Comprehensive evaluation of fusion transcript detection algorithms and a meta-caller to combine top performing methods in paired-end RNA-seq data

Supplementary Tables

Table S1. Summary of computational tools published since 2010.

Tools	Fusion point	Platform	Published
MapSplice	No restriction	Illumina	Wang et al., Nucleic Acids Res. (2010)
Trans-ABySS	Canonical splicing pattern	Illumina	Robertson et al., Nat Methods. (2010)
FusionSeq	No restriction	Illumina/SOLiD	Sboner et al., Genome Biol. (2010)
ShortFuse	No restriction	Illumina/SOLiD	Kinsella et al., Bioinformatics. (2011)
Comrad	No restriction	Illumina	McPherson et al., Bioinformatics. (2011b)
FusionHunter	Canonical splicing pattern	Illumina/SOLiD	Li et al., Bioinformatics. (2011)
FusionMap	Canonical splicing pattern	Illumina	Ge et al., Bioinformatics. (2011)
deFuse	No restriction	Illumina	McPherson et al., PLoS Comput Biol. (2011a)
SnowShoes-FTD	Within exon boundary	Illumina	Asmann et al., Nucleic Acids Res. (2011)
chimerascan	No restriciton	Illumina	Iyer et al., Bioinformatics. (2011)
FusionCatcher	No restriction	Illumina/SOLiD	Edgren et al., Genome Biol. (2011), Nicorici et al., bioRxiv. (2014)
TopHat-Fusion	No restriction	Illumina	Kim et al., Genome Biol. (2011)
BreakFusion	No restriction	Illumina	Chen et al., Bioinformatics. (2012)
FusionAnalyser	Within exon boundary	Illumina	Piazza et al., Nucleic Acids Res. (2012)
LifeScope		SOLiD	Sakarya et al., PLoS Comput Biol. (2012)
Bellerophontes	No restriction	Illumina	Abate et al., Bioinformatics. (2012)
FusionFinder	Within exon boundary	Illumina	Francis et al., PLoS One. (2012)
EricScript	Within exon or exon boundary	Illumina	Benelli et al., Bioinformatics. (2012)
SOAPfuse	No restriction	Illumina	Jai et al., Genome Biol. (2013)
FusionQ	No restriction	Illumina/SOLiD	Liu et al., BMC Bioinformatics. (2013)
SOAPfusion	Within exon or exon boundary	Illumina	Wu et al., Bioinformatics. (2013)
PRADA	Within exon boundary	Illumina	Torres-Garcia et al., Bioinformatics. (2014)
JAFFA	Within exon or exon boundary	Illumina	Davidson et al., Genome Medicine. (2015)

Table S2. Description of fifteen fusion detection tools and their default (or available) detection and filtering parameters.

Tool (version)	Anchor length filter	Read-through transcript filter	Supported reads filter (spanning / split)	PCR artifact filter	Homology based filter	Alignment tools	Assembly (#) / Machine learning (o)	Fusion transcript database information	Description of installation
MapSplice (V 2.1.9)	N	N	N	N	N	bowtie	N	N	Python script. Easy to install.
ShortFuse (V 0.2)	N	N	Y	N	N	bowtie	N	Y	C++ script. Easy to install.
FusionHunter (V 1.4)	10	Y	3/1	Y	Y	bowtie	N	Y	Perl script. Easy to install.
FusionMap (V 2015-03-31)	Y*	Y	Y (different definition)	Y	Y	N	o (GSNAP)	Y	Excutatable file. Easy to install.
deFuse (V 0.6.2)	10	Y	3/1	N	Y	bowtie/ BLAT	N	N	C++ script. Easy to install.
chimerascan (V 0.4.5)	10	Y	4 (total)	N	N	bowtie/ BWA	N	Y	Python script. Easy to install.
FusionCatcher (V 0.99.4b)	10	Y	3/1	N	Y	bowtie/ STAR/ BLAT/ bowtie2	# (velvet)	Y	Python script. easy to install and need to construct database
TopHat - Fusion (V 2.0.14)	10	Y	3/1	N	Y	bowtie	N	Y	Python script. Easy to install.
BreakFusion (V 1.0.1)	N	N	N	N	N	BWA/ BLAT	# (TIGRA-SV)	N	Need to install several supporing tools.
EricScript (V 0.5.1)	N	Y	3/1 (self filter)	Y	Y	BWA/ BLAT	o	N	Perl and R script. Easy to install.
SOAPfuse (V 1.26)	10	N	3/1	N	N	Soap2/ BWA/ BLAT	N	N	Perl script. Easy to install and database is available.
FusionQ (V 5)	10	N	3/1	N	Y	bowtie	# (TIGRA-SV)	N	Perl script.
SnowShoes-FTD (V 2.0)	N	Y	2/N	Y	Y	bowtie/ BWA	N	Y	Perl script. Easy to install. Reference for different read lengths.
PRADA (V1.1)	N	N	N	N	N	BWA/ BLAST	N	N	Python script. Easy to install.
JAFFA (V1.06)	N	Y	3/1 (self filter)	N	Y	bowtie2/ BLAT	N	N	Java script (groovy). Easy to install.

