Practical significance of some full-text indices for read mapping

UNIVERSITY OF MARYLAND

Bowtie2

nature methods

Brief Communication | Published: 04 March 2012

Fast gapped-read alignment with Bowtie 2

Ben Langmead ✉ & Steven L Salzberg

Nature Methods **9**, 357–359 (2012) | Download Citation ↓

☆ Cited by 15867 Related articles All 14 versions Web of Science: 10788 as of 09/18/2019

Based on FM-index for seed finding

Novel strategy / heuristic for seed scoring and exploration

Makes use of SIMD-accelerated alignment DP

Capable of global (end-to-end) or local alignment

No spliced alignment (i.e. for DNA-seq or RNA-seq -> txome)

Bowtie2

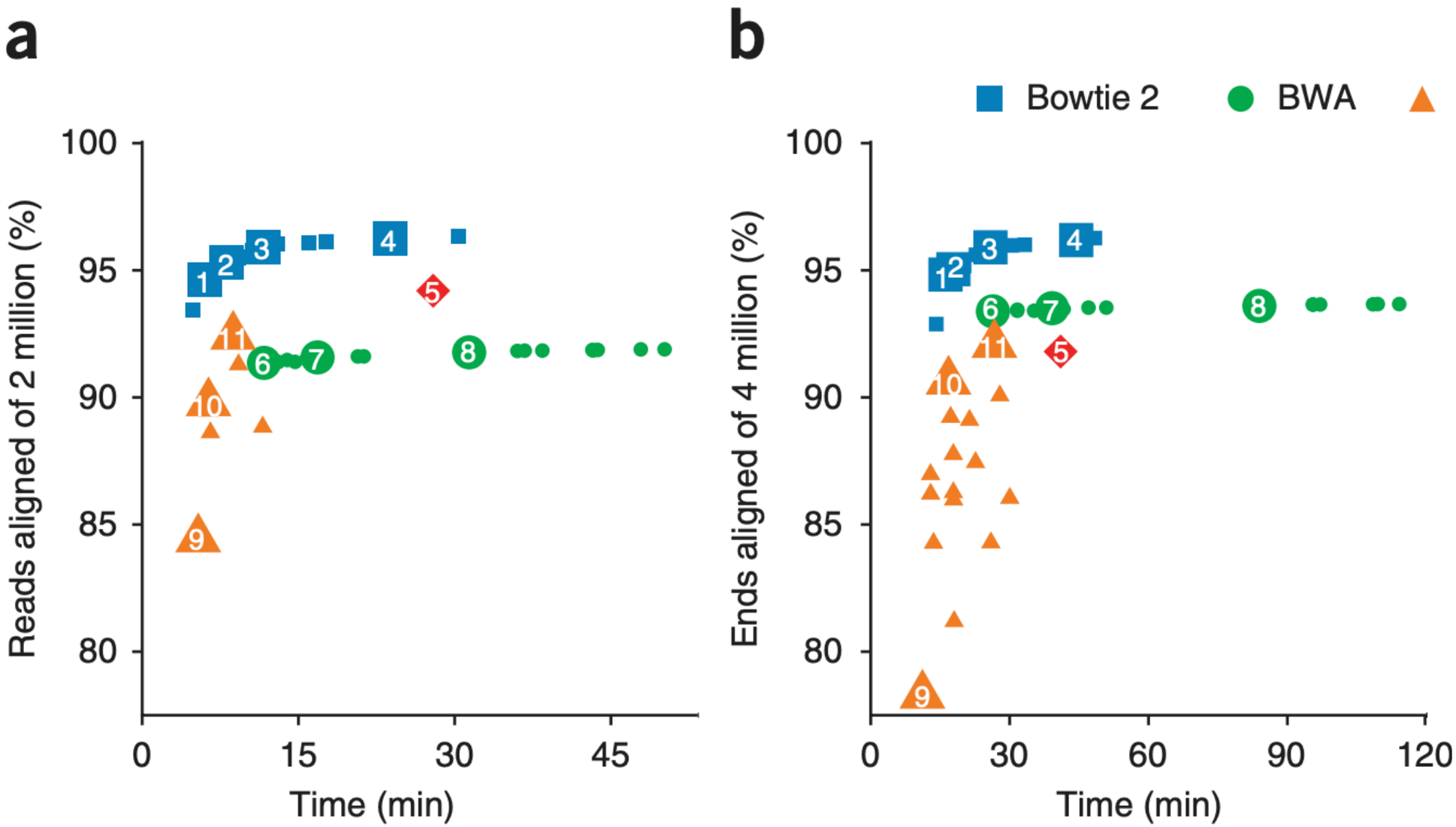


Figure 1 | Alignment comparison using HiSeq 2000, 454 and Ion Torrent reads. (**a**–**d**) Bowtie 2, BWA, SOAP2 and Bowtie were used to align two million 100 nt × 100 nt paired-end HiSeq 2000 reads from a resequencing study[11]. Shown are results for unpaired alignment of end 1 (**a**), paired-end alignment (**b**), Bowtie 2 and BWA-SW alignment of 1 million 454 reads from the 1000 Genomes Project Pilot[12] (**c**), and Bowtie 2 and BWA-SW to align one million Ion Torrent reads from the G. Moore resequencing project[13] (**d**). Plotted is the percentage of reads for which at least one alignment was found. Each numbered point is data obtained using command-line parameters shown in **Supplementary Table 1**.

