



Harmonizing Transcriptomic Data To Discover Clinically Relevant Overexpressed Genes In Pediatric Cancer Patients

By: A. Geoffrey Lyle
September 13th, 2021

UNIVERSITY OF CALIFORNIA
SANTA CRUZ | Genomics
Institute



Outline

1. Treehouse Childhood Cancer Initiative
2. RNA-Seq in the Clinic
3. Case Analysis Process
4. Publicly Available Resources
 - Treehouse Compendia
 - UCSC TumorMap
 - UCSC Xena Viewer

Treehouse Childhood Cancer Initiative

- Started in 2015 at University of California, Santa Cruz by Olena Morozova Vaske under the guidance of David Haussler.



Olena Morozova Vaske, PhD, CGMBS, DABMGG
Colligan Presidential Chair in Pediatric Genomics



David Haussler, PhD
HHMI Investigator



Treehouse Childhood Cancer Initiative

To defeat every child's cancer
by harnessing and sharing
the world's data





Treehouse Childhood Cancer Initiative

- Initial pilot project consortium was created to determine the feasibility of performing comparative RNA-Seq analysis for prospective clinical patients.

Consortium Members



UNIVERSITY OF CALIFORNIA
SANTA CRUZ



PACIFIC PEDIATRIC
NEURO-ONCOLOGY
CONSORTIUM



Stanford
MEDICINE



Treehouse Childhood Cancer Initiative

- Initial pilot project consortium was created to determine the feasibility of performing comparative RNA-Seq analysis for prospective clinical patients.
 - Initial funding provided by California Initiative to Advance Precision Medicine (CIAPM). Extended to a 5 year project supported by the St. Baldrick's Foundation.



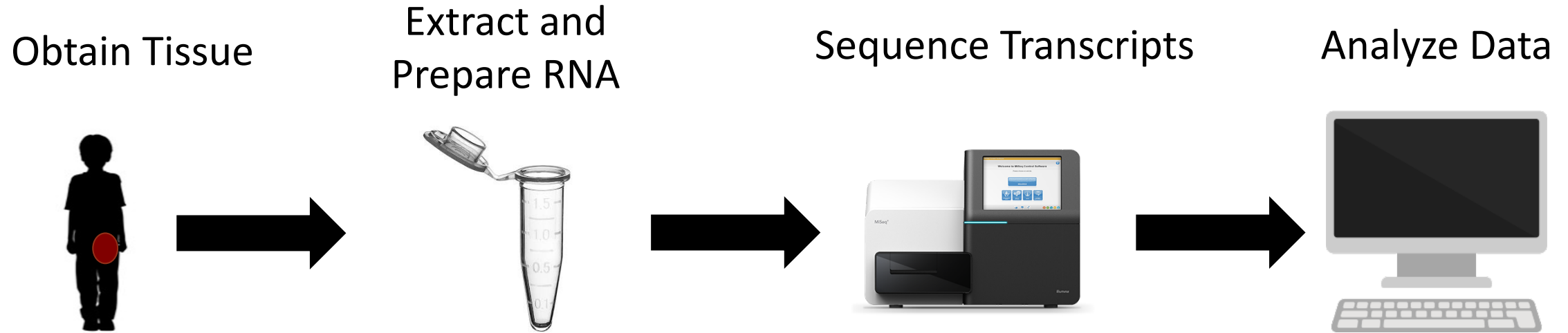
California Initiative to Advance
Precision Medicine



St. Baldrick's
FOUNDATION

Conquer Childhood Cancers

Generating RNA-Seq Data



Obtain Tissue

Extract and Prepare RNA

Sequence Transcripts

Analyze Data

Tumor biopsy or resection from patient

- Flash Frozen
- Formalin Fixed Paraffin Embedded (FFPE)

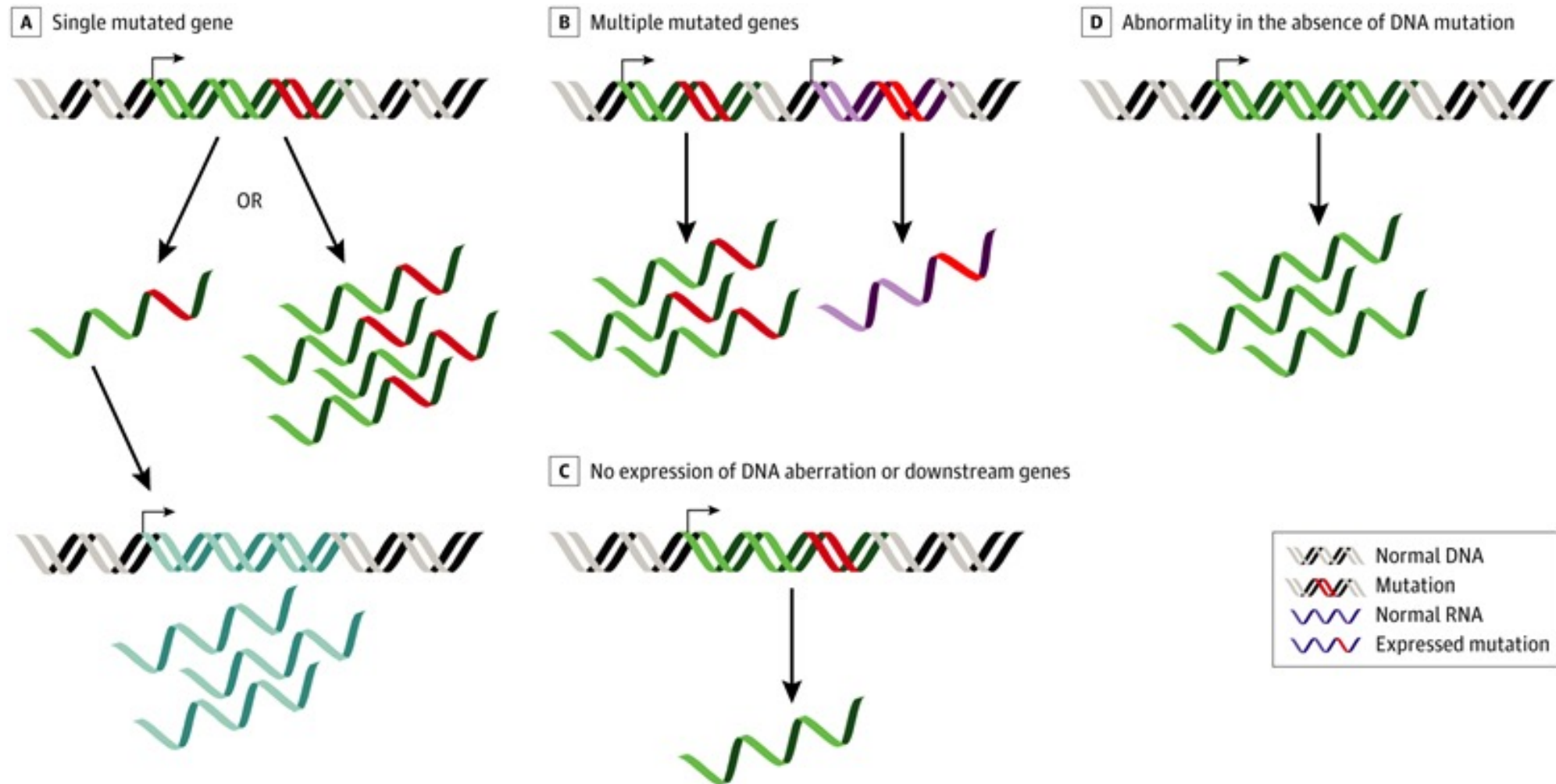
RNA must be extracted from patient sample and enriched before sequencing