* **FusionMap Anchor length filter:** Min(25, Max(17,floor (Read-Length/3)))

Table S3. 150 designed fusions in the synthetic data.

Gene1ID	Gene2ID	Gene1Name	Gene2Name
ENSG00000008282	ENSG00000075790	SYPL1	BCAP29
ENSG00000011405	ENSG00000244165	PIK3C2A	P2RY11
ENSG00000022556	ENSG00000100504	NLRP2	PYGL
ENSG00000028310	ENSG00000173273	BRD9	TNKS
ENSG00000055609	ENSG00000100364	MLL3	KIAA0930
ENSG00000062038	ENSG00000044115	CDH3	CTNNA1
ENSG00000069493	ENSG00000102024	CLEC2D	PLS3
ENSG00000070501	ENSG00000196873	POLB	CBWD3
ENSG00000072041	ENSG00000182944	SLC6A15	EWSR1
ENSG00000077549	ENSG00000143549	CAPZB	TPM3
ENSG00000084710	ENSG00000026103	EFR3B	FAS
ENSG00000095539	ENSG00000181090	SEMA4G	EHMT1
ENSG00000100239	ENSG00000105810	PPP6R2	CDK6
ENSG00000100744	ENSG00000142541	C14orf129	RPL13A
ENSG00000101138	ENSG00000115946	CSTF1	PNO1
ENSG00000101945	ENSG00000116703	SUV39H1	PDC
ENSG00000105568	ENSG00000126070	PPP2R1A	EIF2C3
ENSG00000106991	ENSG00000105856	ENG	HBP1
ENSG00000109072	ENSG00000139372	SEBOX	TDG
ENSG00000110786	ENSG00000146013	PTPN5	GFRA3
ENSG00000112414	ENSG00000100276	GPR126	RASL10A
ENSG00000114120	ENSG00000171307	SLC25A36	ZDHHC16
ENSG00000114626	ENSG00000169439	ABTB1	SDC2
ENSG00000115392	ENSG00000173614	FANCL	NMNAT1
ENSG00000115935	ENSG00000163430	WIPF1	FSTL1
ENSG00000115966	ENSG00000115446	ATF2	UNC50
ENSG00000116809	ENSG00000140577	ZBTB17	CRTC3
ENSG00000118557	ENSG00000051382	PMFBP1	PIK3CB
ENSG00000119139	ENSG00000188554	TJP2	NBR1
ENSG00000119772	ENSG00000130640	DNMT3A	TUBGCP2
ENSG00000121577	ENSG00000205981	POPDC2	DNAJC19
ENSG00000124596	ENSG00000198363	C6orf130	ASPH
ENSG00000125122	ENSG00000186231	LRRC29	KLHL32
ENSG00000125630	ENSG00000119535	POLR1B	CSF3R
ENSG00000128512	ENSG00000186952	DOCK4	TMEM232
ENSG00000130038	ENSG00000150527	EFCAB4B	CTAGE5
ENSG00000130844	ENSG00000104325	ZNF331	DECR1
ENSG00000132321	ENSG00000170236	IQCA1	USP50
ENSG00000132323	ENSG00000173715	ILKAP	C11orf80
ENSG00000134375	ENSG00000140854	TIMM17A	KATNB1
ENSG00000135269	ENSG00000134940	TES	ACRV1