Bowtie2

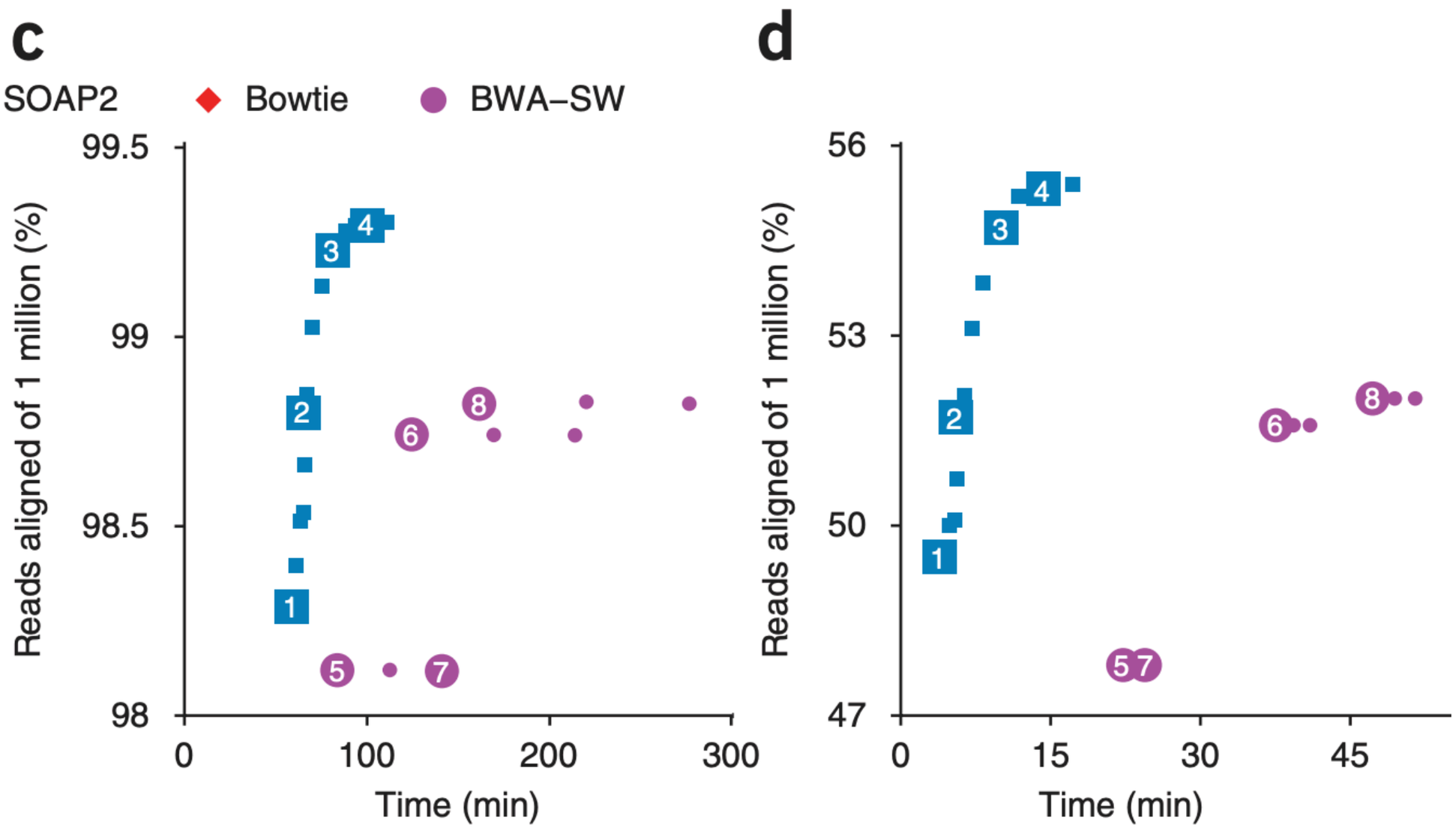
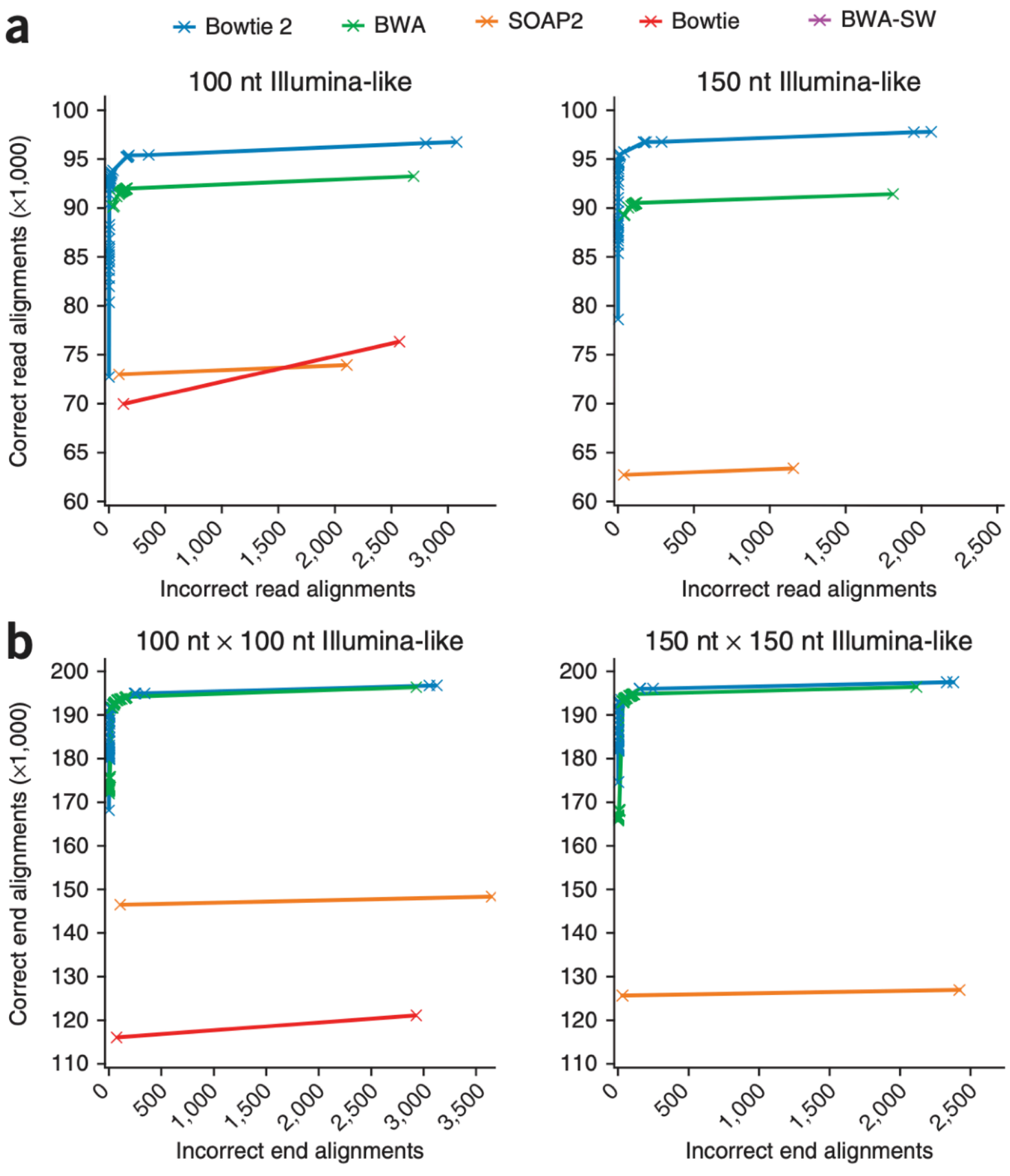


Figure 1 | Alignment comparison using HiSeq 2000, 454 and Ion Torrent reads. (**a**–**d**) Bowtie 2, BWA, SOAP2 and Bowtie were used to align two million 100 nt × 100 nt paired-end HiSeq 2000 reads from a resequencing study[11]. Shown are results for unpaired alignment of end 1 (**a**), paired-end alignment (**b**), Bowtie 2 and BWA-SW alignment of 1 million 454 reads from the 1000 Genomes Project Pilot[12] (**c**), and Bowtie 2 and BWA-SW to align one million Ion Torrent reads from the G. Moore resequencing project[13] (**d**). Plotted is the percentage of reads for which at least one alignment was found. Each numbered point is data obtained using command-line parameters shown in **Supplementary Table 1**.

