Research Experiences for Teachers

Comparison of circRNA from CML using TERRACE

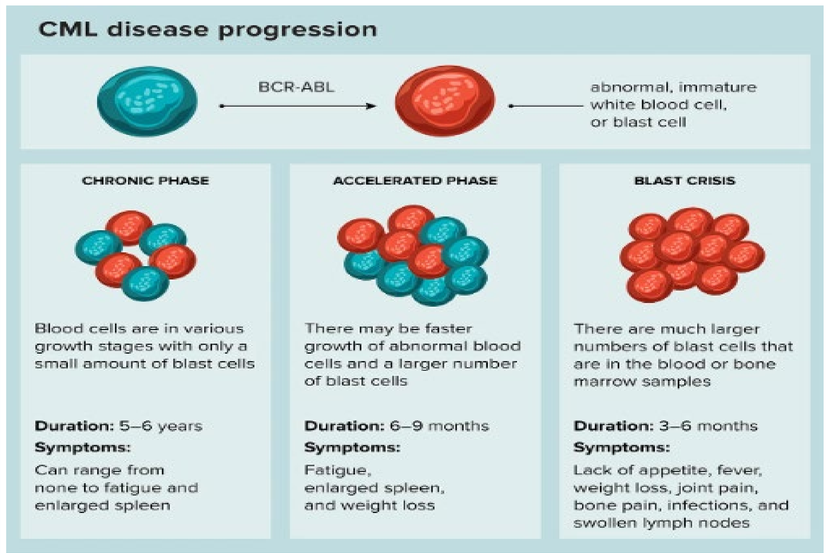
Jason Econome[1], Tasfia Zahin[2], and Dr. Mingfu Shao[2]

[1]Stuyvesant HS, NYC, NY 10282; [2]Electrical Engineering and Computer Science, Pennsylvania State University, University Park, PA 16802

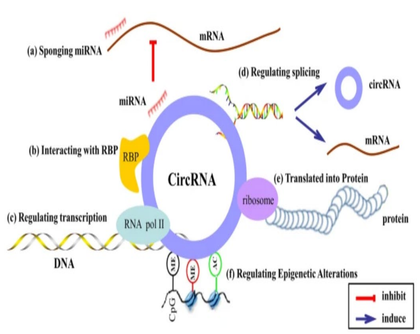
Introduction

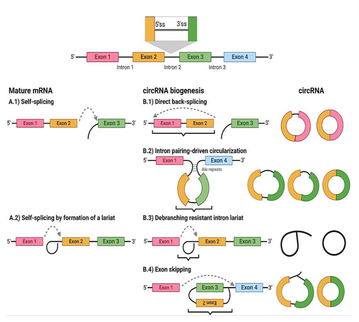
Scientists now can more accurately characterize a cell's functional behavior based on the genes it expresses. RNA sequencing (RNA-seq) allows for the identification and quantification of a cell's entire transcriptome. Many of these genes are alternatively spliced, producing both coding and noncoding variants that affect a cell's overall performance.

One potential application of this tool is in the field of oncology; particularly Chronic Myelogenous Leukemia (CML). The blood cancer's initial-chronic phase is often overlooked because patients rarely show any obvious symptoms. The typical signs of fatigue and achiness are more so associated with mid-advanced or late-blast phases where the prognosis is poorer and relapse is almost guaranteed.



The ability to assay for an early-stage biomarker would improve treatments and one's chance for full recovery. The circular RNA (circRNA) is of interest because they are stable, long lasting and detectable via PCR amplification.

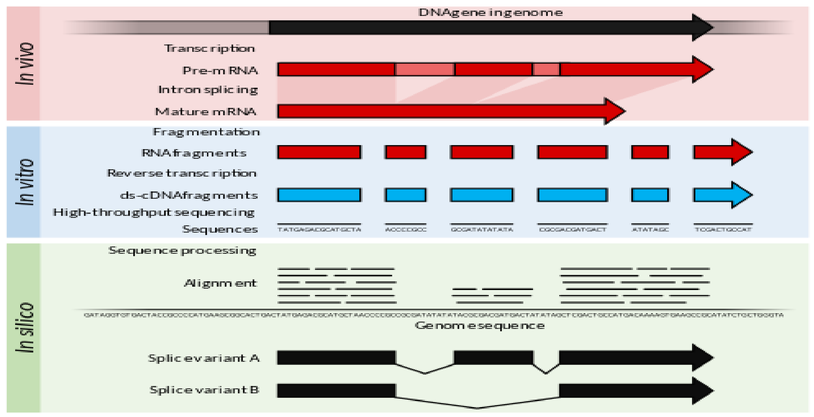




https://en.wikipedia.org/wiki/Circular_RNA

Materials & Methods

I. Single Cell RNA-seq Process



1 RNA extracted, RNA fragmented (<300bp)

2 cDNA synthesized, amplified, sequenced (Illumina)

