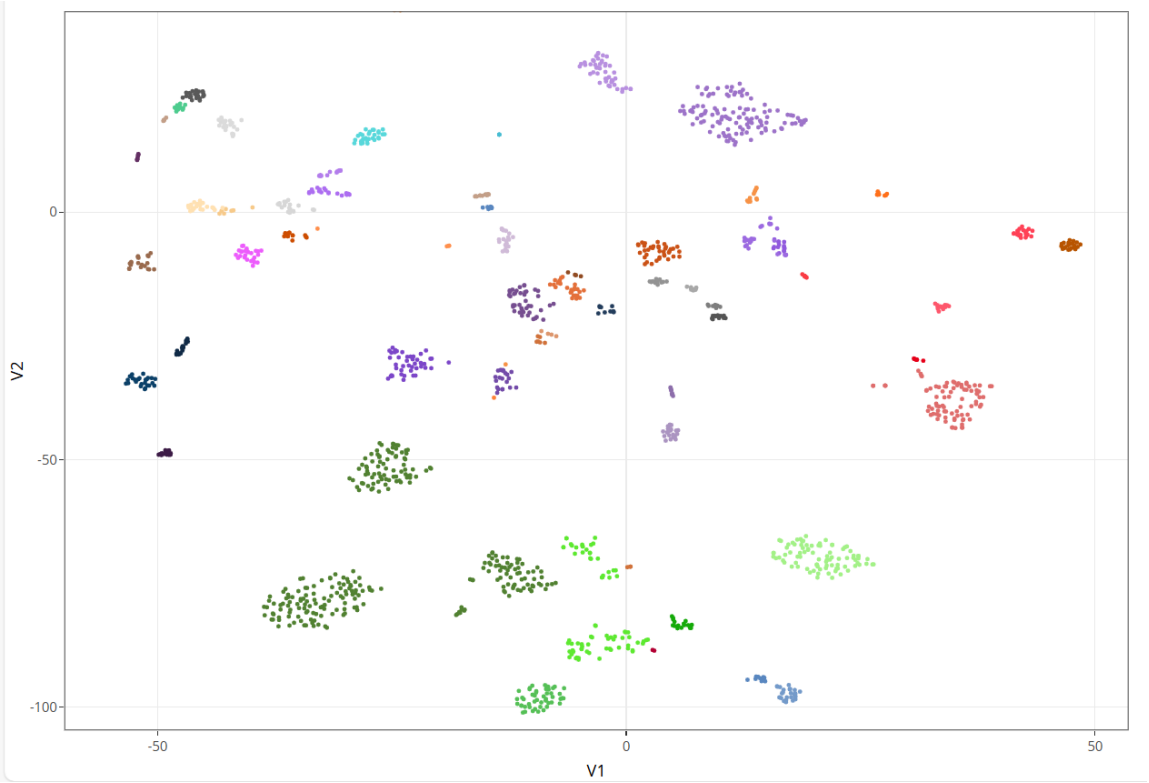
Copy Number Derived from Tumor Exome Analysis