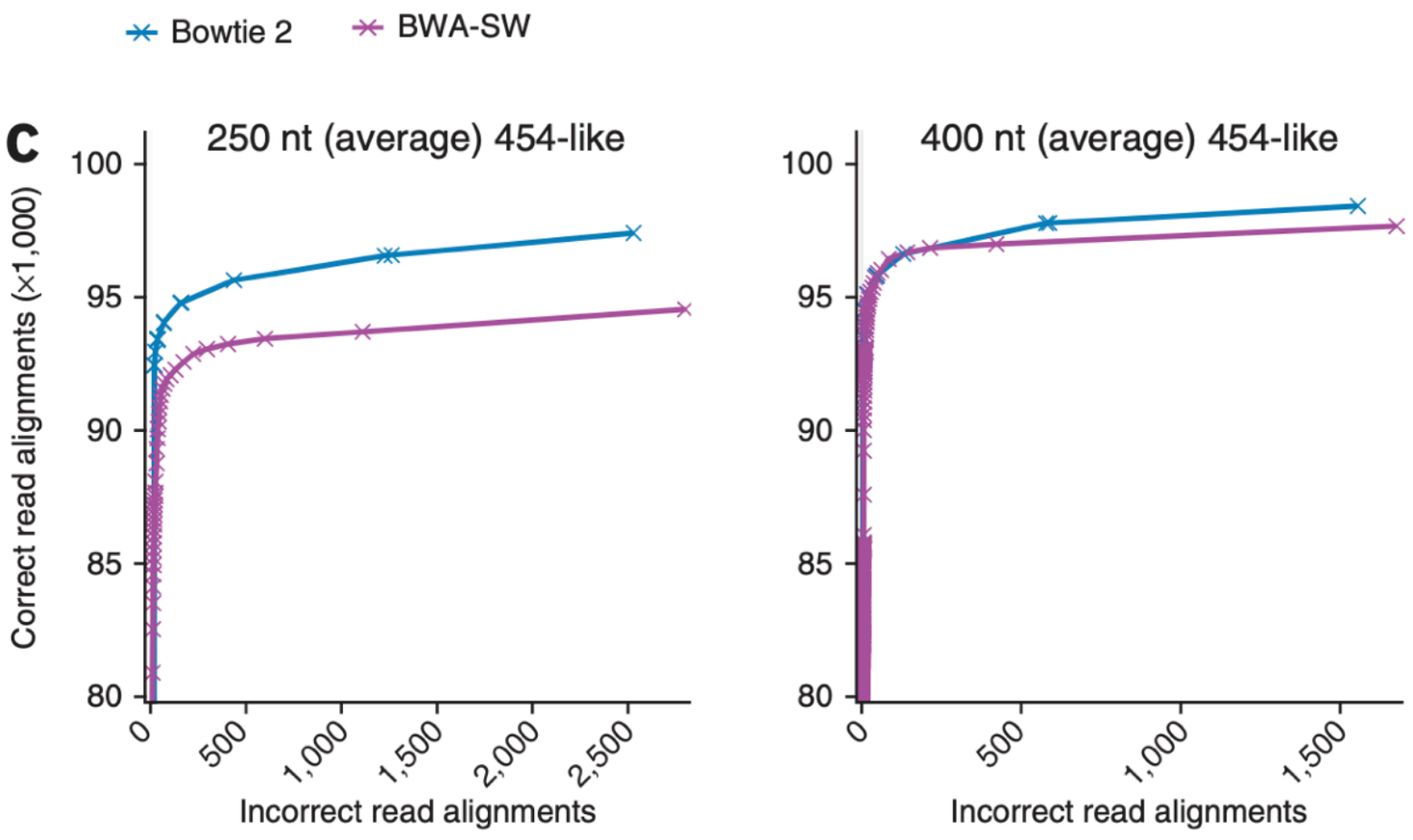
Bowtie2

Results on simulated data



Bowtie2

Results on simulated data



Bowtie2

Dataset	Bowtie 2 versus	Reads or ends aligned by neither	Reads or ends aligned by only Bowtie 2	Reads or ends aligned by only other tool	Reads or ends aligned by both
Unpaired HiSeq 2K	BWA	79,842 (3.99%)	84,136 (4.21%)	449 (0.09%)	1,834,243 (91.71%)
Paired HiSeq 2K	BWA	154,799 (3.87%)	99,852 (2.50%)	9,137 (0.23%)	3,736,212 (93.41%)
454	BWA-SW	7,458 (0.75%)	11,344 (1.13%)	266 (0.03%)	988,390 (98.84%)
Ion Torrent	BWA-SW	450,602 (45.06%)	71,423 (7.14%)	2,270 (0.23%)	475,705 (47.57%)

Bowtie2

Dataset	Bowtie 2 versus	Reads or ends aligned by neither	Reads or ends aligned by only Bowtie 2	Reads or ends aligned by only other tool	Reads or ends aligned by both
Unpaired HiSeq 2K	BWA	79,842 (3.99%)	84,136 (4.21%)	449 (0.09%)	1,834,243 (91.71%)
Paired HiSeq 2K	BWA	154,799 (3.87%)	99,852 (2.50%)	9,137 (0.23%)	3,736,212 (93.41%)
454	BWA-SW	7,458 (0.75%)	11,344 (1.13%)	266 (0.03%)	988,390 (98.84%)
Ion Torrent	BWA-SW	450,602 (45.06%)	71,423 (7.14%)	2,270 (0.23%)	475,705 (47.57%)

BWA-MEM

Quantitative Biology > Genomics

Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM

Heng Li

(Submitted on 16 Mar 2013 (v1), last revised 26 May 2013 (this version, v2))