- Polyadenylation (PolyA)
- Ribosomal RNA Depletion (RiboD)
- Exome Capture

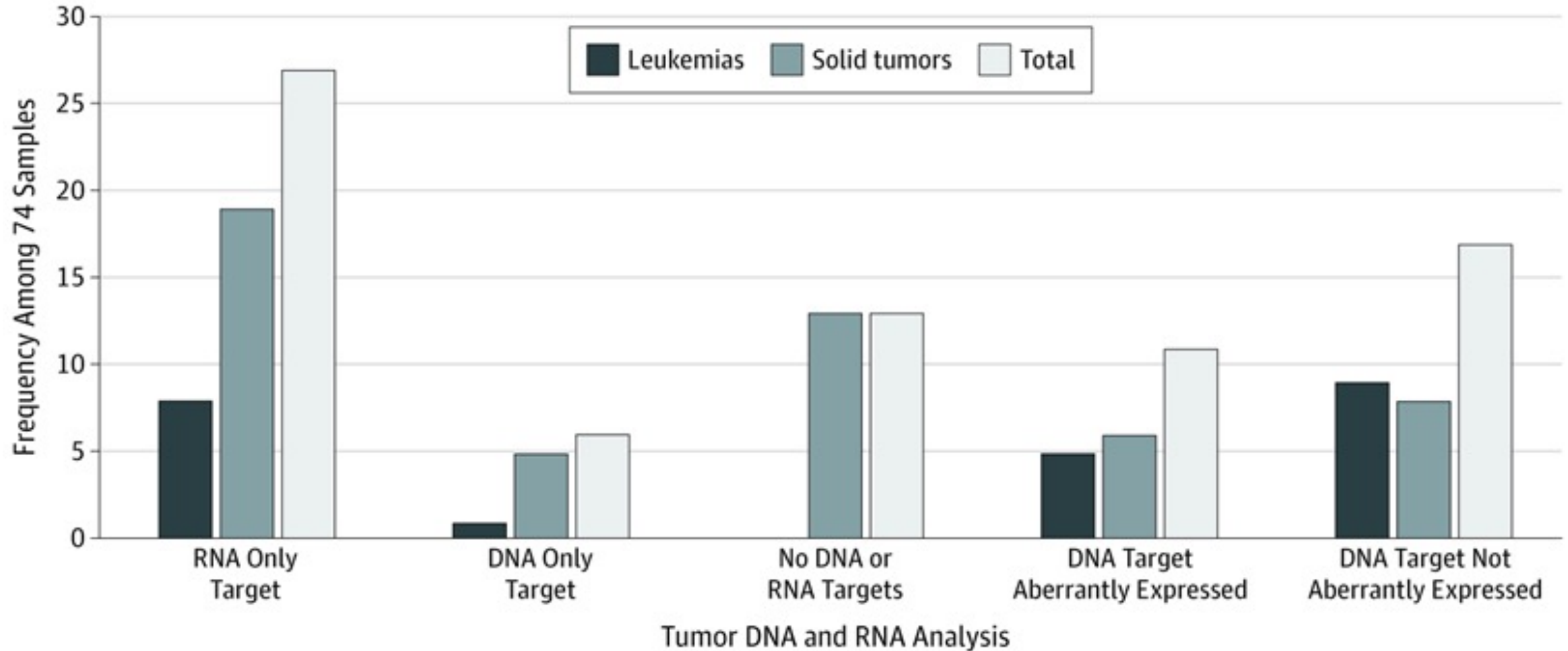
Via next generation sequencing technology, RNA transcripts are converted to millions of reads, providing information about protein expression.

- Align reads to human genome
- Quantify data
- Normalize data
- Variant and Fusion detection

RNA-Seq enhances DNA findings



RNA-Seq enhances DNA findings





Case Study of a Pediatric Patient

Patient 1 benefits from comparative RNA-Seq analysis

Patient 1 – Initial Diagnosis

- Patient 1 was diagnosed at 8 years of age with a left tentorial-based CNS sarcoma after a 2-week history of nausea, lethargy, and diplopia.
- The patient received six cycles of induction chemotherapy— ifosfamide, carboplatin, and etoposide—followed by autologous stem-cell transplantation with a high-dose preparative regimen of carboplatin, thiotepa, and etoposide as well as 54 Gy of focal radiation to the location of the original tumor.
- After a 2-year remission, the tumor recurred with numerous pulmonary lesions in all lobes.

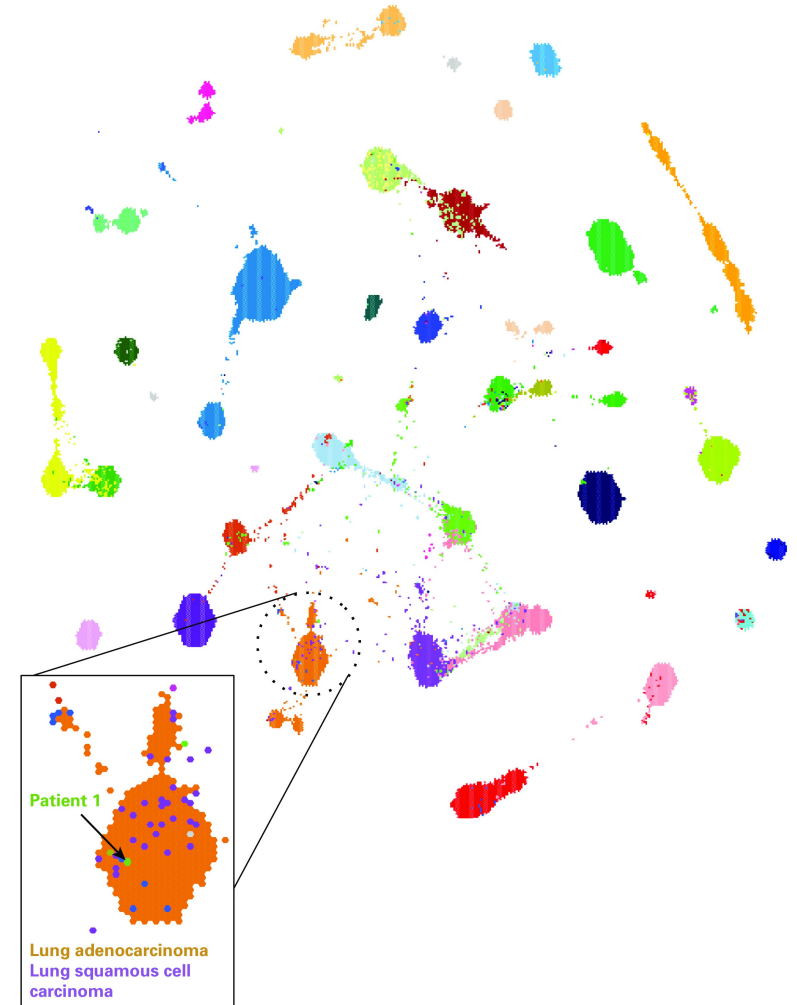
Patient 1 – Recurrence

- WGS and RNA-Seq was performed on lung biopsy.
- *EWSR1-ATF1* fusion was discovered as well as 3 variants of uncertain significance in *PDGFRA*, *PRKCB*, and *SVIL*.
- RNA-Seq expression was compared to 10,668 samples from TCGA (The Cancer Genome Atlas) and TARGET (Therapeutically Actionable Research to Generate Effective Treatments).

