Childhood Cancer Data Initiative (CCDI) Webinar Series

MCI Update: Pediatric Soft Tissue Sarcomas

Jack Shern, M.D.

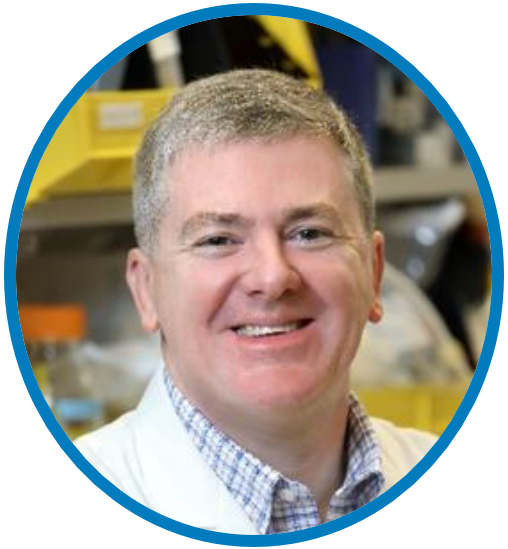
NIH NATIONAL CANCER INSTITUTE

December 10, 2024

Agenda

1. *Molecular Characterization Initiative (MCI): Creating a National Strategy*
2. *MCI Data Flow: Enabling Access to Data for Research and Analysis*
3. *MCI Soft Tissue Sarcoma Data: Clinical Impact*
4. *Q&A*

Today's Speaker

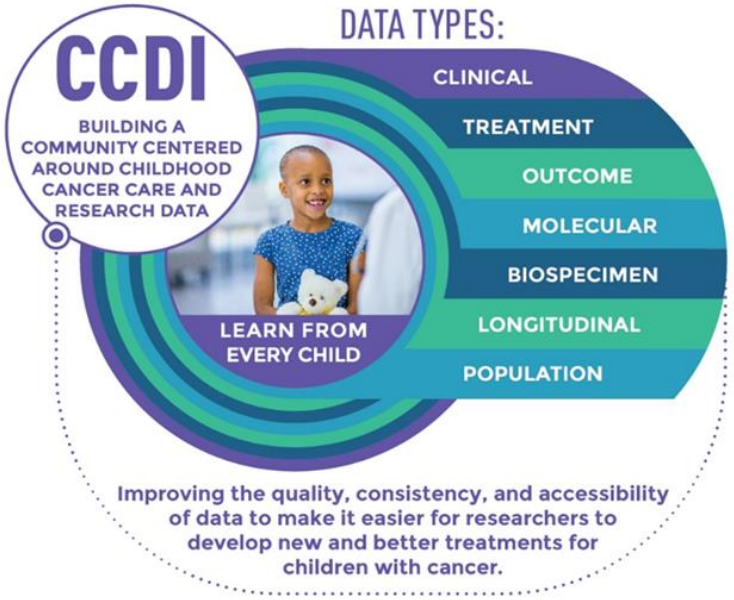


Jack Shern, M.D.

- Lasker Clinical Research Scholar
- Head of the Tumor Evolution and Genomics Section, National Cancer Institute

CCDI Goals

- Gather **data from every child, adolescent, and young adult** diagnosed with a childhood cancer
- **Create a national strategy** of molecular characterization to inform diagnosis and treatment
- **Develop a platform and tools** to bring together clinical and research data that will inform new insights in biology/etiology to improve preventive measures, treatment, quality of life, and survivorship for childhood cancers
- **Engage the entire** childhood cancer care and research community



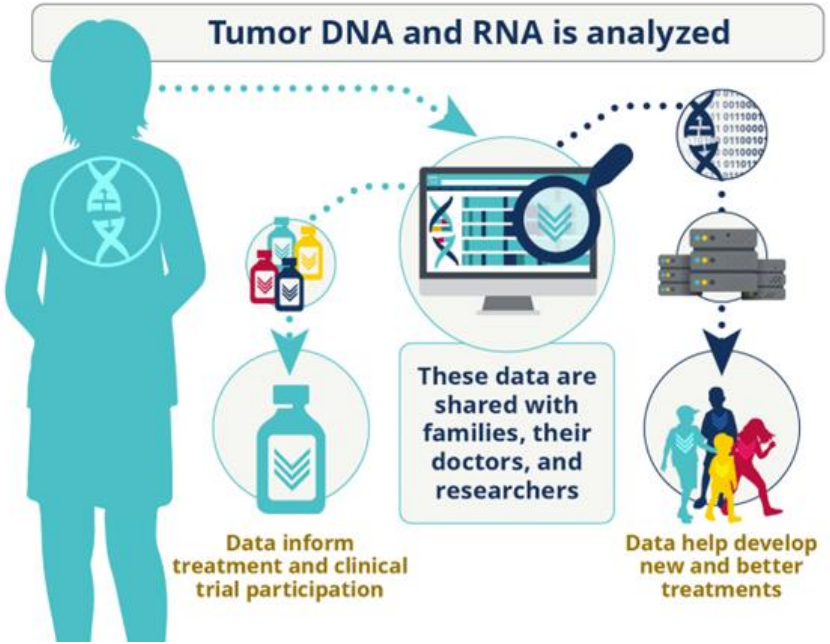
*Flores-Toro JA et al., J Clin Oncol, 2023 (*PMID:37267580); *Jagu S et al., Pediatr Blood Cancer, 2024* (PMID: 37889049)

Molecular Characterization Initiative

Creating a National Strategy

CCDI Molecular Characterization Initiative (MCI)

- Launched in partnership with Children's Oncology Group's (COG) Project:EveryChild
- State-of-the-art molecular characterization at diagnosis (WES, fusions, methylation) in a CLIA-testing environment at no cost to participants
- Results returned to participants and treating physicians within 14-21 days
- Enrolled more than 4,500 participants from all 50 states, Canada, Australia, and New Zealand
- Learn more: ccdi.cancer.gov/MCI

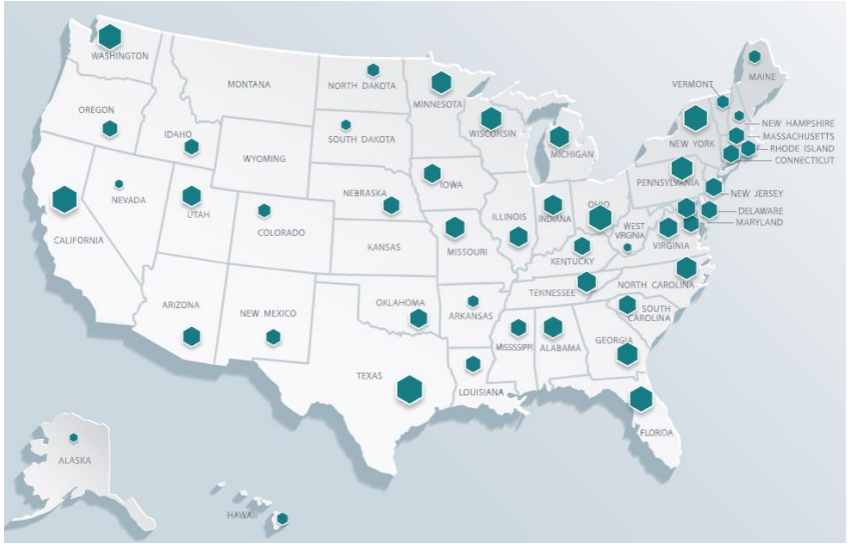


MCI Enrollment Metrics

Enrollments by Diagnosis

Type	Introduced	Number
CNS tumors	March 2022	3503
Soft tissue sarcomas	May 2022	1061
Rare tumors	September 2022	497
High-risk Neuroblastoma	February 2024	265

Enrollment Counts by State



Enrollment metrics as of November 12, 2024

Expanding MCI

- Plan to add Ewing sarcoma and relapsed or refractory tumors
- Expand to include AYAs outside of COG for the Coordinated Pediatric, Adolescent, and Young Adult Rare Cancer Initiative
- Prioritize diseases for **research characterization** and determine which assays (WGS, RNA Seq, single cell, epigenetics, proteomics/metabolomics) are appropriate to deepen our understanding of cancer biology
  - Partnering with COG disease-specific scientific committees and other subject matter experts

