

# deSALT: fast and accurate long transcriptomic read alignment with de Bruijn graph-based index

## Supplementary Material

Bo Liu<sup>1,2</sup>, Yadong Liu<sup>1,2</sup>, Junyi Li<sup>1</sup>, Hongzhe Guo<sup>1</sup>, Tianyi Zang<sup>1,\*</sup>, and Yadong Wang<sup>1,\*</sup>

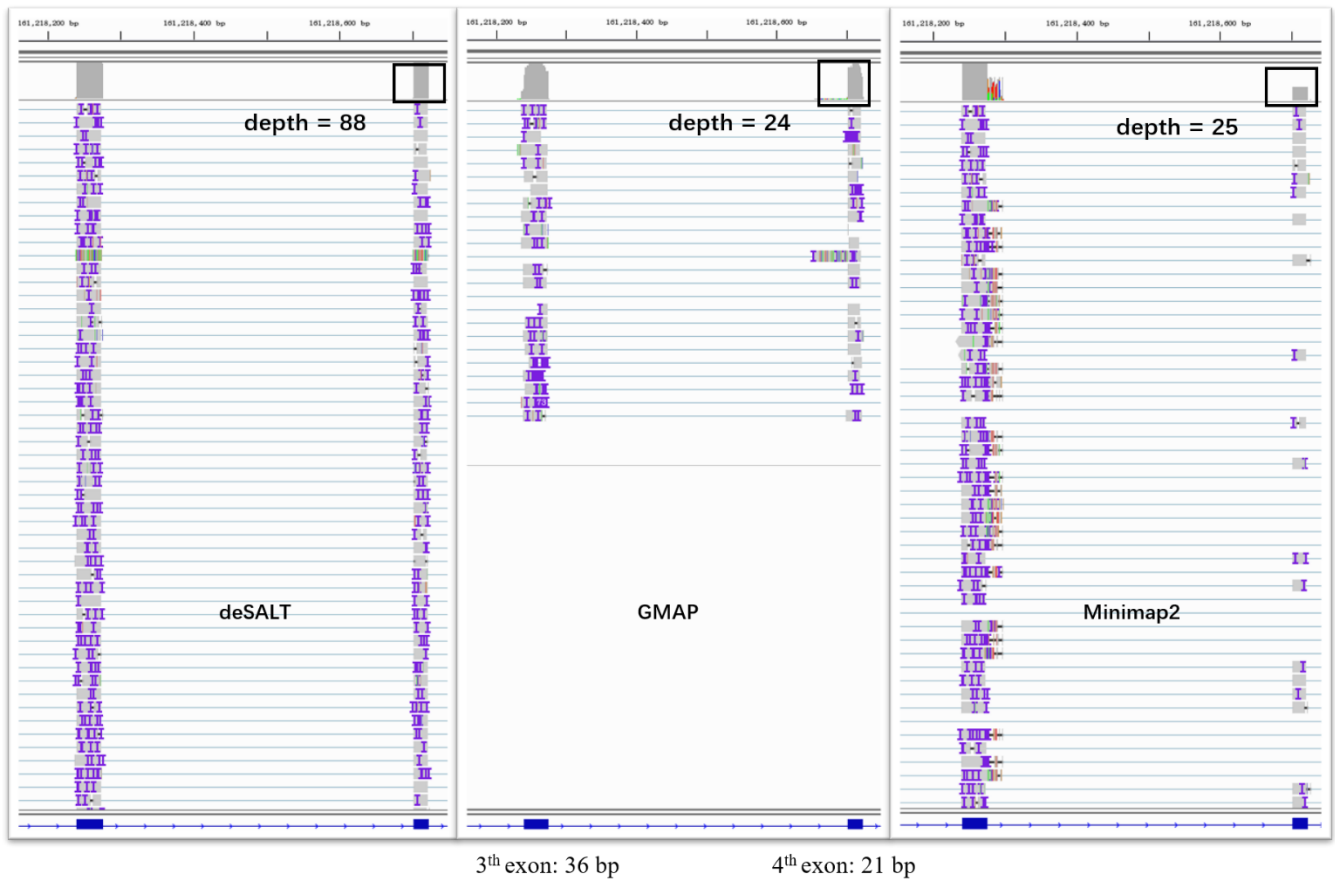
<sup>1</sup>Center for Bioinformatics, School of Computer Science and Technology, Harbin Institute of Technology, Harbin, Heilongjiang 150001, China

<sup>2</sup> These authors contributed equally to this work.

Contact: [bo.liu@hit.edu.cn](mailto:bo.liu@hit.edu.cn), [ydlu@hit.edu.cn](mailto:ydlu@hit.edu.cn), [lijunyi@hit.edu.cn](mailto:lijunyi@hit.edu.cn), [hzguo@hit.edu.cn](mailto:hzguo@hit.edu.cn), [tianyi.zang@hit.edu.cn](mailto:tianyi.zang@hit.edu.cn), and [ydwang@hit.edu.cn](mailto:ydwang@hit.edu.cn)

