PRISSMM Data Model

Katherine A. Janeway, MD, MMSc

NCCR Cancer Center Supplemental Data Summit

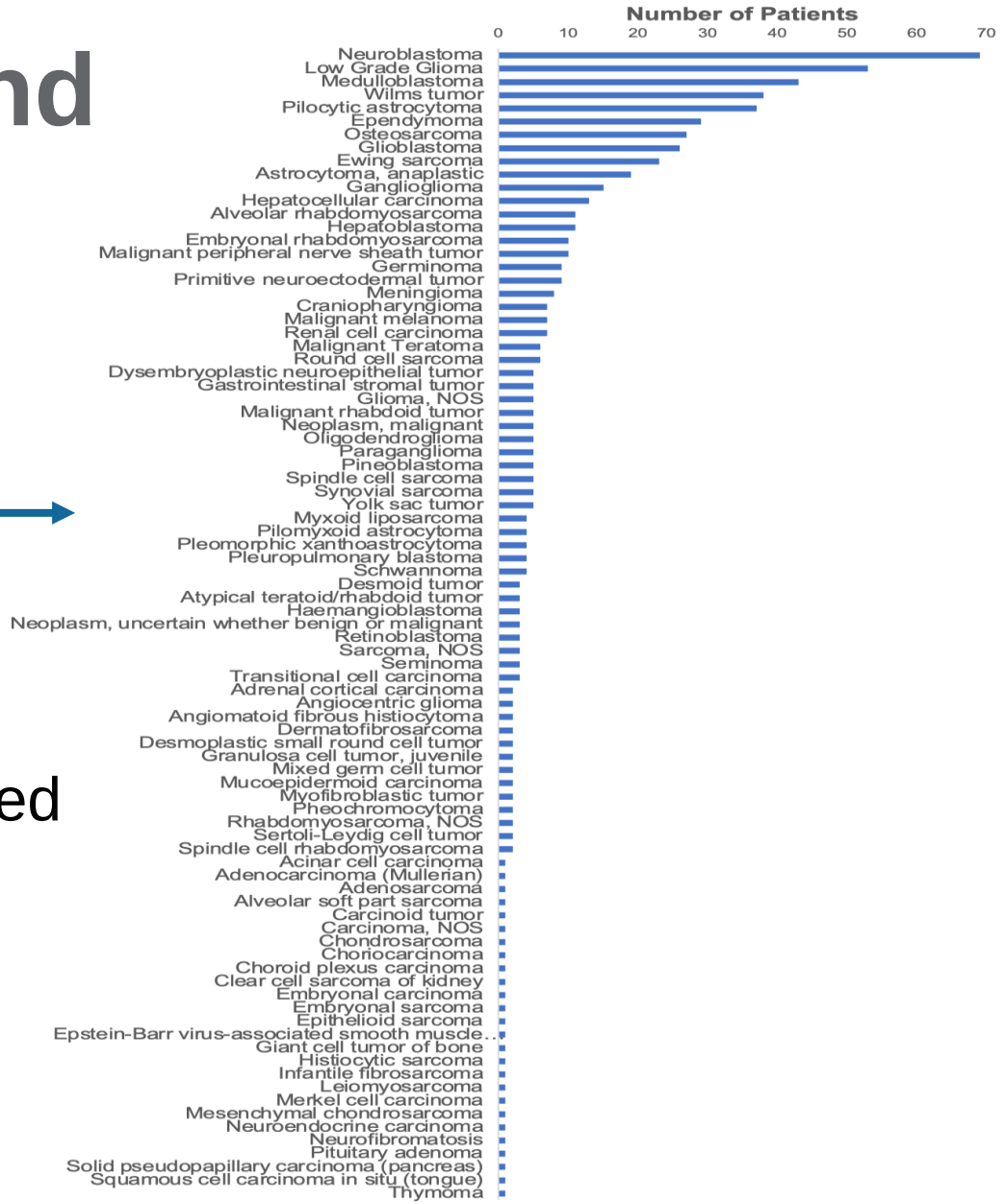
February 8, 2021

Dana-Farber/Boston Children's Cancer and Blood Disorders Center

Background

PROFILE Cancer Research Study

- > 30,000 patient precision oncology cohort study
  - 1,000 pediatric patients
- Intervention: panel sequencing
- Example outcomes of interest
  - Identify extra-ordinary responses to targeted therapy matched to molecular aberrations
  - Identify molecular subgroups with prognostic impact
- Requires sharing longitudinal treatment and response data from the EMR