MCI Data Flow

Enabling Access to Data for Research and Analysis

CCDI Data Ecosystem: A Connected Network of Resources

GDC & cBioPortal:
Visualization Tools

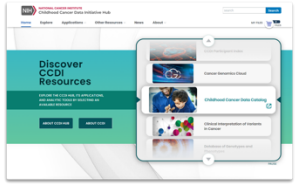
Federation & CPI APIs:
Data Discovery & Integration

Cancer Genomics Cloud:
Data Integration &
Analysis

C3DC & NCCR:
Clinical & Real World Data

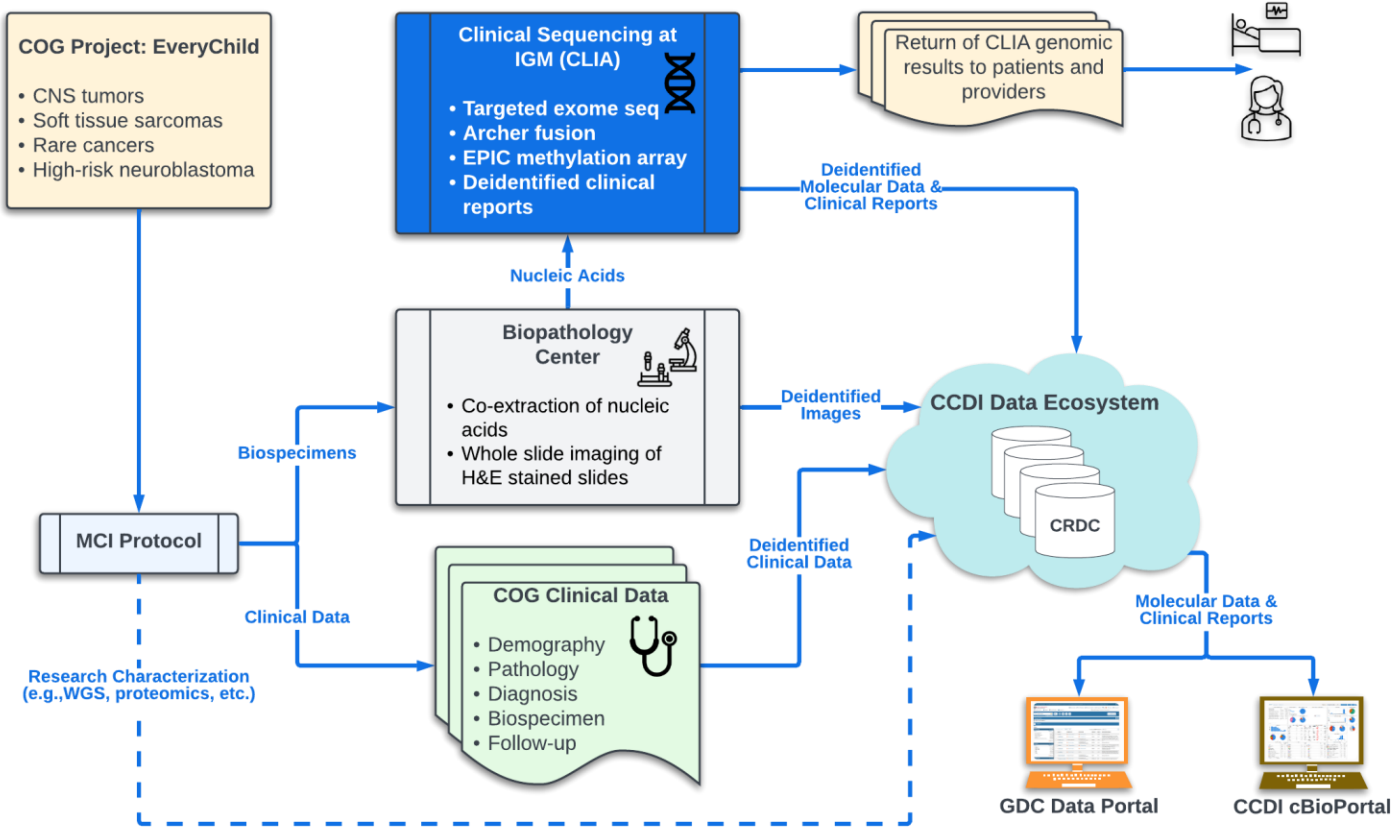
Catalog, MTP, & CIViC:
Knowledgebases

Hub Explore: Build Cohorts



Hub: Entry Point

MCI Data Flow



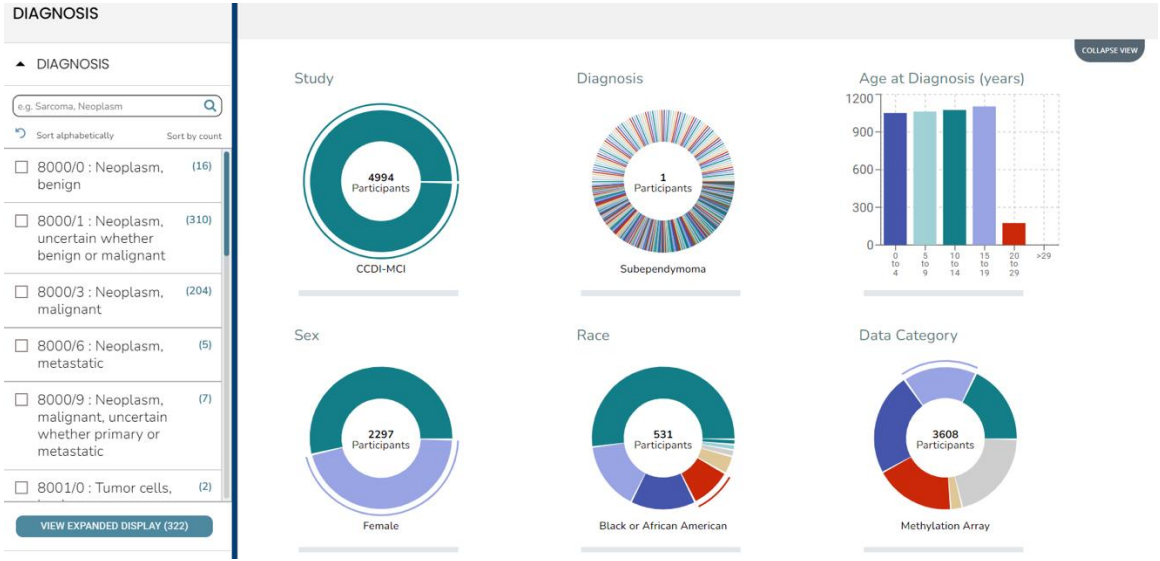
MCI: Assay Types & Data File Formats

Sample	Assay	Analytics	Result Type	File Description	File Format
Tumor + normal DNA	Enhanced WES **Note: > 10-fold less input DNA requirement for exome**	250x Churchill alignment to GRCh38; IGMseq Pipeline Analysis in AWS	Germline + somatic SNVs, INDELs, CNV, and LOH **Note: Clinical report will include tumor mutational burden (TMB)**	Germline and Somatic Read Alignments	.FASTQ, .cram, .crai;
				Germline and Somatic VCFs	.vcf.gz
				PII Redacted Report	.pdf, .json
Tumor RNA	Archer Dx Pan-Solid FusionPlex **Note: RNAseq will replace Archer panel for fusion calling**	Archer Analysis v. 6.0 in AWS	Fusion/ITD detection	Targeted RNA Read Alignment	.cram, .crai
				Archer Fusion Results	.txt
				PII Redacted Report	.pdf, .json
Tumor DNA	Illumina 850K EPIC methyl array	DKFZ Classifier v. 12.5	Disease classification	Illumina EPIC Array Intensity File	.idat
				DKFZ Classification	.html
				PII Redacted Report	.pdf, .json
				Biospecimen and clinical data	.json

