

Comparative Analysis of Fusion Detection Algorithms

Amy Wang, Computational Biology Honors Thesis

First Reader: Ece Uzun, PhD

Second Reader: Eric Morrow, MD, PhD

ABSTRACT

Background: Gene fusions are genomic alterations which function as diagnostic biomarkers for certain cancer types and are thus valuable subjects of research for understanding oncogenesis. Given the constant development of new software and sequencing technologies to identify genomic variations including fusions, bioinformaticians need guidance to select among the available programs. In this study, we conducted a comparative analysis of 8 fusion detection algorithms.

Methods: We reviewed primary literature and selected 8 open-source programs developed during or after 2014: Arriba, deFuse, FusionCatcher, FuSeq, InFusion, JAFFA, pizzly, and STAR-Fusion. We installed and evaluated those programs for ease of installation, efficient performance and accuracy. An existing RNA-seq dataset from patients with multiple tumor types was used to test the programs.

Results: We found that the evaluated metrics varied widely across programs. JAFFA was both the most sensitive and least specific among all programs tested. FuSeq, deFuse and InFusion detected the least number of expected fusions, and STAR-Fusion provided the best balance between sensitivity, specificity, and efficiency.

Conclusions: This study provides guidance on choosing the optimum fusion detection program for certain datasets and/or purposes.

BACKGROUND

Gene fusions occur when two previously independent genes are merged into one or more hybrid genes, often as products of somatic mutations such as inversions or translocations¹. Such gene fusions have often been found to be associated with tumor growth, with certain fusions recognized as diagnostic or prognostic biomarkers for certain cancer types— as in the case of *BCR-ABL1*, which is the “Philadelphia chromosome,” a mutant chromosome associated chronic myelogenous leukemia². Thus, gene fusions are important for cancer diagnosis as well as research, and the identification of gene fusions is an important step towards understanding oncogenesis. Since the advent of large-scale genome sequencing, many programs have been developed to analyze genomic data, including dozens of software packages that aim to discover potentially gene fusions. Both in clinical and research settings, this software can prove useful in discovering candidate fusions associated with disease. However, the available fusion detection packages (FDPs) require input files in different formats, utilize different amounts of memory and runtime, and perform with different levels of sensitivity and specificity. Given the number and lack of uniformity of competing FDPs, usage cases vary depending on the clinical or research setting; hence there is a need for a holistic summary and comparison of these programs.

Several studies have attempted to provide such holistic evaluations of multiple FDPs. In 2016, Kumar et al. analyzed the performance of 12 FDPs on simulated positive, negative and mixed datasets, along with a test dataset comprising RNA-Seq data from human prostate cells³. In 2019, Haas et al. conducted a similar evaluation on 23 FDPs, using RNA-Seq data from simulated test sets as well as stem cell lines⁴. Although these publications serve an important purpose in the field of genetic research, there remains a need for a comparative review of FDPs as applied to genomic data obtained from patients. In this study, we aim to provide an objective evaluation of multiple recent FDPs on human tumor data. We reviewed primary literature on fusion detection algorithms and selected 8 open-source programs most recently updated during or after 2014: Arriba⁵, deFuse⁶, FusionCatcher⁷, FuSeq⁸, InFusion⁹, JAFFA¹⁰, pizzly¹¹, and STAR-Fusion¹². This study reports overall patterns in program performance on a dataset collected from a previous RNA-Seq pipeline study¹³, as well as relative accuracy and efficiency and any consequent implications for clinical or scientific use.

RESULTS

We initially planned to test 15 fusion detection programs for this study. Our original program list was later changed, with certain programs excluded because they did not fit the specifications of the project, and/or issues with program installation. In addition to the 8 programs in our final list, we attempted to test the following software, but due to time constraints and difficulties installing and running, were unable to include them in our final list: FusionScan¹⁴, GeneFuse¹⁵, GFusion¹⁶, IDP-Fusion¹⁷, Integrate¹⁸, PRADA¹⁹, QueryFuse/FusionQuery²⁰.

We found through the installation process that some programs were easier to install and run than others, due to the existence of bootstrap methods (FusionCatcher), conda or Docker methods of installation (FusionCatcher, STAR-Fusion), and active forums for installation help (Arriba, InFusion). Furthermore, output formats varied: some programs produced easily viewable TSV or CSV files, whereas others' output was difficult to parse—for example, pizzly formats its output as a JSON file that must be read via JSON editor, where individual genes. Output interpretability also relied significantly on program specificity, as sorting through fusions was simpler when fewer false positives were reported.

Positive dataset

All of the programs we tested detected a majority of the known fusions recorded by Winters et al (Table 1, Figure 1). Across all programs tested, JAFFA demonstrated the highest sensitivity, detecting 91%, or 22 of the 25 known fusion(s), and deFuse, FuSeq, and InFusion demonstrated the lowest sensitivity, detecting 60% of the known fusion(s) (Table 2). Positive predictive value was low among all programs, with InFusion achieving an overall maximum precision of 52%, and FusionCatcher performing with an overall minimum precision of 32%.

Table 1: Known fusions detected by each software package, in comparison to those detected by Winters et al.

