Comparison of computational fusion detection methods for short-read RNA-seq data

Lucas P. Oliveira[1], **Victor Rigatto**[1], **Natacha A. Migita**[2], **José A. Yunes**[2], **João Meidanis**[1]

[1] Instituto de Computação – Universidade Estadual de Campinas (UNICAMP)
Av. Albert Einstein, 1251 – 13083-852 – Campinas – SP – Brazil

[2]Centro Infantil Boldrini
R. Márcia Mendes, 619 – 13083-884 – Campinas – SP – Brazil

Abstract. *Gene fusions are abnormal genetic events often correlated with oncogenesis. Hence, detecting them from RNA-seq data using bioinformatics methods is an important task in cancer research. Several tools have been developed for this task, but current benchmarks are inconclusive regarding their accuracy and are difficult to reproduce with new data. In this paper, we propose a computational pipeline that gathers fusion detection tools and compares them using standard classification metrics. It can also be used as an ensemble method to detect gene fusions using several tools. This pipeline was applied to simulated and real data, and supplements current benchmarks in the literature towards aiding the users in choosing the tools for their analyses.*

Resumo. *Fusões gênicas são eventos genéticos anormais frequentemente correlacionados com a oncogênese. Por isso, detectá-los a partir de dados de RNA-seq usando métodos de bioinformática é uma tarefa importante na pesquisa do câncer. Várias ferramentas foram desenvolvidas para esta tarefa, mas os benchmarks atuais são inconclusivos quanto à precisão das mesmas e são difíceis de reproduzir com novos dados. Neste artigo, propomos um pipeline computacional que reúne ferramentas de detecção de fusão e as compara usando métricas padrão de classificação. Este também pode ser usado como um método agregado para detectar fusões gênicas usando diversas ferramentas. Esse pipeline foi aplicado a dados simulados e reais, e complementa os benchmarks atuais da literatura para auxiliar os usuários na escolha das ferramentas para suas análises.*

1. Introduction

A major characteristic of oncogenesis is the presence of chromosomal rearrangements that often lead to gene fusions. A gene fusion occurs when previously separated genes join together and form a hybrid gene with potentially altered function compared to the original genes. This genetic event is an important biomarker to diagnose and treat various types of cancer. The detection of gene fusions from short-read RNA-seq data is a well studied area, with dozens of tools published in the last decade. Several of them have already been benchmarked in simulated and real data, and their results often agree in simulated data [Carrara et al. 2013, Liu et al. 2016, Kumar et al. 2016, Singh and Li 2021, Creason et al. 2021, Haas et al. 2019]. For real data, however, most reviews and benchmarks are inconclusive. Hence, a common approach that is employed is running a

number of tools and calling gene fusions based on a voting scheme, mostly by consensus [LaHaye et al. 2021, Vicente-Garcés et al. 2023, Apostolides et al. 2021].

However, ensemble approaches rarely take into account the techniques employed by each tool, and hence run the risk of losing fusions detected by the minority vote. In addition, a tool's performance may vary depending on the dataset, and choosing the best tool for one's data remains a challenge. Furthermore, as new tools are released, benchmarks need to be updated. Some papers make available code to run their benchmarks, but reproducing them with new data can be a difficult task because they were often designed with the original data in mind.

In this paper, we propose a computational pipeline that gathers the top-performing tools from previous benchmarks [Singh and Li 2021, Liu et al. 2016, Haas et al. 2019]: Arriba [Uhrig et al. 2021], STAR-Fusion [Haas et al. 2019], FusionCatcher [Nicorici et al. 2014], FuSeq [Vu et al. 2018], Pizzly [Melsted et al. 2017], and ChimeRScope [Li et al. 2017]. We also included CICERO [Tian et al. 2020], a newer alignment-based tool that leverages a local assembly subroutine to detect non canonical fusions, whose performance has not been assessed. The pipeline compares the tools based on standard classification metrics and can also be used as an ensemble approach for fusion detection, but allows the user to choose the set of tools for the analysis. We applied it to the simulated dataset used in the benchmark of [Haas et al. 2019] and to a real dataset comprising several samples sequenced in a recent targeted RNA-seq experiment at Boldrini Children's Center [Migita et al. 2023].