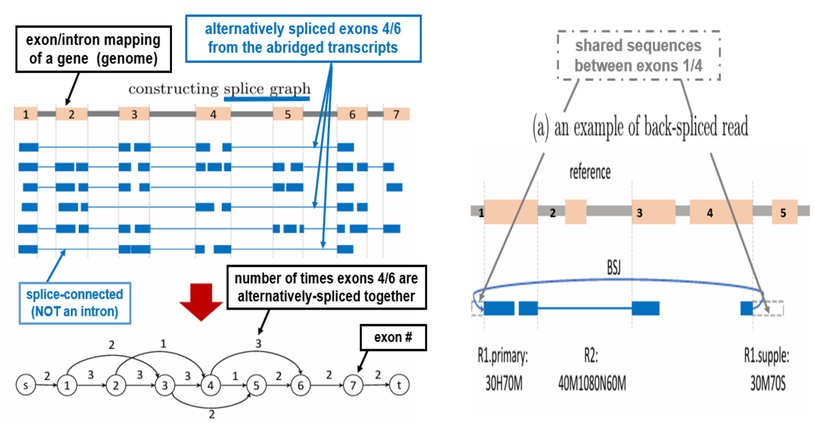
3 RNA transcripts assembled against sequenced genome

4 Computer model, "Accurate Assembly of CircRNAs Using Bridging and Machine Learning" (TERRACE), on Leukemia DB an established data-bank of sequenced, transcripts depo-sited by researchers from the CML community

Output was compared with that of another commonly used model, CIRI-full to confirm reliability. A threshold of 20 sample counts, high score, and matching chromosomal loci between the two models' outputs, suggest that circRNA molecules identified were authentic

II. Identify host genes and circRNAs

TERRACE accurately assembles circRNAs via recognition of back-spliced reads, represented by overlapping sequences, shared ALU elements between distanced exons



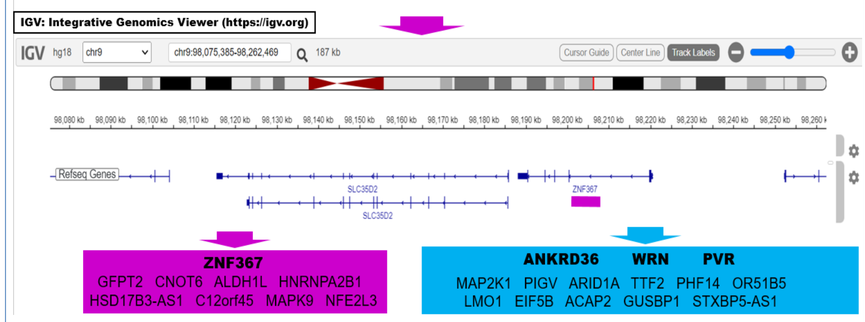
Zahin, Shi, Zang, Shao Accurate Assembly of Circular RNAs with TERRACE bioRxiv preprint Feb 2024 https://doi.org/10.1101/2024.02.09.579380

Results

TERRACE's output was compared with a commonly used model, CIRI-full to confirm reliability. A threshold of 20 sample counts, high score, and matching chromosomal loci between the two models' outputs, suggest that circRNA molecules identified were authentic

chr	src	feature	start	end	score	s-count
9	CIRI-full	circRNA	98,011,285	98,011,389	38	30
9	CIRI-full	exon	98,011,285	98,011,389	38	30
9	CIRI-full	circRNA	98,011,289	98,012,472	153	26

chr	src	feature	start	end	score	s-count
9	TERRACE	circRNA	98,011,271	98,012,472	140	70
9	TERRACE	exon	98,011,271	98,011,389	140	70
9	TERRACE	exon	98,012,421	98,012,472	140	70



ZNF367
GFPT2 CNOT6 ALDH1L HNRNPA2B1 HSD17B3-AS1 C12orf45 MAPK9 NFE2L3

ANKRD36 WRN PVR
MAP2K1 PIGV ARID1A TTF2 PHF14 OR51B5 LMO1 EIF5B ACAP2 GUSBP1 STXBP5-AS1

Conclusions

TERRACE RNA-seq assembly demonstrated great sensitivity and reliability in identifying 23 circRNAs, from CML line, K562

At >20 sample counts threshold, nine circRNA were shared with CIRI-full with four being documented in the literature

CML-specific circRNAs (ZNF367, ANKRD36, WRN, PVR) were identified and their host genes confirmed in the literature, making them strong candidates as potential biomarkers

Future Work

Further characterize host gene-derived circRNA molecules (ZNF367, ANKRD36, WRN, PVR) in order to determine specific phase and possible causal role in CML progression

Supplement TERRACE assembly-identification of CML-specific circRNA with earlier computer model, SCALLOP-LR, identifying linear coding or noncoding transcripts to confirm specific CML-phases (e.g., Philadelphia chromosome)

Develop primers and optimize PCR amplification of circRNA molecules (ZNF367, ANKRD36, WRN, PVR) as a reliable assay for biomarkers in CML phase detection

References

Zahin, Shi, Zang, Shao Accurate Assembly of Circular RNAs with TERRACE bioRxiv preprint Feb 2024 https://doi.org/10.1101/2024.02.09.579380

Mei Luo, A comprehensive landscape of transcription profiles and data resources for human leukemia Blood Advances 25 JULY 2023 • VOL 7, No 14

Li Q, Ren X, Wang Y. 2023. CircRNA: a rising star in leukemia. PeerJ 11:e15577

The RET for Teachers is funded by the National Science Foundation Grant DBI: 2145171.