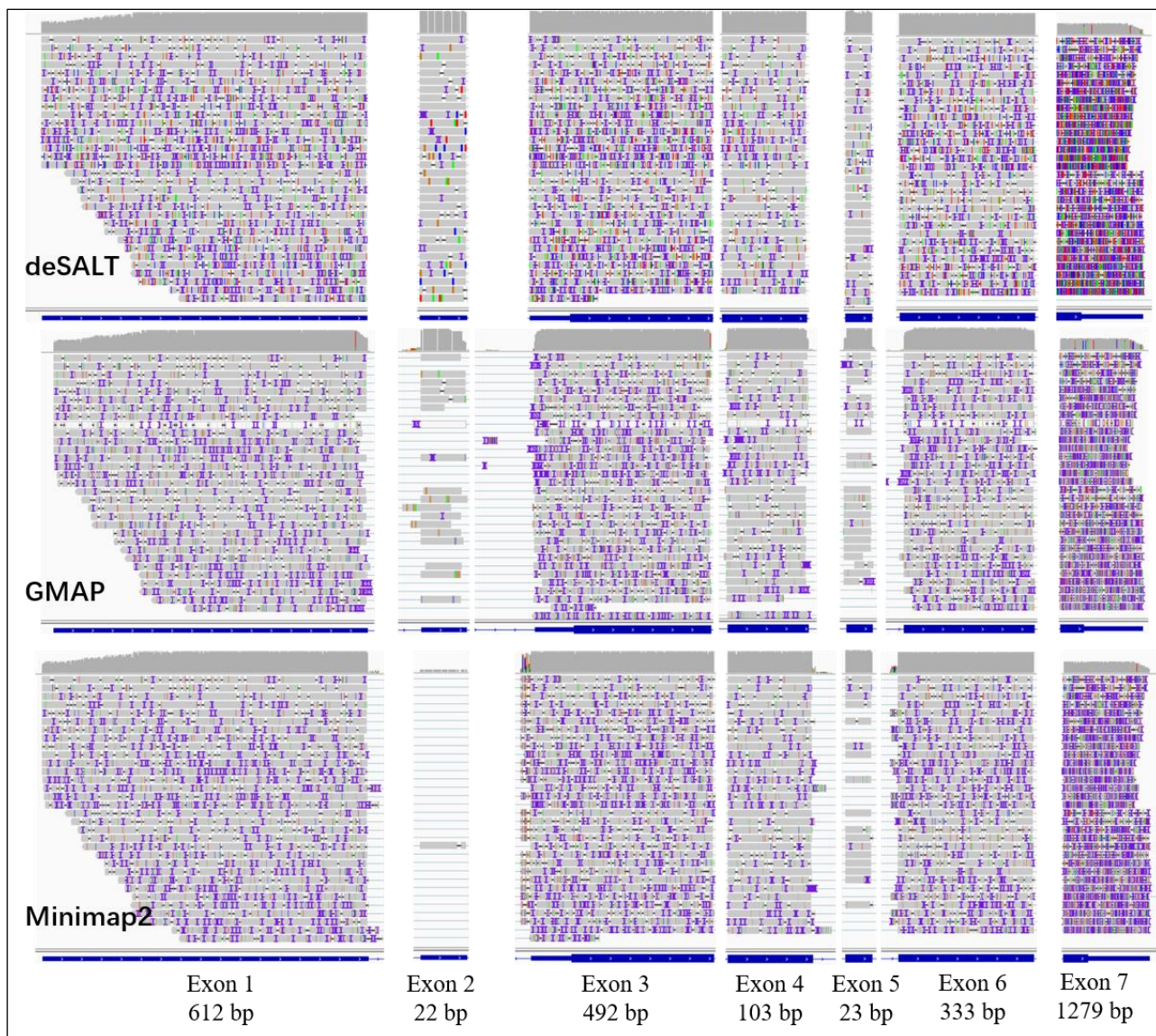
## Contents

Supplementary Figure 1. An example of the alignment of simulated reads around short exons .....	3
Supplementary Figure 2. An example of the alignment of simulated reads from transcripts with multiple exons .....	4
Supplementary Figure 3. An example of the alignment of error-prone reads by deSALT with and without gene annotations .....	5
Supplementary Figure 4. An example of the alignment of real sequencing reads from transcripts with small exons .....	6
Supplementary Figure 5. An example of the homogeneous alignment of real sequencing reads by deSALT .....	7
Supplementary Figure 6. An example of the reads aligned to intron regions .....	8
Supplementary Figure 7. An example of the reads aligned to intergenic regions .....	9
Supplementary Table 1. List of simulated datasets .....	10
Supplementary Table 2. Overall benchmark results on the simulated datasets <sup>a</sup> .....	12
Supplementary Table 3. Benchmark results on the simulated reads with short exons <sup>a</sup> .....	19
Supplementary Table 4. Benchmark results on the simulated reads from the transcripts with various numbers of exons <sup>a</sup> .....	26
Supplementary Table 5. Benchmark results on the simulated reads from the genes with one/multiple isoforms <sup>a</sup> .....	44
Supplementary Table 6. Alignment speed on simulated human datasets with various numbers of threads <sup>a</sup> .....	57
Supplementary Table 7. Benchmark results on the real sequencing datasets .....	59
Supplementary Table 8. Numbers of exactly matched exons of various lengths on the real sequencing datasets <sup>a</sup> .....	60
Supplementary Table 9. Numbers of in silico transcript sequences from various sets of genes <sup>a</sup> .....	61
Supplementary Notes .....	62
1. Alignment skeleton generation (first-pass alignment) .....	62
1.1 Genome indexing .....	62
1.2 Match-block generation .....	62
1.3 SDP-based alignment skeleton generation .....	62
2. Exon inference .....	63
2.1 Draft exon generation .....	63
2.2 Exon refinement .....	64
3. Refined alignment (second-pass alignment) .....	65
3.1 Construction of a local spliced reference sequence (LSRS) .....	65
3.2 Alignment of the read part against the LSRS .....	65
4. Availability of datasets .....	66
4.1 The availability of the human ONT datasets .....	66
4.2 The availability of the mouse PacBio dataset (Accession Number: SRR6238555) .....	66
4.3 The availability of the simulated datasets .....	66
5. Availability of references .....	67
5.1 The availability of the reference genomes .....	67
5.2 The availability of the gene annotations .....	67
6. Implementation of benchmarking .....	67
6.1 Command lines of deSALT (version 1.1) .....	67
6.2 Command lines of Minimap2 (version 2.14-r883) .....	68
6.3 Command lines of GMAP (version 2017-10-12) .....	68
References .....	68