Institutional reimbursement for completion of data submission

CCDI Hub Explore Dashboard: Inventory of Data Sets

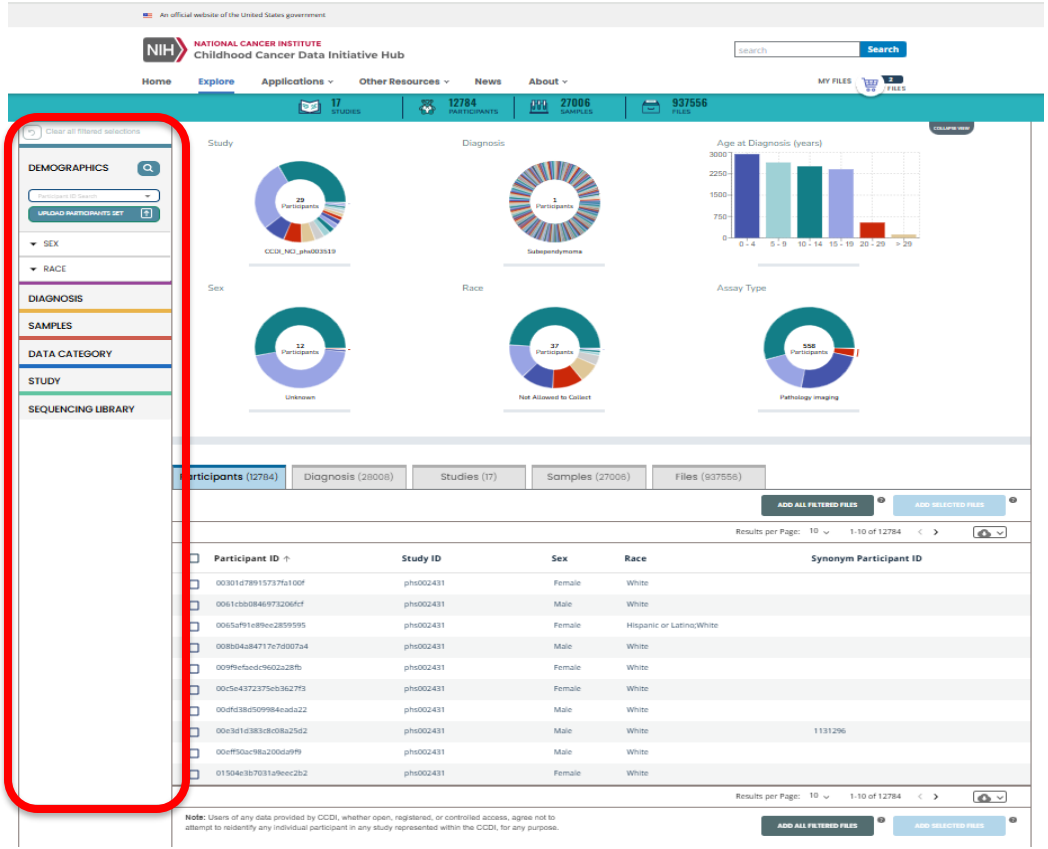
- An inventory of CCDI-managed childhood cancer data
- Provides:
  - Faceted search
  - Visualization of search results
  - Export results for further analysis
- Facilitates data discovery across studies, enabling the building of cohorts



CCDI Hub Explore: Faceted Filtering

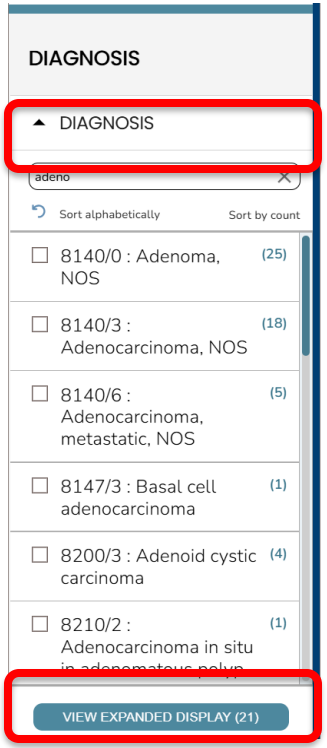
- Filter available data by various annotations of interest
  - Diagnosis
  - Demographics
  - Treatment
  - Treatment response
  - Survival
  - Samples
  - Data category
  - Study

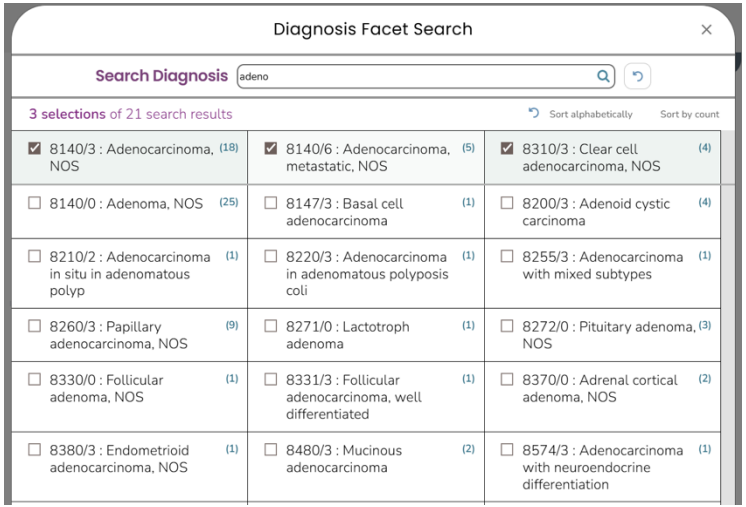
- Facet selections dynamically reload visualization dashboard and tabular lists



CCDI Hub Explore: Diagnosis Search

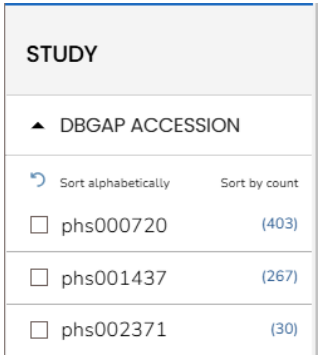
- By default, first several diagnoses are listed sorted by numeric code
- Entering a search term will reduce the space of listed diagnoses to the first several matches
- Clicking the “View Expanded Display” button will show all possible matches
- Checking the associated checkbox will update the visual dashboard and table lists accordingly