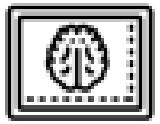
Characteristic	Clinical Trial Data	Real World EMR Data	PRISSMM Solution
Treatment and duration	Defined by trial	Variability in schedule & drugs	Definition treatment regimen
Response endpoint	Standards	No standards Inability to use RECIST	Creates standard
Data collection	Prospective	Retrospective	Consistent directives Methods facilitate QC
Proportion of cancer journeys	Minority	Majority	Capture each treatment course
Future goal	Share & harmonize	Natural language processing	Provides gold standard for training dataset

- Developed by consensus with AACR Project GENIE institutions
- AACR Project GENIE Biopharma collaboration
  - Will use PRISSMM to annotate 8,000 adult cancers with structured phenotype data
- Given how rare pediatric cancer are, real world data will be essential

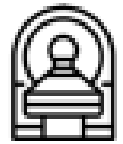
  → Pediatric adaptation PRISSMM

PRISSMM™:
A Taxonomy for Defining Cancer Outcomes

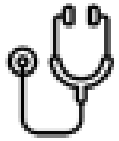
Pathologic evidence of locoregional or distant evidence of tumor



Radiographic evidence of locoregional recurrent or persistent tumor

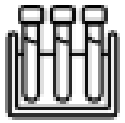


Imaging evidence of distant/disseminated tumor beyond the primary site



Symptoms of tumor on physical exam or symptoms that can be attributed to tumor

Signs of cancer on physical exam or symptoms that can be attributed to tumor



Tumor **M**arker evidence of persistent or recurrent tumor

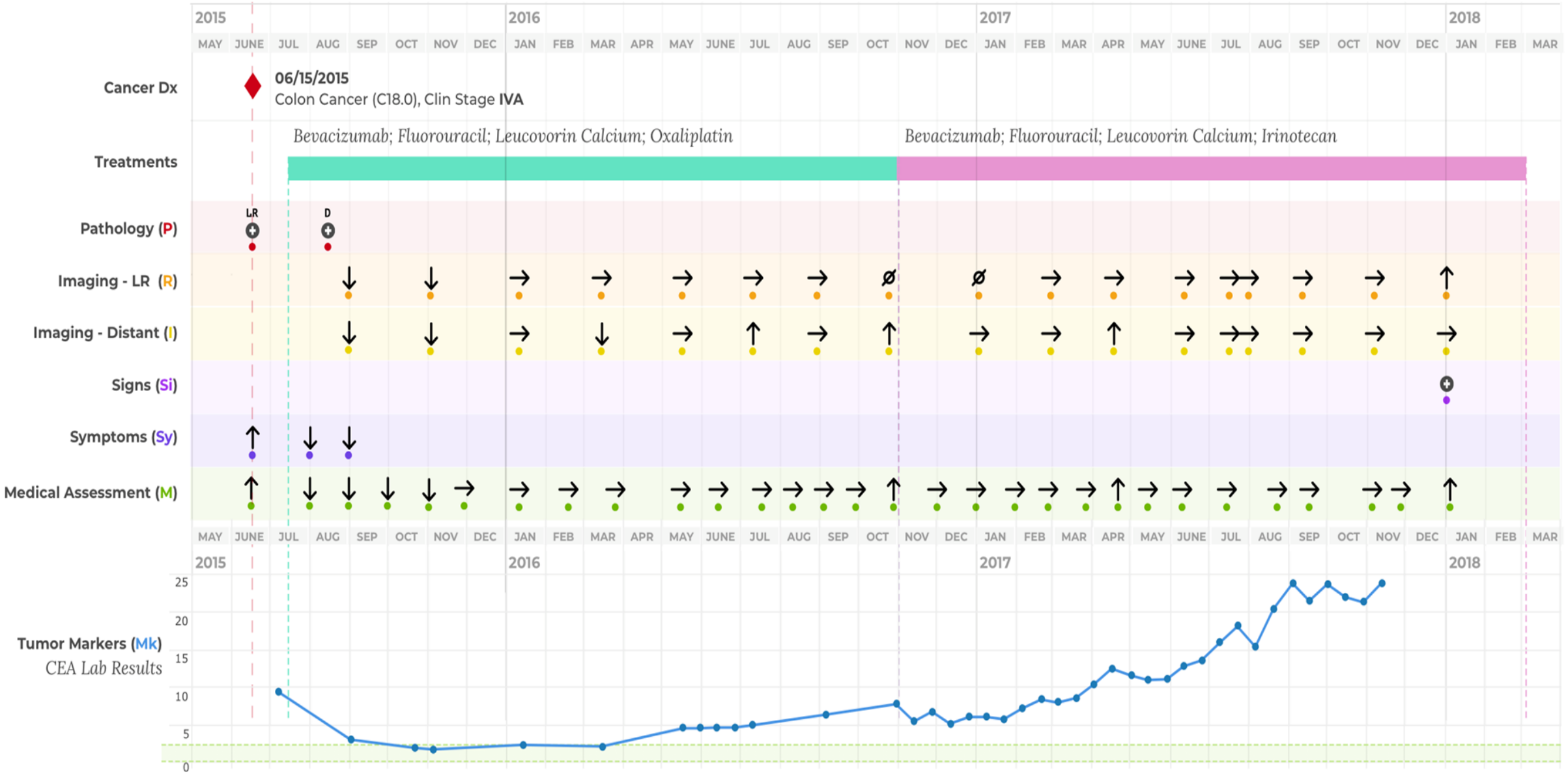
Oncology **M**edical Provider assessment