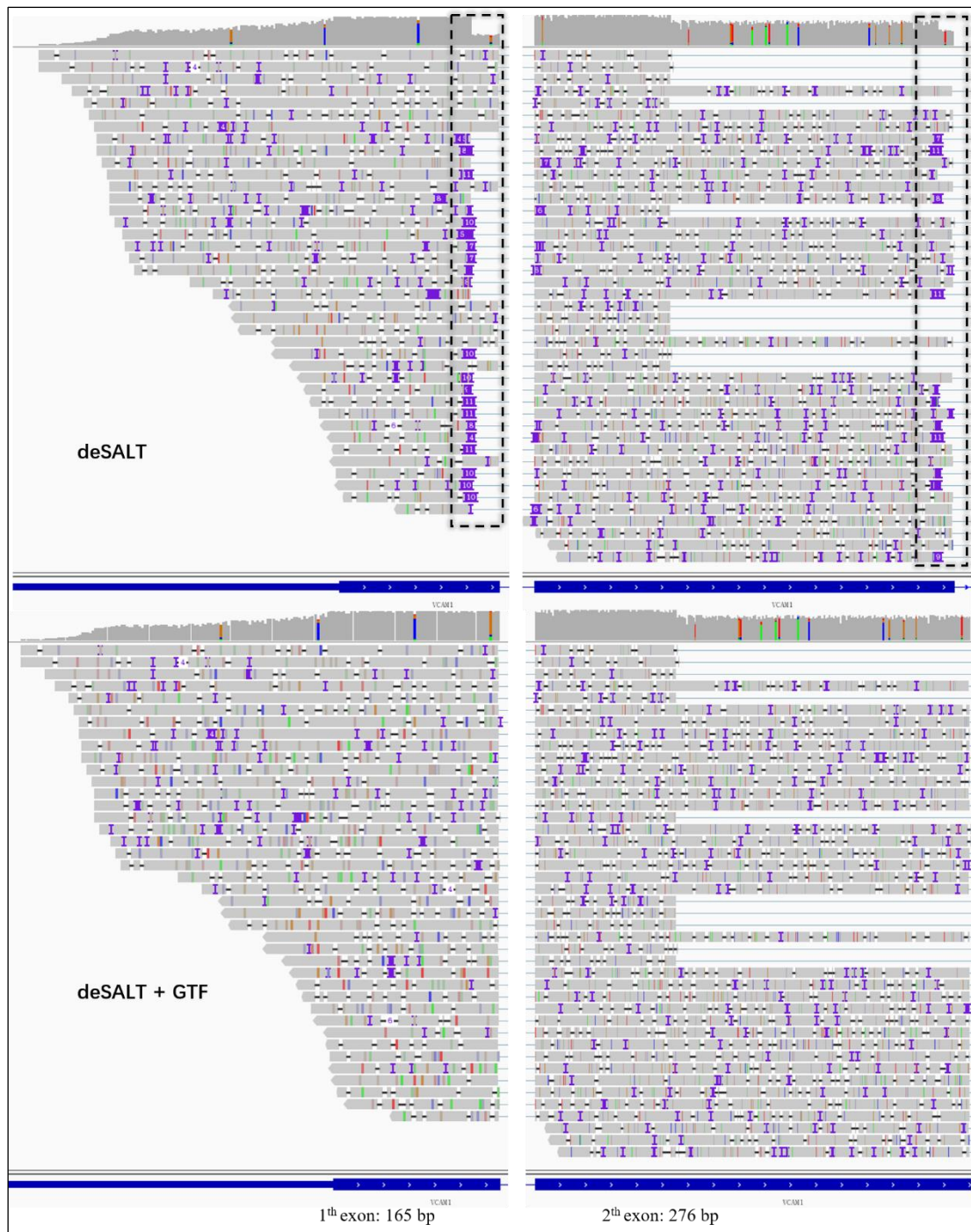


**Supplementary Figure 1. An example of the alignment of simulated reads around short exons**

This figure represents the Integrative Genome Viewer (IGV) [1] snapshots of the alignments of the reads from the simulated 30X PacBio subread human dataset around the FECR1G gene (Chr1: 16125297–16129248) of reference GRCh38. FECR1G has 5 exons, and the snapshots focus on the third and fourth exons, whose lengths are 36 bp and 21 bp, respectively. According to the ground truth, there are 88 reads spanning these two short exons, and deSALT correctly aligns all of them. However, GMAP only correctly aligns a proportion (i.e., 24) of these reads to these two exons. Minimap2 also aligns most of the reads to the third exon; however, it only aligns a few (i.e., 25) reads to the fourth exon. This case suggests the outstanding ability of deSALT to handle the short exon parts of reads, which is beneficial to the production of accurate alignments.