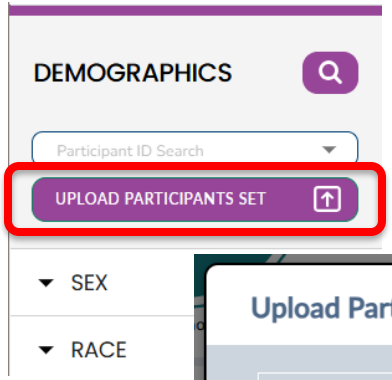




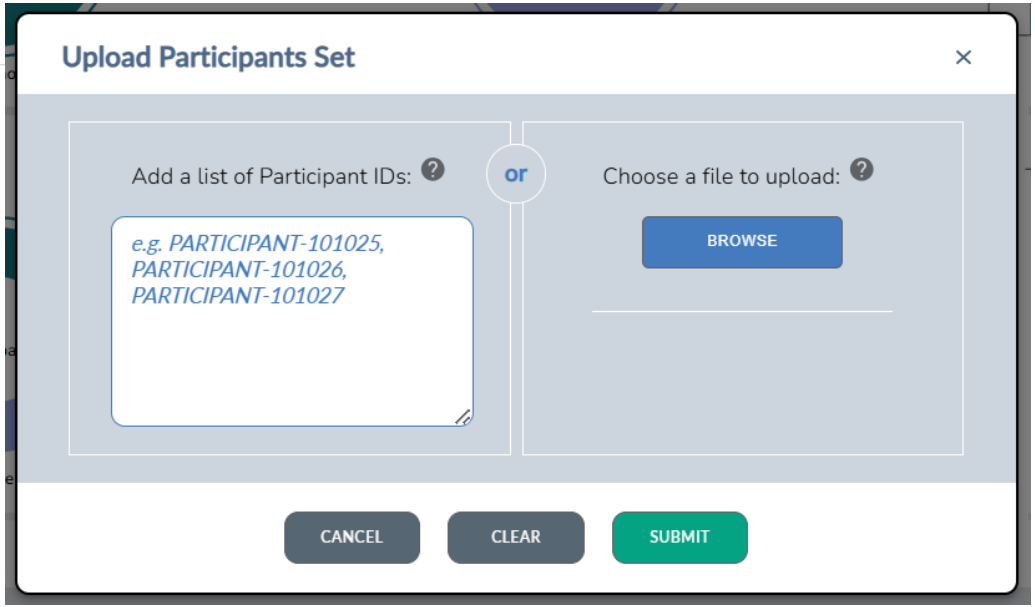
Additional Facet Types

Checkbox selection

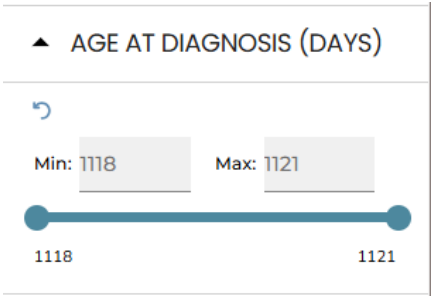


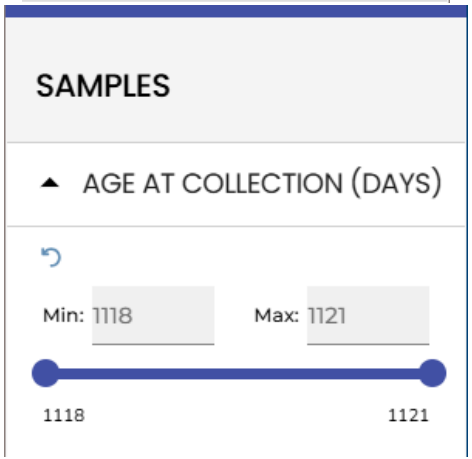


Participant ID upload



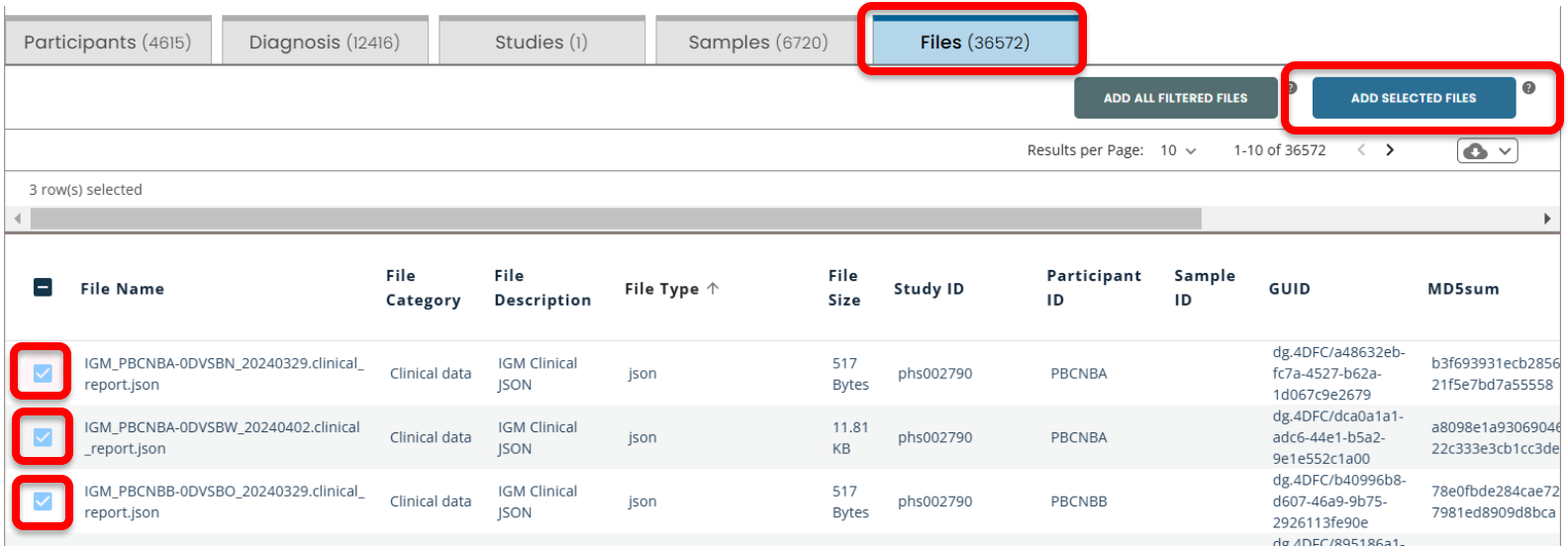
Age sliders





CCDI Hub Explore: File-based Access

- Files can be added to the My Files cart from any of the tab lists
- Options differ by table
  - **Example 1:** add all the files for a Participant from the “Participants” tab
  - **Example 2:** add single files from “Files” tab
- Download study metadata from Studies tab



Obtain Controlled-access Files

1. The database of Genotypes and Phenotypes (dbGaP) access is given using eRA Commons accounts.

- Go to the eRA Commons site and create an account under your organization or institution.

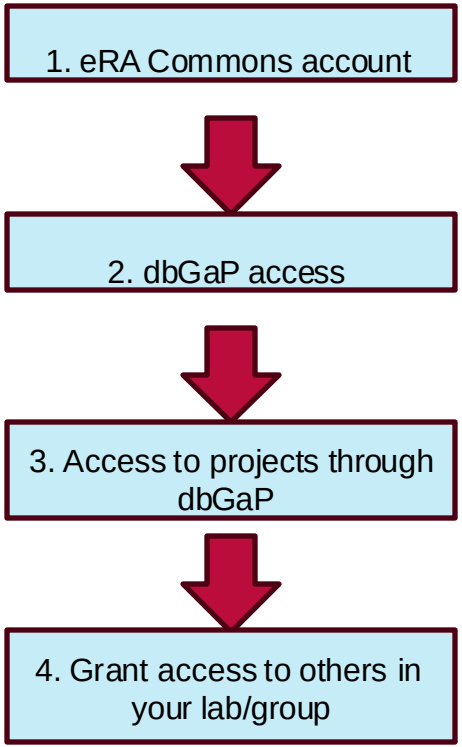
2. Go to the dbGaP Controlled Access Data section and select Authorized Access. Log in with your eRA Commons account.

3. Create a Research Project.

- Select the projects you would like controlled access to.
- Create a Research Use Statement explaining the need for the projects.
- Confirm project structure and send off for review to the Data Access Committee.

4. Go to the My Requests tab to see all current access that is linked to your eRA Commons Account.

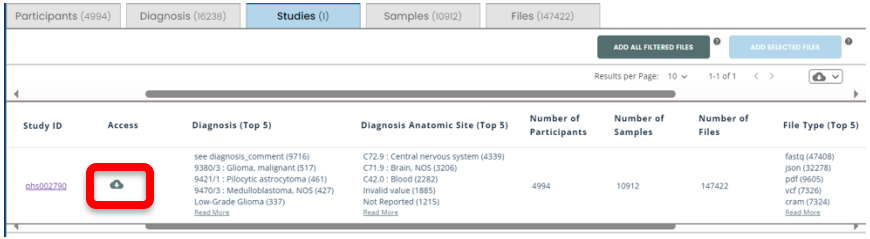
- Go to the "Downloaders" tab and search for other members in your lab/group and add them to the selected Research Projects



How to Apply for Controlled Access on dbGaP: https://www.youtube.com/watch?v=m0xp_cCO7kA

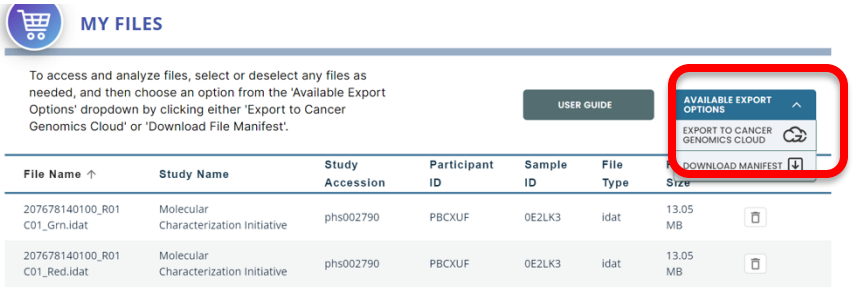
How Do I Access MCI FASTQ Files for Rhabdomyosarcoma?

- **Go to ccdi.cancer.gov/explore**
  - Select "Molecular Characterization Initiative" under Study Name or choose “phs002790” under Study dbGaP Accession
  - Download the Study Manifest
  - Find the tabs ending with "_file" (look for FASTQ, JSON, etc.)



- **Create a Data Repository Service Manifest for Cloud Analysis**

- **Export the manifest to Cancer Genomics Cloud (CGC):**
  - In CGC, go to your Project → Files → Add Files → GA4GH Data Repository Service (DRS)
  - Files will load into your project for analysis



MCI Data Access: FAQs

- **Does the lab data exist in a structured format, or is it expected that we use NLP on the PDF documents to make it computable?**
  - Deidentified clinical sample reports are available in both PDF and JSON formats.