Sample	Histological Type	Known Fusions	Winters et. al	Arriba	defuse	FusSeq	FusionCatcher	Infusion	JAFPA	pizzly	STAR-Fusion
v1-11	Acute myeloid leukemia	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1	RUNX1-RUNX1T1
v11-153	Non small cell lung cancer	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1	SLC34A2-ROS1
v12-161	chondrosarcoma	EWSR1-NR4A3	HEY1-NCOA2	HEY1-NCOA2	HEY1-NCOA2	HEY1-NCOA2	HEY1-NCOA2	HEY1-NCOA2	HEY1-NCOA2	HEY1-NCOA2	HEY1-NCOA2
v12-165	Prostate adenocarcinoma	REK1-PIK3CD	REK1-PIK3CD	REK1-PIK3CD	None	REK1-PIK3CD	REK1-PIK3CD	REK1-PIK3CD	REK1-PIK3CD	REK1-PIK3CD	REK1-PIK3CD
v1-26	Acute promyelocytic leukemia	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA
v3-40	B-cell lymphoma	LP1-FOXP1	None	None	None	None	None	None	None	None	None
v3-42	Acute promyelocytic leukemia	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA	PML-RARA
v3-65	Anaplastic large T-cell lymphoma	NPM1-ALK	NPM1-ALK	NPM1-ALK	NPM1-ALK	NPM1-ALK	NPM1-ALK	None	NPM1-ALK	NPM1-ALK	NPM1-ALK
v6-83	Chronic myeloid leukemia	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3	BCR-ABL1 NUP214-XKR3
v6-84	Cholangiocarcinoma	NUP214-XKR3 FGFR2-BICC1	NUP214-XKR3 FGFR2-BICC1	FGFR2-BICC1	FGFR2-BICC1	FGFR2-BICC1	FGFR2-BICC1	FGFR2-BICC1	FGFR2-BICC1	FGFR2-BICC1	FGFR2-BICC1
v6-86	Prostate adenocarcinoma	TMPRSS2-ERG	TMPRSS2-ERG	None	None	None	TMPRSS2-ERG	None	TMPRSS2-ERG	None	None
v6-89	Diffuse histiocytic lymphoma	NPM1-ALK	NPM1-ALK	NPM1-ALK	NPM1-ALK	None	NPM1-ALK	None	NPM1-ALK	NPM1-ALK	NPM1-ALK
v7-93	Liposarcoma	None	MDM2-TMPO	MDM2-TMPO	MDM2-TMPO	MDM2-TMPO	MDM2-TMPO	MDM2-TMPO	MDM2-TMPO	MDM2-TMPO	MDM2-TMPO
v8-116	Large cell immunoblastic lymphoma	NPM1-ALK	NPM1-ALK	NPM1-ALK	None	NPM1-ALK	NPM1-ALK	None	NPM1-ALK	NPM1-ALK	NPM1-ALK
v8-117	Rhabdomyosarcoma	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1	PAX3-FOXO1
v8-118	Osteosarcoma	TP53-VAV1	TP53-VAV1	TP53-VAV1	None	None	TP53-VAV1	TP53-VAV1	TP53-VAV1	None	TP53-VAV1
v9-123	Synovial sarcoma	SS18-SSX1	SS18-SSX1 PIEZO1-CBFA2T3	SS18-SSX1 PIEZO1-CBFA2T3	SS18-SSX1	SS18-SSX1	SS18-SSX1 PIEZO1-CBFA2T3	SS18-SSX1 PIEZO1-CBFA2T3	SS18-SSX1	SS18-SSX1	SS18-SSX1 PIEZO1-CBFA2T3
v9-124	Colon adenocarcinoma	GRHL2-MAP2K2	None	None	None	None	GRHL2-MAP2K2	None	GRHL2-MAP2K2	None	GRHL2-MAP2K2
v9-125	Bladder transitional cell carcinoma	FGFR3-TACC3	FGFR3-TACC3	FGFR3-TACC3	FGFR3-TACC3	None	FGFR3-TACC3	FGFR3-TACC3	FGFR3-TACC3	FGFR3-TACC3	FGFR3-TACC3
v11-151	Squamous cell carcinoma	BRD4-C15orf55	BRD4-C15orf55 GTF2I-RD1-CLIP2	BRD4-C15orf55 GTF2I-RD1-CLIP2	BRD4-C15orf55	BRD4-C15orf55	Failed	None	BRD4-C15orf55 GTF2I-RD1-CLIP2	BRD4-NUTM1	BRD4-C15orf55
# Fusions Successfully Found		21	22	20	15	15	20	15	22	17	20

Figure 1: Number of fusions detected per sample by each software package, versus total “known” fusions per sample. “Known” fusions were defined as the number of predetermined “known” fusions listed in Winters et al, in addition to fusions newly found and validated by Winters et al.

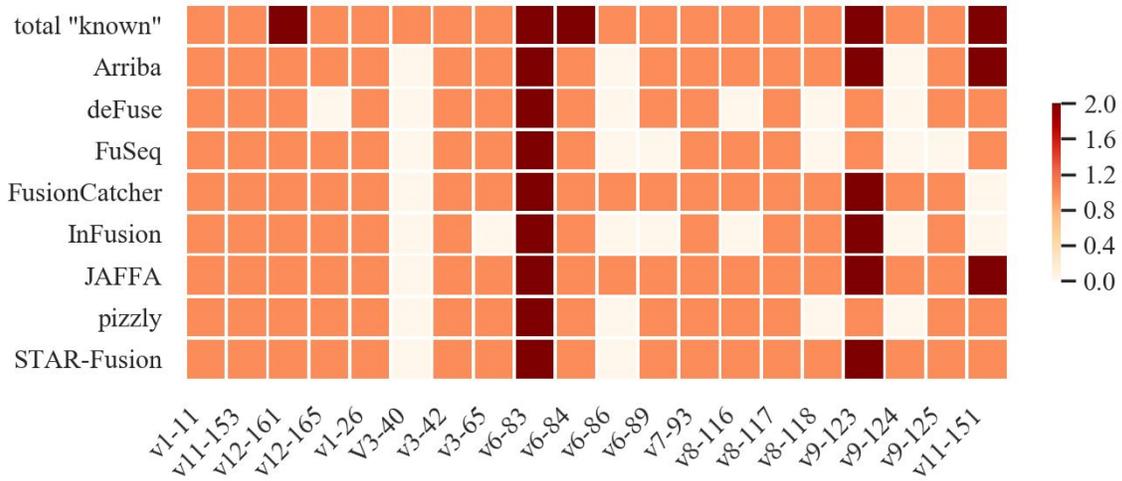


Figure 2: In comparison to the findings of Winters et al, number of samples in which known fusions were successfully detected by each algorithm. Sensitivity (%) = (True Positives/Total Fusions)

