

Discovering age-specific multi-omic patterns in heterogenous pediatric clinical datasets

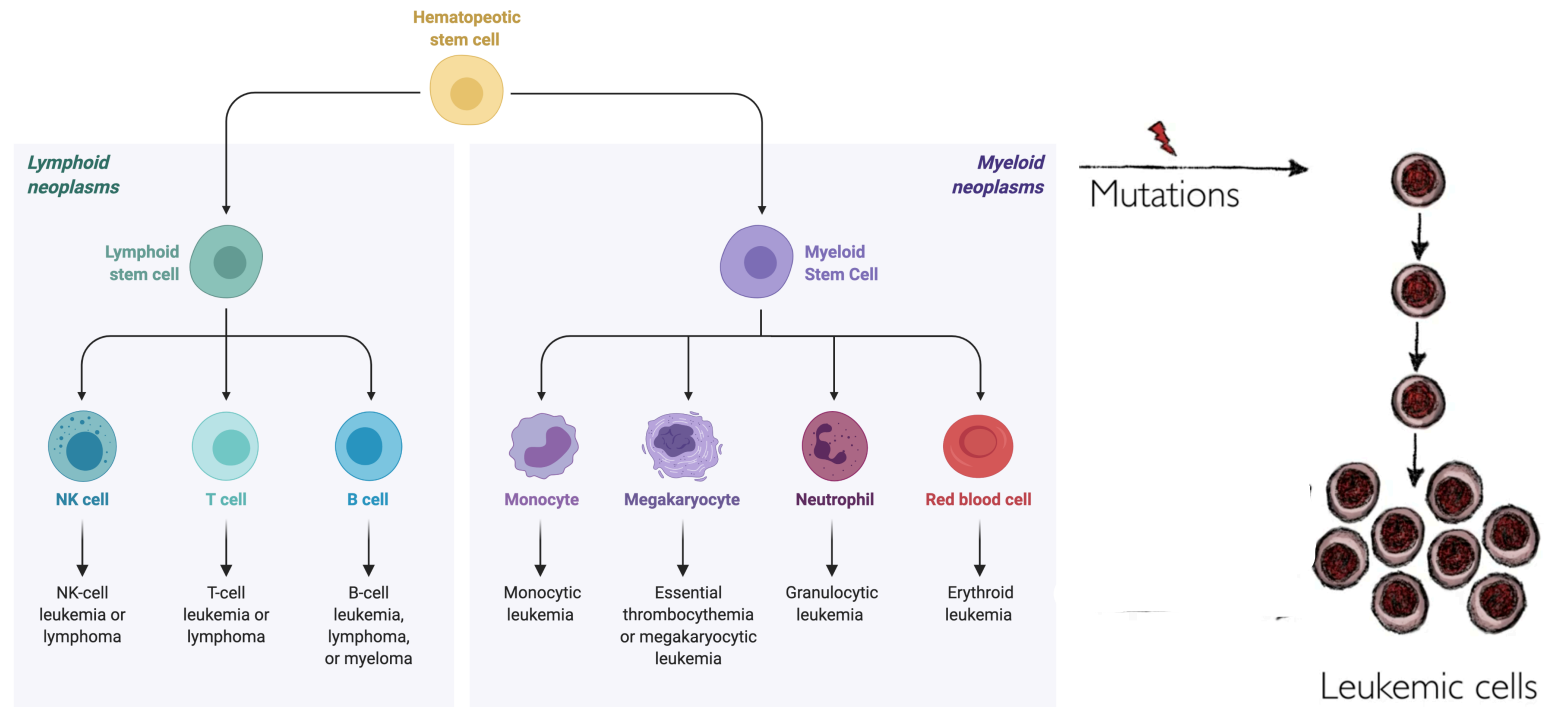
Jenea Adams

Xing Lab Roundtable

May 11, 2021

Acute myeloid leukemia (AML) is the most fatal of childhood cancers with no good treatments

- Affects 25% of children with leukemia
- Accumulation of immature myeloid cells in bone marrow
- Lacks treatment options comparable to ALL
- AML treatments needed that preserve healthy myeloid cells