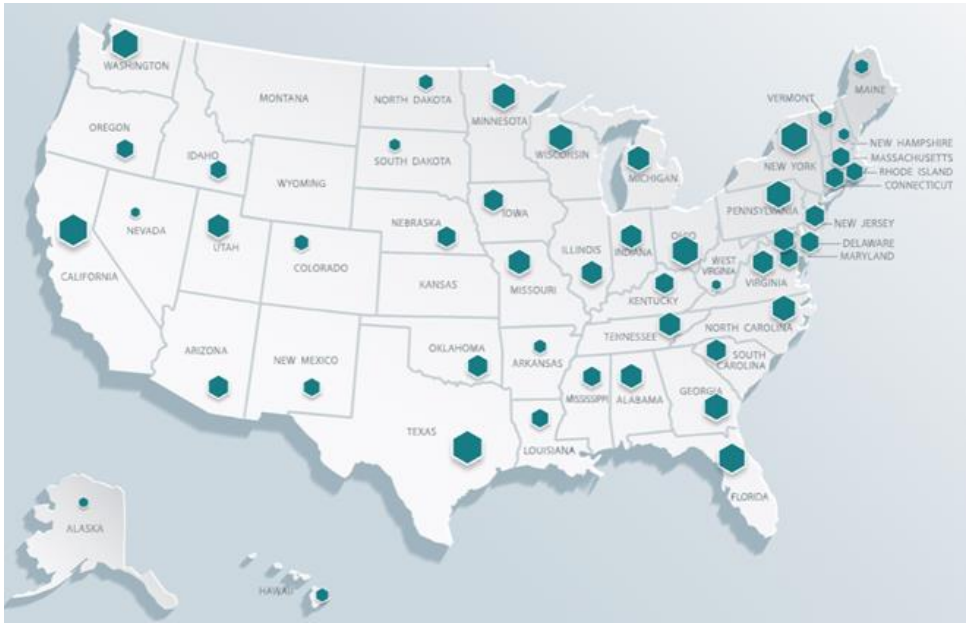
- **Are there institutional IDs attached to the data so we can organize it by institutions in the platform?**
  - Institutional IDs are not attached to the data to reduce the risk of re-identification, especially for extremely rare pediatric cancers.

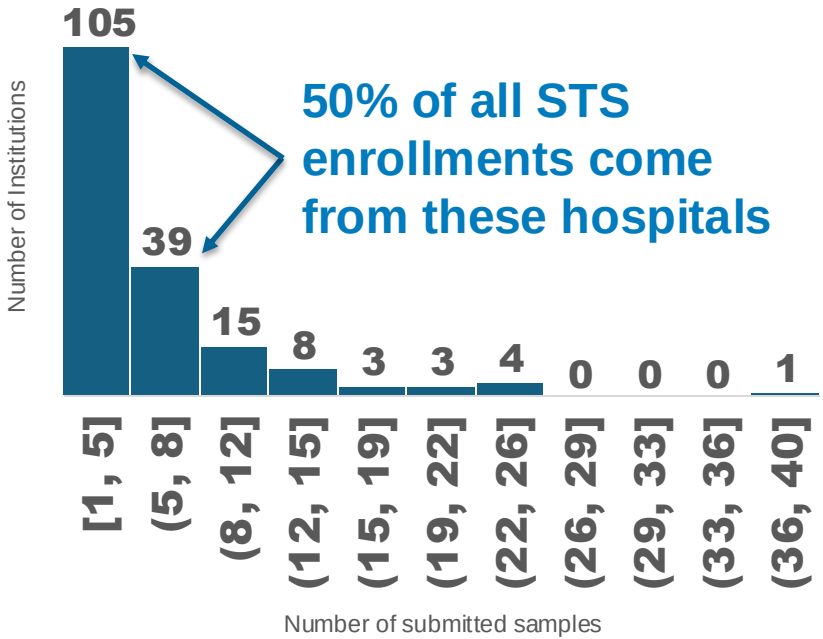
- **Is it possible to download the de-identified summary reports directly without going through the dbGaP approval process?**
  - The summary reports were not provided in this manner due to potential inclusion of sensitive information.

MCI-Soft Tissue Sarcoma Data

Clinical Impact

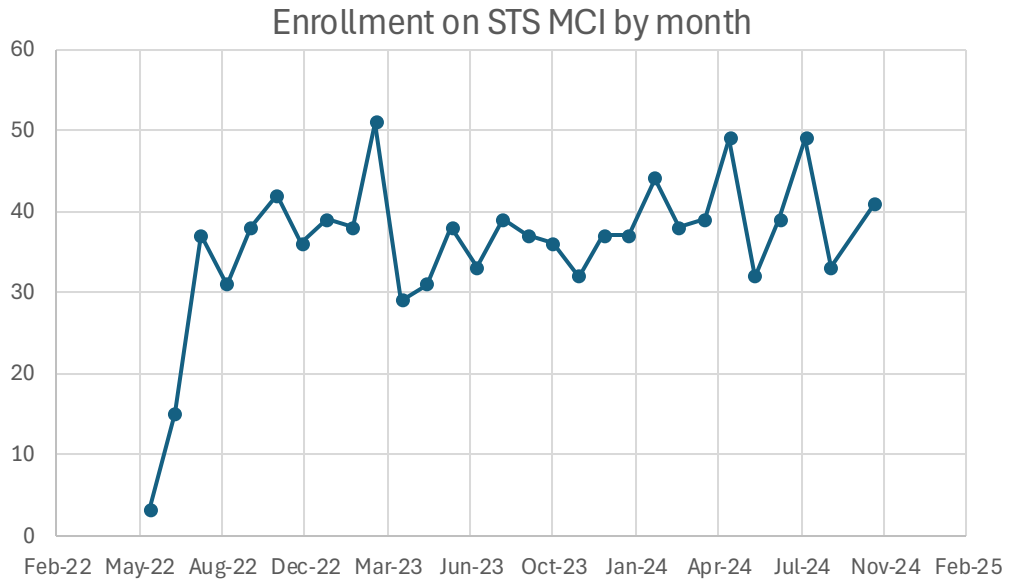
Distribution of Enrollment of Soft Tissue Sarcoma (STS) Patients