Copy Number Derived from Tumor Exome Analysis- Chr 22



tSNE: t-distributed stochastic neighbor embedding Visualization of CNS Tumors



MCI CNS Methylation Array Results

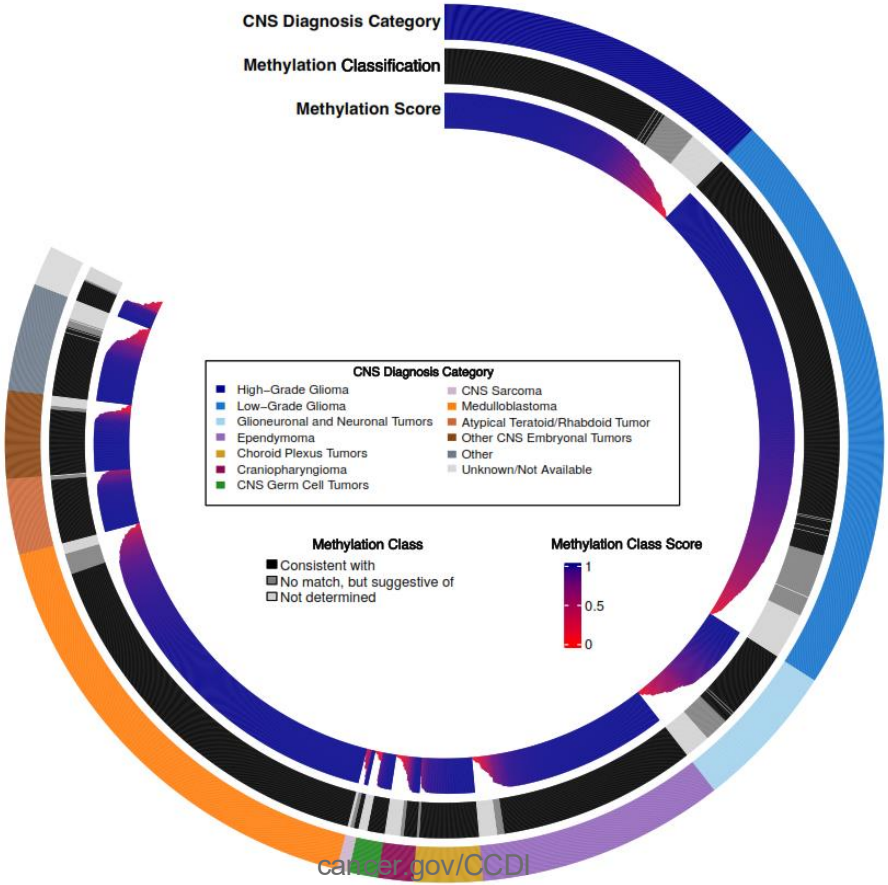
Disease Group	Total Cases Signed Out (excluding cancelled orders)	Positive Results (classified to a methylation group)
CNS	3466	2969 (85.7%)

1. IDAT files
2. HTML files for Classifier Output (IGM v1 CNS)
3. JSON files for Classifier Output (IGM v1 CNS)
4. PDF for Redacted Clinical Report
5. JSON files for Redacted Clinical Report

Data are deposited into the CCDI!

Data courtesy of Kareesma Parbhoo

CNS Tumor Classification by Category MCI N=2223



Plot courtesy of Dr. Kathleen Schieffer; Data courtesy of Drs. Sarah Leary and Diana Thomas