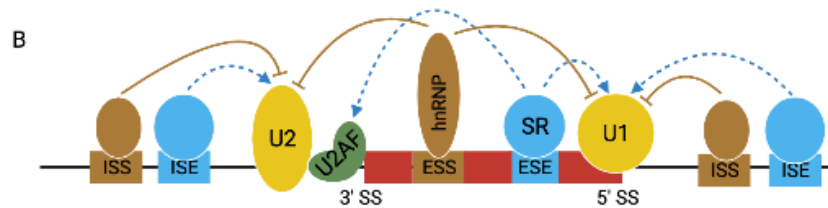


van Galen *et al.*, 2019, BioRender.com

Establishing a multi-omic perspective on the presentation of AML by age group



AML treatments
are harsh on
young bodies →
new therapies
needed



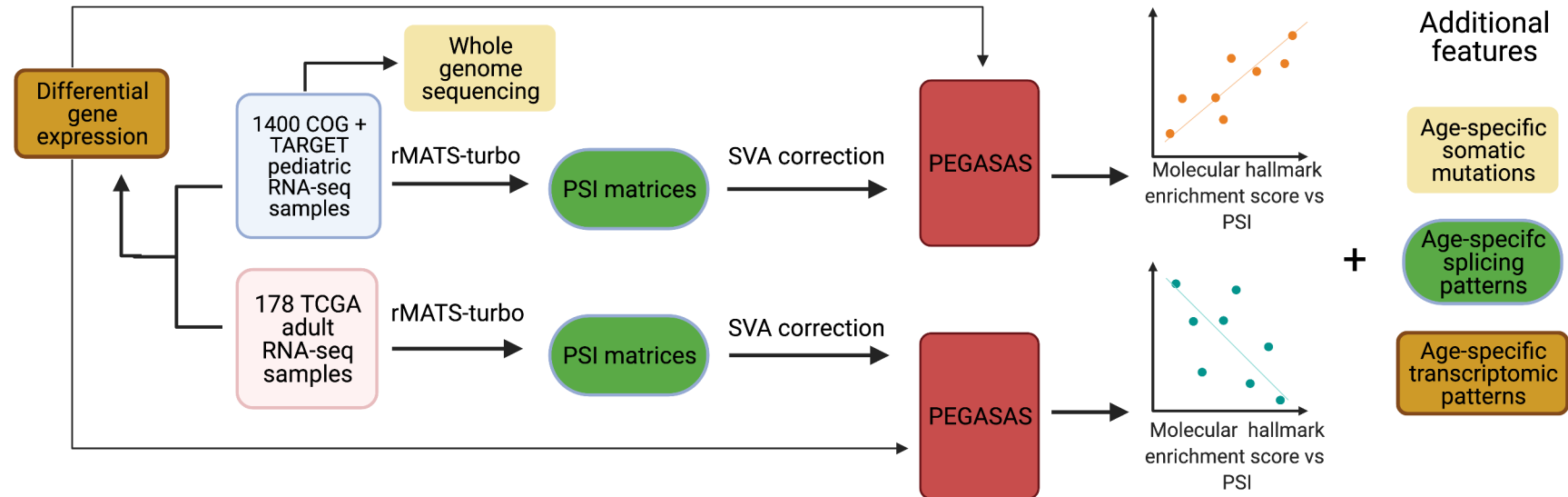
Splicing =
targetable avenue
of disease
progression in
AML

Aim 1: to improve the analysis of
splicing in large heterogeneous
RNA-seq datasets

Aim 2: to discover age-specific,
pathway-dependent alternative splicing
patterns in pediatric AML RNA-seq data

Does splicing provide a parallel mechanism of regulation in pediatric AML?