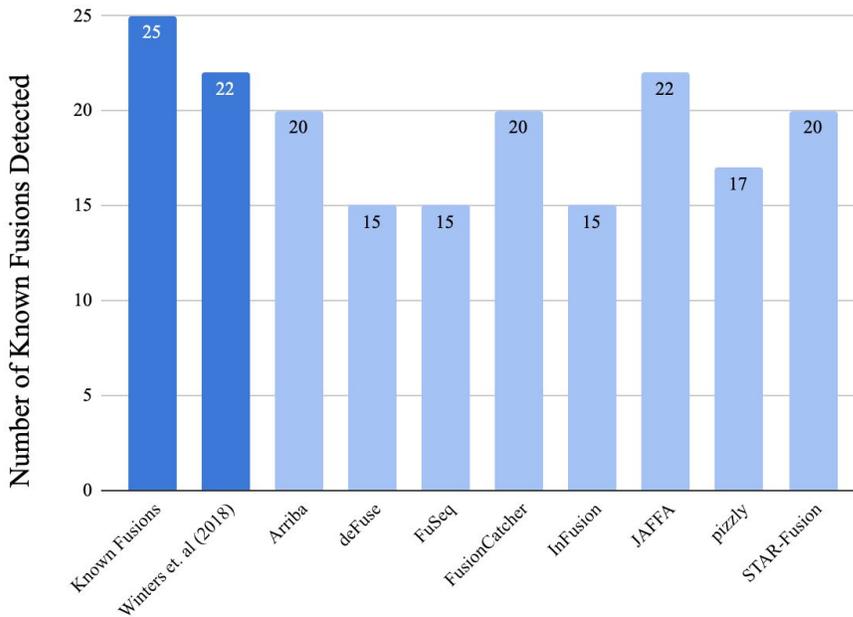


Table 2: Accuracy metrics— Positive predictive value (PPV) and sensitivity of each program.

Program	True Pos.	False Pos.	Positive Predictive Value (Precision)	Sensitivity
Arriba	20	36	0.36	0.80
deFuse	15	19	0.44	0.60
FuSeq	15	21	0.42	0.60
FusionCatcher	20	42	0.32	0.80
Infusion	15	14	0.52	0.60
JAFFA	22	42	0.34	0.88
pizzly	17	31	0.35	0.68
STAR-Fusion	20	36	0.36	0.80

Unreported fusion candidates were identified in many positive samples, with unreported fusion candidates defined as any fusion candidate with supporting reads greater than or equal to the minimum number of supporting reads for a known fusion, using per-program supporting read parameters (Methods, Table 4). The initial pool of unreported fusion candidates was filtered down to consider only fusions that had been detected by multiple programs and previously reported in literature. Such fusions were found in several samples, suggesting the presence of heretofore unidentified tumorigenic fusions (Table 3).

In total, we identified the following unreported fusion candidates, not reported by Winters et al: *MIPOL1-DGKB*²¹, *HPR-MRPS10*²², *KANSL1-ARL17A/B*²³, *TPM4-KLF2*²⁴, *PAIP2-MATR3*²⁵, *FBXO11-MAP2K5*²⁶, *MARS-AVIL*²⁶, *MCMBP-PDCD1*²⁶, *SETD5-NUDT9*²⁶, *DNER-ELL2*²⁷, *MROH1-PARP10*²⁷, *MTAP-BNC2*²⁷, *SLC20A2-SFRP1*²⁸.

Program efficiency

Of all programs tested, Pizzly's runtimes were shortest, and STAR-Fusion used the least memory. STAR-Fusion performed well with respect to both average runtime and average memory use, with an average runtime only slightly less than that of Pizzly. JAFFA took the longest to run, and deFuse consumed the most memory (Figure 3, Table 6).

Figure 3: Plot of runtime (hours) versus memory usage (AveRSS, in units of K) for three fusion detection programs. As a metric of memory usage, we used the average slurm-reported value AveRSS—which denotes the average resident set size of all tasks in a given slurm job—across all samples.

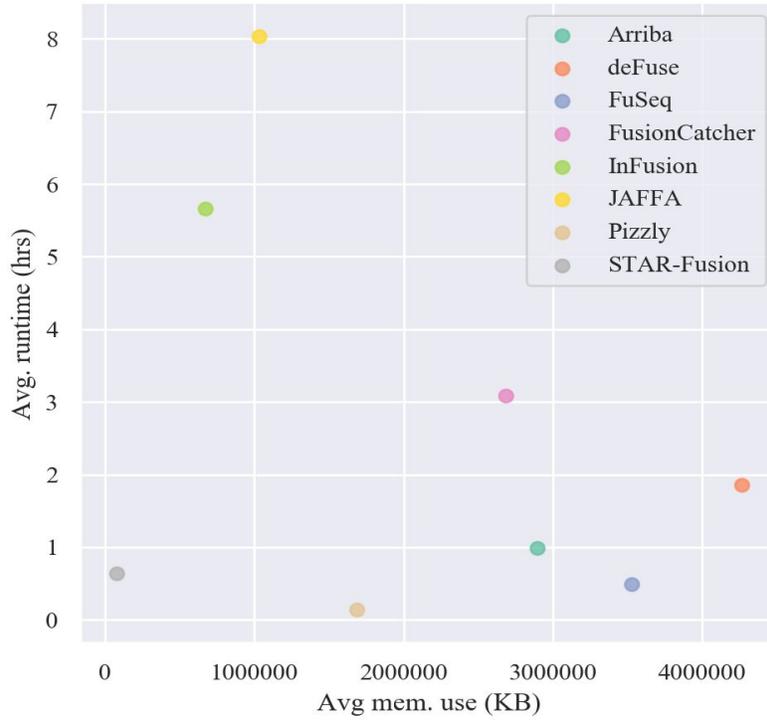


Table 6: Runtime and memory usage for the 8 detection programs.

Program name	Avg mem. use (GB)	Avg. runtime (hrs)
Arriba	2.89	0.99
deFuse	4.26	1.86
FuSeq	3.53	0.49
FusionCatcher	2.70	3.09
InFusion	0.67	5.66
JAFFA	1.03	8.04
Pizzly	1.69	0.14
STAR-Fusion	0.08	0.64

Table 3: Commonly observed false positive or unreported fusion candidates in the positive dataset. Starred fusions have appeared in literature and are possible unreported fusions.