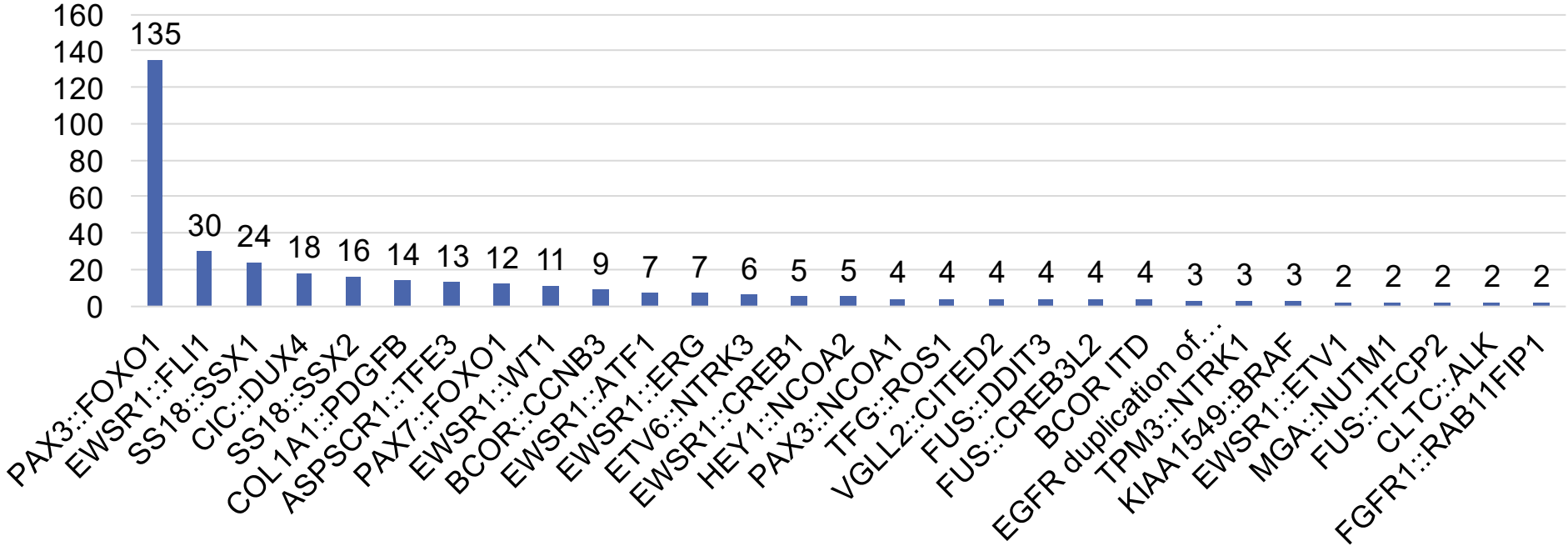
Beyond CNS Classification

Sarcoma **M**ethylation **A**rray **C**lassification **D**iagnostic **A**nd **D**ata **I**ntegration (sMACDADI)

Model employed for research use, gathering data for clinical validation

Soft Tissue Sarcoma Fusions: MCI

Total Cases	Positive Results
1003	418 (41.7%)



Fusions found in at least 2 tumors

3/31/22-7/31/25

Challenges in Classification

- Rare tumors not represented in the dataset
- Samples with limited tumor involvement
  - Minimum 60% content
- Tissue specific methylation can influence tumor classification, particularly among the same tumor entities arising out of different organ systems
- Other variables
  - Microenvironment
  - Genetic drivers
  - Age

Thank you!

cancer.gov

cancer.gov/espanol

Q&A

Childhood Cancer Data Initiative Virtual Symposium Series

CCDI Pediatric Cancer Diagnosis Ontology

Michael Rusch

Overview

1. Illustrate the Challenges
2. Introduce the Working Group
3. Progress to Date

Background: CCDI Data Federation

- Sites providing pediatric cancer data implement a common API to query/report on available datasets
- Participant and sample metadata elements presented
  - In their original “unharmonized” form
  - Harmonized to a standard vocabulary selected by the group and written into the API specification

Challenges Illustrated: Orthography and Logic

Source	Harmonized?	Diagnosis
StJude	No/Original	Acute Myeloid Leukemia with GLIS Family Rearrangement
KidsFirst	No/Original	Acute Megakaryoblastic Leukemia (Fab M7)
Treehouse	No/Original	acute megakaryoblastic leukemia

- French-American-British (FAB) classifications for AML
- World Health Organization Classification of Tumors ("WHO Blue Books") deprioritized FAB starting in the 3rd ed.

Challenges Illustrated: Standards Are Imperfect

Source	Harmonized?	Diagnosis
StJude	No/Original	Acute Myeloid Leukemia with GLIS Family Rearrangement
KidsFirst	No/Original	Acute Megakaryoblastic Leukemia (Fab M7)
Treehouse	No/Original	acute megakaryoblastic leukemia
Treehouse	ICD-O-3.2	9910/3 : Acute megakaryoblastic leukemia

- International Classification of Diseases for Oncology (ICD-O) version 3.2 uses FAB classification.

Challenges Illustrated: Standards Are Changing

Standard	Versions
WHO Blue Books	5 editions
ICD-O	3 editions + 2 revisions
Disease Ontology	>150 releases
OncoTree	>26 releases
Cancer Classifications for Kids (CC4K)	6 releases

- Change is necessary to reflect latest knowledge.
- Changing standards or versions requires large-scale reclassification efforts.