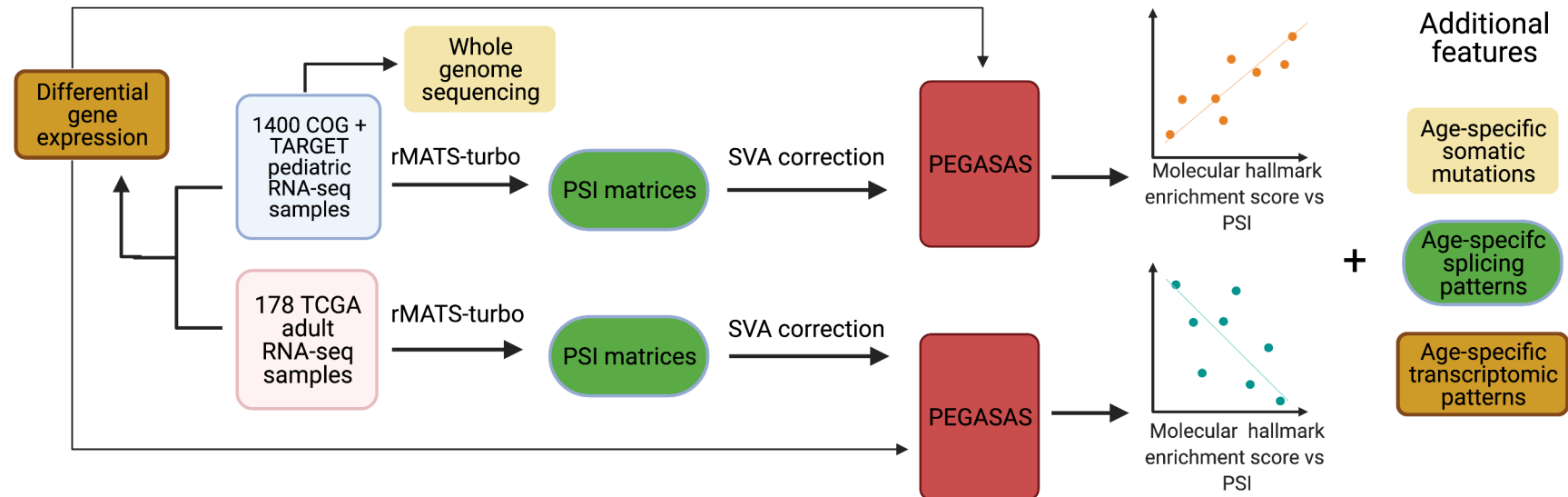
- Low mutation burden in pediatric cancer → different genomic signatures?
 - *DNMT3A*, *NPM1*, and certain structural variations, are in low frequency or even absent in pediatric AML cases
- Multi-omic analysis on large clinical datasets



Does splicing provide a parallel mechanism of regulation in pediatric AML?