**Supplementary Figure 2. An example of the alignment of simulated reads from transcripts with multiple exons**  
 This figure represents the Integrative Genome Viewer (IGV) [1] snapshots of the alignments of the reads from the simulated 30X ONT 2D (1D<sup>2</sup>) fruit fly dataset around the CG42747 gene (Chr3L: 6400590–6485964) of reference BDGP6 (It is worth noting that the intron regions are removed in the figure). The region contains 7 exons of gene CG42747 (their lengths are given at the bottom of the figure), and there are 32 reads in total in this gene region according to the ground truth. It is observed from the figure that deSALT can correctly produce full-length alignments for all the reads spanning the multiple exons of the CG42747 gene. However, for GMAP and Minimap2, some of the reads are not correctly aligned. GMAP misses the short exon parts for a proportion of the reads, and for two of the reads it only aligns a few bases. Minimap2 correctly aligns only one of the reads for its exon2 part and nearly 50% of the reads for their exon5 parts. This case suggests the ability of deSALT to handle the reads from transcripts with multiple exons.



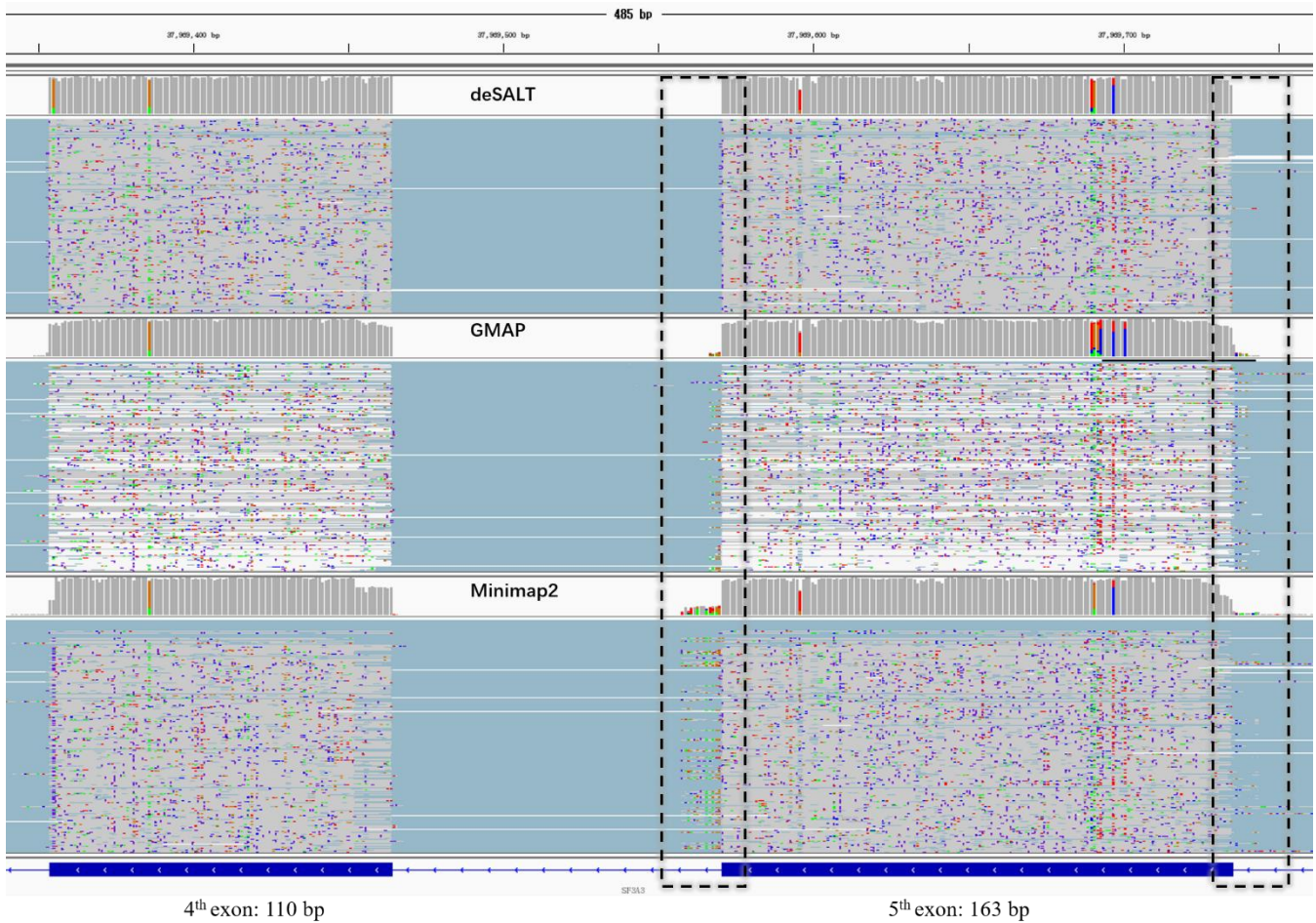
**Supplementary Figure 3. An example of the alignment of error-prone reads by deSALT with and without gene annotations**

This figure represents the Integrative Genome Viewer (IGV) [1] snapshots of the alignments of the reads from the simulated 10X ONT 1D human dataset around the VCAM1 gene (Chr1: 100719640–100739046) of reference GRCh38. The VCAM1 gene has 9 exons, and the snapshots focus on the first two exons; moreover, it is also worth noting that the second exon has two alternative splicing sites (which are respectively at Chr1: 100720565 and Chr1: 100720751). According to the ground truth, there are 43 reads in total in this gene region, and the alignments of deSALT with (termed “deSALT+GTF” in the figure) and without (termed “deSALT” in the figure) annotations are represented in the lower and upper parts of the figure. It is observed that the serious sequencing errors (error rate: 25%) affect the alignment of the reads, especially the bases near the exon boundaries (marked by dashed rectangles in the upper part of the figure). However, the bases can be more confidently aligned with the help of the gene annotation, and the alignment is obviously improved.