PRISSMM Establishes Standard Directives for Curation of Cancer Treatment Outcomes from EHRs

PRISSMM 1.0	Generic Framework and Data Provenance
P	Pathology reports
R	Imaging reports of primary site
I	Imaging reports of distant sites
S_i	Signs on physical exam in oncologist notes
S_y	Symptoms noted in HPI, impression, interval history
M_k	Biomarkers
M	Impression/plan from clinician notes (oncology MD, RN, RNP, PA)

Taxonomy and Standard Nomenclature facilitate interpretation and communication about Real World Data

Cancer Status: ##	Cancer Treatment Response: ##
Time anchor is diagnosis: # months	Time anchor is user specified: # months
+ means present	↘ means responding
- means absent	→ means no change/stable
x means not evaluated	↗ means not responding
$P_{+}R_{-}I_{+}S_{-}S_{-}M_{x}M_{+}$: 36	$P_{+}R_{x}I_{\rightarrow}S_{-}S_{-}M_{x}M_{\searrow}$: 3
At 36 months from diagnosis, the patient has pathologically confirmed tumor evident based on imaging and clinician's assessment	Compared to 3 months prior, the patient's tumor is stable on distant imaging and responding according to the clinician's assessment.


2015
2016
2017
2018
MAY JUNE JUL AUG SEP OCT NOV DEC JAN FEB MAR APR MAY JUNE JUL AUG SEP OCT NOV DEC JAN FEB MAR APR MAY JUNE JUL AUG SEP OCT NOV DEC JAN FEB MAR
Cancer Dx
06/15/2015
Colon Cancer (C18.0), Clin Stage IVA
Treatments
Bevacizumab; Fluorouracil; Leucovorin Calcium; Oxaliplatin
Bevacizumab; Fluorouracil; Leucovorin Calcium; Irinotecan
Pathology (P)
LR
D
Imaging - LR (R)
Imaging - Distant (I)
Signs (Si)
Symptoms (Sy)
Medical Assessment (M)
Tumor Markers (Mk)
CEA Lab Results
25
20
15
10
5
0

