# Transcriptomic Analysis of COG Pediatric AML Data

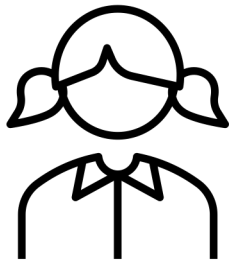
A Prequel

Jenea Adams

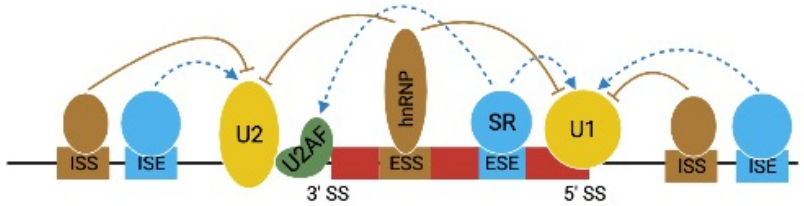
Xing Lab Round Table

September 17, 2021

# Establishing a multiomic perspective of acute myeloid leukemia (AML) presentation by age group



AML treatments are harsh on young bodies
+
AML treatments are based on adult patient data
∴
New pediatric AML therapies are needed




Splicing + Molecular pathways =

**Targetable avenues of disease progression**

Aim 1: Improve analysis of splicing in large, heterogeneous RNA-seq datasets

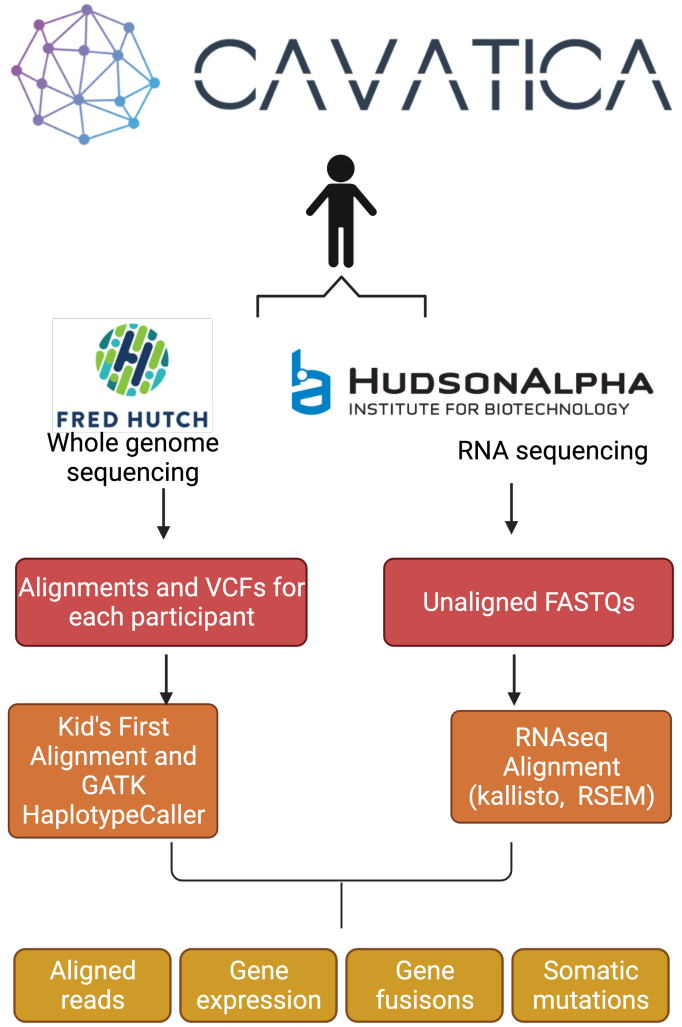
Aim 2: Discover age-specific, pathway-dependent alternative splicing patters in pediatric AML RNA-seq data