ENSG00000136541	ENSG00000048740	ERMN	CELF2
ENSG00000136546	ENSG00000171109	SCN7A	MFN1
ENSG00000137404	ENSG00000184708	NRM	EIF4ENIF1
ENSG00000142494	ENSG00000164941	SLC47A1	INTS8
ENSG00000146263	ENSG00000158552	MMS22L	ZFAND2B
ENSG00000146909	ENSG00000174236	NOM1	REP15
ENSG00000147044	ENSG00000185305	CASK	ARL15
ENSG00000148296	ENSG00000123700	SURF6	KCNJ2
ENSG00000151726	ENSG00000012048	ACSL1	BRCA1
ENSG00000154556	ENSG00000143774	SORBS2	GUK1
ENSG00000156502	ENSG00000106686	SUPV3L1	SPATA6L
ENSG00000157600	ENSG00000108961	TMEM164	RANGRF
ENSG00000157601	ENSG00000124181	MX1	PLCG1
ENSG00000158055	ENSG00000160310	GRHL3	PRMT2
ENSG00000161835	ENSG00000165138	GRASP	ANKS6
ENSG00000162174	ENSG00000104936	ASRGL1	DMPK
ENSG00000163281	ENSG00000164896	GNPDA2	FASTK
ENSG00000165219	ENSG00000148660	GAPVD1	CAMK2G
ENSG00000165275	ENSG00000153029	TRMT10B	MR1
ENSG00000165802	ENSG00000137207	NELF	YIPF3
ENSG00000166333	ENSG00000129219	ILK	PLD2
ENSG00000167822	ENSG00000164032	OR8J3	H2AFZ
ENSG00000168000	ENSG00000141564	BSCL2	RPTOR
ENSG00000168214	ENSG00000149571	RBPJ	KIRREL3
ENSG00000170525	ENSG00000188859	PFKFB3	FAM78B
ENSG00000171055	ENSG00000183474	FEZ2	GTF2H2C
ENSG00000172939	ENSG00000169239	OXSR1	CA5B
ENSG00000173482	ENSG00000156990	PTPRM	RPUSD3
ENSG00000173638	ENSG00000139233	SLC19A1	LLPH
ENSG00000175691	ENSG00000120910	ZNF77	PPP3CC
ENSG00000178882	ENSG00000165799	FAM101A	RNASE7
ENSG00000186431	ENSG00000048828	FCAR	FAM120A
ENSG00000186523	ENSG00000154429	FAM86B1	CCSAP
ENSG00000188493	ENSG00000182934	C19orf54	SRPR
ENSG00000196104	ENSG00000185627	SPOCK3	PSMD13
ENSG00000203685	ENSG00000148426	C1orf95	C10orf47
ENSG00000205327	ENSG00000256061	OR6C68	DYX1C1
ENSG00000215454	ENSG00000100918	KRTAP10-4	REC8
ENSG00000255501	ENSG00000001084	CARD18	GCLC
ENSG00000007264	ENSG00000122642	MATK	FKBP9
ENSG00000007350	ENSG00000137501	TKTL1	SYTL2
ENSG00000011347	ENSG00000136247	SYT7	ZDHHC4
ENSG00000015133	ENSG00000205583	CCDC88C	STAG3L1