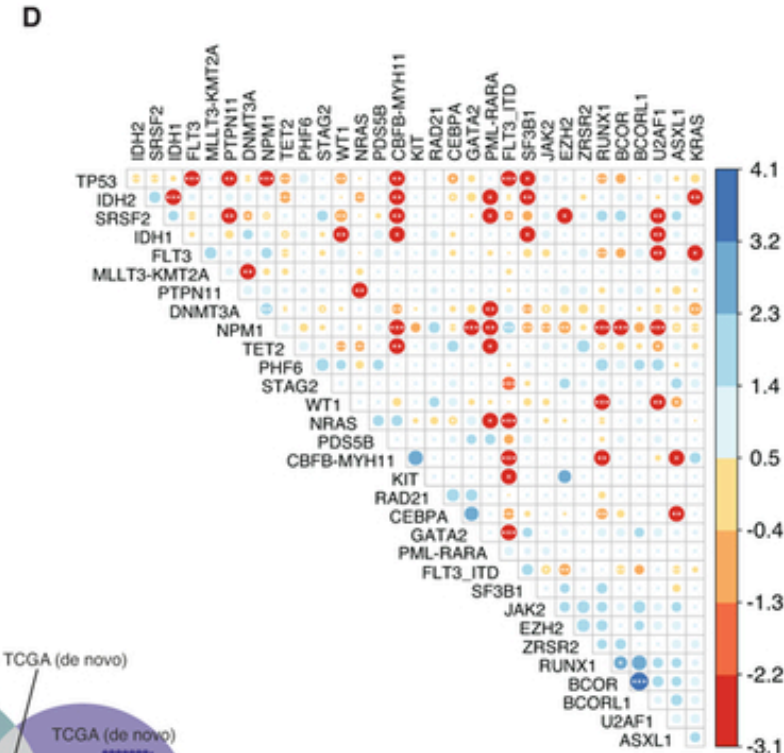
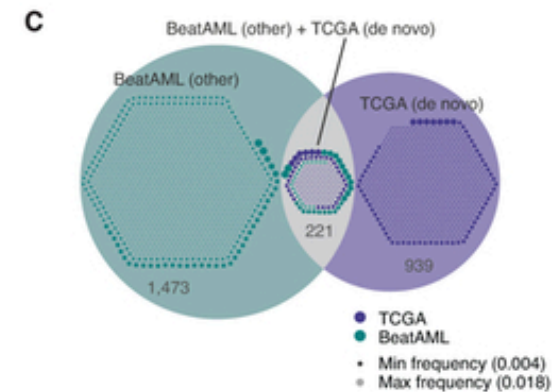
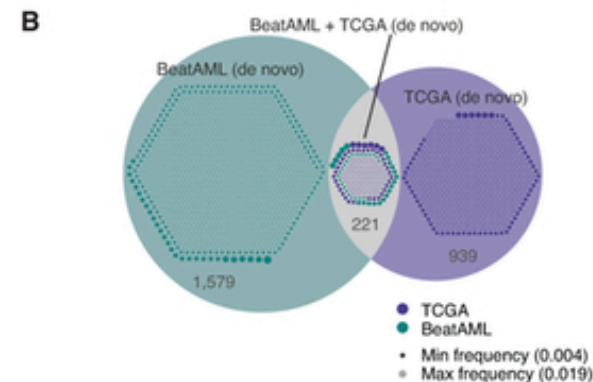
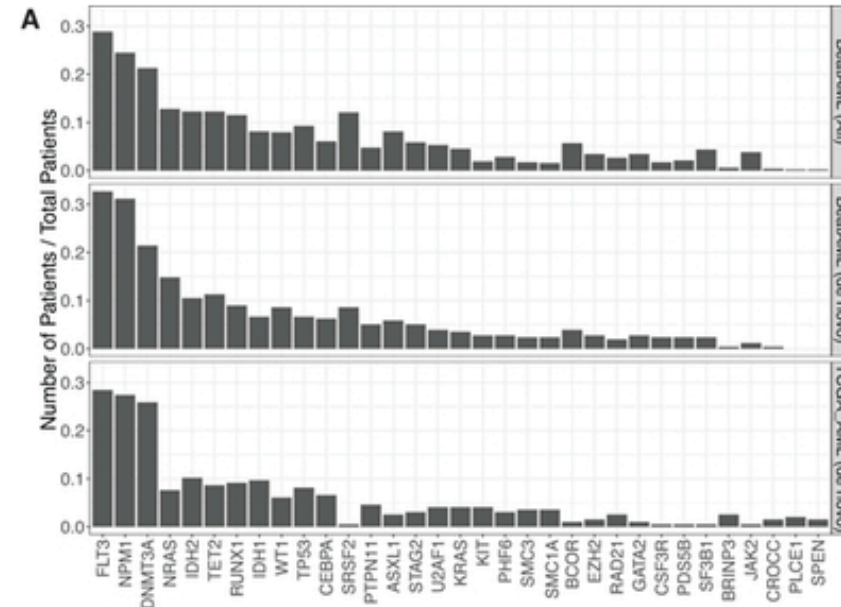
Research Design and Methods

- 1200 Children's Oncology Group Clinical Trial (AAML1031) RNA-seq + Whole Genome sequencing + >200 TARGET samples
- **178 TCGA AML**
- Control: GTEx Normal Blood
- Compare molecular hallmark enrichment score vs PSI correlations between pediatric and adult cases



Beat AML vs TCGA

- < 200 TCGA samples
- >600 Beat AML samples
- Beat AML has more data enriched with splicing factor mutations
- Beat AML has patient-matched WGS data