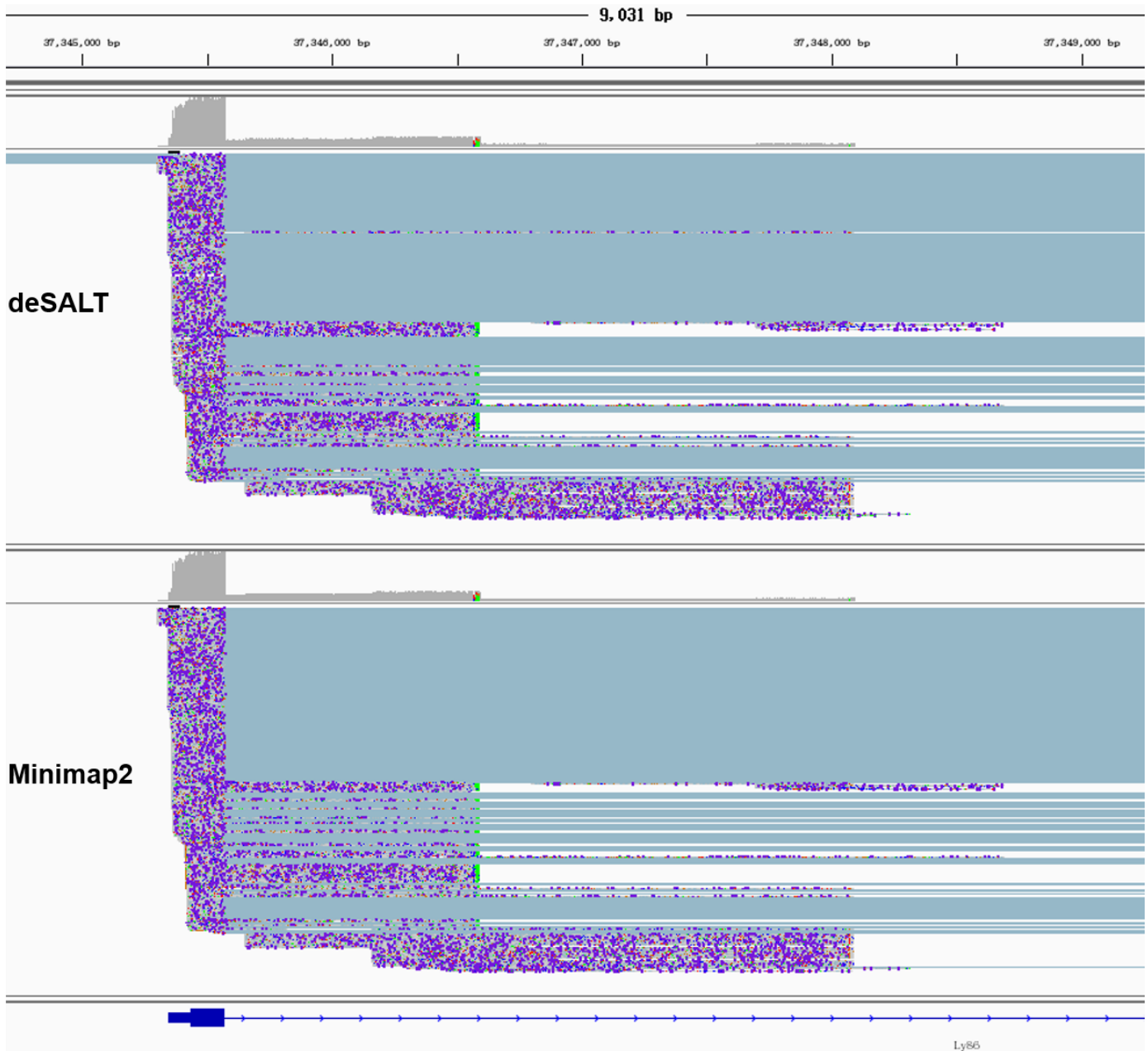
**Supplementary Figure 4. An example of the alignment of real sequencing reads from transcripts with small exons**

This figure represents the Integrative Genome Viewer (IGV) [1] snapshots of the alignments of the reads from the real mouse PacBio dataset around the Hn1 gene (Chr11: 115497353–115514370) of reference GRCm38. The Hn1 gene has one isoform containing 5 exons (according to Ensembl gene annotation; the lengths of the exons are marked at the bottom of the figure), and the length of exon2 is very short (i.e., 19 bp). In the alignment results, deSALT, GMAP, and Minimap2 respectively map 816, 863, and 857 reads to the Hn1 gene region. However, the number of reads being aligned to exon2 by GMAP (i.e., 502 reads) and Minimap2 (i.e., 54 reads) is much less than that of deSALT (710 reads). Moreover, the #ReadGA statistics of the aligners in the Hn1 region are 704 (deSALT), 467 (GMAP), and 48 (Minimap2). This case indicates that deSALT has better ability to handle small exons in real sequencing reads. Furthermore, this ability is helpful in producing full-length alignments for more reads.