2. Methods

2.1. Fusion detection software packages

Table 1 summarizes the fusion detection methods used in this study. We selected four alignment-based methods — Arriba [Uhrig et al. 2021], STAR-Fusion [Haas et al. 2019], FusionCatcher [Nicorici et al. 2014], and CICERO [Tian et al. 2020] — and three alignment-free methods — FuSeq [Vu et al. 2018], Pizzly [Melsted et al. 2017], and ChimeRScope [Li et al. 2017]. Notice, however, that FuSeq and Pizzly are not considered alignment-free: the former is based on *quasi-mapping* from RapMap [Srivastava et al. 2016] and the latter on pseudo-alignment from kallisto [Bray et al. 2016]. For the purpose of this study, we include them in this category because both avoid the dynamic programming-based alignment step found in most aligners.

The alignment-free methods share the idea of indexing the k-mers of a reference transcriptome, in which the k-mers are associated with the genomic features (often transcripts) they originated from. Raw reads are then classified as fusion supporting or not based on their k-mer composition. On the other hand, alignment-based methods map the reads against a reference genome to identify chimeric alignments that may indicate a fusion event. The main signatures that are looked after are junction reads and spanning reads: the former occurs when a read maps to two different regions in the genome, while the latter, in the case of paired-end reads, have each read map to a different region. The presence of these signatures does not necessarily imply a gene fusion; hence, these alignments undergo extensive filtering steps based on a number of heuristics to remove

artifacts and false positives. In the case of Pizzly and FuSeq, these signatures can also be identified through their respective approaches.

Pizzly pseudoaligns reads against a reference transcriptome using kallisto and flags spanning reads that may indicate a fusion event. FuSeq employs a similar approach, but instead of pseudoalignment, it *quasi-maps* the reads against a reference transcriptome using RapMap and then flags junction reads. In both methods, the selected spanning reads are evaluated to filter false positives. ChimeRScope is a fully alignment-free tool that identifies gene fusions by scoring the raw reads through k-mer counting. This score is maximized when a read contains enough k-mers from distinct transcripts.

Arriba and STAR-Fusion are alignment-based tools that have been consolidated as the top-performing in most simulated short-read RNA-seq datasets [Haas et al. 2019, Creason et al. 2021, Singh and Li 2021]. Both optimize the parameters of the STAR aligner [Dobin et al. 2013] so that alignments of junction reads and spanning reads are reported. These alignments then undergo extensive filtering steps to remove artifacts and false positives [Uhrig et al. 2021, Haas et al. 2019]. Arriba, in particular, employs multiple knowledge-based filters, including a blacklist that the user provides containing artifacts to be ignored [Uhrig et al. 2021]. FusionCatcher employs a similar strategy, but, in addition to STAR, it also uses Bowtie2 [Langmead and Salzberg 2012] and BLAT [Kent 2002], although the user is able to skip STAR or BLAT.

CICERO is a newer alignment-based tool that also uses the STAR aligner to identify chimeric alignments for downstream analysis [Tian et al. 2020]. In addition to the conventional alignment step, it also locally assembles a subset of soft-clipped reads using CAP3 [Huang and Madan 1999] and maps them against the reference genome using BLAT [Kent 2002]. It has been reported that this subroutine improves accuracy and allows the detection of non canonical fusion events not detectable by other methods [Tian et al. 2020].

2.2. Simulated dataset

We used simulated data generated by the Fusion Transcript Simulation Toolkit[1] [Haas et al. 2019]. It consists of 10 datasets, 5 with paired-end reads of length 50bp (available at `https://zenodo.org/records/13354907`) and 5 with paired-end reads of length 101bp (available at `https://zenodo.org/records/13359589`). All of them have 500 fusions known and annotated beforehand.

2.3. Targeted RNA-seq dataset

We assessed fusion prediction accuracy using real data from a targeted RNA-seq experiment comprising a cohort of B-cell precursor acute lymphoblastic leukemia (BCP-ALL) cases [Migita et al. 2023]. Most BCP-ALL can be classified into well-known subtypes based on the presence of characteristic rearrangements, e.g. translocation between chromosomes 9 and 22, resulting in the gene fusion *BCR-ABL1*. These are referred to as classical BCP-ALL, and account for approximately 70% of BCP-ALL cases [Migita et al. 2023]. The remaining 30% are called "B-other" because they do not have the characteristic rearrangements at diagnosis, and the underlying driver events were

[1] `https://github.com/FusionSimulatorToolkit/FusionSimulatorToolkit/wiki`

Table 1. Fusion detection tools for short-read RNA-seq data.