Patient 1 – TumorMap Analysis

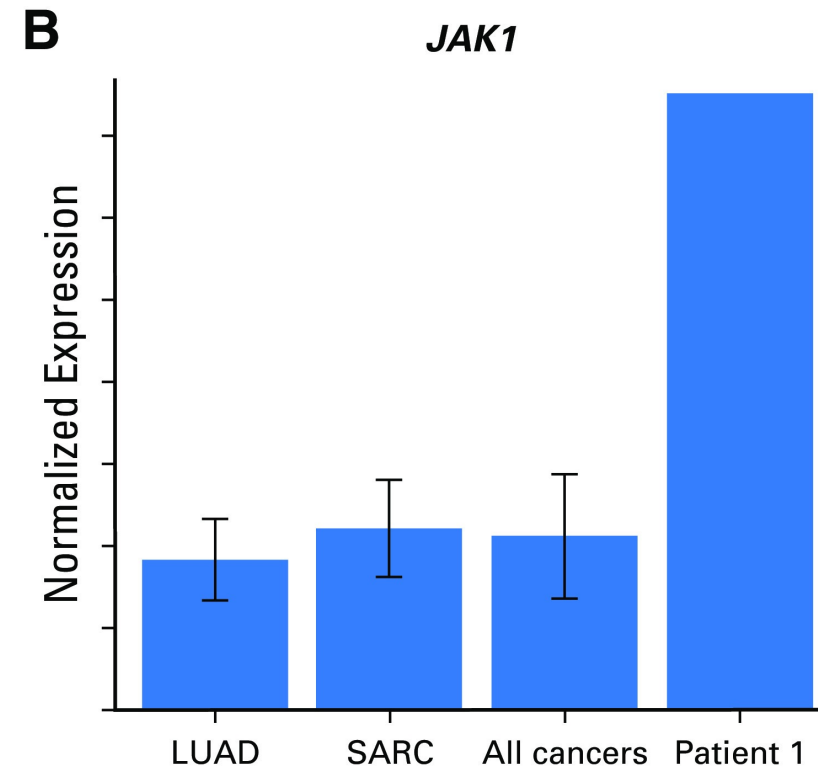
- Comparative analysis found that Patient 1's sample was most similar to Lung Adenocarcinoma (LUAD).

Patient 1's tumor RNA-Seq profile in the context of 38 tumor types



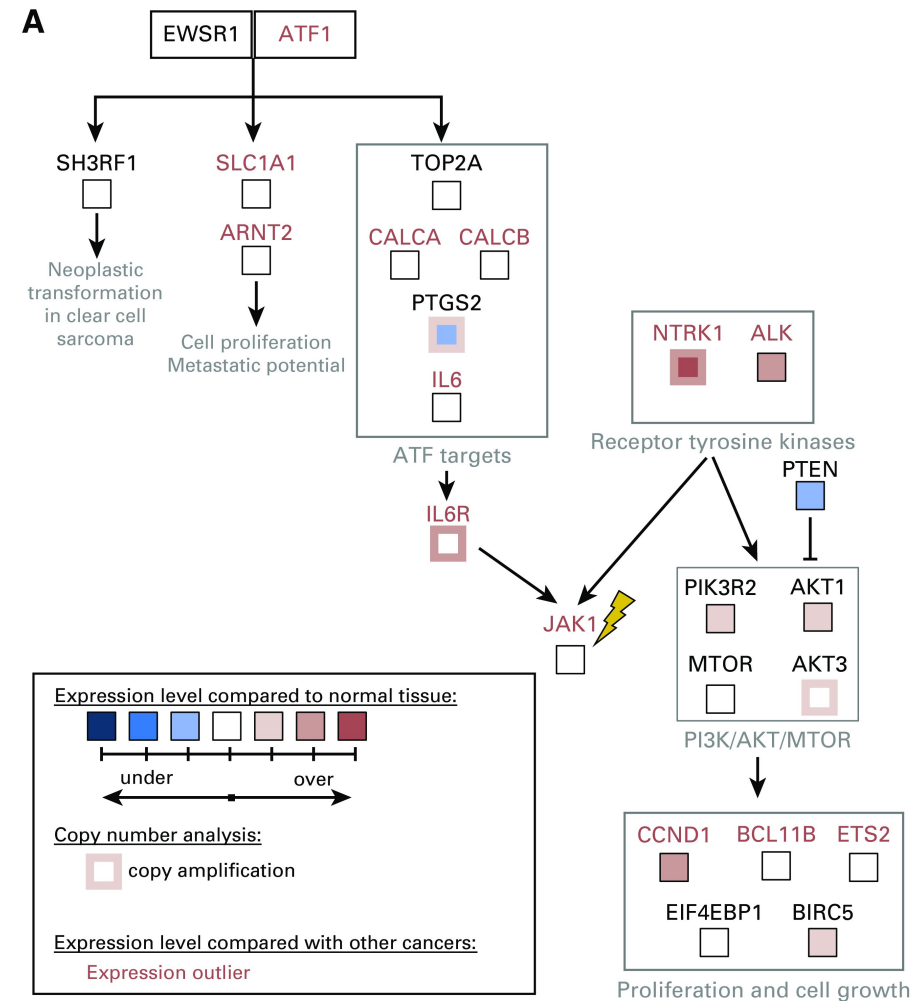
Patient 1 – Comparative Analysis

- Comparative analysis found that Patient 1's sample was most similar to Lung Adenocarcinoma (LUAD).
- Expression analysis was also performed against all cancer tumors as well as against sarcomas.
- *JAK1*, *ALK*, *NTRK1*, and *CCND1*, emerged as overexpression outliers in all analyses.