### Supplementary Figure 5. An example of the homogeneous alignment of real sequencing reads by deSALT

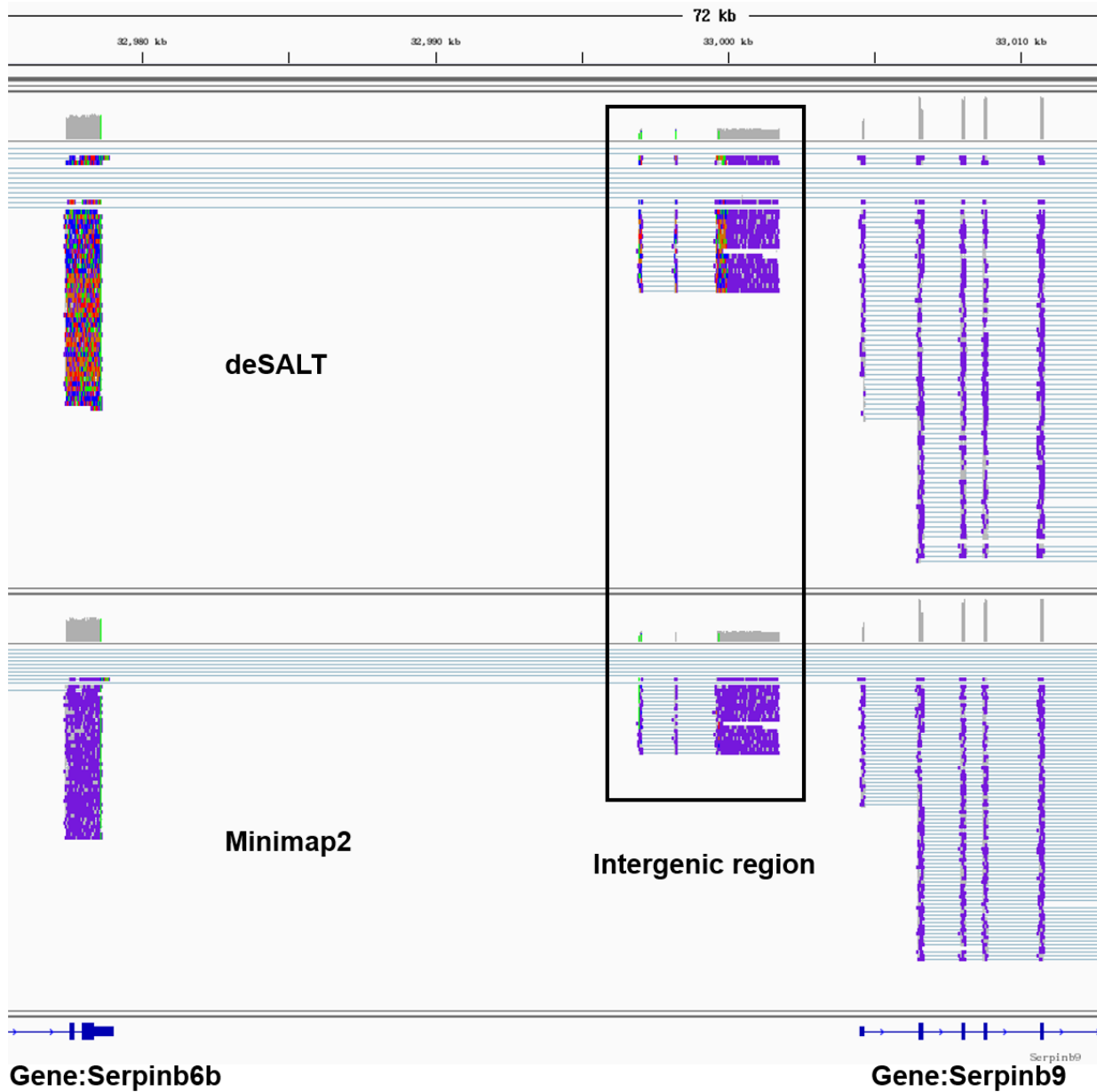
This figure represents the Integrative Genome Viewer (IGV) [1] snapshots of the alignments of the reads from the real human ONT dataset around the SF3A3 gene (Chr1: 37956980–37990110) of reference GRCh38. The SF3A3 gene has 17 exons (according to Ensembl gene annotation), and the snapshots focus on the fourth and fifth exons, whose lengths are 110 bp and 163 bp, respectively. In the alignment results, there are 658, 426, and 633 reads mapped to this region by deSALT, GMAP, and Minimap2, respectively. It is observed that the alignments produced by deSALT are highly homogeneous. However, the alignments produced by GMAP are more heterogeneous for both the inner regions and boundaries of the two exons. The alignments produced by Minimap2 are even more heterogeneous at the exon boundaries, although they are not as heterogeneous as those produced by GMAP in the inner regions. Moreover, the #ReadGA statistics of the aligners in the SF3A3 region are respectively 411 (deSALT), 109 (GMAP) and 173 (Minimap2), and the ratios #BaseGA/#BaseT of the aligners are respectively 89.82% (deSALT), 86.92% (GMAP), and 83.7% (Minimap2), where #BaseT is the total number of bases aligned to the SF3A3 region by the corresponding aligner. This case indicates that deSALT enables the production of homogeneous read alignments; furthermore, this ability has the potential to improve the overall alignment accuracy.



**Supplementary Figure 6. An example of the reads aligned to intron regions**

This figure represents the Integrative Genome Viewer (IGV) [1] snapshots of the alignments of the reads from the real mouse PacBio dataset around the Ly86 gene (Chr13: 37345345–37419036) of reference GRCm38. The Ly86 gene has 5 exons (according to Ensembl gene annotation), and the snapshots focus on the intron region between the first and second exons. It is observed from the alignment results that both deSALT and Minimap2 align 430 reads to this region. Moreover, similar numbers of reads are aligned to the intron region between the first and second exons, and their detailed alignments (i.e., CIGARs) produced by the two aligners are also similar. Considering the numerous aligned reads and the similar alignments independently produced by various aligners, there could be unannotated splicing events in this region.