Tool	Overview
Arriba [Uhrig et al. 2021]	STAR is used to map reads against a reference genome, followed by application of extensive knowledge-based filters to the resulting alignments to remove artifacts and false positives.
STAR-Fusion [Haas et al. 2019]	STAR is used to map reads against a reference genome, filters are applied to the resulting alignments and richer annotations are added.
FusionCatcher [Nicorici et al. 2014]	Reads are mapped against reference genome using Bowtie2, STAR, and BLAT aligners.
CICERO [Tian et al. 2020]	STAR is used to map reads against a reference genome; a subset of soft-clipped reads are assembled into contigs by CAP3, which are mapped against the reference by BLAT to find non canonical fusions.
FuSeq [Vu et al. 2018]	Reads are *quasi-mapped* against a reference transcriptome using RapMap; then, fusion supporting reads ensue statistical tests for filtering.
Pizzly [Melsted et al. 2017]	Reads are pseudo-aligned against reference transcriptome using kallisto.
ChimeRScope [Li et al. 2017]	Alignment-free approach that indexes the k-mers of the reference and identifies fusion events through fingerprints in the raw reads.

(until recently) unknown [Migita et al. 2023]. The dataset we used comprises 40 samples of classical BCP-ALL and 47 samples of B-other BCP-ALL.

Due to space constraints, we show results for a subset of the samples, namely, 31/40 classical samples and 17/47 B-other samples. In the extended version of this paper, we will show results for the full set. The classical samples are divided into three groups according to the following defining gene fusions: *TCF3-PBX1* (3), *ETV6-RUNX1* (24), and *BCR-ABL1* (4). The same applies to the B-other samples, but by using the following (recently described) gene fusions: *ZNF384*-rearranged (9), *MEF2D*-rearranged (4), and *NUTM1*-rearranged (4). Different from the classical samples, the genes *ZNF384*, *MEF2D*, and *NUTM1* in the B-other samples can have several partners. In particular, [Migita et al. 2023] discovered novel gene partners for *ZNF384* (*NCOA3-ZNF384*, *SPI1-ZNF384*), *MEF2D* (*MEF2D-PYGO2*), and *NUTM1* (*KAT6A-NUTM1*), whose samples (4

of the 17 samples, one for each gene partner) were included in this study. All those fusions were validated, and for each of those groups, we used them as ground truth for their respective samples. We note, however, that some tools might have identified other fusions not validated *in vitro*, and these were not taken into account.

2.4. Data analysis

The experiments were run on a machine running Ubuntu version 22.04 with a 2.10 GHz Intel(R) Xeon(R) Silver 4110 CPU (32 cores) and 132GB of RAM. The tools shown in Table 1 were executed with parameters recommended by the respective authors, including STAR parameters (STAR is used by all alignment-based tools). It was not possible to reuse a single execution of STAR for all these tools, because each tool has different recommended parameters for it, with the exception of CICERO; hence, we used Arriba's STAR parameters for it. For the alignment-based tools, read mapping was done against reference genome GRCh38, except for CICERO, for which we mapped the reads against the reference genome GRCh37. Although its documentation states that it is compatible with GRCh38, we faced operational issues to run CICERO with this version, while with GRCh37 all the analyses ran smoothly. As for the alignment-free tools, we used the corresponding transcriptome reference GRCh38.

After the fusion calling step, we gathered the output of all tools, specifically the fusion calls and the number of supporting reads (SR) for those predictions. The notion of SR varies depending on the tool, but most of them report the number of junction and spanning reads that support a prediction. Recall that Pizzly and FuSeq count just one of these quantities. Thus, we consider the sum of junction and spanning reads as the number of SRs for a given prediction, when both are available; otherwise, we use one of the two according to the tool at hand.

After parsing the output of the tools, we mapped the gene coordinates to the GRCh37 reference standard following the Gencode v19 [Frankish et al. 2022] annotation set so that the coordinates were consistent. This allows a more amenable comparison among tools because many of them come with their own bundle of annotation resources. Then, we enriched the annotations of the gene fusions using FusionAnnotator [Haas et al. 2023], which adds metadata as provided by several databases. These include whether a gene is an oncogene, whether the fusion is interchromosomal or intrachromosomal, the encoded protein, etc, which are useful for downstream analysis and validation of the predictions. The next step leverages this enriched annotation to filter gene fusions of ambiguous origin, such as neighbor genes or genes of mitochondrial origin.