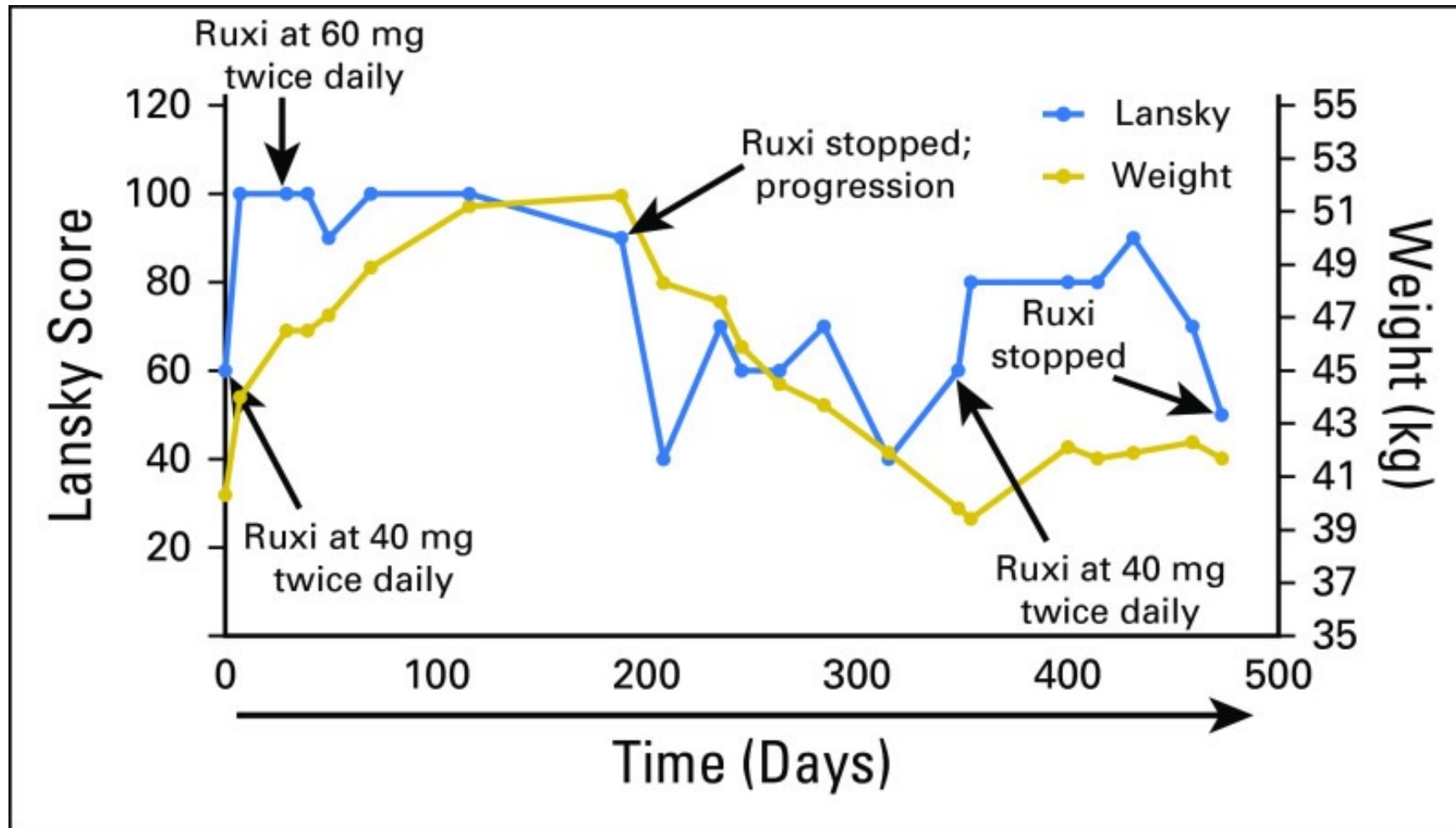


Patient 1 – Pathway Analysis

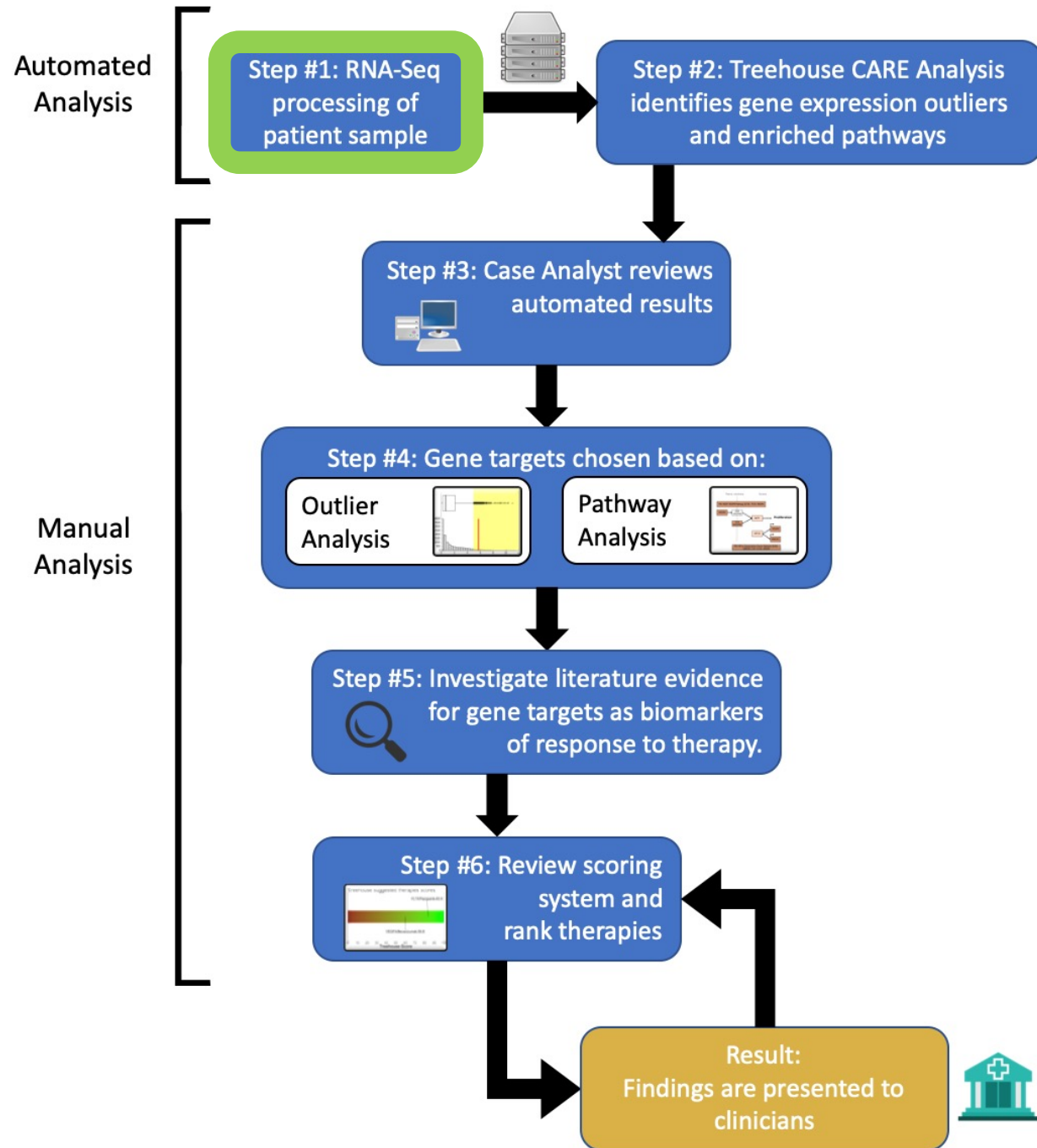
- The *EWSR1-ATF1* fusion as well as receptor tyrosine kinases *NTRK1* and *ALK* can activate the JAK-STAT pathway.
- Decision was made to target JAK1 using ruxolitinib.
 - Consideration was given to targeting ALK with crizotinib.