Challenges Illustrated: Characteristics Are Added/Removed

Source	Versions
oncotree_2025_04_08	Therapy-Related Myeloid Neoplasms (TMN)
CC4K v0.2	Therapy-Related Myeloid Neoplasms (TMN)
WHO 5th ed.	Myeloid neoplasm post cytotoxic therapy
Umeda, et al.	*[Classify based on biomarker]*
CC4K v0.3	[AML] with KMT2A Rearrangement *post_cytotoxic_therapy = True*

- *Umeda M, Ma J, …, Klco JM. A new genomic framework to categorize pediatric acute myeloid leukemia. Nat Genet. 2024*

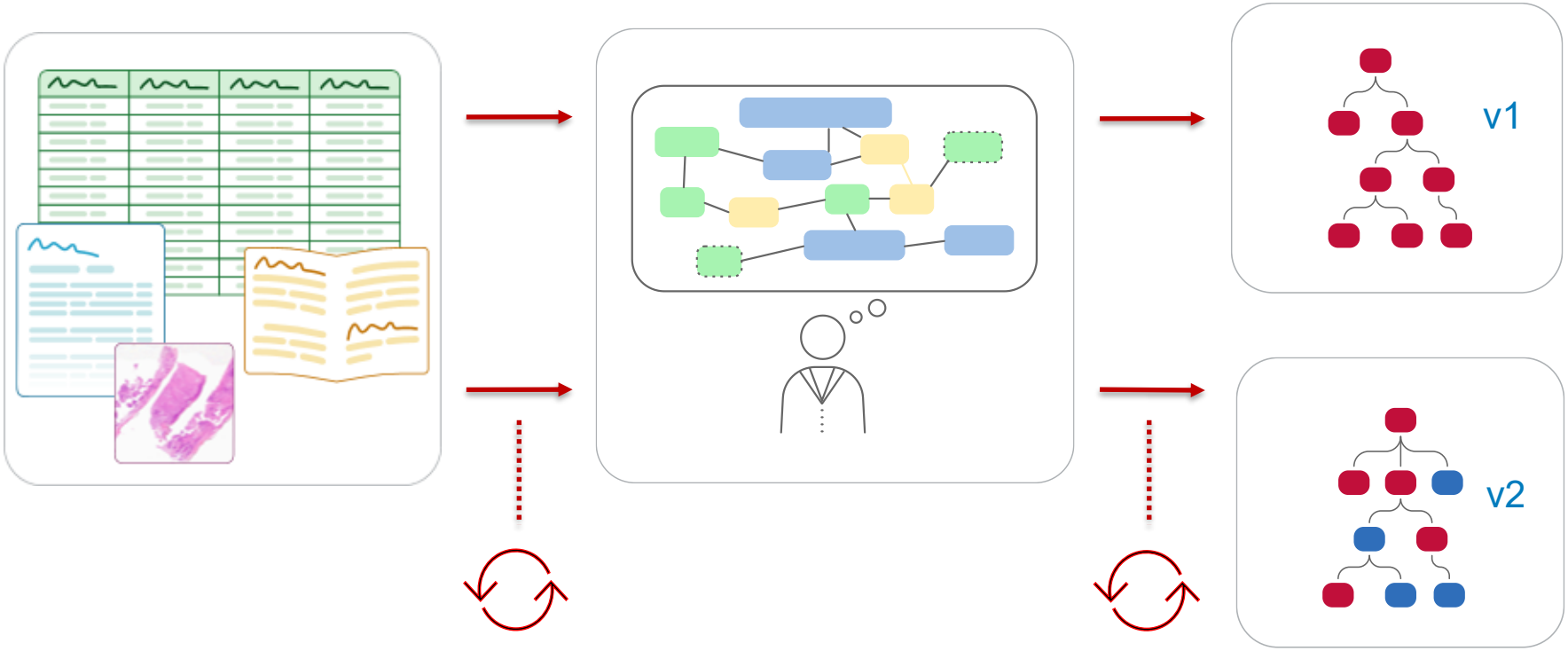
CCDI Pediatric Diagnosis Ontology Task Force