Finally, we measured the accuracy of the tools using the metrics *recall* and *precision*, defined as follows:

$$\textit{Recall} = \frac{TP}{TP + FN}, \qquad \textit{Precision} = \frac{TP}{TP + FP}.$$

(TP, FP, and FN are, respectively, the number of true positives, false positives, and false negatives.) These metrics are adequate for measuring gene fusion detection because one needs to take into account both TP and FP fusions reported by a tool. In this context, recall measures the proportion of correctly detected fusions with respect to all fusions

present in a dataset, while precision measures the proportion of correctly detected fusions with respect to all fusions retrieved by a tool. The desirable behavior of an accurate fusion detection method is to maximize both recall and precision. For each dataset with its accompanying ground truth fusions, we ran each tool and computed the Precision-Recall curve (with recall in the x-axis and precision in the y-axis) along with its area under the curve (AUC). The threshold used was the minimum number of supporting reads to call the fusions. As this threshold is decreased, a tool with good overall performance will report more fusions without necessarily harming the AUC.

We summarize these results for simulated and real data in Sections 3.2 and 3.3, respectively. All the steps described were incorporated in a computational pipeline (Section 3.1) that automates the data analysis. It parses the fusion calls of each tool in a standardized format for downstream analysis and compares the tools based on the metrics described.

3. Results

3.1. Computational pipeline

We developed a pipeline that incorporates many steps used in the benchmark of [Haas et al. 2019], but with important differences. It consists of a set of modules that perform well-defined tasks using standard utilities available in UNIX environments. Specifically, each module consists of a Makefile that automates a step in the pipeline. The `make` program allows one to work upon previous analyses without rerunning them by automatically handling dependency management. For reproducibility, we also provide Dockerfiles that encapsulate the environments to run the tools and scripts to build them.

Figure 1 depicts a diagram of the pipeline in which each square represents a module numbered by their order of execution. We refer to the numbering in the squares to explain how each module was implemented:

(1) Each fusion detection tool (Table 1) has a dedicated Makefile to orchestrate its execution. These can also serve as standalone scripts to automate a fusion calling analysis with a single tool.

(2) Each fusion detection tool has an associated script that processes its output and collects the predictions and the number of reads supporting those predictions. These results are gathered in a standardized format for further processing.

(3) Fusion partners are mapped to the GRCh37 reference standard following the Gencode v19 annotation set. The gene coordinates are freely available at `https://github.com/fusiontranscripts/FusionBenchmarking/tree/master/resources`. The mapping is efficiently computed through an interval tree as implemented by [Haas et al. 2019].

(4) Fusions are annotated by the software FusionAnnotator, which enriches the annotations with metadata for downstream analysis.

(5) We consider ambiguous fusions those whose genes are neighbors, are sequence-similar, have a gene of mitochondrial origin, or an HLA gene. These can be discarded in this step or carried on. At this point, the pipeline also fulfills the role of an ensemble detection tool.

(6)-(7) These modules are specialized for plotting and measuring run-time. We modified scripts from [Haas et al. 2019] and implemented some of our own to plot PR curves and compute their AUC.

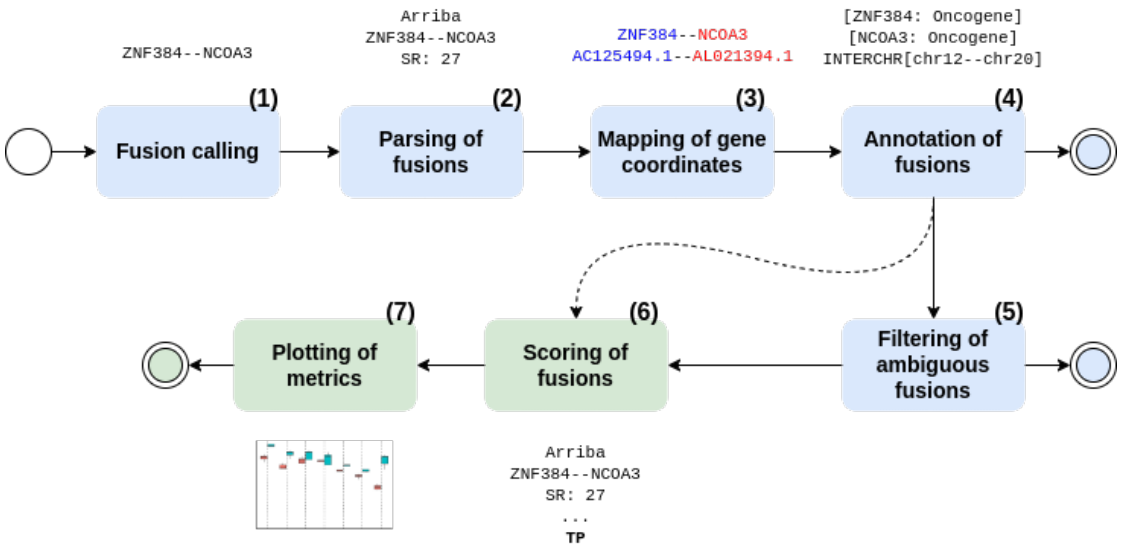

Figure 1. Computational pipeline to detect fusions and benchmark tools. Each square represents a module, the white circle represents the starting point, and filled circles represent ending points. The blue modules, from steps (1) to (4)-(5), constitute an ensemble gene fusion detection pipeline, which may include step (5) for filtering. Green modules add the benchmark utilities that score the tools and plot the results. For the sake of example, the modules are labelled with a simplified view of their outputs for the fusion *ZNF384-NCOA3* retrieved by Arriba.