☆ Cited by 2411 Related articles » as of 09/18/2019

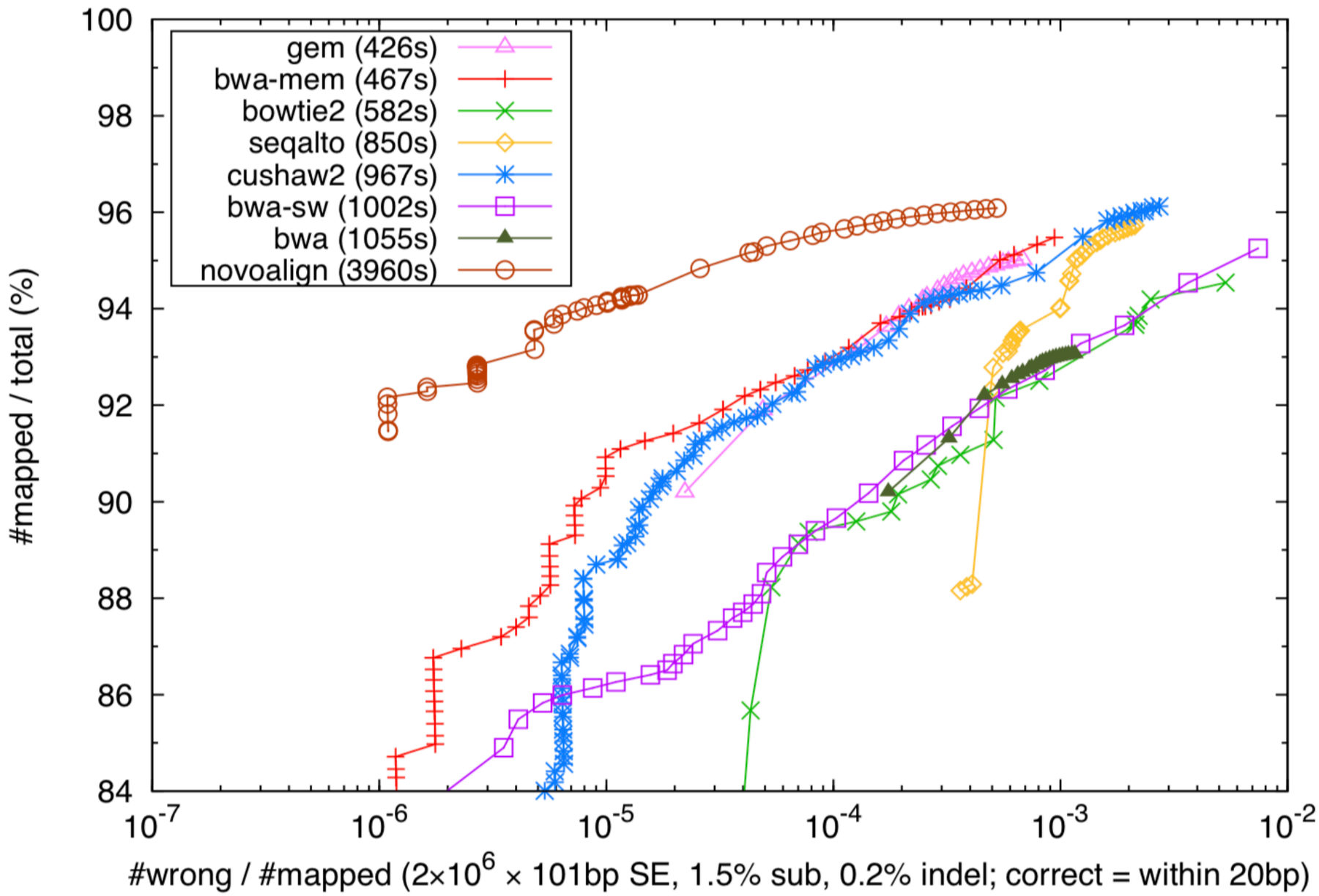
Based on FMD-index for seed finding

Novel “chaining” strategy to find potential alignment loci

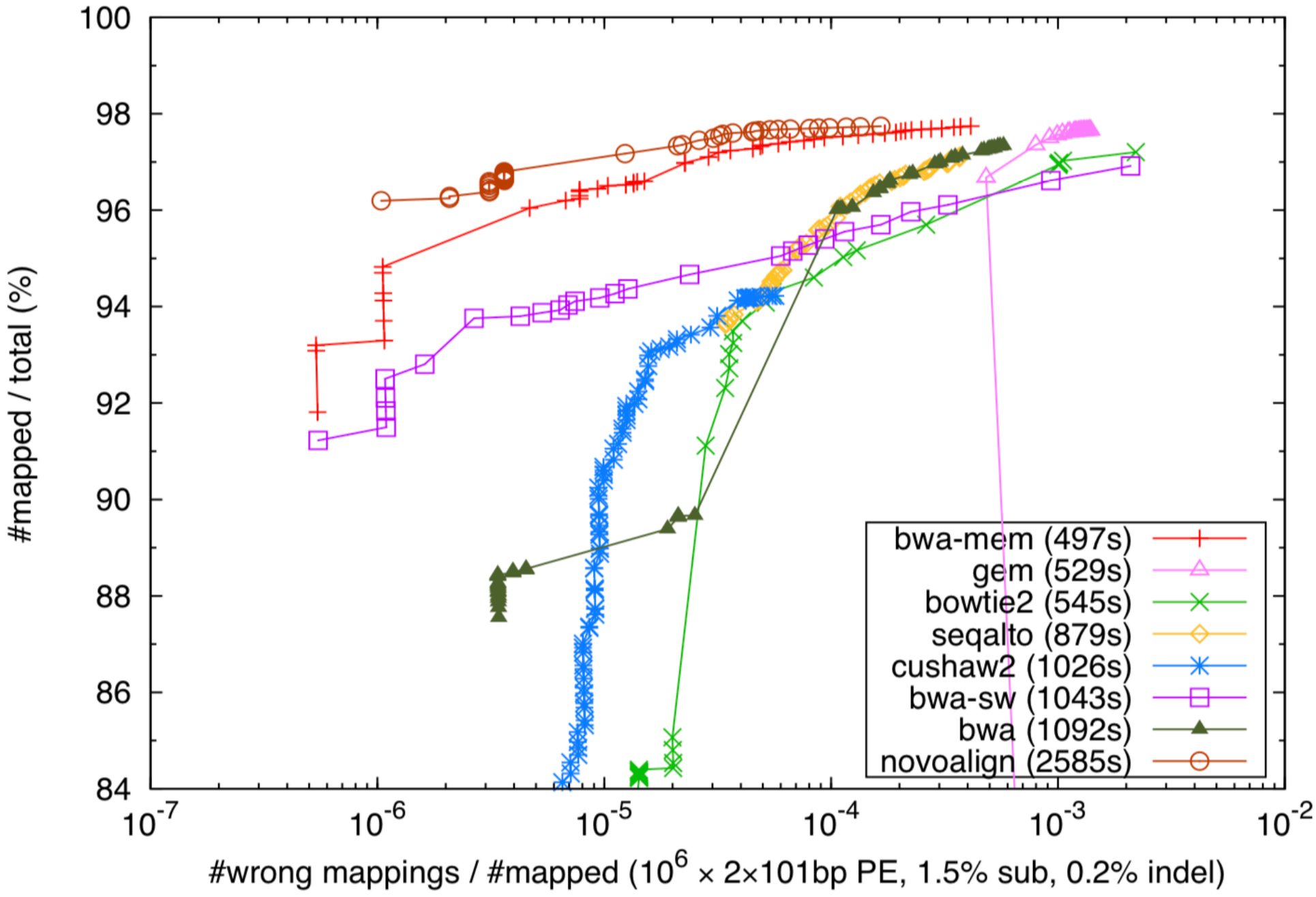
No spliced alignment (i.e. for DNA-seq or RNA-seq -> txome)

Note: The BWA-MEM “paper” is this arXiv pre-print. The manuscript itself was never “published” in a traditional journal. This is a great example of software with *huge* impact that was nonetheless never published.

BWA-MEM


gem (426s)
bwa-mem (467s)
bowtie2 (582s)
seqalto (850s)
cushaw2 (967s)
bwa-sw (1002s)
bwa (1055s)
novoalign (3960s)
#mapped / total (%)
100
98
96
94
92
90
88
86
84
10-7
10-6
10-5
10-4
10-3
10-2
#wrong / #mapped (2×10^6 × 101bp SE, 1.5% sub, 0.2% indel; correct = within 20bp)

BWA-MEM


100
98
96
94
92
90
88
86
84
#mapped / total (%)
10-7
10-6
10-5
10-4
10-3
10-2
#wrong mappings / #mapped (10^6 × 2×101bp PE, 1.5% sub, 0.2% indel)
bwa-mem (497s)
gem (529s)
bowtie2 (545s)
seqalto (879s)
cushaw2 (1026s)
bwa-sw (1043s)
bwa (1092s)
novoalign (2585s)

STAR

BIOINFORMATICS ORIGINAL PAPER

Vol. 29 no. 1 2013, pages 15–21
doi:10.1093/bioinformatics/bts635

Sequence analysis

Advance Access publication October 25, 2012

STAR: ultrafast universal RNA-seq aligner

Alexander Dobin[1,*], Carrie A. Davis[1], Felix Schlesinger[1], Jorg Drenkow[1], Chris Zaleski[1], Sonali Jha[1], Philippe Batut[1], Mark Chaisson[2] and Thomas R. Gingeras[1]