PRISSMM for Natural Language Processing

Signs and Symptoms

- Trained curators curate the first outpatient medical oncology physician note for each month, starting from the date of cancer diagnosis. If no MD available then NP. Review only the assessment and plan.
  - Cancer present Y/N
  - If Y, “improving/responding”, “stable/no change”, “mixed”, “progressing/worsening/enlarging”, “not stated/indeterminate”
- To evaluate the reproducibility of human curation, a random 10% subset of curated records was selected for repeat curation and assessment of inter-rater variability

PISSMM for Natural Language Processing

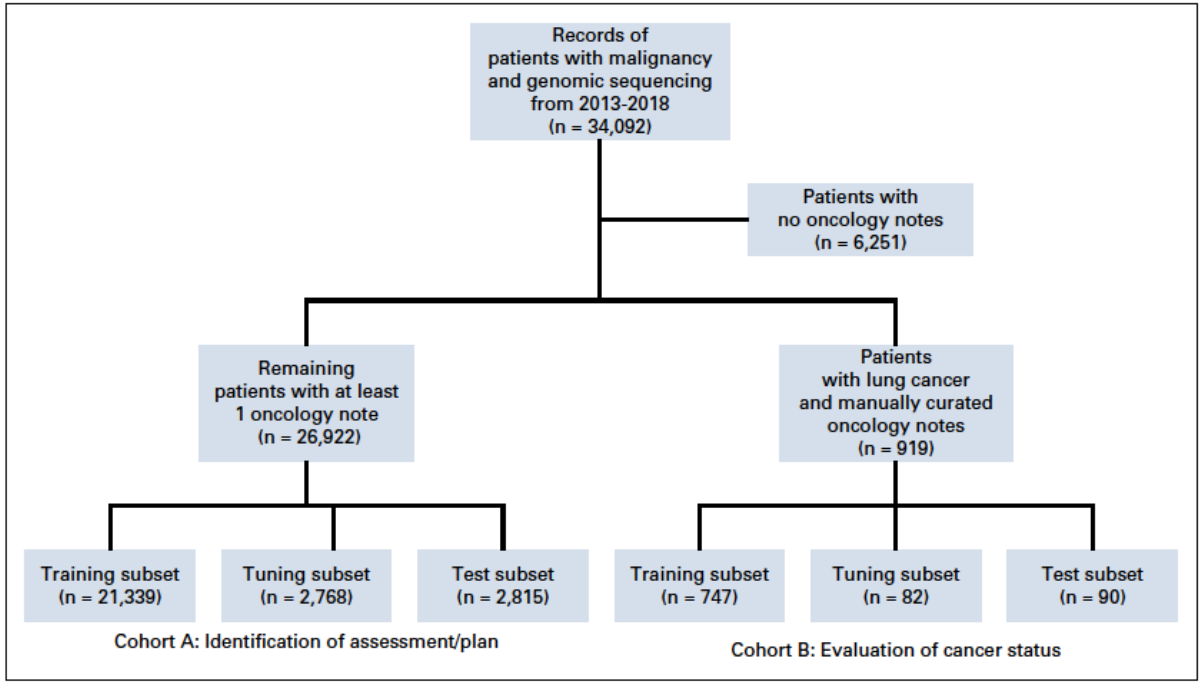


TABLE 3. Performance Characteristics in Cohort B for NLP Model Prediction of Cancer Status

	Tuning Set				Test Set			
Cancer Status	AUROC	Prevalence	AUPRC	Best F1	AUROC	Prevalence	AUPRC	Best F1
Cancer present	0.94	0.80	0.98	0.95	0.94	0.77	0.97	0.94
Progression/worsening	0.89	0.20	0.73	0.67	0.86	0.20	0.65	0.62
Response/improvement	0.90	0.12	0.61	0.61	0.90	0.12	0.57	0.64

