An Open Pediatric Brain Tumor Atlas

Authors

- **John Doe**
  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ◯ [johndoe](#) · 🐦 [johndoe](#)
  Department of Something, University of Whatever · Funded by Grant XXXXXXXX

- **Jane Roe**
  ⓘD [XXXX-XXXX-XXXX-XXXX](#) · ◯ [janeroe](#)
  Department of Something, University of Whatever; Department of Whatever, University of Something

Abstract

Introduction

Introduction will go here.

Materials and Methods

Children's Brain Tumor Tissue Consortium Data

Pediatric Pacific Neuro-oncology Consortium Data

DNA Alignment

Germ Line Single Nucleotide Variant Calling

Somatic Single Nucleotide Variant Calling

Somatic Copy Number Variant Calling

Somatic Structural Variant Calling

Gene Expression Abundance Estimation

RNA Fusion Calling and Prioritization

Gene fusion detection

We set up Arriba v1.1.0 and STAR-Fusion 1.5.0 [1] fusion detection tools using CWL on CAVATICA. For both these tools we used aligned BAM and chimeric SAM files from STAR as inputs and GRCh38_gencode_v27 GTF for gene annotation. We ran STAR-Fusion with default parameters and annotated all fusion calls with GRCh38_v27_CTAT_lib_Feb092018.plug-n-play.tar.gz provided in the STAR-fusion release. For Arriba, we used a blacklist file (blacklist_hg38_GRCh38_2018-11-04.tsv.gz) from the Arriba release tarballs to remove recurrent fusion artifacts and transcripts present in healthy tissue. We also provided Arriba with strandedness information or set it to auto-detection for polyA samples.

Fusion prioritization

We built a fusion prioritization pipeline to filter and annotate fusions. We considered all inframe and frameshift fusion calls with 1 or more junction reads and fused genes expressed with TPM greater than one to be true calls. If a fusion call had large number of spanning fragment reads compared to junction reads (spanning fragment minus junction read greater than ten) or if either $5^{'}$ or $3^{'}$ genes fused to more than five different genes we removed these calls as a potential false positive. We also removed fusions if the $5^{'}$ or $3^{'}$ ends were the same gene, and these were tagged as non-canonical splicing or duplication. We used a list of curated fusion calls for each histology to capture each occurrence of the fusion as a putative driver fusion. We prioritized a union of fusion calls as true calls if the fused genes were detected by both callers, the same fusion was recurrent in histology (>2 samples) or the fusion was specific to the broad histology. We annotated putative driver fusions and prioritized fusions lists with kinases, oncogenic, tumor suppressor, transcription factor, fused genes and known TCGA fusions from curated datasheets. We also added chimerDB [2] annotations to both driver and prioritized fusion list.

Clinical Data Harmonization

Results

Results section stub.

Conclusions

Stub in conclusions section

References

1. **STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq**
Brian J. Haas, Alex Dobin, Nicolas Stransky, Bo Li, Xiao Yang, Timothy Tickle, Asma Bankapur, Carrie Ganote, Thomas G. Doak, Nathalie Pochet, … Aviv Regev
Cold Spring Harbor Laboratory (2017-03-24) https://doi.org/gf5pc5
DOI: 10.1101/120295

2. **OUP accepted manuscript**
Nucleic Acids Research
(2016) https://doi.org/gf6bx9
DOI: 10.1093/nar/gkw1083 · PMID: 27899563 · PMCID: PMC5210563