[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, USA and [2]Pacific Biosciences, Menlo Park, CA, USA

Associate Editor: Inanc Birol

☆ Cited by 7979 Related articles Web of Science: 5224 as of 09/18/2019

Based on suffix array + prefix-table for seed finding

Custom “chaining” & between match alignment strategy

Capable of both contiguous and spliced alignment, behavior is *highly* configurable via parameters (DNA-seq or RNA-seq alignment directly to the genome)

STAR

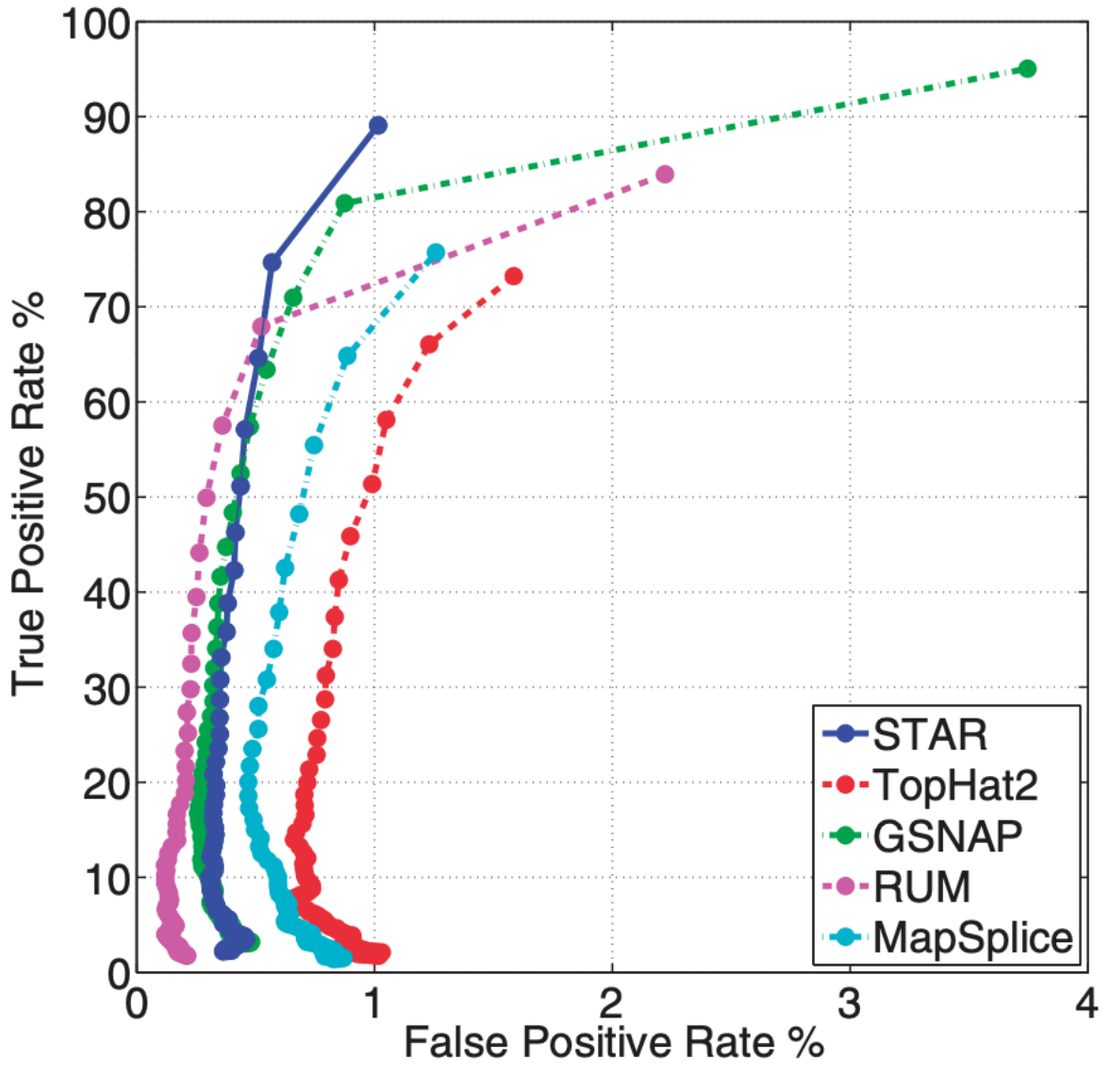
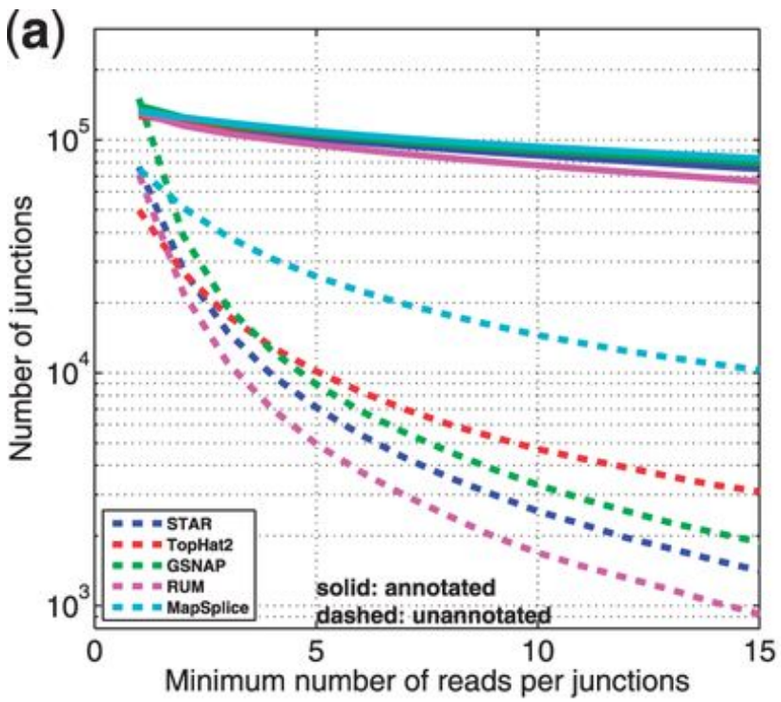
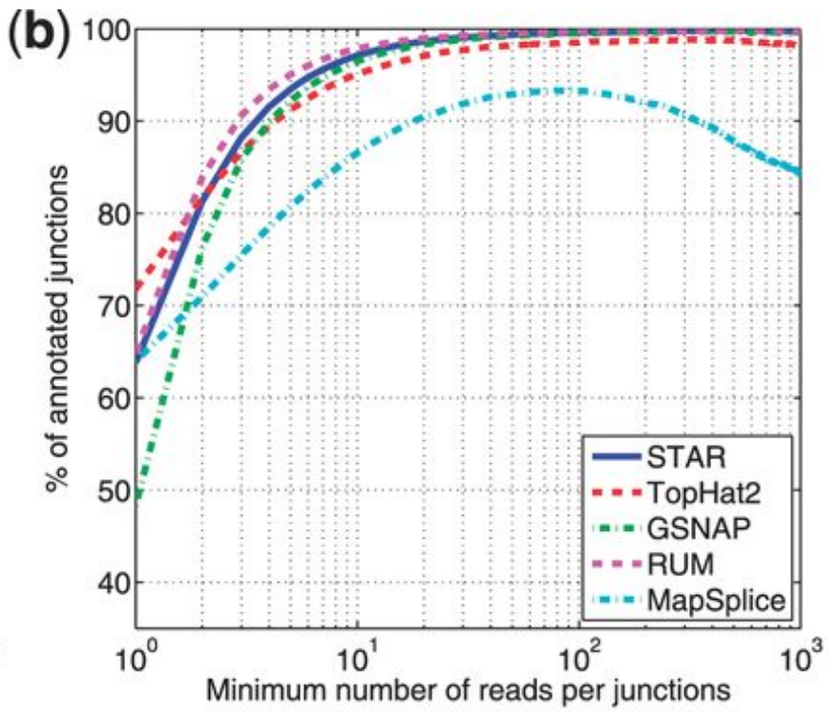
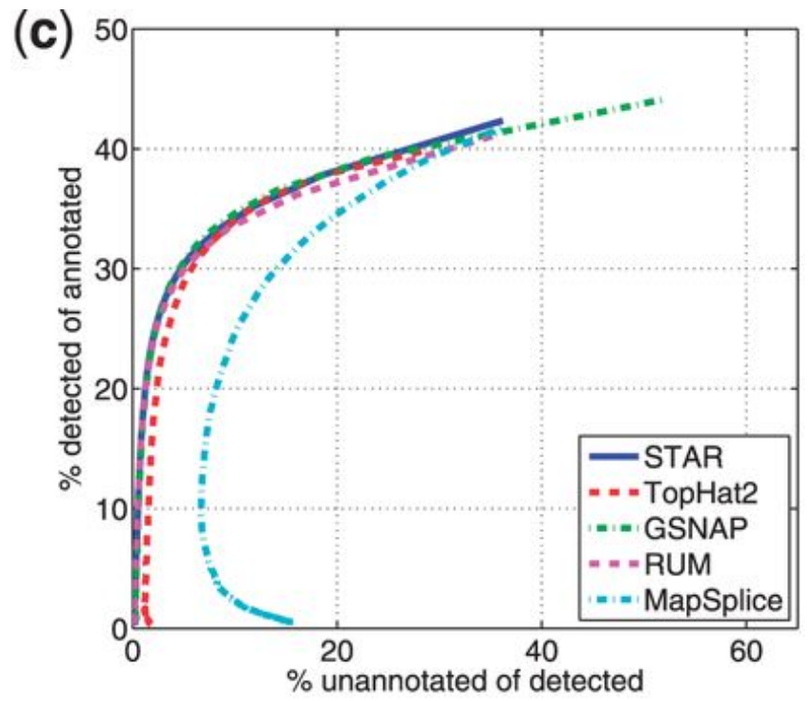