Diagnosis Classification Standard Practice

Multi-Modal Sample Data

Expert Curation

Classification Mapping

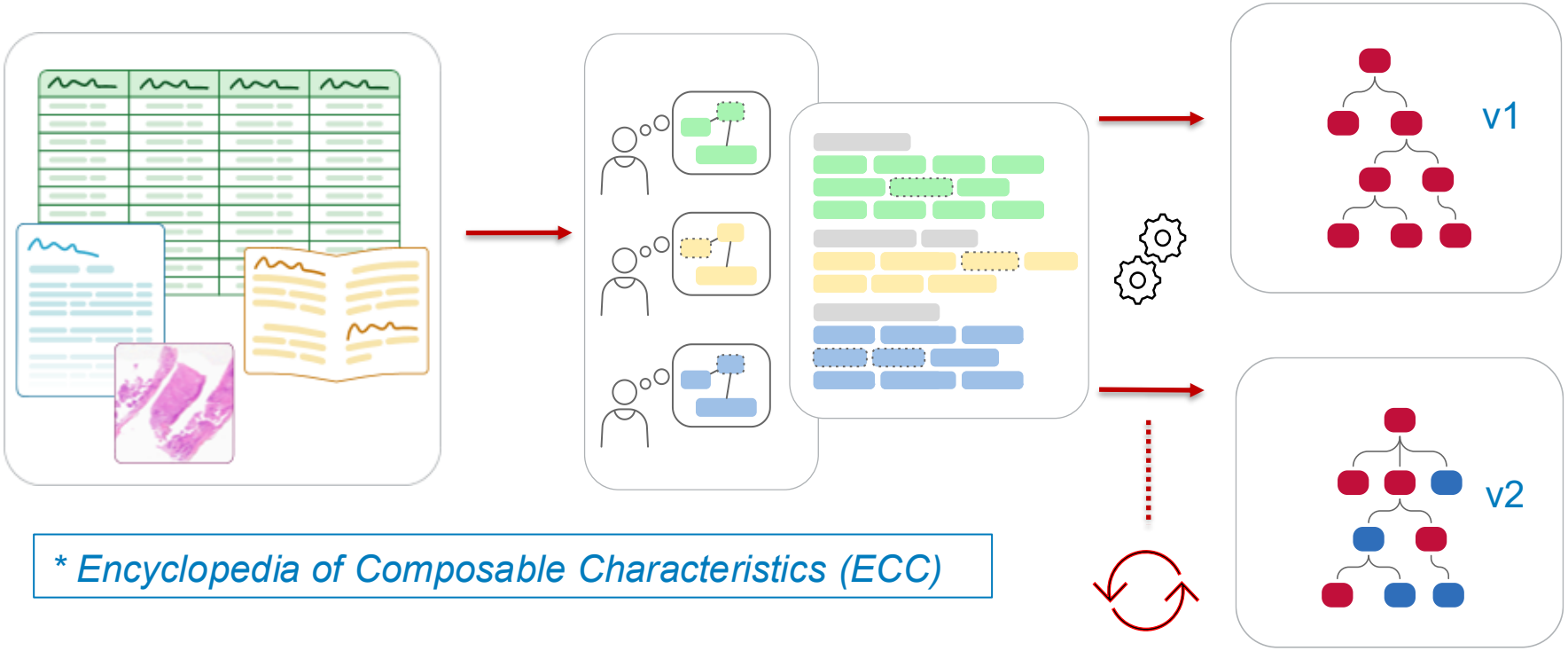


Diagnosis Classification with Composable Characteristics

Multi-Modal Sample Data

ECC* Characteristic Assignment

Classification Mapping



* *Encyclopedia of Composable Characteristics (ECC)*

Tool Mockup: Start

SAMPLE0001234
Subtype Classification

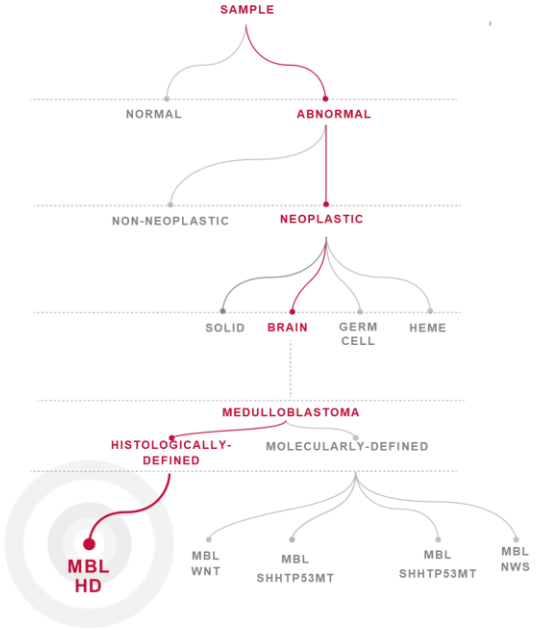
Assigned Characteristics

- Patient-derived Sample
- Abnormal Tissue
- Neoplastic Tumor
- CNS-derived Tumor
- Blue-round Cell
- Necrosis/Apoptosis
- ...
- Classic MBL
- Cerebellar Vermis