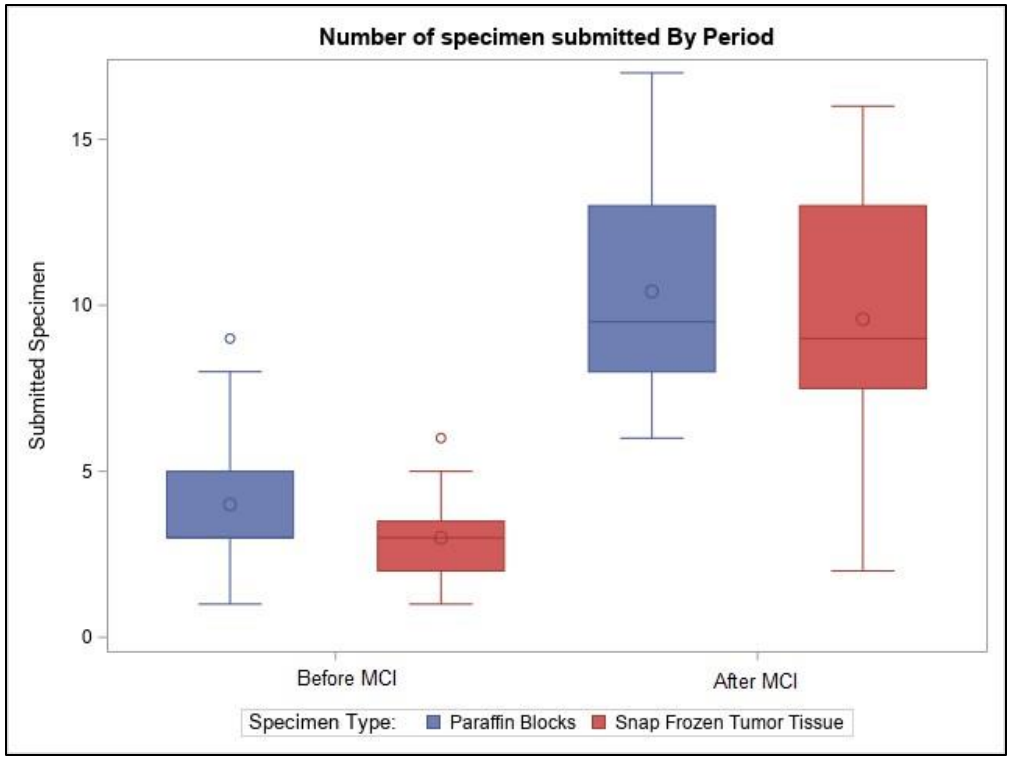


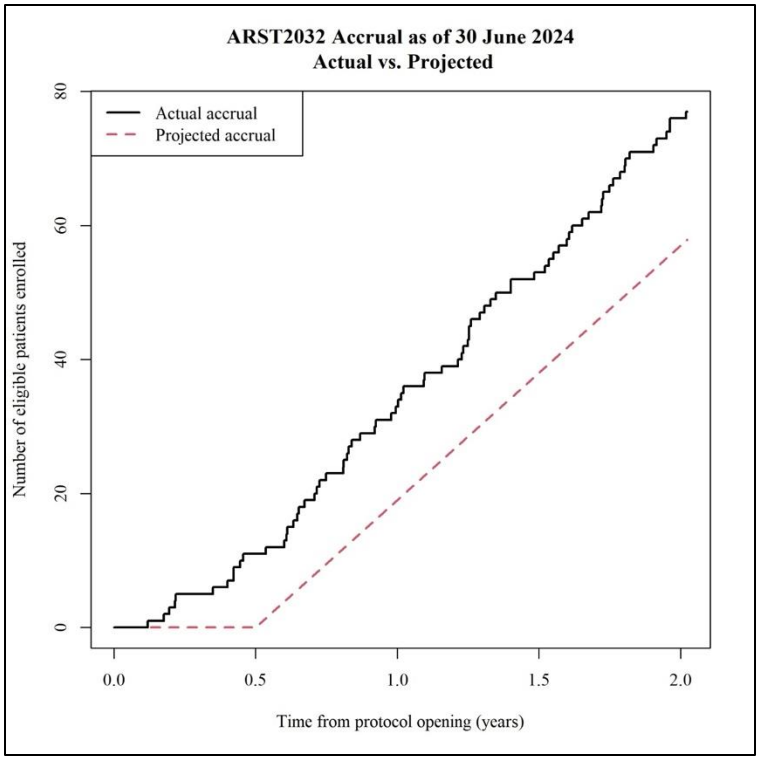
MCI - STS Study: As of 10/29/2024

- 1,043 enrolled patients
- 851 (82%) of enrollees have a submitted paired blood/tumor specimen
- 777 (75%) pass QC and have completed profiling
- 2,170 completed assays



Increased Submission of Samples with Opening of MCI

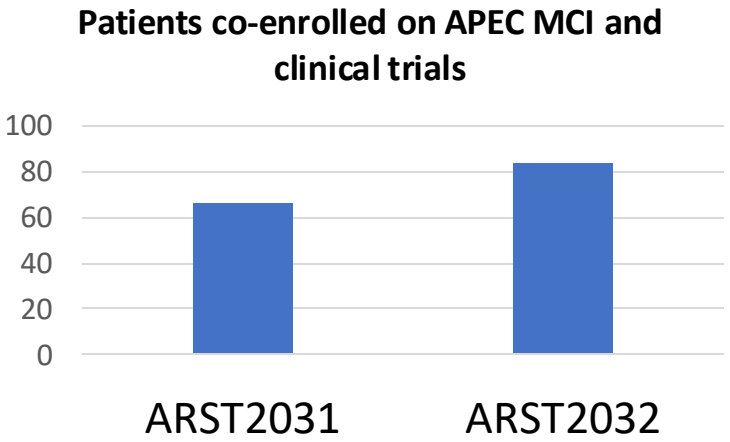




Sapna Oberoi; Wei Xu COG

Co- enrollment on Soft Tissue Sarcoma Clinical Trials

- MCI is required for eligibility on ARST2032 for low-risk rhabdomyosarcoma
- Transition to “Regimen M” for patients with *MYOD1 or TP53* mutant tumors allowed at Week 3 or 6



2. SOMATIC CANCER-ASSOCIATED SEQUENCE VARIATION IN THE TUMOR

Variant Information	Genomic Change (GRCh38)	Etiology	Germline Variant Allele Fraction (%)	Tumor Variant Allele Fraction (%)	Variant Classification (AMP/ASCO/CAP)
NRAS (NM_002524.5) c.35G>C p.Gly12Ala	chr1:114716126 C>G	Somatic	not detected	66%	Tier I (Level B)
TP53 (NM_000546.6) c.404G>T p.Cys135Phe	chr17:7675208 C>A	Somatic	not detected	2%**	Tier I (Level B)
NF1 (NM_001042492.3) c.6420_6427+18del p.?	chr17:31336907 _31336932del	Somatic	not detected	24%*	Tier I (Level B)

*The variant allele frequency/fraction for this variant may be underestimated due to its nature as a deletion event.

3. PERTINENT NEGATIVES

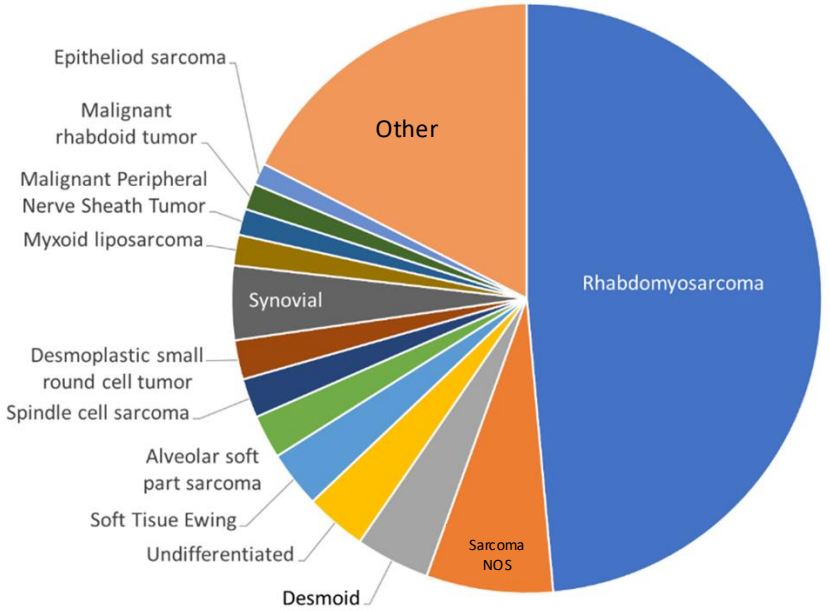
***TP53* Germline Sequence Variants** - No pathogenic or likely pathogenic germline single nucleotide variants or small insertion-deletion events were detected.

***TP53* Copy Number** - No biallelic loss detected. However, a single copy loss of *TP53* was detected on 17p in the tumor specimen.

Soft Tissue Sarcoma patients enrolled in the first year of MCI

Total Number of patients is 425

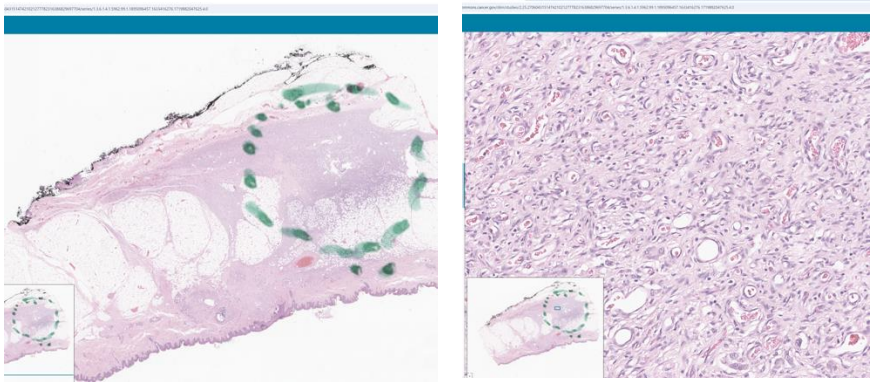
Median age at diagnosis	11.2 (0.07, 25.86)	
Variable	**Frequency**	**Percentage**
Age at diagnosis		
≤ 18 years	382	89.9
> 18 years	43	10.1
Sex		
Female	196	46.1
Male	229	53.9
Race		
American Indian or Alaska Native	5	1.2
Asian	18	4.2
Black or African American	61	14.4
Multiple Races	7	1.7
Native Hawaiian or other Pacific Islander	6	1.4
White	270	63.5
Not Reported or UKN	58	13.7
Ethnicity		
Hispanic or Latino	79	18.6
Not Hispanic or Latino	301	70.8
Not Reported or UKN	45	10.6