Created with BioRender.com

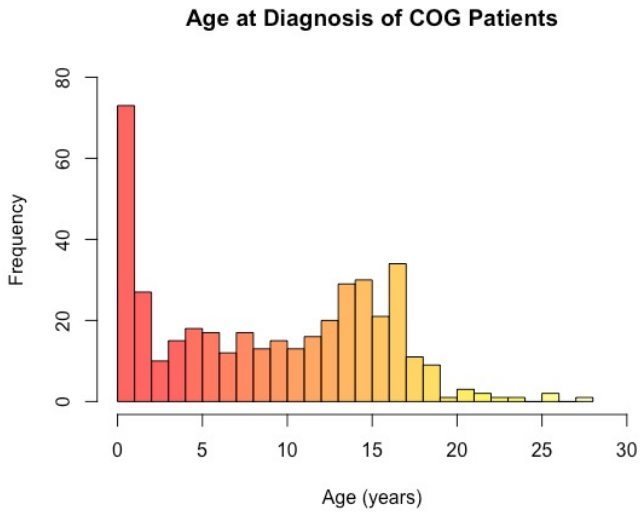
# CAVATICA: CHOP data-analysis platform of raw and harmonized multiomic data

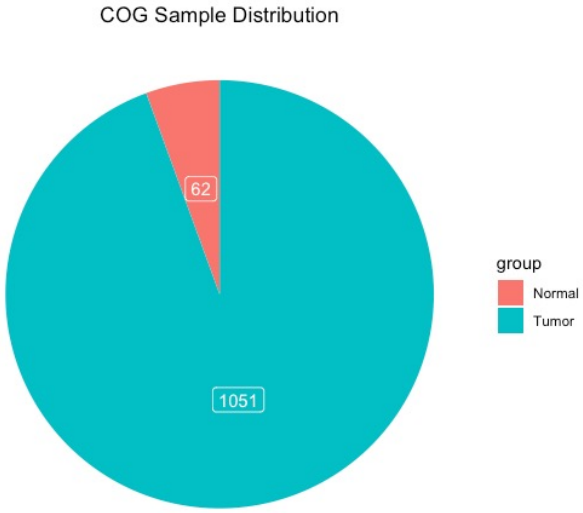
**Overview of data used for this analysis**

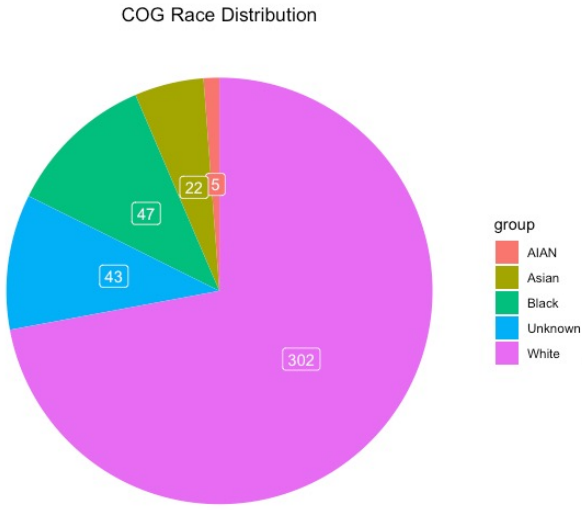
- De-novo AML, DS-AML, APL-AML
- Data from Children's Oncology Group Clinical Trial (No. AAML1031)
- 1,113 RNA-seq files (aligned and quantified)
  - Both kallisto and RSEM were used
  - Gene fusions also quantified
- 408 whole-genome sequencing (WGS) files