Sample ID	Histological Type	Fusion	Detected By
v1-11	Acute myeloid leukemia	CHP1-FAM189A1	FusionCatcher, JAFFA, pizzly, STAR-Fusion
v11-151	Squamous cell carcinoma	SENP6-KHDC1	Arriba, FuSeq, JAFFA, STAR-Fusion
v11-153	Non small cell lung cancer	GMEB1-RCC1	Arriba, FuSeq, FusionCatcher, JAFFA, pizzly
		PEX3-AIG1	Arriba, FuSeq, FusionCatcher, InFusion, JAFFA, pizzly, STAR-Fusion
v12-165	Prostate adenocarcinoma	CCSER2-CYP2C19	Arriba, FuSeq, FusionCatcher, InFusion, JAFFA, pizzly, STAR-Fusion
		FAM117B-BMPR2	Arriba, FusionCatcher, InFusion, JAFFA, pizzly, STAR-Fusion
		GPS2-MPP2	Arriba, deFuse, FusionCatcher, JAFFA, pizzly
		MIPOL1-DGKB*	Arriba, FuSeq, FusionCatcher, InFusion, JAFFA, pizzly, STAR-Fusion
		MRPS10-HPR*	Arriba, FusionCatcher, pizzly, STAR-Fusion
		SNX9-CYP2C19	Arriba, FusionCatcher, JAFFA, STAR-Fusion
v3-40	B-cell lymphoma	ACACB-HVCN1	Arriba, FusionCatcher, JAFFA, STAR-Fusion
		KANSL1-ARL17A*	FuSeq, FusionCatcher, JAFFA, pizzly
v3-42	Acute progranulocytic leukemia	KANSL1-ARL17B*	deFuse, FusionCatcher, JAFFA, pizzly
		TPM4-KLF2*	FuSeq, FusionCatcher, JAFFA, pizzly
v6-83	Chronic myeloid leukemia	BAG6-SLC44A4	Arriba, deFuse, FuSeq, FusionCatcher, JAFFA, pizzly, STAR-Fusion
		C16orf87-ORC6	Arriba, FusionCatcher, JAFFA, STAR-Fusion
		KANSL1-ARL17A*	deFuse, FuSeq, FusionCatcher, JAFFA, pizzly
		KIAA1958-HSDL2	Arriba, FusionCatcher, JAFFA, pizzly, STAR-Fusion
v6-86	Prostate adenocarcinoma	TMEFF2-FASTKD2	Arriba, FusionCatcher, JAFFA, STAR-Fusion
v6-89	Diffuse histiocytic lymphoma	HSD17B7-RGS4	Arriba, FusionCatcher, JAFFA, STAR-Fusion
		MEGF8-CNFN	Arriba, deFuse, FusionCatcher, JAFFA, STAR-Fusion
		OAZ1-FKBP8	Arriba, FusionCatcher, JAFFA, STAR-Fusion
		PAIP2-MATR3*	Arriba, deFuse, InFusion, JAFFA, STAR-Fusion
		PFKFB3-RBM17	Arriba, FusionCatcher, JAFFA, pizzly, STAR-Fusion
		SPOP-MYT1	Arriba, deFuse, FuSeq, FusionCatcher, JAFFA, pizzly, STAR-Fusion
		TOP1-NCOA3	Arriba, FusionCatcher, JAFFA, STAR-Fusion
		VMP1-FLCN	deFuse, Arriba, FusionCatcher, InFusion, JAFFA, pizzly, STAR-Fusion
v7-93	Liposarcoma	DAZAP1-AVIL	Arriba, deFuse, FuSeq, FusionCatcher, InFusion, JAFFA, STAR-Fusion
		SCAF8-LYZ	Arriba, deFuse, FuSeq, FusionCatcher, InFusion, JAFFA, pizzly, STAR-Fusion
		TMBIM4-CYP4F22	FuSeq, FusionCatcher, InFusion, JAFFA
v8-116	Large cell immunoblastic lymphoma	KANSL1-ARL17A*	deFuse, FuSeq, FusionCatcher, JAFFA, pizzly
v8-117	Rhabdomyosarcoma	FBXO11-MAP2K5*	FusionCatcher, InFusion, JAFFA, pizzly
		MARS-AVIL*	Arriba, deFuse, FuSeq, FusionCatcher, InFusion, JAFFA, pizzly, STAR-Fusion
		MCMBP-PDCD11*	Arriba, FuSeq, FusionCatcher, InFusion, STAR-Fusion, deFuse
		RNF219-AKAP11	Arriba, FusionCatcher, JAFFA, STAR-Fusion
		SETD5-NUDT9*	Arriba, FusionCatcher, JAFFA, pizzly, STAR-Fusion
v8-118	Osteosarcoma	DNER-ELL2*	Arriba, FuSeq, FusionCatcher, JAFFA, pizzly, STAR-Fusion, deFuse
		ELL2-TRIP12	Arriba, FuSeq, FusionCatcher, JAFFA, pizzly, STAR-Fusion
		MROH1-PARP10*	Arriba, FusionCatcher, JAFFA, pizzly, STAR-Fusion
		MTAP-BNC2*	Arriba, deFuse, FuSeq, FusionCatcher, JAFFA, pizzly, STAR-Fusion
		TFG-GPR128*	deFuse, InFusion, JAFFA, pizzly, STAR-Fusion
v9-124	Colon adenocarcinoma	ADAP1-PRKAR1B	Arriba, deFuse, FuSeq, FusionCatcher, JAFFA, pizzly, STAR-Fusion
		FZR1-CREB3L3	Arriba, deFuse, InFusion, pizzly, STAR-Fusion
		KANSL1-ARL17A*	deFuse, FusionCatcher, JAFFA, pizzly
v9-125	Bladder transitional cell carcinoma	PPIE-ADCY10	Arriba, FuSeq, FusionCatcher, pizzly, STAR-Fusion
		SLC20A2-SFRP1*	Arriba, FusionCatcher, JAFFA, STAR-Fusion

Negative dataset

Table 4: Commonly observed false positives in the negative dataset.