**Supplementary Figure 7. An example of the reads aligned to intergenic regions**

This figure represents the Integrative Genome Viewer (IGV) [1] snapshots of the alignments of the reads from the real mouse PacBio dataset in the intergenic region between the Serpinb6b gene (Chr13: 32965513–32979037) and Serpinb9 gene (Chr13: 33004541–33017955) of reference GRCm38. It is observed from the alignment results that deSALT and Minimap2 align 25 and 27 reads to this region, respectively, and the alignments of the reads are also highly similar. This indicates that there could be unannotated transcripts in this region.

**Supplementary Table 1. List of simulated datasets**

No.	Type <sup>a</sup>	Depth	Average length(bp)	Sequencing error rate	Error ratio <sup>b</sup>	#of reads	#of bases	Reference/ Annotations
1	Human PacBio ROI	4X	6000	2%	75:5:20	46783	57203247	GRCh38, version 94
2		10X	6000	2%	75:5:20	103339	141769367	
3		30X	6000	2%	75:5:20	292117	424010883	
4	Human ONT 2D (1D <sup>2</sup> )	4X	7800	13%	41:23:36	46594	56604793	
5		10X	7800	13%	41:23:36	102920	141283445	
6		30X	7800	13%	41:23:36	299397	423898156	
7	Human PacBio subreads	4X	7800	15%	1:12:2	46537	56809800	
8		10X	7800	15%	1:12:2	102533	141326345	
9		30X	7800	15%	1:12:2	289375	423841890	
10	Human ONT1D	4X	7800	25%	48:15:37	46644	56517199	
11		10X	7800	25%	48:15:37	103532	141296228	
12		30X	7800	25%	48:15:37	309135	423843578	
13	Mouse PacBio ROI	4X	6000	2%	75:5:20	46812	66151862	GRCm38, version 94
14		10X	6000	2%	75:5:20	103498	164193767	
15		30X	6000	2%	75:5:20	292784	491418994	
16	Mouse ONT 2D (1D <sup>2</sup> )	4X	7800	13%	41:23:36	46591	65584331	
17		10X	7800	13%	41:23:36	102879	163762584	
18		30X	7800	13%	41:23:36	299413	491329071	
19	Mouse PacBio subreads	4X	7800	15%	1:12:2	46481	65756249	
20		10X	7800	15%	1:12:2	102433	163803693	
21		30X	7800	15%	1:12:2	289221	491278535	
22	Mouse ONT1D	4X	7800	25%	48:15:37	46600	65508866	
23		10X	7800	25%	48:15:37	103647	163775669	
24		30X	7800	25%	48:15:37	309276	491282118	
25	Fruit fly PacBio ROI	4X	6000	2%	75:5:20	45296	102311172	BDGP6, version 94
26		10X	6000	2%	75:5:20	101290	254849528	
27		30X	6000	2%	75:5:20	288355	763849030	
28	Fruit fly ONT 2D (1D <sup>2</sup> )	4X	7800	13%	41:23:36	44445	101867685	
29		10X	7800	13%	41:23:36	98980	254601058	
30		30X	7800	13%	41:23:36	288337	763814989	
31	Fruit fly PacBio subreads	4X	7800	15%	1:12:2	44169	101945130	
32		10X	7800	15%	1:12:2	97910	254613734	

33		30X	7800	15%	1:12:2	277541	763798783
34	Fruit fly ONT1D	4X	7800	25%	48:15:37	44546	101842292
35		10X	7800	25%	48:15:37	100127	254608663
36		30X	7800	25%	48:15:37	297821	763802581

- a) The reads are simulated by PBsim; also refer to the Methods section in the main text and the Supplementary Notes for a more detailed description of the generation of the simulated datasets and the PBsim command lines used.
- b) The error models of various platforms, i.e., the total sequencing error rates and the ratios of the sequencing errors (represented as mismatches: insertions: deletions) are configured by referring to previous studies. More precisely, the error models of the PacBio ROI datasets, ONT 2D (1D<sup>2</sup>) datasets, and ONT1D datasets are configured by referring to the previous study conducted by Weirather *et al.* [2], and the error model of the PacBio subread datasets are configured by referring to the previous study conducted by Carneiro *et al.* [3].