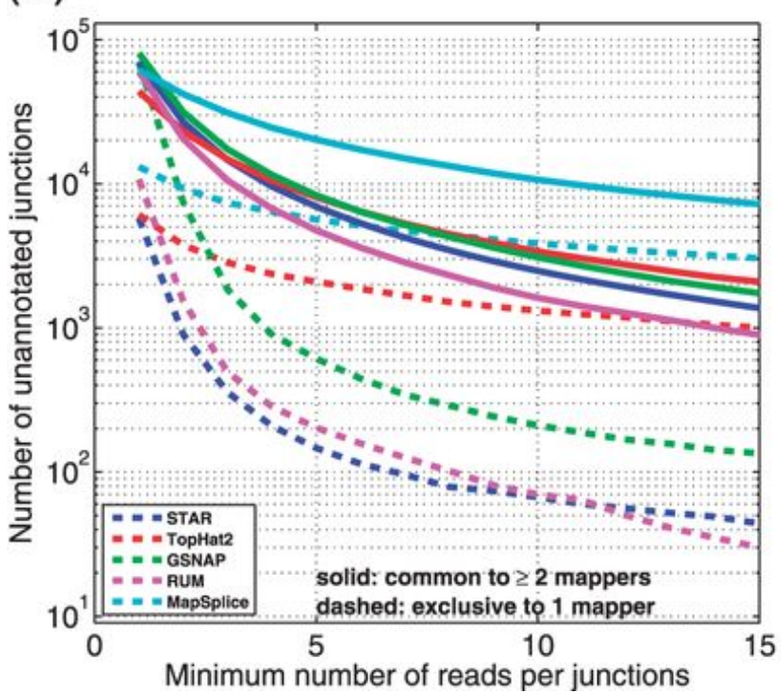
Fig. 2. True-positive rate versus false-positive rate (ROC-curve) for simulated RNA-seq data for STAR, TopHat2, GSNAP, RUM and MapSplice