Sample ID	Histological Type	Fusion	Detected By
v10-144	Breast adenocarcinoma	BRIP1-VMP1	InFusion, JAFFA, pizzly, FusionCatcher
		CDYL-CDKAL1	Arriba, InFusion, JAFFA, pizzly, STAR-Fusion, FusionCatcher
		CMTM4-AKR1C4	Arriba, FuSeq, InFusion, JAFFA, pizzly, STAR-Fusion, FusionCatcher
		DDX42-PITPNC1	Arriba, FuSeq, InFusion, JAFFA, pizzly, STAR-Fusion, FusionCatcher
		EZR-OSTCP1	Arriba, deFuse, JAFFA, STAR-Fusion
		MAP1LC3B-ZNF821	Arriba, FuSeq, InFusion, JAFFA, STAR-Fusion, FusionCatcher
		RAB22A-PHACTR3	Arriba, JAFFA, STAR-Fusion, FusionCatcher
		SUPT4H1-CEP112	Arriba, FuSeq, InFusion, JAFFA, pizzly, STAR-Fusion, FusionCatcher
		TANC2-MTMR4	Arriba, FuSeq, InFusion, JAFFA, pizzly, STAR-Fusion, FusionCatcher
		TANC2-PSMD12	Arriba, FuSeq, InFusion, JAFFA, STAR-Fusion, FusionCatcher
		TMEM104-CDK12	FuSeq, InFusion, JAFFA, pizzly, FusionCatcher
		TOX3-GNAO1	Arriba, JAFFA, STAR-Fusion, FusionCatcher
		v6-81	Epithelioid sarcoma
v9-126	Breast adenocarcinoma	CD151-BLVRB	InFusion, JAFFA, STAR-Fusion, FusionCatcher
		ESR1-CCDC170	Arriba, deFuse, InFusion, JAFFA, pizzly, STAR-Fusion, FusionCatcher
		RNF187-OBSCN	Arriba, InFusion, JAFFA, FusionCatcher
		SPAG9-NGFR	JAFFA, FusionCatcher, Arriba, InFusion, pizzly, STAR-Fusion

Control dataset

Table 5: Commonly observed false positives in the control dataset.

Sample ID	Histological Type	Fusion	Detected By
v10-133	Normal (thyroid)	KANSL1-ARL17A	FuSeq, JAFFA, pizzly, FusionCatcher
		KANSL1-ARL17B	deFuse, FuSeq, pizzly, FusionCatcher
		TFG-GPR128	deFuse, InFusion, JAFFA, pizzly, STAR-Fusion

DISCUSSION

This study analyzes the performance of 8 open-source fusion detection software packages on RNA-Seq data from tumor cells with known oncogenic fusions. All programs demonstrated true positive detection rates of >60% among the total fusions demarcated as “known,” demonstrating reasonable sensitivity across the board. However, positive predictive value was low among all programs, with the highest PPV at only 52%. These sensitivity values fall into a similar range as those reported in previous studies, but the PPV results noticeably differ: both Haas et al. and Kumar et al. reported PPV values close to 100% for most programs tested^{3,4}. Because both previous papers used simulated data or cancer cell line data, we speculate that reduced levels of

noise in their datasets may have incurred fewer false positives than if they had used patient tumor cell data. This explanation is supported by multiple program-associated publications that found significant discrepancies in program specificity when running on real tumor datasets rather than simulated data^{11, 12}.

Only FusionCatcher and JAFFA successfully detected the known fusion *TMPRSS-ERG* in a sample with prostate adenocarcinoma. Similarly, in sample v9-124, only FusionCatcher, JAFFA, and STAR-Fusion detected the known fusion *GRHL2-MAP2K2*, which additionally went undetected by Winters et al. On the other hand, neither Winters et al, nor any of the tested programs, was able to detect *LPP-FOXP1* in sample v3-40. While the *FOXP1* gene is well-documented in B-cell lymphoma²⁹, and *LPP* fusions have been previously associated with tumorigenesis³⁰, there exists very little documentation of the *LPP-FOXP1* fusion. This is most likely due to the panel used in the sequencing, which might not have captured *LPP* and *FOX1P* well enough to be detected by the tested programs.

In samples containing more than two known fusions, programs frequently detected only one of the two. Furthermore, certain fusions appeared in multiple samples, and failed to be detected by a given program in every one of such samples, thereby potentially skewing the reported accuracy of that program. For example, InFusion failed to detect *NPM1-ALK* in any of the three samples that contained this fusion, suggesting that there is some attribute of the fusion for which InFusion's algorithm fails to account. Successful detection of this particular fusion across all three samples would heighten the reported sensitivity of InFusion from 60% to 72%. A dataset representing a greater variety known fusions would thus provide more a comprehensive reflection of program sensitivity.

Additionally, we identified a number of fusion candidates, not reported by Winters et al. Among these candidates, *KANSL1-ARL17A/B* (detected in multiple samples) and *MTAP-BNC2* (detected in the osteosarcoma sample) appeared most frequently in literature. *KANSL1-ARL17A/B* results from an inversion on chromosome 17 that fuses *KANSL1*, which codes for a protein involved in histone H4 acetylation, and *ARL17A* is involved in several carcinogenic pathways. The fusion has been found to confer cancer predisposition in individuals with European ancestry, suggesting that while these unreported fusion candidates may be germline fusions²³. Likewise, *MTAP-BNC2*, an inversion on chromosome 9 resulting in the fusion of two genes respectively encoding a methylthioadenosine phosphorylase and zinc finger protein, has been reported to be involved in osteosarcoma carcinogenesis²⁷.

In comparing runtime, half of the tested programs completed fusion calls in under 1 hour on average, with deFuse, FusionCatcher, InFusion and JAFFA taking longer. Running JAFFA consumed significantly more time than any other program—however, this fits with the

observation that JAFFA attained the highest sensitivity among all programs tested. Of all the programs, Pizzly ran the fastest, in accordance with the results reported by Pizzly's developers¹¹. Memory usage varied widely, with programs like FuSeq and Arriba using 3-4 GB of memory on average per run, and others like STAR-Fusion and InFusion using less than 1 GB. It is worth noting because each program required slightly different amounts of memory and time to run, certain parameters in batch scripts—cores, nodes, allocated memory and allocated runtime—were specific to the programs being tested (Table 4, Methods). This inconsistency in parameters is likely to skew runtime comparisons. As such, for a more accurate comparison, we recommend running all the programs again using the same number of cores to evaluate efficiency.

We found that overall, program performance varied depending on the input and the metrics of evaluation. Due to this variation, it is necessary to tailor program selection to clinical or research use. For example, 60% may not be a sufficiently high sensitivity score for clinical cancer detection purposes, as the risks of not detecting an extant fusion are very high. As such, users in clinical settings opt for programs with comparatively high sensitivities, such as JAFFA or FusionCatcher. However, although JAFFA demonstrated the highest sensitivity among all tested programs, it also detected a high number of false positives and exhibited relatively long runtimes. For more precise results and a quicker turnaround rate, other programs may be more suitable. Of all programs tested, STAR-Fusion and Arriba appeared to have the best balance of sensitivity and specificity, each detecting 20 out of the 25 known fusions and achieving PPVs that lay the middle of the reported PPV range. Moreover, both programs ran fairly quickly, with average runtimes of under an hour. These findings correlate with those of Haas et al⁴. It is worth noting, finally, that we ran each program using default parameters. In practice these parameters may be adjusted, yielding higher specificities or sensitivities than our findings suggest.