ENSG00000037474	ENSG00000189308	NSUN2	LIN54
ENSG00000065150	ENSG00000110077	IPO5	MS4A6A
ENSG00000074319	ENSG00000082516	TSG101	GEMIN5
ENSG00000078618	ENSG00000076554	NRD1	TPD52
ENSG00000085377	ENSG00000196642	PREP	RABL6
ENSG00000087460	ENSG00000140299	GNAS	BNIP2
ENSG00000092421	ENSG00000167173	SEMA6A	C15orf39
ENSG00000102125	ENSG00000049618	TAZ	ARID1B
ENSG00000102243	ENSG00000102271	VGLL1	KLHL4
ENSG00000103363	ENSG00000172366	TCEB2	FAM195A
ENSG00000103591	ENSG00000075426	AAGAB	FOSL2
ENSG00000109339	ENSG00000179698	MAPK10	KIAA1875
ENSG00000111271	ENSG00000174175	ACAD10	SELP
ENSG00000112624	ENSG00000107438	KIAA0240	PDLIM1
ENSG00000116254	ENSG00000083454	CHD5	P2RX5
ENSG00000116761	ENSG00000116830	CTH	TTF2
ENSG00000118729	ENSG00000113312	CASQ2	TTC1
ENSG00000118997	ENSG00000151779	DNAH7	NBAS
ENSG00000119042	ENSG00000137948	SATB2	BRDT
ENSG00000119844	ENSG00000152413	AFTPH	HOMER1
ENSG00000119986	ENSG00000115685	AVPI1	PPP1R7
ENSG00000122145	ENSG00000127993	TBX22	RBM48
ENSG00000122741	ENSG00000102359	DCAF10	SRPX2
ENSG00000123552	ENSG00000106078	USP45	COBL
ENSG00000126107	ENSG00000013288	HECTD3	MAN2B2
ENSG00000130787	ENSG00000002726	HIP1R	ABP1
ENSG00000133026	ENSG00000163328	MYH10	GPR155
ENSG00000133985	ENSG00000181135	TTC9	ZNF707
ENSG00000134324	ENSG00000088305	LPIN1	DNMT3B
ENSG00000134343	ENSG00000101158	ANO3	TH1L
ENSG00000134627	ENSG00000204580	PIWIL4	DDR1
ENSG00000137070	ENSG00000088298	IL11RA	EDEM2
ENSG00000138069	ENSG00000144908	RAB1A	ALDH1L1
ENSG00000138386	ENSG00000158639	NAB1	PAGE5
ENSG00000138395	ENSG00000123130	CDK15	ACOT9
ENSG00000140285	ENSG00000100353	FGF7	EIF3D
ENSG00000144820	ENSG00000108292	GPR128	MLLT6
ENSG00000146828	ENSG00000168275	SLC12A9	C1orf31
ENSG00000147133	ENSG00000127989	TAF1	MTERF
ENSG00000149636	ENSG00000134453	DSN1	RBM17
ENSG00000150477	ENSG00000075914	KIAA1328	EXOSC7
ENSG00000150995	ENSG00000142920	ITPR1	ADC
ENSG00000151136	ENSG00000177106	BTBD11	EPS8L2

ENSG00000154134	ENSG00000131503	ROBO3	ANKHD1
ENSG00000160867	ENSG00000134986	FGFR4	NREP
ENSG00000162909	ENSG00000168569	CAPN2	TMEM223
ENSG00000163541	ENSG00000163166	SUCLG1	IWS1
ENSG00000168509	ENSG00000102158	HFE2	MAGT1
ENSG00000169230	ENSG00000158517	PRELID1	NCF1
ENSG00000169914	ENSG00000173077	OTUD3	1-Dec
ENSG00000170166	ENSG00000092203	HOXD4	TOX4
ENSG00000172007	ENSG00000141540	RAB33B	TTYH2
ENSG00000174899	ENSG00000139182	C3orf55	CLSTN3
ENSG00000176641	ENSG00000148985	RNF152	PGAP2
ENSG00000177311	ENSG00000164129	ZBTB38	NPY5R
ENSG00000183963	ENSG00000153237	SMTN	CCDC148
ENSG00000184497	ENSG00000099204	FAM70B	ABLIM1
ENSG00000185189	ENSG00000239713	NRBP2	APOBEC3G
ENSG00000185303	ENSG00000125676	SFTPA2	THOC2
ENSG00000188282	ENSG00000162191	RUFY4	UBXN1
ENSG00000188343	ENSG00000104824	FAM92A1	HNRNPL
ENSG00000189157	ENSG00000073417	FAM47E	PDE8A
ENSG00000197561	ENSG00000125944	ELANE	HNRNPR
ENSG00000198130	ENSG00000166428	HIBCH	PLD4
ENSG00000205726	ENSG00000188130	ITSN1	MAPK12
ENSG00000239857	ENSG00000180822	GET4	PSMG4

Table S4. The read numbers of type-1A and type-1B synthetic datasets.