3.2. Performance on simulated dataset

Figure 2 depicts the AUC of the PR curves computed for each tool after running on each group of datasets (paired-end reads of length 50bp and 101bp). We noticed that read length affected the performance of all tools, although the alignment-free ones proved to be less sensitive to read length. Since longer reads carry more information, it is expected that any tool would produce better results with increased read length. Overall, the alignment-based tools FusionCatcher, STAR-Fusion, and Arriba had superior performance. However, FusionCatcher demands significantly more computational resources when compared to STAR-Fusion and Arriba: the former performs multiple alignment steps (see Table 1), while the latter produces comparable results with a single alignment step.

Although the accuracy of CICERO greatly improved with increased read length, it was the worst performer among the alignment-based methods. We noted that the local assembly subroutine benefits the recall of CICERO, but hampers its precision because many false positives are generated. The discrepancy in accuracy between the two groups of datasets is most likely due to the degradation of the assembly routine of CAP3 in shorter reads.

We remark that, although the alignment-free tools performed worse than the alignment-based ones, their execution time is far better. In our experiments, the processing of many datasets finished in the order of minutes, while with the alignment-based ones, most took hours. In this regard, Pizzly is the best performer because all of its analyses did not take more than 1 hour. Here, we do not take into account the execution time to

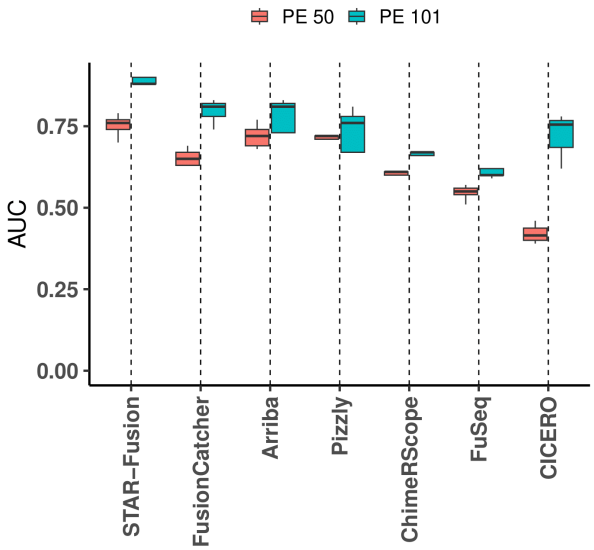

Figure 2. Area under the curve of PR curves for each tool.

build the indexes used by the tools, which is a one-time endeavor. Finally, a drawback we observed with ChimeRScope is the excessive memory peak it requires ($\approx$ 40GB), which can impair its usage in commodity hardware.

3.3. Performance on targeted RNA-seq dataset

Figures 3 and 4 depict the average AUC of the tools after running on the classical and B-other samples, respectively. FusionCatcher, STAR-Fusion, and FuSeq were the best performers, while CICERO was the worst.

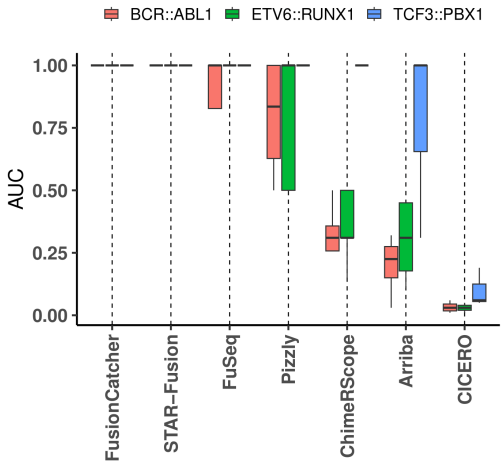

Figure 3. Classical samples.