CONCLUSIONS

Fusion detection software fills an important role in genomics research and the expansion of medical knowledge. This study provides an overview and evaluation of several more recent packages, geared towards scientists who may be interested in using them. Among the 8 programs tested in this study, we identified JAFFA as the most sensitive option, Pizzly as the fastest, and STAR-Fusion as the program providing the best balance across sensitivity, PPV, runtime and memory usage. Fusion detection programs continue to be developed; we therefore anticipate that other review studies will be needed. We recommend that future studies expand the size of the tested dataset, as well as the diversity of fusions and tumor types represented.

METHODS

All programs were tested using bash scripts on Brown University Center for Computation and Visualization's Oscar system, using the GRCh37 genome assembly as a reference genome. In all

cases, we used the default parameter settings for each program in our fusion calls, so as to maintain consistency across the programs' performance results.

Software packages

Arriba

v1.1.0

Program design centered around runtime and sensitivity. Filters for artifacts characteristic of gene fusions among fusion candidates generated by the STAR aligner.

Available from: <https://github.com/suhrig/arriba>

DeFuse

v0.6.2

Resolves ambiguous discordant reads using a novel algorithm, uses discordant pairs to analyze split reads, then applies an Adaboost classifier to discriminate between false positives and fusion candidates.

Available from: <https://github.com/amcpherson/defuse>

FuSeq

v1.1.2

Uses Rapmap's quasi-mapping to generate mapped reads and split reads, from which fusion gene candidates are determined. Applies filtering, statistical tests, and de novo assembly validation to produce final candidates.

Available from: <https://github.com/nghiavtr/FuSeq>

FusionCatcher

v1.10

Specifically applied to somatic fusion detection in diseased samples. Aligns reads to transcriptome with Ensembl genome annotation and Bowtie aligner, then uses an ensemble of four aligners to identify fusion junctions.

Available from: <https://github.com/ndaniel/fusioncatcher>

Note: We initially ran FusionCatcher version 1.10, released April 10, 2019. Using this version, two samples, SRR6796300 and SRR6796350, failed to complete. We decided to rerun the program, but in the time between our initial run of FusionCatcher and our rerun attempt, our initial installation had been deleted and the software had been upgraded to version 1.20. Upon installing and running v1.20, we found that the program underperformed significantly in comparison to v1.10 (sensitivity: 0.44) and failed on 4 samples. Thus we decided to conduct our analyses using the v1.10 results, excluding samples SRR6796300 and SRR6796350. Due to time constraints, we were unable to re-install v1.10, and as such, certain samples have been excluded from the calculations of average runtime and memory.

InFusion

v0.8

Developed for chimeric RNA detection, including but not limited to gene fusions. Performs detection and clustering of “split” and “bridge” reads from the Bowtie2 aligner, then filters fusion candidates.

Available from: <https://bitbucket.org/kokonech/infusion/wiki/Home>

JAFFA

v1.09

Fusion-detection pipeline consisting of three different modes, variable depending on read length: “Direct,” “Assembly,” and “Hybrid.” Each method uses Bowtie to align reads, then maps these reads to the transcriptome using a combination of BLAT and Oases.

Available from: <https://github.com/Oshlack/JAFFA/wiki>

Pizzly

v0.37.3

Utilizes kallisto to identify possible fusion junctions through pseudo-aligning reads to the reference transcriptome. False positives are then removed through additional filtering.

Available from: <https://github.com/pmelsted/pizzly>

STAR-Fusion

v1.8.1

Identifies chimeric junctions using split reads from the STAR-aligner, validates these junctions by counting supporting junction reads and spanning fragments, and utilizes a series of filters to narrow down candidate genes.

Available from: <https://github.com/STAR-Fusion/STAR-Fusion/wiki>

Datasets

The 30 total samples utilized in this project were taken from a 2018 study conducted by Winters et al, in which an RNA sequencing assay of human tissue samples was collected for the validation of a fusion detection pipeline¹⁴. Our dataset was compiled in an attempt to be as diverse as possible with regard to tissue and known fusion type. Each sample was obtained in the format of an SRA file from the NCBI Sequence Read Archive and converted into the appropriate input format for the given programs (FASTQ, FASTQ.gz, FASTA, etc). We considered pairs of reciprocal fusions (e.g. PML-RARA and RARA-PML) to be one fusion. We followed the definition of a “known fusion” given by Winters et al.: that is, one of 571 reportable genes previously described in fusions or with verified oncogenic potential¹⁴. All such known fusions were verified by Winters et al using other methods such as RT-PCR.

Positive dataset

Table 1: List of positive tumor samples tested, along with histological type, previously known and newly detected fusions discovered in each sample. We considered “positive samples” samples that contain either fusion(s) known prior to the Winters et al. study, and/or fusion(s) detected and validated by Winters et al. As such, sample v7-93 is considered a positive sample. Note that BRD4-C15orf55 is recognized by some fusion detection algorithms as BRD5-NUTM1.

Sample Number	Histological Type	Known Fusions	Detected by Winters et al	SRA ID
v1-11	Acute myeloid leukemia	RUNX1-RUNX1T1	RUNX1-RUNX1T1	SRR6796292
v11-151	Squamous cell carcinoma	BRD4-C15orf55	BRD4-C15orf55 GTF2IRD1-CLIP2	SRR6796300
v11-153	Non small cell lung cancer	SLC34A2-ROS1	SLC34A2-ROS1	SRR6796304
v12-161	chondrosarcoma	EWSR1-NR4A3	HEY1-NCOA2	SRR6796318
v12-165	Prostate adenocarcinoma	RERE-PIK3CD	RERE-PIK3CD	SRR6796326
v1-26	Acute progranulocytic leukemia	PML-RARA	PML-RARA	SRR6796328
v3-40	B-cell lymphoma	LPP-FOXP1	None	SRR6796332
v3-42	Acute progranulocytic leukemia	PML-RARA	PML-RARA	SRR6796334
v3-65	Anaplastic large T-cell lymphoma	NPM1-ALK	NPM1-ALK	SRR6796340
v6-83	Chronic myeloid leukemia	BCR-ABL1, NUP214-XKR3	BCR-ABL1, NUP214-XKR3	SRR6796352
v6-84	Cholangiocarcinoma	NUP214-XKR3, FGFR2-BICC1,	NUP214-XKR3, FGFR2-BICC1,	SRR6796354
v6-86	Prostate adenocarcinoma	TMPRSS2-ERG	TMPRSS2-ERG	SRR6796356
v6-89	Diffuse histiocytic lymphoma	NPM1-ALK	NPM1-ALK	SRR6796358
v7-93	Liposarcoma	None	MDM2-TMPO	SRR6796360
v8-116	Large cell immunoblastic lymphoma	NPM1-ALK	NPM1-ALK	SRR6796374
v8-117	Rhabdomyosarcoma	PAX3-FOXO1	PAX3-FOXO1	SRR6796376
v8-118	Osteosarcoma	TP53-VAV1	TP53-VAV1	SRR6796378
v9-123	Synovial sarcoma	SS18-SSX1	SS18-SSX1, PIEZO1-CBFA2T3	SRR6796380
v9-124	Colon adenocarcinoma	GRHL2-MAP2K2	None	SRR6796382
v9-125	Bladder transitional cell carcinoma	FGFR3-TACC3	FGFR3-TACC3	SRR6796384