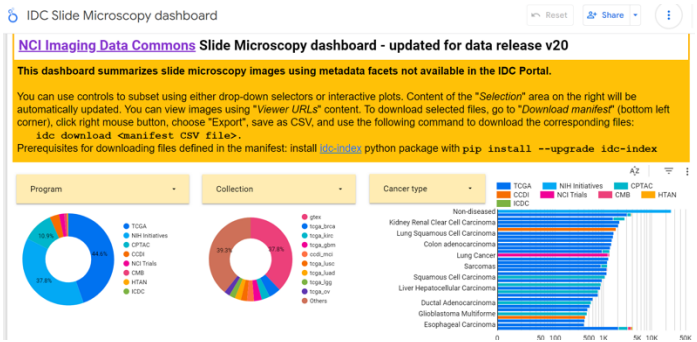
MCI H&E Images Now Available in Imaging Data Commons

- Hematoxylin and eosin (H&E) stained images in Digital Imaging and Communications in Medicine (DICOM) format are now accessible as open-access through the Imaging Data Commons (IDC)
- Additional images associated will be released in the coming months, new images will be periodically added as participants are enrolled
- Links to images on IDC are accessible through CCDI Hub

Imaging Data Commons Collections

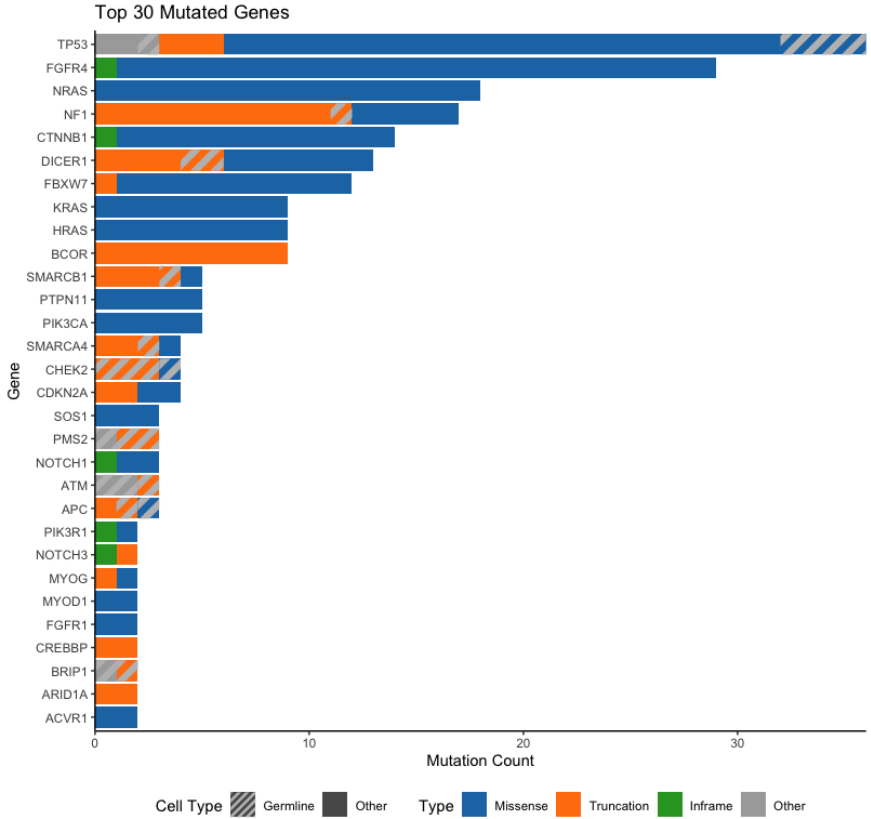


IDC Slide Microscopy Dashboard



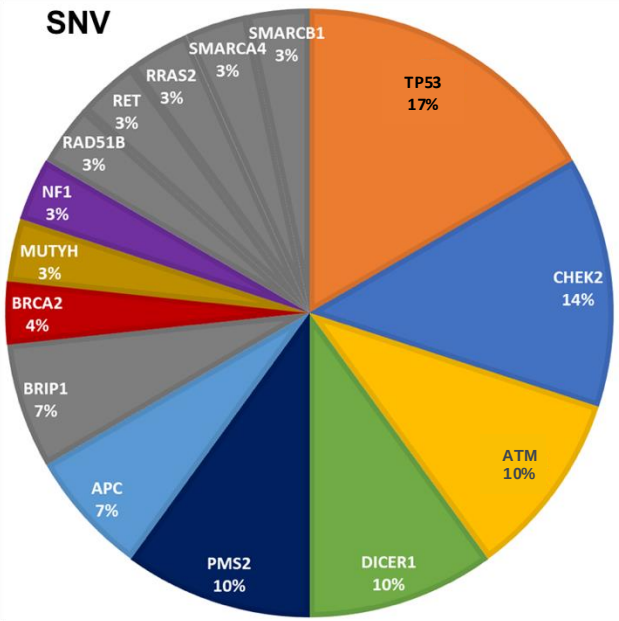
Soft Tissue Sarcoma Mutational Summary

Recurrently Mutated Genes in the STS cohort



- Recurrent gene rank list is largely driven by fusion negative RMS
- Mutation of *FGFR4* is common in fusion negative RMS
- *MYOD1* L122R mutations are uncommon (at least at the time of diagnosis)

11% of profiled STS patients have a reportable germline finding



*Single nucleotide variants reported in 1 or more patients

1. GERMLINE CANCER-ASSOCIATED SEQUENCE VARIATION

Gene (Transcript ID)	Genomic Change (GRCh38)	Nucleotide Change	Etiology/Zygosity	Predicted Protein Change	Associated Disease/Condition	Variant Interpretation (ACMG/AMP Evidence)
RET (NM_020975.6)	chr10:43114598 G>C	c.1998G>C	Het	p.Lys666Asn	(AD) Medullary thyroid carcinoma (OMIM: 155240) (AD) Multiple endocrine neoplasia IIA (OMIM: 171400) (AD) Multiple endocrine neoplasia IIB (OMIM: 162300) (AD) Pheochromocytoma (OMIM: 171300)	Pathogenic (PS1, PS3_Moderate, PS4_Moderate, PM2, PM5)

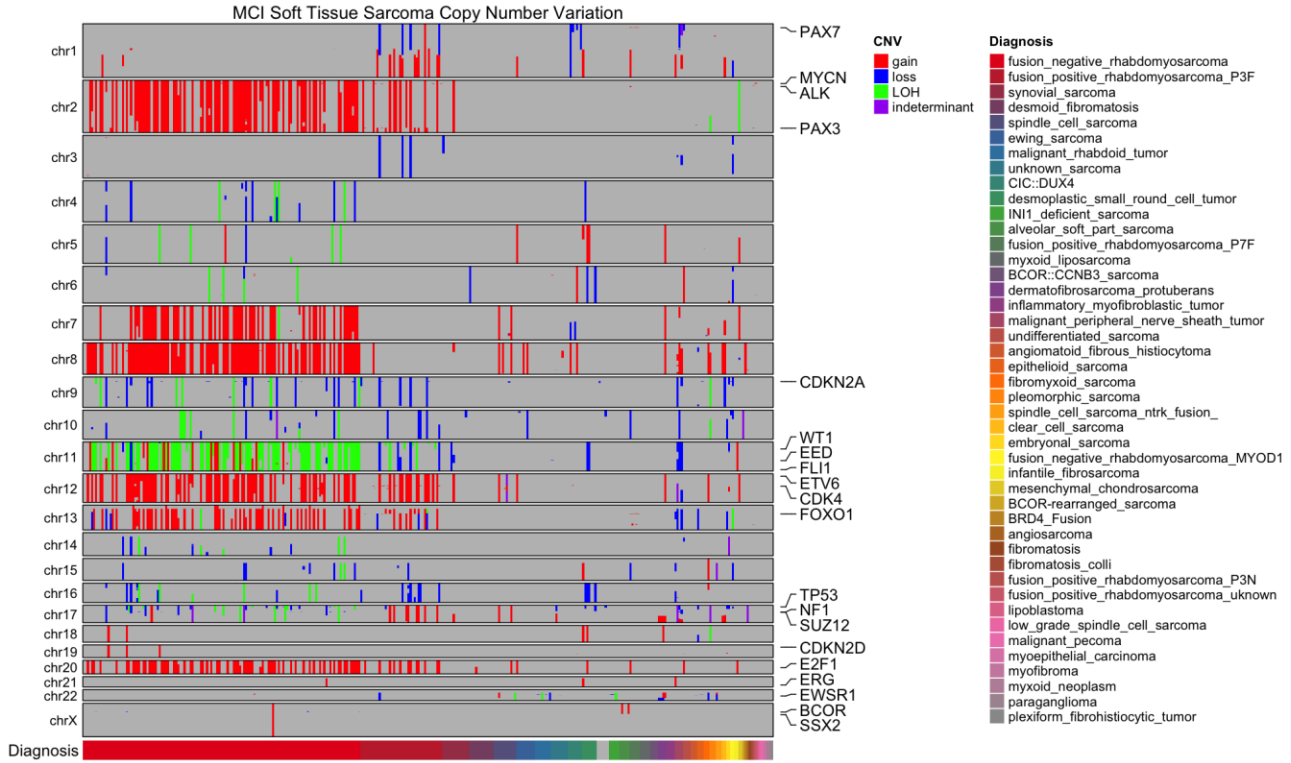
NM_020975.6(RET):c.1998G>C p.(Lys666Asn)