Coverage	50 bp	75 bp	100 bp
5X	16,832	11,221	8,416
20X	67,328	44,885	33,664
50X	168,320	112,213	84,160
100X	336,641	224,427	168,320
200X	673,282	448,854	336,641

Table S5. Details for type-2, type-3A and type-3B synthetic datasets: (A) read numbers; (B) insert sizes.

A

Type	Composition	50 bp	75 bp	100 bp
Type-2	Background data*	2,000,000	2,000,000	2,000,000
Type-3A	Synthetic data	336,641	224,426	168,284
Type-3B	Synthetic + Background data	2,336,641	2,224,426	2,168,284

*Background data are the real tissue data: lung, parathyroid, skeletal myocyte, bladder and T cell sample respectively.

B

Tissue	SRA ID*	Insert size mean value	Insert size SD value
Lung	SRR349695	164 bp	48 bp
Parathyroid	SRR479053	192 bp	85 bp
Skeletal myocyte	SRR1693845	353 bp	116 bp
Bladder	SRR400342	248 bp	26 bp
T cell	SRR1909130	290 bp	120 bp

*SRA: Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra)

Table S6. Data description for three real datasets.

Cancer	Sample	Read length (bp)	Insert size* (bp)	Read number	Validated fusions
Breast Cancer (Edgren, et al., 2011)	BT474	50	180 ± 80	21,423,697	ACACA - STAC2, RPS6KB1 - SNF8, VAPB - IKZF3, ZMYND8 - CEP250, RAB22A - MYO9B, SKA2 - MYO19, DIDO1 - KIAA0406, STARD3 - DOK5, LAMP1 - MCF2L, GLB1 - CMTM7, CPNE1 - PI3
	KPL4	50	180 ± 80	6,796,443	BSG - NFIX, PPP1R12A - SEPT10, NOTCH1 - NUP214,
	MCF7	50	180 ± 80	8,409,785	BCAS4 - BCAS3, ARFGEF2 - SULF2, PRPS6KB1 - TMEM49
	SKBR3	50	180 ± 80	18,140,246	TATDN1 - GSDMB, CSE1L - ENSG00000236127, RARA - PKIA, ANKHD1 - PCDH1, CCDC85C - SETD3, SUMF1 - LRRFIP2, WDR67 - ZNF704, CYTH1 - EIF3H, DHX35 - ITCH, NFS1 - PREX1
Melanoma (Berger, et al., 2010)	501Mel	50	351 ± 139	14,857,046	CCT3 - C1orf61, GNA12 - SHANK2, SLC12A7 - C11orf67, PARP1 - MIXL1
	M000216	50	393 ± 115	13,868,165	KCTD2 - ARHGEF12
	M000921	50	627 ± 564	14,468,771	TMEM8B - TLN1, RECK - ALX3
	M010403	50	374 ± 159	8,168,750	SCAMP2 - WDR72
	M980409	50	334 ± 119	15,768,555	GCN1L1 - PLA2G1B
	M990802	50	355 ± 125	16,066,999	ANKHD1 - C5orf32, RB1 - ITM2B
Prostate Cancer (Yu, et al., 2014)	158T	100	140 ± 27	221,206,388	SLC45A3 - ELK4, MTOR - TP53BP1
	159T	100	153 ± 33	159,766,465	TRMT11 - GRIK2, MAN2A1 - FER, KDM4B - AC011523.2, CCNH - C5orf30
	165T	100	158 ± 40	243,191,643	SLC45A3 - ELK4, TMEM135 - CCDC67
	171T	100	143 ± 30	118,742,381	TMPRSS2 - ERG
	49T	100	158 ± 37	250,071,864	SLC45A3 - ELK4, TMPRSS2 - ERG, LRRC59 - SLC22A10

*Insert size is estimated by alignment tool BWA.