Negative dataset

Table 2: List of negative samples tested. We define a negative sample as a tumor sample containing no known tumorigenic fusions.

Sample Number	Histological Type	Known Fusions	Detected by Winters et al	SRA ID
v10-144	Breast adenocarcinoma	None	None	SRR6796290
v6-80	Brain tumor, NOS	None	None	SRR6796348
v6-81	Epithelioid sarcoma	None	None	SRR6796350
v8-113	Lipoblastoma	None	None	SRR6796372
v9-126	Breast adenocarcinoma	None	None	SRR6796386

Control dataset

Table 3: List of control samples tested. We define a control sample as a sample of normal tissue containing no known tumorigenic fusions.

Sample Number	Histological Type	Known Fusions	Detected by Winters et al	SRA ID
v10-133	Normal (thyroid)	None	None	SRR6796280
v10-135	Normal (retroperitoneum)	None	None	SRR6796282
v10-136	Normal (brain)	None	None	SRR6796284
v12-158	Normal (prostate)	None	None	SRR6796312
v3-58	Normal (lung)	None	None	SRR6796338

Analysis of program performance

To gauge the sensitivity of each program, we used the typical calculation for sensitivity:

Sensitivity = True Positives / (True Positives + False Negatives),

Where a true positive was defined as any “known” fusion validated by Winters et al, including both pre-existing known fusions as well as new fusions discovered the Winters et al study, and a false negative was defined as any such “known” fusion that failed to be detected by the program in question.

To gauge the specificity of each program, we followed the method utilized by Kumar et al, using positive predictive value or precision rather than calculating specificity exactly:

Positive predictive value (PPV) = True Positives / (True Positives + False Positives),

Where a true positive was defined as any “known” fusion validated by Winters et al, and a false positive was defined as any additional fusion reported by the given program that met the criteria of a “common false positive” (Table 3). This definition of false positive was employed in order to filter out artifacts from program output, reducing the set of false positives to those that had demonstrated reproducibility. We used PPV rather than the standard specificity calculation, due to difficulty defining and quantifying “true negatives.”

As a metric of memory usage, we used the average slurm-reported value AveRSS—which denotes the average resident set size of all tasks in a given slurm job—across all samples (positive, negative control, and normal) for each program. To measure runtime, we used the average duration of the slurm job across all samples. Note that because each program required slightly different amounts of memory and time to run, certain parameters in batch scripts—cores, nodes, allocated memory and allocated runtime— were specific to the programs being tested. A list of the batch script parameters is provided below.

Table 4: Batch script parameters by program.

Program name	Cores	Nodes	Runtime (hrs)	Memory per node (GB)
Arriba	2	1	9	50
deFuse	8	1	19	60
FuSeq	1	4	9	20
FusionCatcher	4	1	48	60
InFusion	2	1	19	50
JAFFA	5	1	28	50
Pizzly	8	1	3	60
STAR-Fusion	5	1	9	100

Identifying Common False Positives and Unreported Fusions

To identify unreported fusions, we first calculated the minimum number of supporting reads for any known fusions, per program (Table 5). “Supporting reads” are metrics specific to each program; for example, Arriba catalogues the “discordant mates,” “split_reads1,” and “split_reads2” that support a given candidate fusion, whereas JAFFA reports “spanning reads” and “spanning pairs.” We then filtered each program to only fusions whose supporting reads were greater than or equal to the minimum number. We pooled the findings of all programs, and conducted a literature search of all unknown fusions that were detected by four or more programs in the same sample. Fusions that had been documented in at least one fusion database

or scientific article were considered possible unreported fusions. Fusions that had not been recorded in one of the aforementioned methods were considered false positives.

Table 5: Minimum read count criteria for identification of false positives and unreported fusions.

Program	Criteria
Arriba	Discordant mates: 1
deFuse	Split count: 23 Span count: 6
FuSeq	Supporting reads: 9
FusionCatcher	Spanning unique reads: 2 Spanning pairs: 1
InFusion	Split reads: 8 Paired reads: 0
JAFFA	Spanning reads: 1
pizzly	Pair count: 6
STAR-Fusion	Junction read count: 2 Spanning fragments: 1

ACKNOWLEDGMENTS

I would like to thank Dr. Ece Uzun for conceptualizing this project, reviewing my progress and answering my questions throughout this process, and Dr. Eric Morrow for reviewing my manuscript and being my secondary advisor. I would also like to thank my concentration advisor Dr. Sorin Istrail and the Brown Center for Computational and Molecular Biology for sponsoring this thesis; as well as the Brown Center for Computationalization and Visualization and the Brown Computational Biology Core for help with program installation and debugging.

SUPPLEMENTARY MATERIAL

For a full list of false positives with ≥ 3 supporting programs, see this file: [<All Unreported Fusions Candidates>](#)

Program run results, analysis scripts, and more detailed documentation of methodology available upon request.