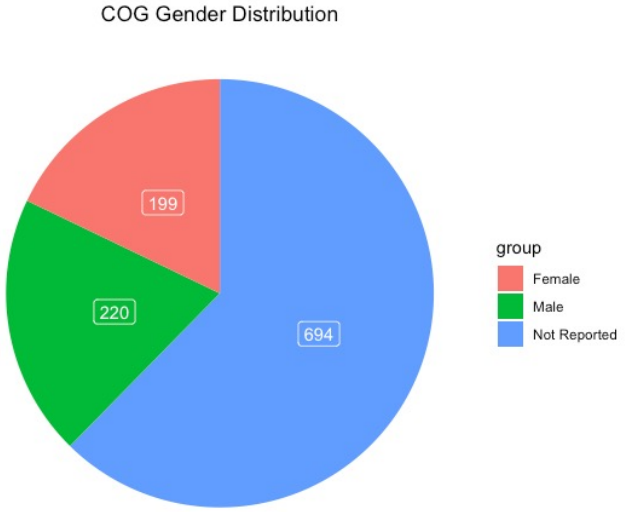


# CAVATICA allows easy metadata extraction



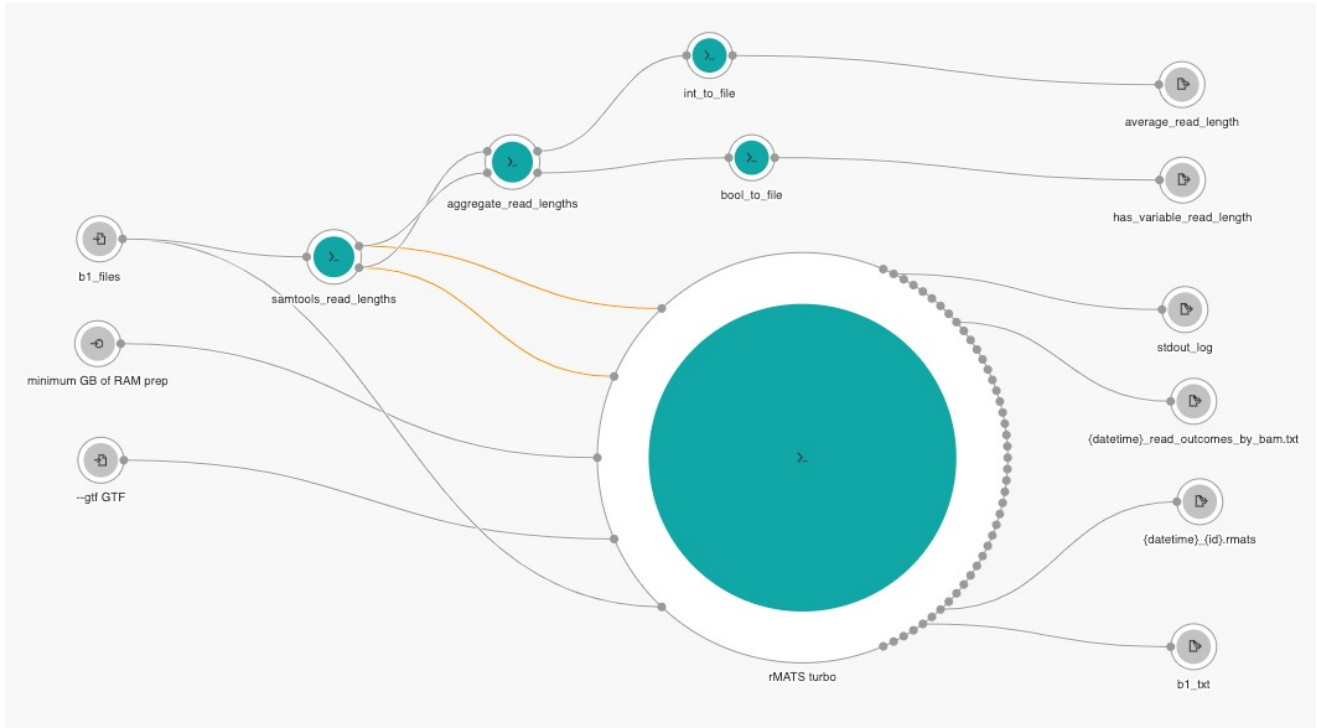



# rMATS on the cloud

- Cost effective
- Shareable, reproducible
- Scalable (as more data are uploaded)




ULTRA-FAST SPLICING QUANTIFICATION

rMATS - TURBO

CAVATICA APP

- Apps = collections of tools and workflows
- Housed within a “project”

# rMATS-turbo analysis (so far) provides low-cost, fast output

Time

| # BAMs | CAVATICA | | HPC | |
|---|---|---|---|---|
| | **-- prep** | **--post** | **--prep** | **--post** |
| 1 | 30 min | 5 min | | |
| 200 | 2 h, 12 min | 24 min | | |
| 1,113 | 9 h, 49 min | TBD | | |

Cost

| # BAMs | CAVATICA | |
|---|---|---|
| | **-- prep** | **--post** |
| 1 | $0.14 | $0.02 |
| 200 | $34.07 | $0.13 |
| 1,113 | $199.35 | TBD |

# In Progress

- Documenting rMATS-turbo cloud use
- Documenting multi-use API functions for handling large amounts of output data
- Analysis: Sourcing metadata
  - Treatment history
  - Use of whole genome seq data for ancestry association with molecular subtypes and disease progression
  - RNA-seq data for gene expression analysis of molecular subtypes