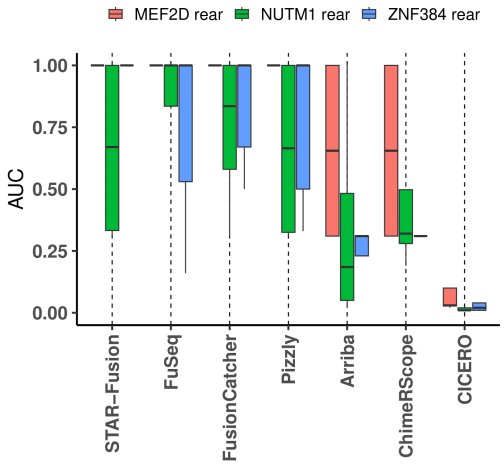

Figure 4. B-other samples.

CICERO retrieved all relevant gene fusions, but also many others with higher evidence that we did not consider as true positives. In addition, for a gene X, we noticed

that many fusions were of the form *X-X*, which suggests that they are actually pseudogenes. These were not validated *in vitro*, and for the moment were considered false positives. Again, this indicates that the local assembly subroutine hampers precision, although helpful to discover non canonical fusions.

In contrast to its accuracy on the simulated dataset, FuSeq performed remarkably well in real data. Although it was the worst performer in our simulated data compared to the other alignment-free tools, [Singh and Li 2021] reported that FuSeq was a top-performer in their comprehensive review. On the other hand, ChimeRScope was consistently behind the other alignment-free tools in our experiments and in the benchmark of [Singh and Li 2021]. Finally, the results we obtained with Pizzly were consistent with what has been reported previously [Haas et al. 2019, Singh and Li 2021].

Regarding the novel gene partners in the B-other samples, we remark that all tools retrieved the relevant fusions, except for FusionCatcher who missed the *SPI1-ZNF384* fusion.

4. Discussion

A noteworthy pattern we observed in the results is that most gene fusions retrieved by alignment-free tools are true positives, but they often miss other true positives in the dataset. Alignment-based tools, on the other hand, are able to retrieve most true positives, but they also produce many false positives. This is a known limitation of alignment-based methods: due to the aberrant nature of cancer genomes, aligners often produce spurious alignments when mapping these genomes reads against a reference if one does not fine-tune the parameters.

Hence, a fruitful strategy is to perform a first round of fusion calling with alignment-based tools followed by a second round with alignment-free tools to pinpoint potential false positives and validate the true positives from the first round. This is a cheap heuristic because alignment-free tools often have fast execution time. Notice, however, that this does not eliminate validation in a wet lab. We also remark that the alignment-free tool FuSeq had excellent accuracy in our real dataset and requires considerably less computational resources compared to STAR-Fusion. The notable performance of FuSeq has also been reported in a recent review [Singh and Li 2021].

The current version of the pipeline has some limitations. First, adding new tools is a nontrivial task that requires programming knowledge, although done in a few steps: one has to write a wrapper script and an output parser for the tool at hand. As for the current tools, we expect that they will remain compatible with the pipeline as long as their output format and command line interface does not change. Second, our ensemble approach for fusion detection is naive compared to other pipelines that perform a similar task [LaHaye et al. 2021, Vicente-Garcés et al. 2023, Apostolides et al. 2021]: we do not apply a consensus algorithm, but rather keep all the retrieved fusions in a standardized format. Third, the pipeline lacks a run-time and space usage measurement feature, but we will include this functionality for an extended version of this paper.

5. Conclusion

In this paper, we presented a computational pipeline to compare different fusion detection methods for short-read RNA-seq data. It can be used as an ensemble fusion detection

tool and its application to simulated and real data supplements current benchmarks in the literature. In our benchmark, STAR-Fusion was the top-performer in both the simulated and real datasets, while CICERO was the worst. We concluded that the poorer accuracy of the latter was mostly due to its local assembly subroutine, which can be helpful to detect non canonical fusions, but increases the number of false positives. This downside is aggravated with shorter reads as shown in our experiment with simulated data. In addition, we noticed that alignment-free tools exhibited superior precision and faster execution time, which suggests that they can be used separately to complement alignment-based tools for fusion calling.

Future research directions for this work include adding more tools to the pipeline and extending the experiments to more real cancer cell lines. Toward this end, our group is currently working to extend the experiments to include the entirety of our targeted RNA-seq dataset and also to other bulk RNA-seq datasets.

6. Acknowledgements

We thank the anonymous reviewers for their helpful suggestions to improve this paper. This work was supported by grants #2023/05887-5, #2018/00031-7, and #2024/01200-8, São Paulo Research Foundation (FAPESP).