Table S7. Parameter setting for TopHat-Fusion.

Options	Explanation
--no-coverage-search	Disable the coverage based search for junctions
--fusion-min-dist 1000000	Default minimum distance
--fusion-ignore-chromosomes chrM	Ignore chromosome M

*We use bowtie2 hg19 index in TopHat step, and use both bowtie and bowtie2 hg19 index in the Fusion step.

Table S8. Parameter setting for anchor length and spanning/split reads of all the fifteen tools.

Detection tools	**Anchor length**	**Spanning read**	**Split read**	**Note**
MapSplice	-	-	-	User self-filter for spanning+split as 4
ShortFuse	-	-	-	A score to rank the detected fusion transcripts
FusionHunter	10	3	1	
FusionMap	Min(25, Max(17,floor(ReadLength/3)))	Distinct definition	Distinct definition	
deFuse	10	3	1	
chimerascan	10	-	-	Set spanning+split as 4; maximum mismatch is 3
FusionCatcher	10	3	1	
TopHat-Fusion	10	3	1	
BreakFusion	-	-	-	User self-filter for spanning+split as 4
EricScript	-	-	-	User self-filter for spanning/split read on the output
SOAPfuse	10	3	1	
FusionQ	10	3	1	
PRADA	-	-	-	User self-filter for spanning+split as 4
SnowShoes-FTD	-	-	-	Set spanning+split as 4
JAFFA	-	3	1	User self-filter for spanning/split read on the output

*Otherwise default settings are used for the other parameters.

Table S9. Completeness of the fifteen tools on the synthetic and real datasets.

Tools	Type-1 synthetic dataset	Type-2 synthetic dataset	Type-3 synthetic dataset	Breast cancer dataset	Melanoma dataset	Prostate dataset	Validation dataset	Note
MapSplice	Y	Y	Y	Y	Y	49T failed	Y	Running time exceed 14 days and larger than 2TB temp files
ShortFuse	100 bp data failed	75 bp and 100 bp data failed	0 output for most of the datasets	Y	Y	failed	failed	Error with 'parsing the discordant reads'
FusionHunter	failed	failed	failed	Y	Y	Y	Y	Stack smashing for synthetic data
FusionMap	Y	Y	Y	Y	Y	Y		
deFuse	Y	Y	Y	Y	Y	failed	0 output	Running time exceed 14 days
chimerascan	Y	Y	Y	Y	Y	Y	Y	
FusionCatcher	Y	Y	Y	Y	Y	Y	Y	
Tophat-Fusion	Y	Y	Y	Y	Y	Y	Y	Changing parameter setting
BreakFusion	Y	Y	Y	Y	Y	Y	Y	
Ericscript	Y	Y	Y	Y	Y	0 output	Y	
SOAPfuse	Y	Y	Y	Y	Y	Y	Y	
FusionQ	Y	Y	Y	Partially could not complete (CNC)	CNC	CNC	0 output	Tool cannot stop at 3rd step if data are large
PRADA	Y	Y	Y	Y	Y	Y	Y	
SnowShoes-FTD	Y	Y	Y	Y	Y	Y	Y	Trim 100 bp reads to 50 bp
JAFFA	Y	Y	Y	Y	Y	Y	Y	

Table S10. The read numbers of prostate cancer 171T dataset and its subsamples.

Subsample	Read numbers
1/1	118,742,381
1/2	59,371,192
1/4	29,685,596
1/8	14,842,798

Table S11. The summary of the recall rates, precision rates and F-measures for three representative synthetic data: (A) type-1A with read 100 bp & 100X, (B) type-1B with read 100 bp & 100X, and (C) type-3B lung sample with read 50 bp & 100X.