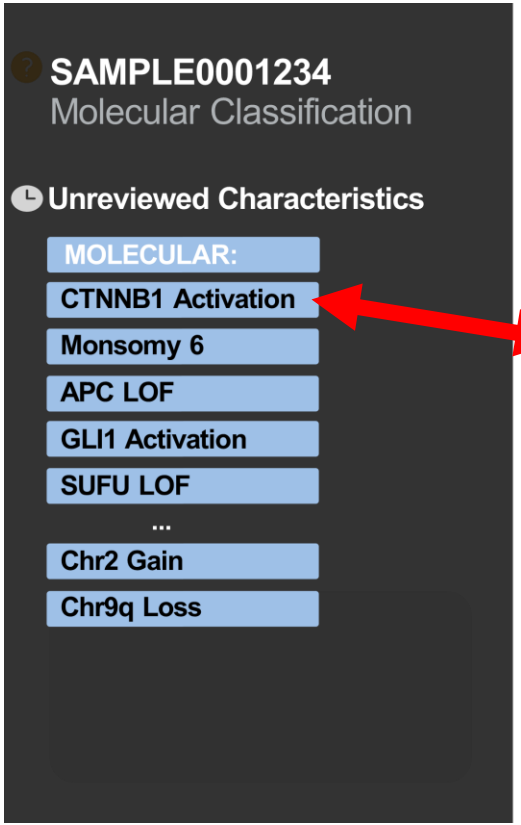
Unreviewed Characteristics

- MOLECULAR

Assignment:
Medulloblastoma, Histologically Defined



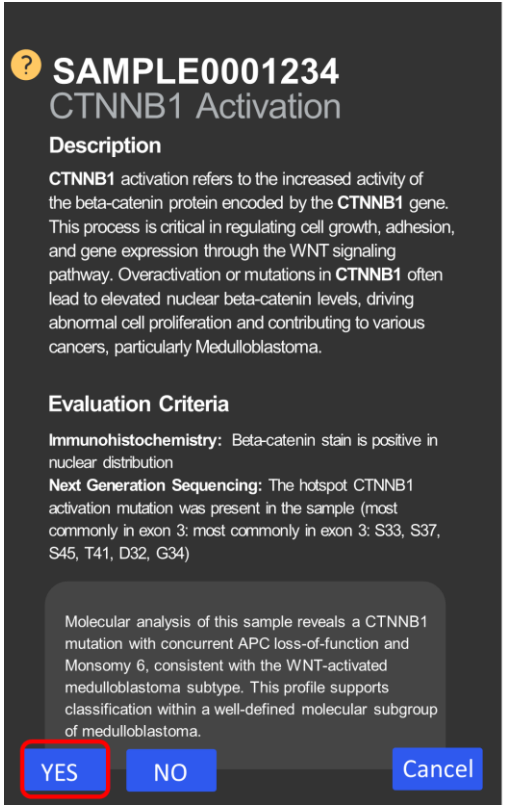
Tool Mockup: Molecular Characteristics



Review: Molecular Classification

GENE	Alteration	WES(%)	RNA(%)
CTNNB1	G34E	48	48
AKT3	L313M	53	23
SPARCL1	E633D	53	13

Tool Mockup: CTNNB1 Characteristic



Review: CTNNB1 Activation

GENE	Alteration	WES(%)	RNA(%)
CTNNB1	G34E	48	48
AKT3	L313M	53	23
SPARCL1	E633D	53	13