References

[Apostolides et al. 2021] Apostolides, M., Jiang, Y., Husić, M., Siddaway, R., Hawkins, C., Turinsky, A. L., Brudno, M., and Ramani, A. K. (2021). MetaFusion: a high-confidence metacaller for filtering and prioritizing RNA-seq gene fusion candidates. *Bioinformatics*, 37(19):3144–3151.

[Bray et al. 2016] Bray, N. L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature Biotechnology*, 34(5):525–527. Number: 5 Publisher: Nature Publishing Group.

[Carrara et al. 2013] Carrara, M., Beccuti, M., Lazzarato, F., Cavallo, F., Cordero, F., Donatelli, S., and Calogero, R. A. (2013). State-of-the-Art Fusion-Finder Algorithms Sensitivity and Specificity. *BioMed Research International*, 2013:e340620. Publisher: Hindawi.

[Creason et al. 2021] Creason, A., Haan, D., Dang, K., Chiotti, K. E., Inkman, M., Lamb, A., Yu, T., Hu, Y., Norman, T. C., Buchanan, A., van Baren, M. J., Spangler, R., Rollins, M. R., Spellman, P. T., Rozanov, D., Zhang, J., Maher, C. A., Caloian, C., Watson, J. D., Uhrig, S., Haas, B. J., Jain, M., Akeson, M., Ahsen, M. E., Zhang, H., Wang, Y., Guan, Y., Nguyen, C., Sugai, C., Jha, A., Li, J. W., Dobin, A., Stolovitzky, G., Guinney, J., Boutros, P. C., Stuart, J. M., and Ellrott, K. (2021). A community challenge to evaluate RNA-seq, fusion detection, and isoform quantification methods for cancer discovery. *Cell Systems*, 12(8):827–838.e5.

[Dobin et al. 2013] Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T. R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*, 29(1):15–21.

[Frankish et al. 2022] Frankish, A., Carbonell-Sala, S., Diekhans, M., Jungreis, I., Loveland, J. E., Mudge, J. M., Sisu, C., Wright, J. C., Arnan, C., Barnes, I., Banerjee, A.,

Bennett, R., Berry, A., Bignell, A., Boix, C., Calvet, F., Cerdán-Vélez, D., Cunningham, F., Davidson, C., Donaldson, S., Dursun, C., Fatima, R., Giorgetti, S., Giron, C. G., Gonzalez, J. M., Hardy, M., Harrison, P. W., Hourlier, T., Hollis, Z., Hunt, T., James, B., Jiang, Y., Johnson, R., Kay, M., Lagarde, J., Martin, F. J., Gómez, L. M., Nair, S., Ni, P., Pozo, F., Ramalingam, V., Ruffier, M., Schmitt, B. M., Schreiber, J. M., Steed, E., Suner, M.-M., Sumathipala, D., Sycheva, I., Uszczynska-Ratajczak, B., Wass, E., Yang, Y. T., Yates, A., Zafrulla, Z., Choudhary, J. S., Gerstein, M., Guigo, R., Hubbard, T. J. P., Kellis, M., Kundaje, A., Paten, B., Tress, M. L., and Flicek, P. (2022). GENCODE: reference annotation for the human and mouse genomes in 2023. *Nucleic Acids Research*, 51(D1):D942–D949.

[Haas et al. 2023] Haas, B. J., Dobin, A., Ghandi, M., Van Arsdale, A., Tickle, T., Robinson, J. T., Gillani, R., Kasif, S., and Regev, A. (2023). Targeted *in silico* characterization of fusion transcripts in tumor and normal tissues via FusionInspector. *Cell Reports Methods*, 3(5):100467.

[Haas et al. 2019] Haas, B. J., Dobin, A., Li, B., Stransky, N., Pochet, N., and Regev, A. (2019). Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biology*, 20(1):213.

[Huang and Madan 1999] Huang, X. and Madan, A. (1999). CAP3: A DNA Sequence Assembly Program. *Genome Research*, 9(9):868–877. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[Kent 2002] Kent, W. J. (2002). BLAT—The BLAST-Like Alignment Tool. *Genome Research*, 12(4):656–664. Company: Cold Spring Harbor Laboratory Press Distributor: Cold Spring Harbor Laboratory Press Institution: Cold Spring Harbor Laboratory Press Label: Cold Spring Harbor Laboratory Press Publisher: Cold Spring Harbor Lab.

[Kumar et al. 2016] Kumar, S., Vo, A. D., Qin, F., and Li, H. (2016). Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Scientific Reports*, 6(1):21597.