Patient 1 – Clinical Response to Ruxolitinib



Treehouse Pipeline

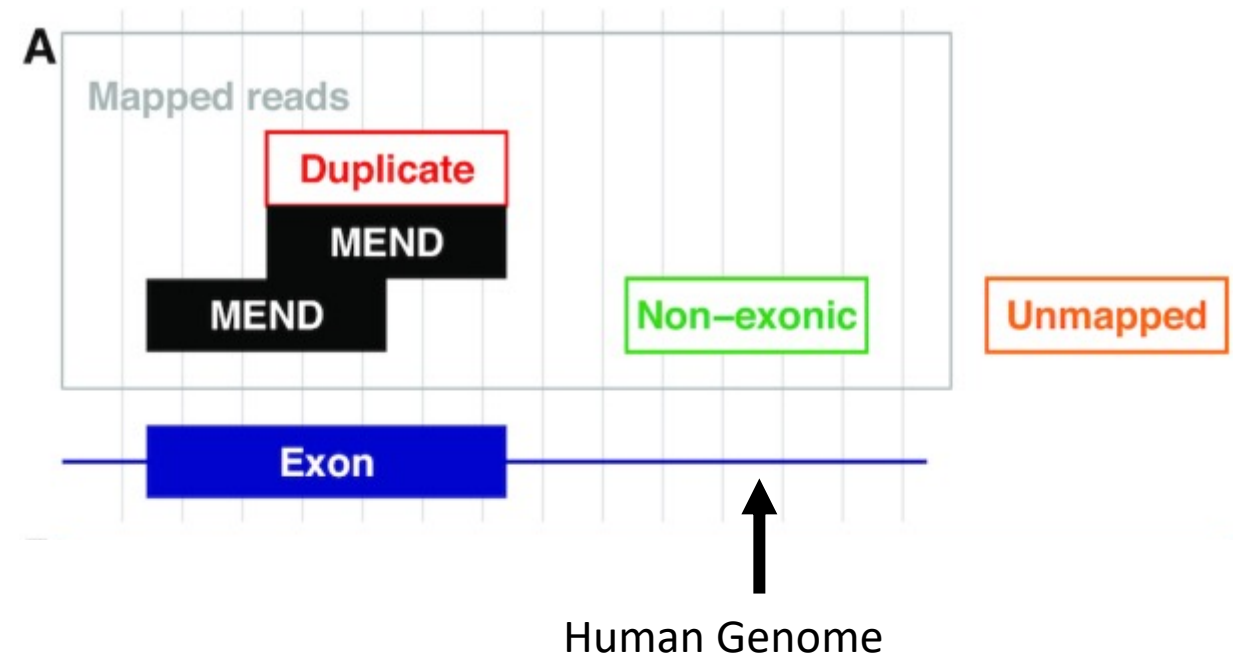


Treehouse RNA-Seq processing is fully consistent and reproducible

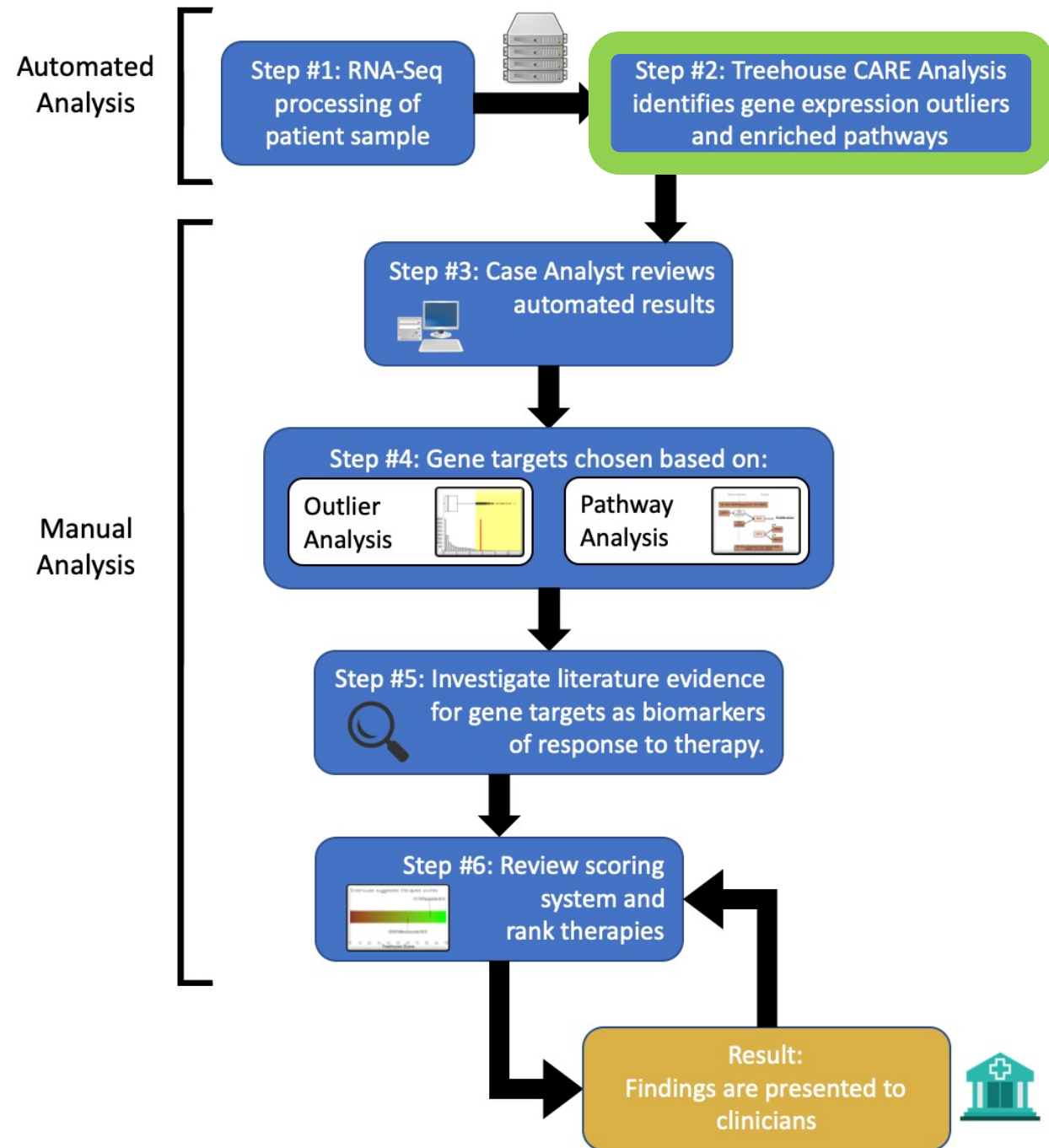
- In order to minimize batch effect of having samples sequenced by different institutions, Treehouse processes all the primary data generated by the sequencer through TOIL.
- TOIL uses STAR to align the sequence reads to the human genome for expression calculations.
- <https://github.com/DataBiosphere/toil>

Treehouse includes high quality samples in compendia

- Several methods for determining quality of data:
 - RIN (RNA Integrity Number)
 - Total reads
- Treehouse uses a 10 million Mapped Exonic Non-Duplicate (MEND) threshold for a sample's inclusion into a compendium.