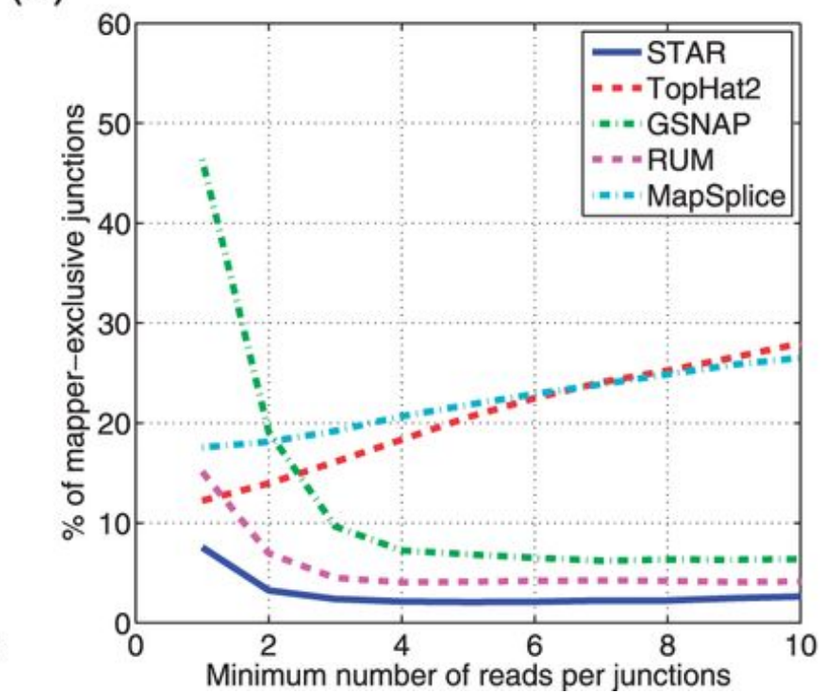
STAR

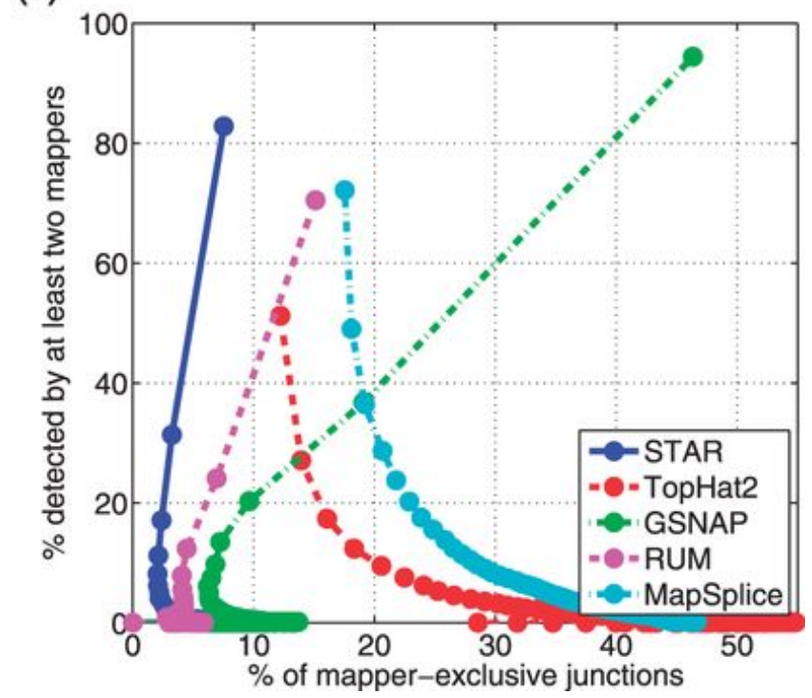

(a)
Number of junctions
STAR
TopHat2
GSNAP
RUM
MapSplice
solid: annotated
dashed: unannotated
Minimum number of reads per junctions


(b)
% of annotated junctions
STAR
TopHat2
GSNAP
RUM
MapSplice
Minimum number of reads per junctions


(c)
% detected of annotated
STAR
TopHat2
GSNAP
RUM
MapSplice
% unannotated of detected


(d)
Number of unannotated junctions
STAR
TopHat2
GSNAP
RUM
MapSplice
solid: common to ≥ 2 mappers
dashed: exclusive to 1 mapper
Minimum number of reads per junctions


(e)
% of mapper−exclusive junctions
STAR
TopHat2
GSNAP
RUM
MapSplice
Minimum number of reads per junctions


(f)
% detected by at least two mappers
STAR
TopHat2
GSNAP
RUM
MapSplice
% of mapper−exclusive junctions

STAR

Table 1. Mapping speed and RAM benchmarks on the experimental RNA-seq dataset

Aligner	Mapping speed: million read pairs/hour		Peak physical RAM, GB	
	6 threads	12 threads	6 threads	12 threads
STAR	309.2	549.9	27.0	28.4
STAR sparse	227.6	423.1	15.6	16.0
TopHat2	8.0	10.1	4.1	11.3
RUM	5.1	7.6	26.9	53.8
MapSplice	3.0	3.1	3.3	3.3
GSNAP	1.8	2.8	25.9	27.0

HISAT2

Article | Published: 02 August 2019

Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype

Daehwan Kim ✉, Joseph M. Paggi, Chanhee Park, Christopher Bennett & Steven L. Salzberg

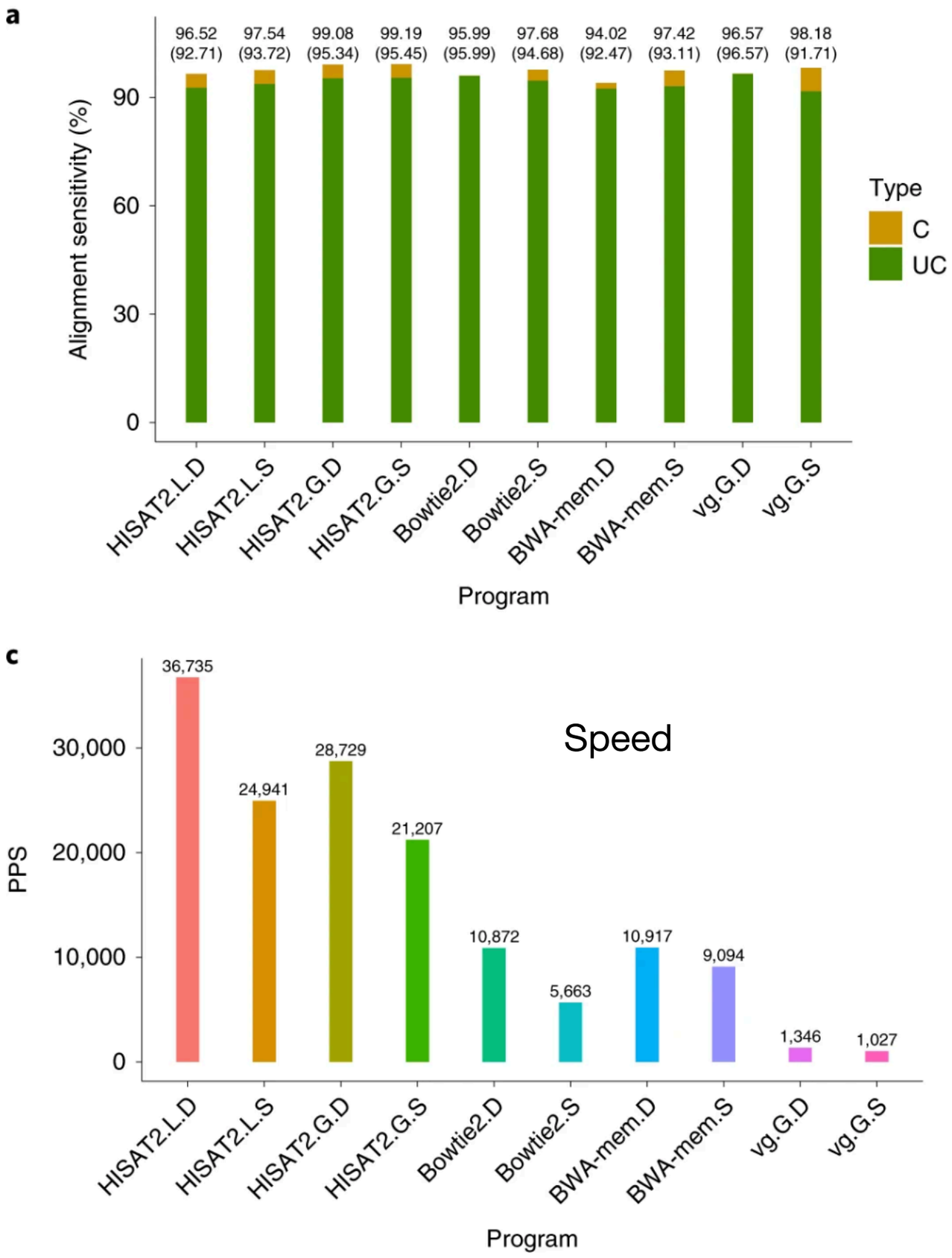
Based on hierarchical graph FM-index for alignment