A

Detection tools	TP	FP	FN	TP+FP	Recall	Precision	F-measure
SOAPfuse	139	26	11	165	0.927	0.842	0.882
FusionCatcher	108	20	42	128	0.720	0.844	0.777
JAFFA	88	16	62	104	0.587	0.846	0.693
EricScript	104	13	46	117	0.693	0.889	0.779
chimerascan	105	30	45	135	0.700	0.778	0.737
PRADA	57	2	93	59	0.380	0.966	0.545
deFuse	73	9	77	82	0.487	0.890	0.630
FusionMap	92	27	58	119	0.613	0.773	0.684
TopHat-Fusion	53	14	97	67	0.353	0.791	0.488
MapSplice	51	8	99	59	0.340	0.864	0.488
BreakFusion	106	44	44	150	0.707	0.707	0.707
SnowShoes-FTD	3	1	147	4	0.020	0.750	0.039
FusionQ	81	18	69	99	0.540	0.818	0.651
FusionHunter	-	-	-	-	-	-	-
ShortFuse	-	-	-	-	-	-	-

B

Detection tools	TP	FP	FN	TP+FP	Recall	Precision	F-measure
SOAPfuse	136	22	14	158	0.907	0.861	0.883
FusionCatcher	110	18	40	128	0.733	0.859	0.791
JAFFA	86	20	64	106	0.573	0.811	0.672
EricScript	111	15	39	126	0.740	0.881	0.804
chimerascan	96	26	54	122	0.640	0.787	0.706
PRADA	57	3	93	60	0.380	0.950	0.543
deFuse	123	15	27	138	0.820	0.891	0.854
FusionMap	95	22	55	117	0.633	0.812	0.711
TopHat-Fusion	64	16	86	80	0.427	0.800	0.557
MapSplice	53	9	97	62	0.353	0.855	0.500
BreakFusion	74	36	76	110	0.493	0.673	0.569
SnowShoes-FTD	3	1	147	4	0.020	0.750	0.039
FusionQ	51	12	99	63	0.340	0.810	0.479
FusionHunter	-	-	-	-	-	-	-
ShortFuse	-	-	-	-	-	-	-

C

Detection tools	TP	FP	FN	TP+FP	Recall	Precision	F-measure
SOAPfuse	138	27	12	165	0.920	0.836	0.876
FusionCatcher	102	19	48	121	0.680	0.843	0.753
JAFFA	88	12	62	100	0.587	0.880	0.704
EricScript	105	15	45	120	0.700	0.875	0.778
chimerascan	91	25	59	116	0.607	0.784	0.684
PRADA	55	2	95	57	0.367	0.965	0.532
deFuse	43	7	107	50	0.287	0.860	0.430
FusionMap	72	11	78	83	0.480	0.867	0.618
TopHat-Fusion	66	19	84	85	0.440	0.776	0.562
MapSplice	55	9	95	64	0.367	0.859	0.514
BreakFusion	114	141	36	255	0.760	0.447	0.563
SnowShoes-FTD	3	1	147	4	0.020	0.750	0.039
FusionQ	37	37	113	74	0.247	0.500	0.331
FusionHunter	-	-	-	-	-	-	-
ShortFuse	-	-	-	-	-	-	-

Table S12. The correlation between five normal tissues by the F-measure of the fifteen tools on the type-3B dataset.

	Lung	Parathyroid	Bladder	Skeletal myocyte	T cell
Lung	1	0.95	0.78	0.94	0.92
Parathyroid	-	1	0.79	0.99	0.79
Bladder	-	-	1	0.87	0.56
Skeletal myocyte	-	-	-	1	0.76
T cell	-	-	-	-	1

Table S13. The summary of the recall rates, precision rates and F-measures for three real datasets (pool samples together): (A) breast cancer data; (B) melanoma data; (C) prostate cancer data.