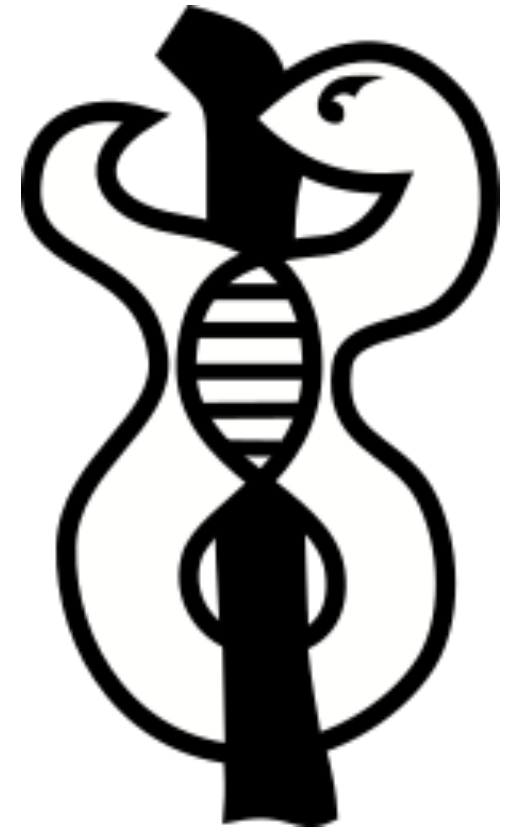


Treehouse Pipeline

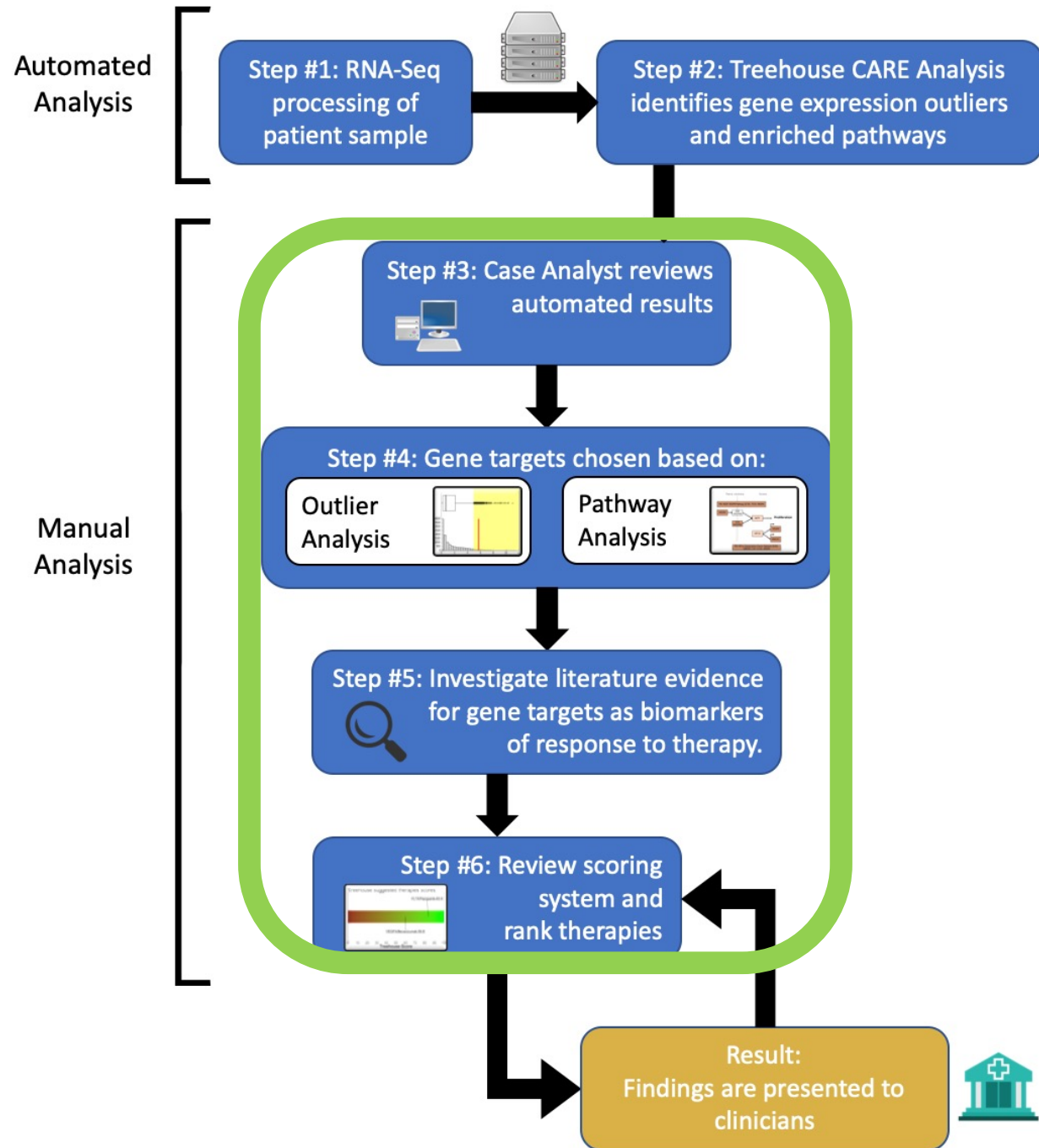


Treehouse CARE enables quick and automated detection of outlier genes and enriched pathways

- Comparative Analysis RNA-Seq Expression (CARE) is an open source tool developed to detect gene outliers in a single sample compared to a background cohort.
- CARE packages the outliers with pathway and drug database data for researchers.
 - Github: <https://github.com/UCSC-Treehouse/CARE>
 - Docker: <https://hub.docker.com/r/ucsctreehouse/care>