REFERENCES

1. Latysheva NS, Babu MM. Discovering and understanding oncogenic gene fusions through data intensive computational approaches. *Dig Nucleic Acids Res.* 2016; doi:10.1093/nar/gkw282
2. Kang ZJ, Liu YF, Xu LZ, et al. The Philadelphia chromosome in leukemogenesis. *Dig Chin J Cancer.* 2016; doi:10.1186/s40880-016-0108-0

3. Kumar S, Vo A, Qin F et al. Comparative assessment of methods for the fusion transcripts detection from RNA-Seq data. *Dig Sci Rep.* 2016; doi:10.1038/srep21597
4. Haas BJ, Dobin A, Li B, et al. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Dig Genome Biol.* 2019; doi:10.1186/s13059-019-1842-9
5. Uhrig S, Fröhlich M, Hutter B, Brors B. Arriba – fast and accurate gene fusion detection from rna-seq data. *Dig ESMO Open.* 2018; doi:10.1136/esmoopen-2018-EACR25.427
6. McPherson A, Hormozdiari F, Zayed A, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. *Dig PLoS Comput Biol.* 2011; doi:10.1371/journal.pcbi.1001138
7. Nicorici D, Şatalan M, Edgren H, et al. FusionCatcher - a tool for finding somatic fusion genes in paired-end RNA-sequencing data. *Dig bioRxiv.* 2014; doi:10.1101/011650
8. Vu TN, Deng W, Trac QT, et al. A fast detection of fusion genes from paired-end RNA-seq data. *Dig BMC Genomics.* 2018; doi:10.1186/s12864-018-5156-1
9. Okonechnikov K, Imai-Matsushima A, Paul L, et al. InFusion: Advancing Discovery of Fusion Genes and Chimeric Transcripts from Deep RNA-Sequencing Data. *Dig PLoS One.* 2016; doi:10.1371/journal.pone.0167417
10. Davidson NM, Majewski IJ, Oshlack A. JAFFA: High sensitivity transcriptome-focused fusion gene detection. *Dig Genome Med.* 2015; doi:10.1186/s13073-015-0167-x
11. Melsted P, Hately S, Joseph IC, et al. Fusion detection and quantification by pseudoalignment. *bioRxiv.* 2017; doi:10.1101/166322
12. Haas BJ, Dobin A, Stransky N, et al. STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. *bioRxiv.* 2017; doi: 10.1101/120295
13. Winters JL, Davila JI, McDonald AM, et al. Development and verification of an RNA sequencing (RNA-Seq) assay for the detection of gene fusions in tumors. *Dig J Mol Diagn.* 2018; doi: 10.1016/j.jmoldx.2018.03.007.
14. Kim P, Jang YE, Lee S. FusionScan: accurate prediction of fusion genes from RNA-Seq data. *Dig Genomics Inform.* 2019; doi:10.5808/GI.2019.17.3.e26
15. Chen S, Liu M, Huang T, et al. GeneFuse: detection and visualization of target gene fusions from DNA sequencing data. *Dig Int J Biol Sci.* 2018; doi:10.7150/ijbs.24626.
16. Zhao J, Chen Q, Wu J, et al. GFusion: an Effective Algorithm to Identify Fusion Genes from Cancer RNA-Seq Data. *Dig Sci Rep.* 2017; doi:10.1038/s41598-017-07070-6
17. Weirather JL, Afshar PT, Clark TA, et al. Characterization of fusion genes and the significantly expressed fusion isoforms in breast cancer by hybrid sequencing. *Dig Nucleic Acids Res.* 2015; doi:10.1093/nar/gkv562
18. Zhang J, White NM, Schmidt HK, et al. INTEGRATE: gene fusion discovery using whole genome and transcriptome data. *Dig Genome Res.* 2016; doi:10.1101/gr.186114.114
19. Wandaliz TG, Zheng S, Sivachenko A, et al. PRADA: pipeline for RNA sequencing data analysis. *Dig Bioinform.* 2014; doi: 10.1093/bioinformatics/btu169
20. Tan Y. Queryfuse is a sensitive algorithm for detection of gene-specific fusions. *bioRxiv.* 2020; doi: 10.1101/2020.03.15.993089
21. Maher CA, Kumar-Sinha C, Cao X, et al. Transcriptome sequencing to detect gene fusions in cancer. *Nature.* 2009; doi:10.1038/nature07638

22. Yorukoglu D, Hach F, Swanson L, Collins CC, Birol I, Sahinalp SC. Dissect: detection and characterization of novel structural alterations in transcribed sequences. *Dig Bioinformatics*. 2012; doi:10.1093/bioinformatics/bts214
23. Zhou JX, Yang X, Ning S, et al. Identification of KANSARL as the first cancer predisposition fusion gene specific to the population of European ancestry origin. *Dig Oncotarget*. 2017; doi:10.18632/oncotarget.16385
24. Kumar A, Kankainen M, Parsons A, Kallioniemi O, Mattila P, Heckman CA. The impact of RNA sequence library construction protocols on transcriptomic profiling of leukemia. *Dig BMC Genomics*. 2017; doi:10.1186/s12864-017-4039-
25. Liang D, Tatomer DC, Luo Z, et al. The Output of Protein-Coding Genes Shifts to Circular RNAs When the Pre-mRNA Processing Machinery Is Limiting. *Dig Mol Cell*. 2017; doi:10.1016/j.molcel.2017.10.034
26. Xie Z, Babiceanu M, Kumar S, et al. Fusion transcriptome profiling provides insights into alveolar rhabdomyosarcoma. *Dig Proc Natl Acad Sci U S A*. 2016; doi:10.1073/pnas.1612734113
27. Lorenz S, Barøy T, Sun J, et al. Unscrambling the genomic chaos of osteosarcoma reveals extensive transcript fusion, recurrent rearrangements and frequent novel TP53 aberrations. *Dig Oncotarget*. 2016; doi:10.18632/oncotarget.6567
28. Souza-Santos PT, Soares Lima SC, Nicolau-Neto P, et al. Mutations, Differential Gene Expression, and Chimeric Transcripts in Esophageal Squamous Cell Carcinoma Show High Heterogeneity. *Transl Oncol*. 2018; doi:10.1016/j.tranon.2018.08.002
29. Gascoyne D, Banham A. The significance of FOXP1 in diffuse large B-cell lymphoma. *Dig Leukemia & Lymphoma*. 2017; doi: 10.1080/10428194.2016.1228932
30. Ngan E, Kiepas A, Brown CM, Siegel PM. Emerging roles for LPP in metastatic cancer progression. *Dig J Cell Commun Signal*. 2018; doi:10.1007/s12079-017-0415-5