This individual's comparator peripheral blood specimen harbors a heterozygous missense variant in the *RET* gene, which is one of the receptor tyrosine kinases that plays a key role in cell proliferation and differentiation (PMID: 16979782). Pathogenic variants in this gene are associated with autosomal dominant disorders; medullary thyroid carcinoma (OMIM: 155240), multiple endocrine neoplasia IIA (OMIM: 171400), multiple endocrine neoplasia IIB (OMIM: 162300) and pheochromocytoma (OMIM: 171300). A germline *RET* variant has also been reported in the setting of relapsed leukemia (PMID: 30936199).

This missense variant (p.Lys666Asn) is a well-described pathogenic variant as it has been reported in the setting of medullary thyroid carcinoma in multiple unrelated individuals in the literature. *In-vitro* studies showed that this variant results in increase phosphorylation of ERK and high kinase activity (ACMG/AMP: PS4_Moderate, PS3_Moderate; PMIDs: 27673361, 15858153, 20103606). This variant is present at an extremely low frequency in gnomAD, a large-scale control population database (ACMG/AMP: PM2) and this same amino acid change resulting from a different nucleotide change (c.1998G>T) has been reported as pathogenic in literature and in ClinVar by multiple clinical laboratories (ACMG/AMP: PS1; ClinVar ID: 24932). Additionally, a different amino acid change at this same residue (Lys666Glu) has been reported as pathogenic in literature and ClinVar (ACMG/AMP: PM5; ClinVar ID: 24931). Together, and in accordance with ACMG/AMP guidelines, this variant is classified as pathogenic.

"Rhabdomyosarcoma NOS"

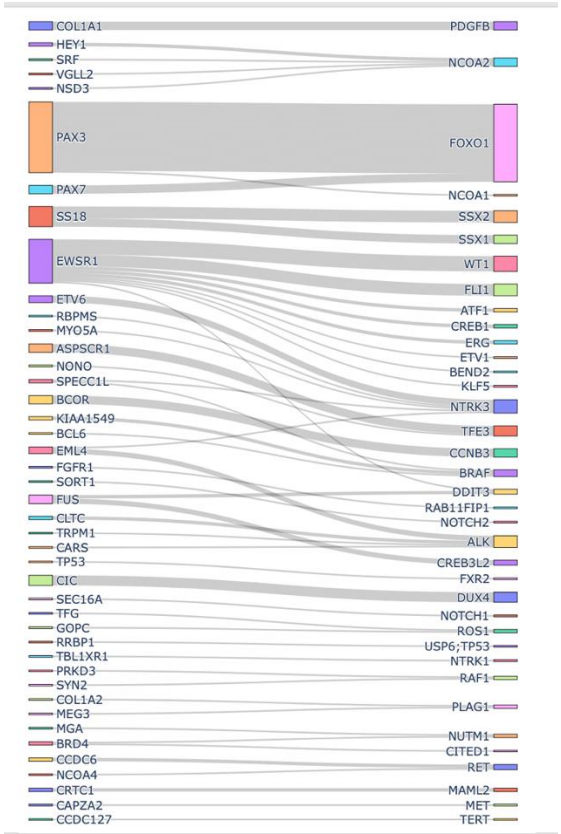
Soft Tissue Sarcoma Copy Number Summary



- Fusion Negative RMS has a very distinct CN profile
- Focal deletion and amplification of *CDKN2A/CDK4* occurs in ~20% of the population
- Multiple tumors especially patients with Li Fraumeni have evidence of chromothripsis

STS Fusion Oncogene Summary – Archer Results

Detected Fusion Oncogenes in MCI STS



"Soft tissue sarcoma NOS"

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

Gene Fusion	5' Fusion Partner	3' Fusion Partner	Classification (AMP/ASCO/CAP)
EWSR1::KLF5	*EWSR1*:NM_001163287.2 |Exon: 9| (GRCh37) chr22:29686457	*KLF5*:NM_001730.4 |exon: 2| (GRCh37) chr13:73636660	Tier I (Level B)

Targeted RT-PCR followed by subsequent Sanger sequencing confirmed the presence of this event.

"Spindle cell neoplasm"

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

Gene Fusion	5' Fusion Partner	3' Fusion Partner	Classification (AMP/ASCO/CAP)
EML4::NTRK3	*EML4:NM_019063.4* |exon: 2| (GRCh37) chr2:42472827	*NTRK3:NM_001012338.2* |exon: 14| (GRCh37) chr15:88576276	Tier I (Level A)

Targeted RT-PCR followed by subsequent Sanger sequencing confirmed the presence of this event.

Resources

Access CCDI resources through the CCDI Hub

CCDI Data Access:

- Access guide: https://ccdi.cancer.gov/static/media/CCDI_Usage_Instructions_Nov2024_v2.5.0.69ea3cd5.pdf
- Information about CCDI study data: datacatalog.ccdi.cancer.gov/resource/CCDI
- Cancer Data Standards Registry and Repository: cadsr.cancer.gov/onedata/Home.jsp

Data Models and Federation API:

- CCDI: github.com/CBIIT/ccdi-model
- CCDI template: github.com/CBIIT/Model_to_Submission
- Childhood Cancer Clinical Data Commons: github.com/CBIIT/c3dc-model
- CCDI Federation API: cbiit.github.io/ccdi-federation-api

Team Members

- Patients and Families
- Advocacy Community
- Local treatment, pathology and surgical teams
- Children's Oncology Group Project EveryChild Team
- Children's Oncology Group Soft Tissue Sarcoma Committee
- Nationwide Biopathology Center Team
- Institute for Genomic Medicine Team
- NCI Childhood Cancer Data Initiative Team

Yassmine Akkari, Institute for Genomic Medicine
Michael A. Arnold, University of Colorado
Natalie Bir, Biopathology Center
Catherine E. Cottrell, Institute for Genomic Medicine
Avery Funkhouser, National Cancer Institute
Doug Hawkins, Seattle Childrens Hospital
Subhashini Jagu, National Cancer Institute
Javed Khan, National Cancer Institute
Corinne M. Linardic, Duke School of Medicine
Elaine R. Mardis, Institute for Genomic Medicine
Mariam Mathew, Institute for Genomic Medicine
Yvonne Moyer, Biopathology Center
Sapna Oberoi, University of Manitoba
Nilsa C. Ramirez, Biopathology Center
Greg Reaman, National Cancer Institute
Erin R. Rudzinski, Seattle Childrens Hospital
Kathleen Schieffer, Institute for Genomic Medicine
Malcolm Smith, National Cancer Institute
Shountea Stover, Biopathology Center
Diana Thomas, Biopathology Center
Rajkumar Venkatramani, Texas Children's Hospital
Wei Xue, University of Florida
Aaron Weiss, Maine Medical Center
Greg Wheeler, Institute for Genomic Medicine

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