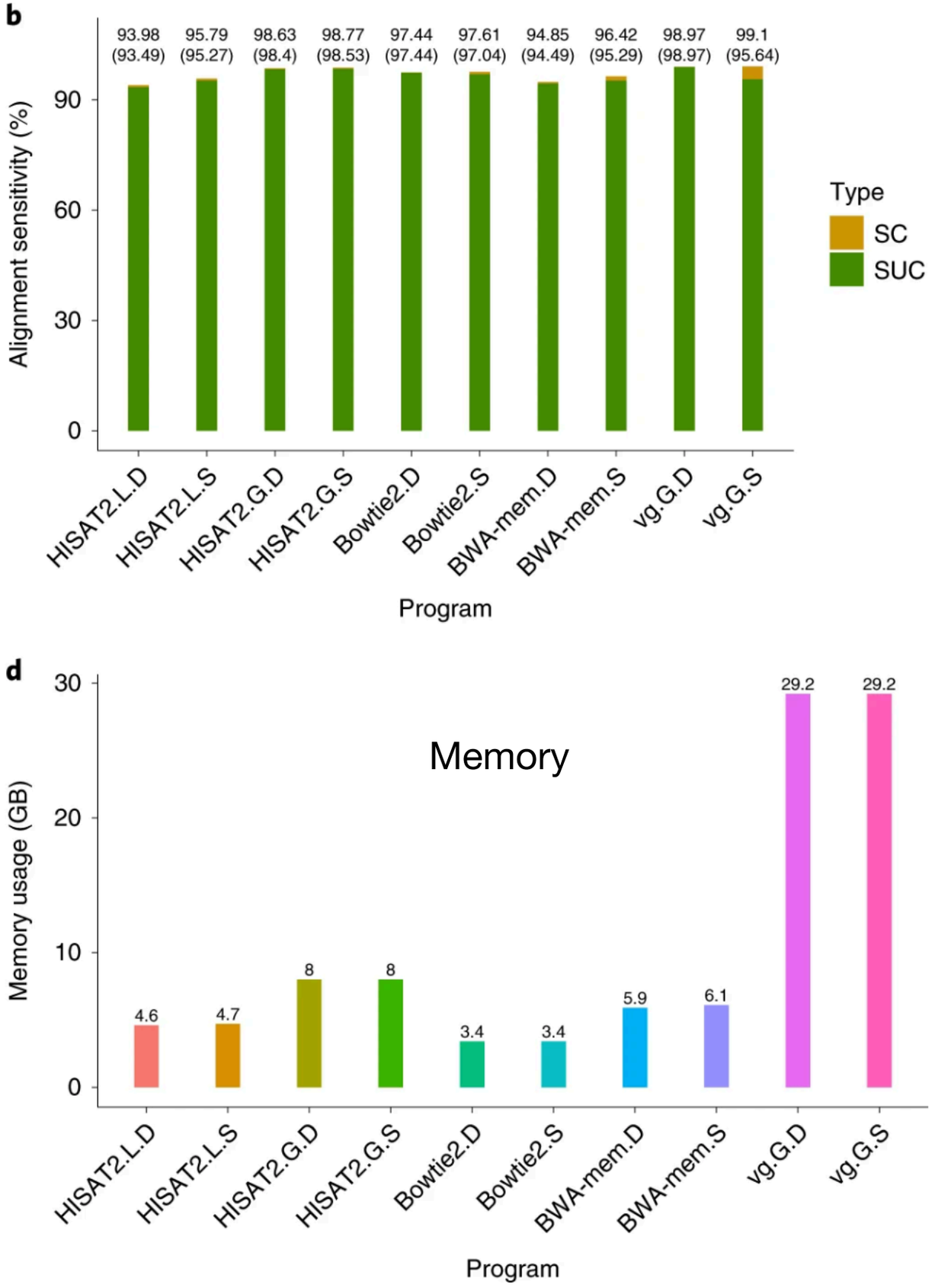
Custom strategy to deal with highly-repetitive regions

Capable of both contiguous and spliced alignment, behavior is *highly* configurable via parameters (built for both DNA-seq and RNA-seq alignment directly to the genome)

Built-in algorithm to do HLA-typing over aligned reads

HISAT2


a
96.52 (92.71)
97.54 (93.72)
99.08 (95.34)
99.19 (95.45)
95.99 (95.99)
97.68 (94.68)
94.02 (92.47)
97.42 (93.11)
96.57 (96.57)
98.18 (91.71)
Alignment sensitivity (%)
90
60
30
0
Type
C
UC
HISAT2.L.D
HISAT2.L.S
HISAT2.G.D
HISAT2.G.S
Bowtie2.D
Bowtie2.S
BWA-mem.D
BWA-mem.S
vg.G.D
vg.G.S
Program


b
93.98 (93.49)
95.79 (95.27)
98.63 (98.4)
98.77 (98.53)
97.44 (97.44)
97.61 (97.04)
94.85 (94.49)
96.42 (95.29)
98.97 (98.97)
99.1 (95.64)
Alignment sensitivity (%)
90
60
30
0
Type
SC
SUC
HISAT2.L.D
HISAT2.L.S
HISAT2.G.D
HISAT2.G.S
Bowtie2.D
Bowtie2.S
BWA-mem.D
BWA-mem.S
vg.G.D
vg.G.S
Program

c
Speed
PPS
36,735
24,941
28,729
21,207
10,872
5,663
10,917
9,094
1,346
1,027
30,000
20,000
10,000
0
HISAT2.L.D
HISAT2.L.S
HISAT2.G.D
HISAT2.G.S
Bowtie2.D
Bowtie2.S
BWA-mem.D
BWA-mem.S
vg.G.D
vg.G.S
Program

d
Memory
Memory usage (GB)
4.6
4.7
8
8
3.4
3.4
5.9
6.1
29.2
29.2
30
20
10
0
HISAT2.L.D
HISAT2.L.S
HISAT2.G.D
HISAT2.G.S
Bowtie2.D
Bowtie2.S
BWA-mem.D
BWA-mem.S
vg.G.D
vg.G.S
Program

Graph alignment improves sensitivity

	10 million read pairs with SNPs and 0.2% per base sequencing error					10 million read pairs with SNPs and no sequencing error				
	C	UC	SC	SUC	PPS	C	UC	SC	SUC	PPS
HISAT2.Linear (default)	96.52%	92.71%	93.98%	93.49%	36,735	97.05%	93.15%	94.65%	94.15%	37,934
HISAT2.Linear (sensitive)	97.54%	93.72%	95.79%	95.27%	24,941	97.83%	93.92%	96.07%	95.55%	27,331
HISAT2.Graph (default)	99.08%	95.34%	98.63%	98.40%	28,729	99.36%	95.54%	98.84%	98.62%	32,096
HISAT2.Graph (sensitive)	99.19%	95.45%	98.77%	98.53%	21,207	99.36%	95.54%	98.84%	98.61%	25,639
Bowtie2 (default)	95.99%	95.99%	97.44%	97.44%	10,872	96.05%	96.05%	97.50%	97.50%	10,575
Bowtie2 (sensitive)	97.68%	94.68%	97.61%	97.04%	5,663	97.85%	94.77%	97.63%	97.07%	5,597
BWA-mem (default)	94.02%	92.47%	94.85%	94.49%	10,917	94.03%	92.49%	94.83%	94.47%	12,110
BWA-mem (sensitive)	97.42%	93.11%	96.42%	95.29%	9,094	97.57%	93.15%	96.40%	95.28%	10,106
VG.Linear (default)	95.56%	95.56%	96.91%	96.91%	1,315	95.34%	95.34%	96.65%	96.65%	1,367
VG.Linear (sensitive)	97.31%	89.74%	97.27%	92.27%	1,012	97.18%	90.31%	97.14%	92.71%	1,028
VG.Graph (default)	96.57%	96.57%	98.97%	98.97%	1,346	96.64%	96.64%	99.02%	99.02%	1,413
VG.Graph (sensitive)	98.18%	91.71%	99.10%	95.64%	1,027	98.37%	91.51%	99.16%	95.40%	1,083