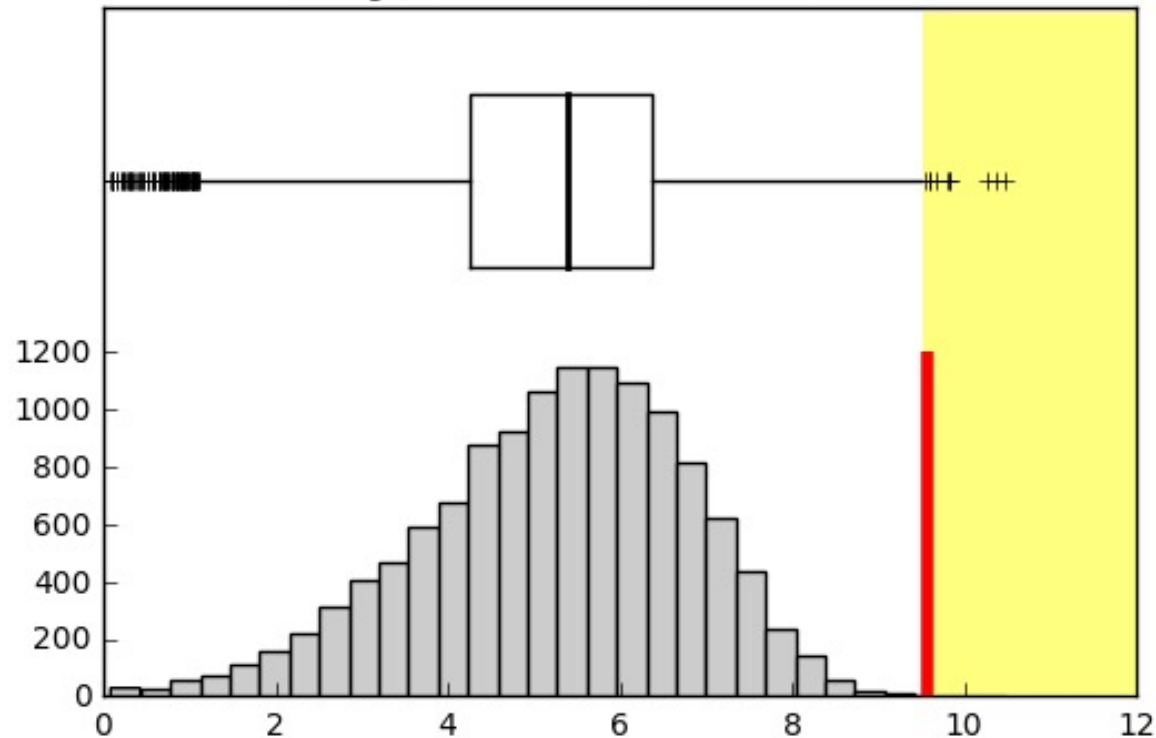


Treehouse Pipeline

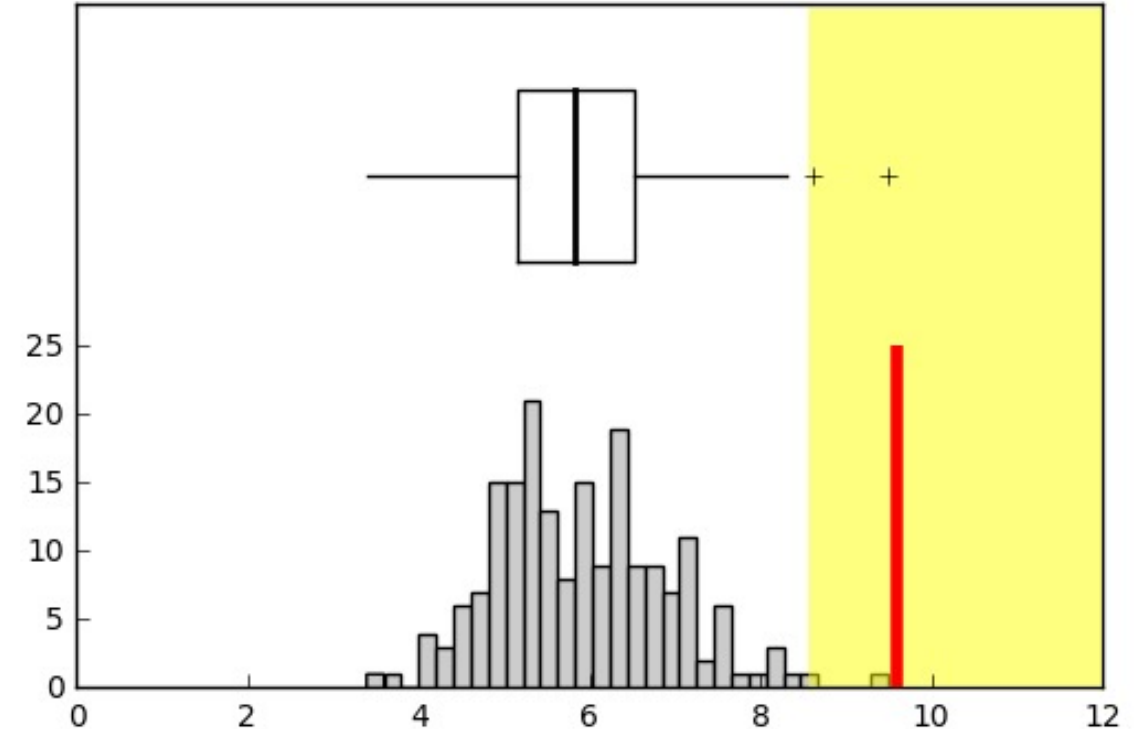


Outlier status is used to determine gene overexpression

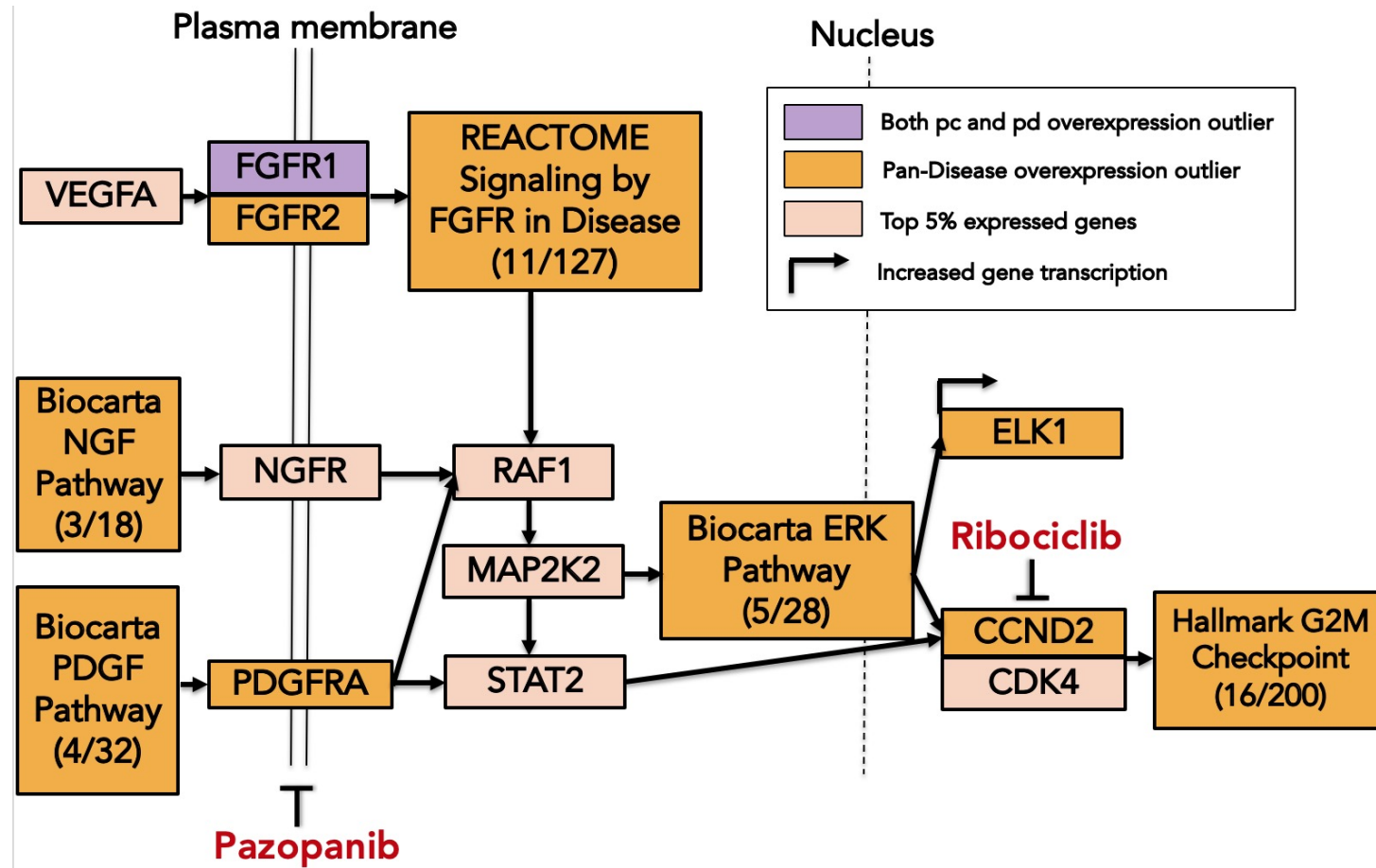
MYC is a Pan-Cancer outlier
n=12,747



MYC is a Pan-Disease outlier
n=181



Enriched pathways enhance understanding of tumor biology



Literature investigation determines strongest gene targets

- Focus on drugs that are FDA approved or in clinical trials.
- Focus on drugs that are approved to treat gene overexpression or gene amplification.
- Focus on studies in pediatric cancer patients.

NIH U.S. National Library of Medicine

ClinicalTrials.gov

PubMed.gov

Google Scholar



Treehouse RNA-Seq comparative analysis results are presented to clinicians

FOR RESEARCH PURPOSES ONLY - PREPARED FOR THE CLINICAL TEAM

Treehouse Treehouse Analysis: Sample SAMPLE_ID (TREEHOUSE_ID)

Diagnosis

Gender **Anatomical Source**

Age at Diagnosis **Sample Collected by Stanford**

Age at Relapse **Sequence Received by Treehouse**

Relevant Clinical Notes

Foundation Medicine
(Any Foundation Medicine results provided by Stanford)

Additional Molecular Testing
(Any other molecular testing results)

Treehouse Primary Categorization

RTK: VEGFR/PDGFR/FGFR

Cell Cycle

JAK/STAT

PI3K/AKT/mTOR

Other:

Comparative Tumor RNA-Seq Analysis Findings			
Molecular Abnormality	Molecular Category	Associated Drugs*	Analyst Summary
GENE1	Other ^{b,e,h}	DRUG1	GENE3 is a pan-cancer up-outlier with pan-cancer and pan-disease pathway enrichment.
GENE2, GENE3	Other ^{a,h}	DRUG2	GENE1 and GENE2 are pan-cancer up-outliers. They both activate the PATHWAY1 pathway, which is expressed in this sample.

a) outlier compared to all tumors (pan-cancer)
 b) outlier compared to similar tumors (pan-disease)
 c) pathway enrichment in pan-cancer outlier list
 d) pathway enrichment in pan-disease outlier list
 e) pathway enrichment in overlap outlier list
 f) pathway enrichment in top 5% expressed genes
 g) tumor similarity analysis
 h) literature evidence
 i) other molecular support (mutation, fusion, amplification, immunohistochemistry)

Associated Drugs have been identified by Stanford treating oncologists and are included in Stanford IRB Protocol #44179. For more information on these drugs, visit cancer.gov and search for the drug name.

Pathway Summary

(Description of Pathway Diagram. Image is an example.)

TumorMap

(Description of TumorMap findings. Image is an example.)