TABLE 5. Association Between Curated Outcomes and Mortality

	Hazard Ratio for Mortality (95% CI)	
Cancer Status	Manual Curation	NLP Models
Progression/worsening	2.93 (2.33 to 3.67)	2.49 (2.00 to 3.09)
Response/improvement	0.70 (0.47 to 1.03)	0.45 (0.29 to 0.67)

Pediatric Adaptation PRISSMM

Selected pediatric cancers

With 2 other pediatric cancer centers

- UCSF and MSKCC
- Select data elements, sources

Incorporated existing or emerging data standards

- Toronto staging guidelines
- PCDC for overlapping diseases (neuroblastoma)

Development of paediatric non-stage prognosticator guidelines for population-based cancer registries and updates to the 2014 Toronto Paediatric Cancer Stage Guidelines

Sumit Gupta, Joanne Aitken*, Ute Bartels, Nickhill Bhakta, Mihaela Bucurenci, James D Brierley, Beatriz De Camargo, Eric Chokunonga, Jessica Clymer, Dana Coza, Chris Fraser, Soad Fuentes-Alabi, Gemma Gatta, Thomas Gross, Zsuzsanna Jakab, Betsy Kohler, Tezer Kutluk, Florencia Moreno, Kayo Nakata, Sari Nur, D M Parkin, Lynne Penberthy, Jason Pole, Jenny N Poynter, Kathy Pritchard-Jones, Oscar Ramirez, Lorna Renner, Eva Steliarova-Foucher, Michael Sullivan, Rajaraman Swaminathan, Liesbet Van Eycken, Tushar Vora, A L Frazier*

Lancet Oncology, 2020

Pediatric Adaptation PRISSMM

Osteosarcoma – Janeway, Shukla, Sweet-Cordero

Pathology:
- Tumor Necrosis, Margins from Local Control procedures, Tumor Grade

Staging:
- Disease specific definitions for metastatic disease

Prognostic Factors:
- Size of primary tumor

Ewing Sarcoma – Janeway, Shukla, Sweet-Cordero

Pathology:
- Fusions Identified from Clinical Testing, CD 99 Expression, Tumor Necrosis, Margins from Local Control procedures

Staging:
- Disease specific definitions for metastatic disease

Prognostic Factors:
- Size of primary tumor

Wilms Tumor– Mullen, Ortiz

Diagnosis:
- Nephroblastomatosis and Nephrogenic Rests, Number and Size of lesions

Pathology:
- Histology (e.g. Anaplasia)

Staging:
- Kidney and overall

Neuroblastoma– Shusterman

Staging:
- INRG Staging

Prognostic Factors:
- COG Risk Classification
- MYNC Status and Ploidy
- Revised INPC Prognostic Group
- Mitosis Karyorrhexis Index (MKI)