[LaHaye et al. 2021] LaHaye, S., Fitch, J. R., Voytovich, K. J., Herman, A. C., Kelly, B. J., Lammi, G. E., Arbesfeld, J. A., Wijeratne, S., Franklin, S. J., Schieffer, K. M., Bir, N., McGrath, S. D., Miller, A. R., Wetzel, A., Miller, K. E., Bedrosian, T. A., Leraas, K., Varga, E. A., Lee, K., Gupta, A., Setty, B., Boué, D. R., Leonard, J. R., Finlay, J. L., Abdelbaki, M. S., Osorio, D. S., Koo, S. C., Koboldt, D. C., Wagner, A. H., Eisfeld, A.-K., Mrózek, K., Magrini, V., Cottrell, C. E., Mardis, E. R., Wilson, R. K., and White, P. (2021). Discovery of clinically relevant fusions in pediatric cancer. *BMC Genomics*, 22(1):872.

[Langmead and Salzberg 2012] Langmead, B. and Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 9(4):357–359. Publisher: Nature Publishing Group.

[Li et al. 2017] Li, Y., Heavican, T. B., Vellichirammal, N. N., Iqbal, J., and Guda, C. (2017). ChimeRScope: a novel alignment-free algorithm for fusion transcript prediction using paired-end RNA-Seq data. *Nucleic Acids Research*, 45(13):e120.

[Liu et al. 2016] Liu, S., Tsai, W.-H., Ding, Y., Chen, R., Fang, Z., Huo, Z., Kim, S., Ma, T., Chang, T.-Y., Priedigkeit, N. M., Lee, A. V., Luo, J., Wang, H.-W., Chung, I.-F., and Tseng, G. C. (2016). Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data. *Nucleic Acids Research*, 44(5):e47.

[Melsted et al. 2017] Melsted, P., Hateley, S., Joseph, I. C., Pimentel, H., Bray, N., and Pachter, L. (2017). Fusion detection and quantification by pseudoalignment.

[Migita et al. 2023] Migita, N. A., Jotta, P. Y., Nascimento, N. P. d., Vasconcelos, V. S., Centoducatte, G. L., Massirer, K. B., Azevedo, A. C. d., Brandalise, S. R., and Yunes, J. A. (2023). Classification and genetics of pediatric B-other acute lymphoblastic leukemia by targeted RNA sequencing. *Blood Advances*, 7(13):2957–2971.

[Nicorici et al. 2014] Nicorici, D., Şatalan, M., Edgren, H., Kangaspeska, S., Murumägi, A., Kallioniemi, O., Virtanen, S., and Kilkku, O. (2014). FusionCatcher – a tool for finding somatic fusion genes in paired-end RNA-sequencing data. Pages: 011650 Section: New Results.

[Singh and Li 2021] Singh, S. and Li, H. (2021). Comparative study of bioinformatic tools for the identification of chimeric RNAs from RNA Sequencing. *RNA Biology*, 18(sup1):254–267.

[Srivastava et al. 2016] Srivastava, A., Sarkar, H., Gupta, N., and Patro, R. (2016). RapMap: a rapid, sensitive and accurate tool for mapping RNA-seq reads to transcriptomes. *Bioinformatics*, 32(12):i192–i200.

[Tian et al. 2020] Tian, L., Li, Y., Edmonson, M. N., Zhou, X., Newman, S., McLeod, C., Thrasher, A., Liu, Y., Tang, B., Rusch, M. C., Easton, J., Ma, J., Davis, E., Trull, A., Michael, J. R., Szlachta, K., Mullighan, C., Baker, S. J., Downing, J. R., Ellison, D. W., and Zhang, J. (2020). CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biology*, 21(1):126.

[Uhrig et al. 2021] Uhrig, S., Ellermann, J., Walther, T., Burkhardt, P., Fröhlich, M., Hutter, B., Toprak, U. H., Neumann, O., Stenzinger, A., Scholl, C., Fröhling, S., and Brors, B. (2021). Accurate and efficient detection of gene fusions from RNA sequencing data. *Genome Research*, 31(3):448–460.

[Vicente-Garcés et al. 2023] Vicente-Garcés, C., Maynou, J., Fernández, G., Esperanza-Cebollada, E., Torrebadell, M., Català, A., Rives, S., Camós, M., and Vega-García, N. (2023). Fusion InPipe, an integrative pipeline for gene fusion detection from RNA-seq data in acute pediatric leukemia. *Frontiers in Molecular Biosciences*, 10.

[Vu et al. 2018] Vu, T. N., Deng, W., Trac, Q. T., Calza, S., Hwang, W., and Pawitan, Y. (2018). A fast detection of fusion genes from paired-end RNA-seq data. *BMC Genomics*, 19(1):786.