A

Detection tools	TP	FP	FN	TP+FP	Recall	Precision	F-measure
SOAPfuse	20	48	7	68	0.741	0.294	0.421
FusionCatcher	19	48	8	67	0.704	0.284	0.405
JAFFA	16	16	11	32	0.593	0.500	0.543
EricScript	16	67	11	83	0.593	0.193	0.291
chimerascan	19	96	8	115	0.704	0.165	0.267
PRADA	15	22	12	37	0.556	0.405	0.469
deFuse	19	116	8	135	0.704	0.141	0.235
FusionMap	6	126	21	132	0.222	0.045	0.075
TopHat-Fusion	15	58	12	73	0.556	0.205	0.300
MapSplice	16	37	11	53	0.593	0.302	0.400
BreakFusion	15	1923	12	1938	0.556	0.008	0.016
SnowShoes-FTD	15	5	12	20	0.556	0.750	0.639
FusionQ	4	453	23	457	0.148	0.009	0.017
FusionHunter	13	10	14	23	0.481	0.565	0.520
ShortFuse	19	24	8	43	0.704	0.442	0.543

B

Detection tools	TP	FP	FN	TP+FP	Recall	Precision	F-measure
SOAPfuse	10	98	1	108	0.909	0.093	0.169
FusionCatcher	3	6	8	9	0.273	0.333	0.300
JAFFA	2	2	9	4	0.182	0.500	0.267
EricScript	3	67	8	70	0.273	0.043	0.074
chimerascan	5	189	6	194	0.455	0.026	0.049
PRADA	3	4	8	7	0.273	0.429	0.334
deFuse	10	189	1	199	0.909	0.05	0.095
FusionMap	2	85	9	87	0.182	0.023	0.041
TopHat-Fusion	4	25	7	29	0.364	0.138	0.200
MapSplice	5	39	6	44	0.455	0.114	0.182
BreakFusion	6	3092	5	3098	0.545	0.002	0.004
SnowShoes-FTD	4	1	7	5	0.364	0.800	0.500
FusionQ	-	-	-	-	-	-	-
FusionHunter	4	4	7	8	0.364	0.500	0.421
ShortFuse	7	30	4	37	0.636	0.189	0.291

C

Detection tools	TP	FP	FN	TP+FP	Recall	Precision	F-measure
SOAPfuse	7	75	5	82	0.583	0.085	0.148
FusionCatcher	11	82	1	93	0.917	0.118	0.209
JAFFA	6	1815	6	1821	0.500	0.003	0.006
EricScript	11	3798	1	3809	0.917	0.003	0.006
chimerascan	8	1665	4	1673	0.667	0.005	0.010
PRADA	0	13	12	13	0	0	0
deFuse	-	-	-	-	-	-	-
FusionMap	7	3759	5	3766	0.583	0.002	0.004
TopHat-Fusion	0	61	12	61	0	0	0
MapSplice	0	4	12	4	0	0	0
BreakFusion	1	3764	11	3765	0.083	0	0
SnowShoes-FTD	5	6	7	11	0.417	0.455	0.435
FusionQ	-	-	-	-	-	-	-
FusionHunter	0	5	12	5	0	0	0
ShortFuse	-	-	-	-	-	-	-

Table S14. The summary of the recall rates, precision rates and F-measures for validation dataset.

Detection tools	TP	FP	FN	TP+FP	Recall	Precision	F-measure
SOAPfuse	4	9	5	13	0.444	0.308	0.364
FusionCatcher	6	17	3	23	0.667	0.261	0.375
JAFFA	5	8	4	13	0.556	0.385	0.455
EricScript	5	358	4	363	0.556	0.014	0.027
chimerascan	6	127	3	133	0.667	0.045	0.084
PRADA	3	4	6	7	0.333	0.429	0.375
deFuse	0	0	0	0	0	0	0
FusionMap	6	324	3	330	0.667	0.018	0.035
TopHat-Fusion	1	5	8	6	0.111	0.167	0.133
MapSplice	4	8	5	12	0.444	0.333	0.381
BreakFusion	2	142	7	144	0.222	0.014	0.026
SnowShoes-FTD	4	3	5	7	0.444	0.571	0.5
FusionQ	0	0	0	0	0	0	0
FusionHunter	2	4	7	6	0.222	0.333	0.266
ShortFuse	-	-	-	-	-	-	-