9 of11 added fields equivalent to PCDC

DFCI PRISSMM DATA

250 Patients

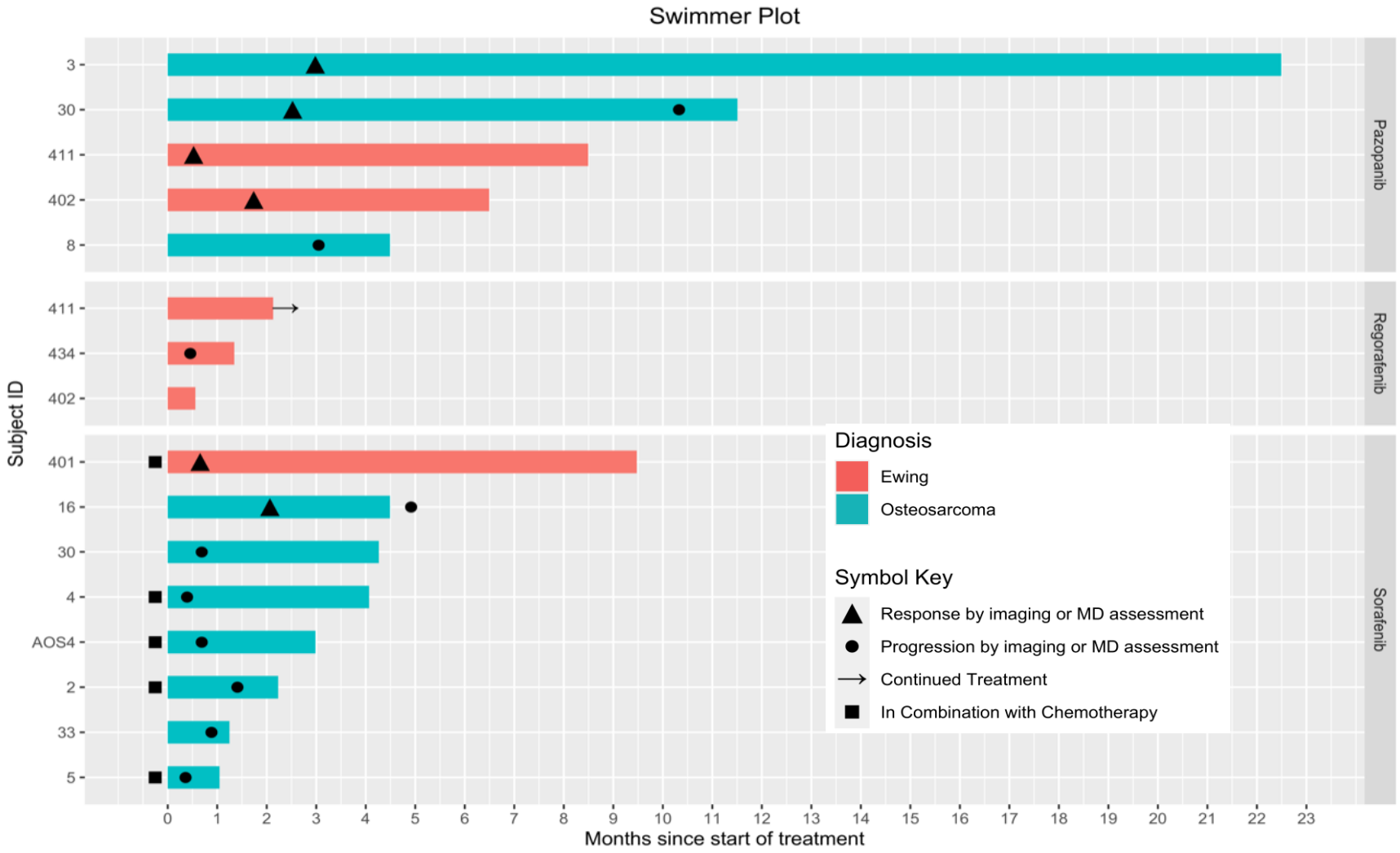
4 Pediatric solid tumors (OS, EWS, WT, NBL)

Average **18** curated imaging reports per patient (range 1-87)

Average **4** curated pathology reports per patient (range 0-24)

Median follow-up **27** months (range 0-263)

Real World Use Multi-TKIs in Bone Sarcomas



Acknowledgements

- Debroah Schrag
- Eva Lepisto
- Pediatric Project Managers
  - Seth Sheffler-Collins, AACR
  - Abbie Ward, DFCI
  - Stephanie Suser, MSKCC
  - Nancy Bouvier, MSKCC
  - Michelle Turski, USCF
- Research Coordinators
  - Michaela O'Connell, DFCI
  - Sidney Benich, DFCI
- Others
  - John Philips, MSKCC

Dana-Farber / Boston Children's Harvard Cancer Center, MA State Registry NCCR Team

- Kira Bona
- Kevin Copson
- Sarah Cronholm
- Suzanne Forrest
- Lindsay Frazier
- Susan Gershman
- Coleman Lindsley
- Wendy London
- Jenny Mack
- Yana Pikman
- Madhu Sridharan
- Kathy Warren
- Aaron Yeo

Disease experts

- Neal Shukla, MSKCC
- Alejandro Sweet-Cordero, UCSF
- Elizabeth Mullen, DFCI
- Suzy Shusterman, DFCI
- Michael Ortiz, MSKCC

Shared data standards

- Sam Volchhenboum