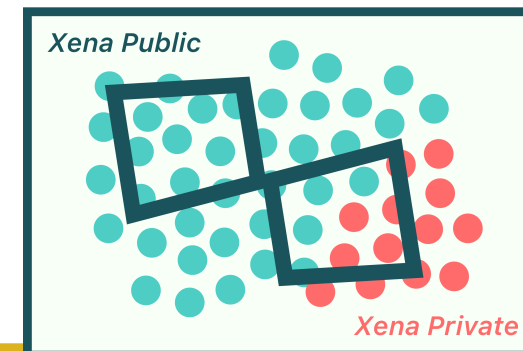
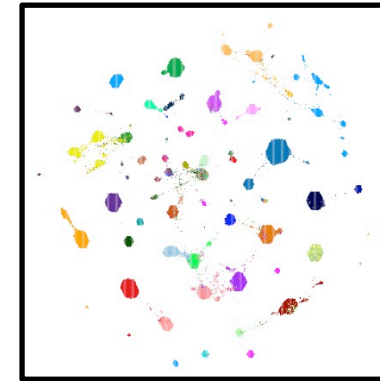


Any Questions?

Next: Publicly Available Resources

Publicly Available Resources

- Treehouse Compendia – Data
- UCSC TumorMap – Visualization
- UCSC Xena – Data & Visualization





Treehouse Compendia

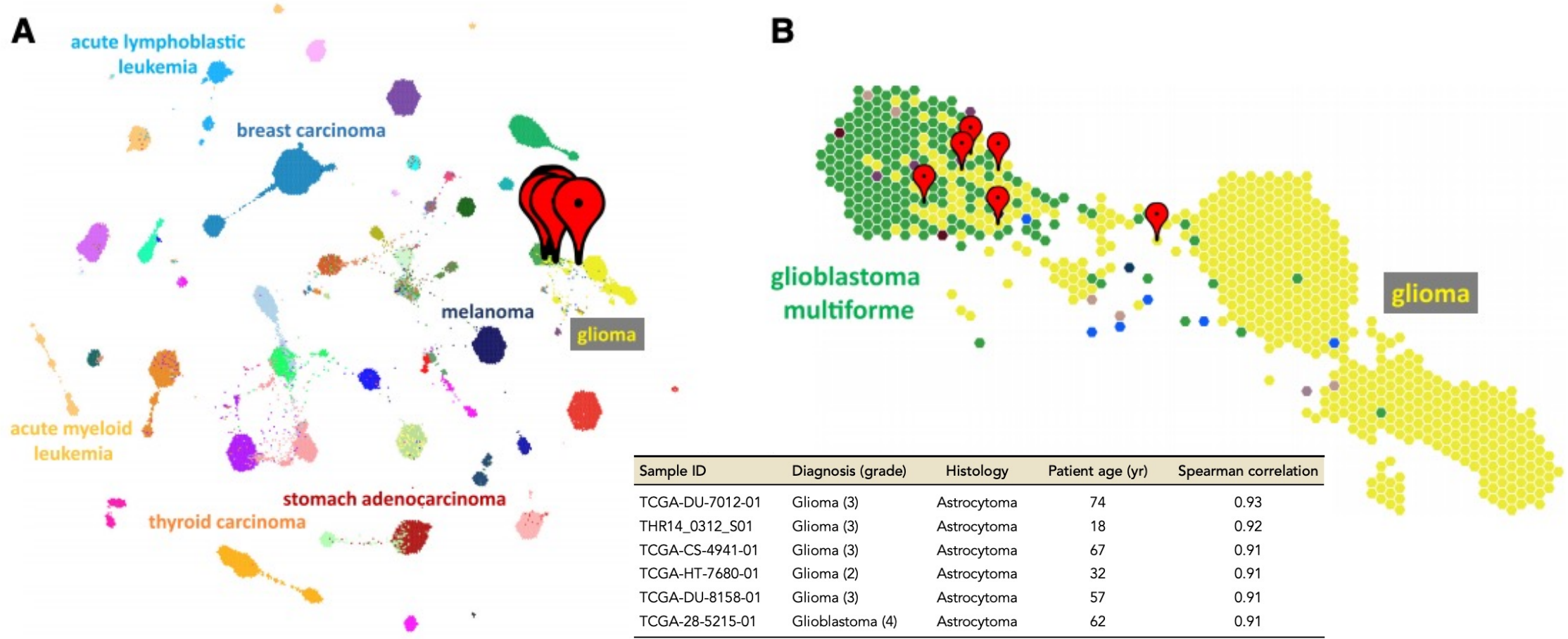
- Treehouse creates different compendia based on sample type and library preparation method.
- Treehouse compendia include:
 - PolyA (poly-adenylated tumor – 12,747)
 - Ribodeplete (ribodepleted tumor – 296 samples)
 - Cell Line PolyA (poly-adenylated cell line – 932 samples)
 - **GTEX** (poly-adenylated normal tissue – 7,862 samples)
 - **Patient Derived Xenograft (PDX)** (poly-adenylated PDX – 60+ samples)
- All Treehouse Compendia are available to the public:
<https://treehousegenomics.soe.ucsc.edu/public-data/>



UCSC TumorMap

- Provides 2-dimensional mapping of high dimensional genomic data.
- Spatial correlation analysis (SCA) enables users to find associations between attributes in their data.
- Users can generate their own map or investigate a pre-built map from a published project: <https://tumormap.ucsc.edu/>

TumorMap analysis aids in diagnosis of rare pediatric tumor



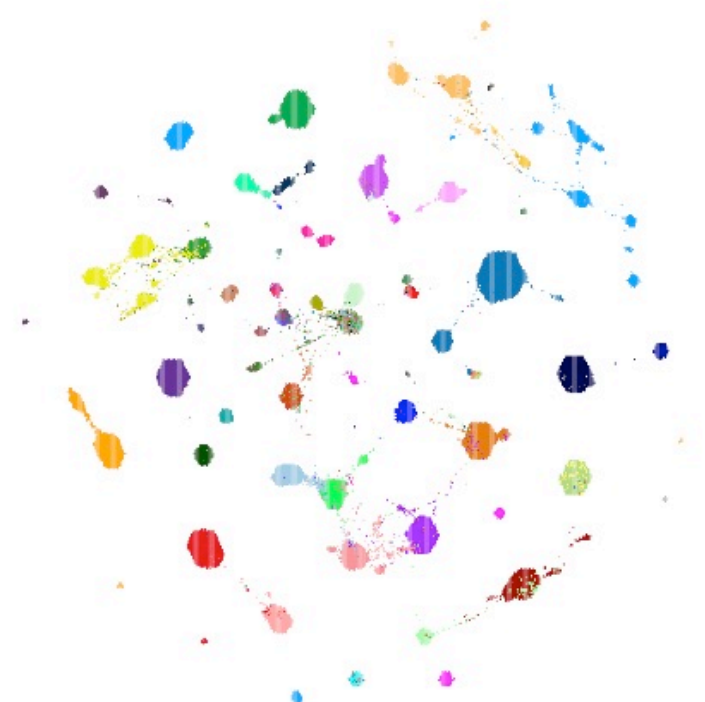
Visualizing Treehouse Compendia using TumorMap



v5 PolyA Compendia
11,258 samples



v9 PolyA Compendia
11,454 samples



v11 PolyA Compendia
12,747 samples



TumorMap Demo

- Demonstration of TumorMap capabilities using the Treehouse v11 public compendium.
- Tumormap: <https://tumormap.ucsc.edu/>

UCSC Xena – Host and Visualize public data

- Supports multi-omic and phenotypic data.
- Users can host and maintain multiple release versions.
- Combine private data with public data.
- Filter, sort and visualize data.



Acknowledgements



Vaske Lab

Olena Vaske, Principal Investigator
Holly Beale, Computational Biologist
Gina Mawla, Postdoctoral Student
Allison Cheney, Graduate Student
Yvonne Vasquez, Graduate Student
Anouk van den Bout, Research Associate
Molly McCabe, Lab Aide

Treehouse

Isabel Bjork, Program Director
Katrina Learned, Data Coordinator
Ellen Kephart, Software Engineer
Sofie Salama, Research Scientist
David Haussler, Scientific Director, Genomic Institute