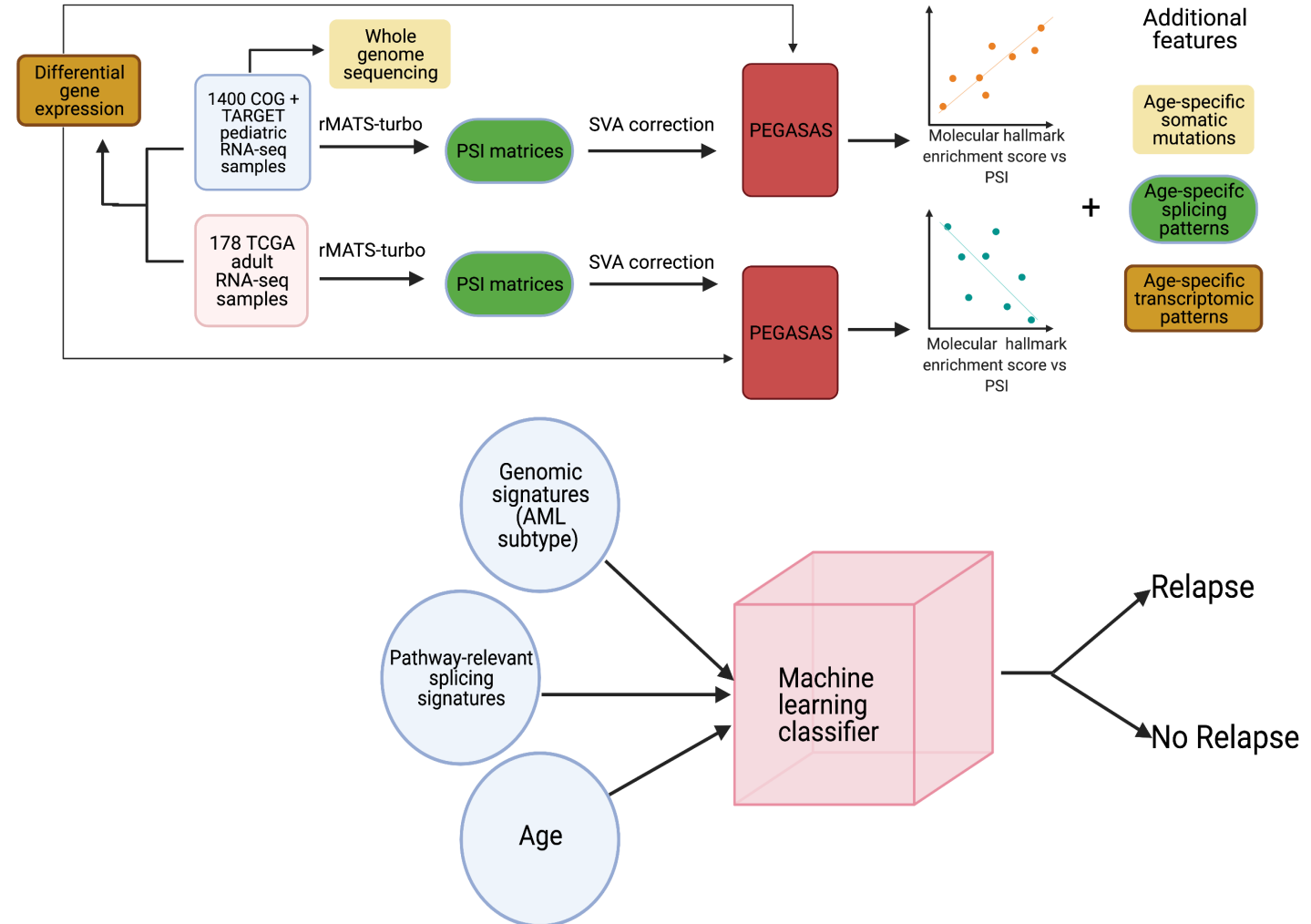


Looking ahead to true data integration

Research Design and Methods

- Data integration
 - Logistic regression, random forest, and SVM
 - 5-fold CV, regularization, hyperparameter sweep (AutoML?)
 - Compare ROC AUC curves

Is this enough?



Looking ahead to true data integration

- Current methods for data integration on this front?
- Is the goal data integration or feature selection?
- How is this evaluated?
- Greater potential for making better use of this large-scale data?