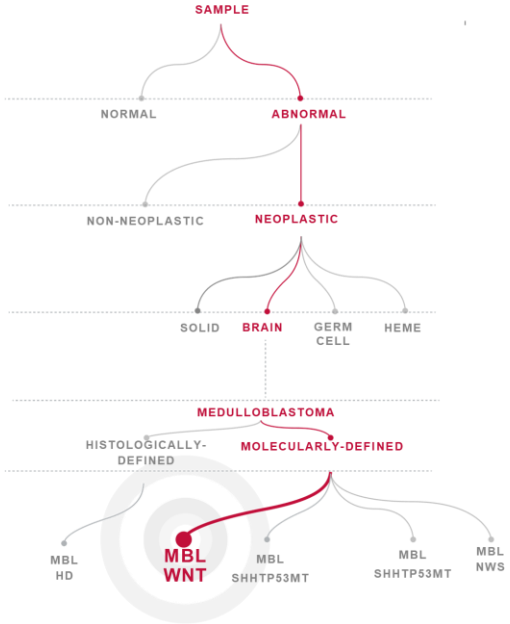
Tool Mockup: Finish

SAMPLE0001234
Subtype Classification

Assigned Characteristics

- Patient-Derived Sample
- Abnormal Tissue
- Neoplastic Tumor
- CNS-derived Tumor
- Blue-Round Cell
- Necrosis/Apoptosis
- ...
- Classic MBL
- Cerebellar Vermis
- CTNNB1 Activation

Assignment:
Medulloblastoma, WNT Activated



ECC Progressive Pilot

Complete

1 **Assess mapping feasibility**: single-site with unstructured characteristics

In Progress

2 **Design characteristics**: evaluate existing terminologies, consider reverse mapping

3
- **End-to-end pilot**: multi-site with fully developed characteristics and mapping

Federation API: Diagnosis Search

Source	ID	Diagnosis
PCDC	PAVVLW	8900/3 : Rhabdomyosarcoma, NOS
PCDC	PAVVGG	8920/3 : Alveolar rhabdomyosarcoma
Treehouse	1157	8920/3 : Alveolar rhabdomyosarcoma
Treehouse	1163	8910/3 : Embryonal rhabdomyosarcoma, NOS
StJude	SJ000004	*Embryonal Rhabdomyosarcoma (ERMS) **
StJude	SJ000008	*Rhabdomyosarcoma, NOS (RMSNOS);Alveolar Rhabdomyosarcoma (ARMS) **
KidsFirst	pt-1fdv89dh	Rhabdomyosarcoma
KidsFirst	pt-2v4gz9cq	Alveolar Rhabdomyosarcoma

https://federation.ccdi.cancer.gov/api/v1/**subject-diagnosis?search=rhabdomyosarcoma**

** indicates unharmonized values*

WHO Blue Book Translation

- WHO Blue Book is not a structured vocabulary
- Duplication of effort to translate to discrete data elements
- Thorough translation had been performed by Cancer Pathology Coding Histology And Registration Terminology (**Cancer PathCHART**)
- Their translation being used to prevent future duplicated effort

Coordination Between OncoTree and CC4K

- CC4K started as a fork of OncoTree
- Significant divergence due to adult vs pediatric focus and priorities of distinct user groups
- Meetings to coordinate efforts, eliminate needless divergence
- Started by comparing hematologic (heme) categories,
  - Structural convergence for B- and T-lineage classifications
  - CC4K updating 39 codes to match OncoTree
  - OncoTree adding 16 terms present in CC4K

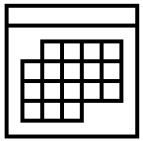
Diagnosis Categories

Improve diagnosis-level search and cohort aggregation by enabling users to query based on high-level groupings

Brain
n=13

Heme
n=6

Solid
n=12



- CCDI Data Model – Updated to include Diagnosis Categories
- CCDI Hub and C3DC Explore dashboards - Q3

Summary

- Task Force effectively balancing quick wins with innovative long-term solutions
- Work continues on each of these efforts
  - Composable characteristics
  - OncoTree/CC4K coordination
  - Diagnosis categories

Thank you!

cancer.gov

cancer.gov/espanol

Q&A

Join Us For Our Upcoming Events!

CCDI Virtual Symposium Series

April 7

April 8

CCDI Webinar

April 14

Learn more and register at **events.cancer.gov/ccdi/webinar**

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