**Supplementary Table 2. Overall benchmark results on the simulated datasets <sup>a</sup>**

Aligner	Parameters <sup>b</sup>	Base% <sup>c</sup>	Exons% <sup>d</sup>	Read80% <sup>e</sup>	Read100% <sup>f</sup>
<b>Simulated PacBio ROI datasets</b>					
<b>Fruit fly PacBio ROI, coverage = 4X, 45296 reads/102311172 bases/230387 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.63	96.47	98.32	88.83
Minimap2	-ax splice	99.55	98.01	98.75	92.78
deSALT	-d 10 -x ccs	99.66	98.67	99.55	94.95
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.66	98.68	99.55	95.03
<b>Fruit fly PacBio ROI, coverage = 10X, 101290 reads/254849528 bases/561473 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.63	96.62	98.11	88.5
Minimap2	-ax splice	99.57	98.33	99.15	93.32
deSALT	-d 10 -x ccs	99.67	98.91	99.59	95.44
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.68	98.91	99.58	95.49
<b>Fruit fly PacBio ROI, coverage = 30X, 288355 reads/763849030 bases/1666790 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.64	96.62	98.01	88.1
Minimap2	-ax splice	99.62	98.44	99.4	93.49
deSALT	-d 10 -x ccs	99.69	99.02	99.61	95.66
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.69	99.02	99.59	95.7
<b>Mouse PacBio ROI, coverage = 4X, 46812 reads/66151862 bases/232811 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.31	97.16	97.96	90.25
Minimap2	-ax splice	98.25	95.88	96.25	89.55
deSALT	-d 10 -x ccs	98.47	97.28	97.67	91.84
deSALT + GTF	-d 10 -x ccs -G anno.gtf	98.48	97.32	97.69	92.01
<b>Mouse PacBio ROI, coverage = 10X, 103498 reads/164193767 bases/564165 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.3	94.68	96	84.27
Minimap2	-ax splice	98.37	96.47	96.99	90.88
deSALT	-d 10 -x ccs	98.57	97.68	97.87	92.74
deSALT + GTF	-d 10 -x ccs -G anno.gtf	98.6	97.73	97.88	92.96
<b>Mouse PacBio ROI, coverage = 30X, 292784 reads/491418994 bases/1669598 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.32	94.8	95.93	84.29

Minimap2	-ax splice	98.38	96.72	97.29	91.52
deSALT	-d 10 -x ccs	98.56	97.88	97.95	93.22
deSALT + GTF	-d 10 -x ccs -G anno.gtf	98.57	97.92	97.96	93.47
<b>Human PacBio ROI, coverage = 4X, 46783 reads/57203247 bases/222656 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.9	95.14	96.97	86.51
Minimap2	-ax splice	98.65	96.48	96.8	90.81
deSALT	-d 10 -x ccs	99.19	97.81	98.38	93.44
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.23	97.9	98.44	93.75
<b>Human PacBio ROI, coverage = 10X, 103339 reads/141769367 bases/537916 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.96	95.39	96.81	86.23
Minimap2	-ax splice	98.77	97.16	97.59	92.2
deSALT	-d 10 -x ccs	99.22	98.23	98.57	94.34
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.26	98.32	98.59	94.63
<b>Human PacBio ROI, coverage = 30X, 292117 reads/424010883 bases/1591417 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.96	95.46	96.79	86.37
Minimap2	-ax splice	98.8	97.36	98	92.95
deSALT	-d 10 -x ccs	99.25	98.39	98.71	94.91
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.28	98.48	98.7	95.29
<b>Simulated ONT 2D (1D<sup>2</sup>) datasets</b>					
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 4X, 44445 reads/101867685 bases/229107 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.26	74.17	74.08	43.78
Minimap2	-ax splice	95.14	92.64	91.47	72.43
deSALT	-d 10 -x ont2d -s 2	95.63	95.28	96.45	80.43
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.68	96.09	97.31	83.55
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 10X, 98980 reads/254601058 bases/560709 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.46	74.86	73.28	42.3
Minimap2	-ax splice	95.45	94.08	95.87	75.59
deSALT	-d 10 -x ont2d -s 2	95.76	96.79	98.1	84.62
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.79	97.26	98.5	86.71
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 30X, 288337 reads/763814989 bases/1670965 exons</b>					

GMAP	-f samse --cross-species -z sense_force	94.45	74.79	72.76	41.83
Minimap2	-ax splice	95.48	94.29	95.9	76.1
deSALT	-d 10 -x ont2d -s 2	95.8	97.28	98.47	86.7
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.82	97.57	98.62	87.85
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 4X, 46591 reads/65584331 bases/232141 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.51	68.52	67.98	40.13
Minimap2	-ax splice	91.92	85.88	81.23	59.36
deSALT	-d 10 -x ont2d -s 2	93.64	91.5	91.1	72.01
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	93.76	92.65	92.3	75.73
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 10X, 102879 reads/163762584 bases/565884 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.84	69.48	67.89	39.59
Minimap2	-ax splice	92.59	88.31	88.24	64.64
deSALT	-d 10 -x ont2d -s 2	94.09	93.77	94.26	77.73
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.15	94.37	94.81	80.08
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 30X, 299413 reads/491329071 bases/1688076 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.92	69.46	67.28	39.42
Minimap2	-ax splice	92.68	88.57	88.57	65
deSALT	-d 10 -x ont2d -s 2	94.08	94.12	94.58	79.06
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.12	94.54	94.91	80.69
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 4X, 46594 reads/56604793 bases/223025 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.7	68.99	67.62	40.72
Minimap2	-ax splice	91.77	85.57	79.95	58.76
deSALT	-d 10 -x ont2d -s 2	94.36	91.81	91.11	72.83
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.48	93.01	92.55	76.51
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 10X, 102920 reads/141283445 bases/543017 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.88	69.22	67.14	40.21
Minimap2	-ax splice	92.57	87.83	87.85	64.07
deSALT	-d 10 -x ont2d -s 2	94.78	93.86	94.84	78.28
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.86	94.64	95.59	81.05
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 30X, 299397 reads/423898156 bases/1615900 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.96	79.05	77.04	50.44
Minimap2	-ax splice	92.69	88.29	88.03	64.53

deSALT	-d 10 -x ont2d -s 2	94.85	94.33	95.06	79.63
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.91	94.92	95.63	81.6
<b>Simulated PacBio subread datasets</b>					
<b>Fruit fly PacBio subread, coverage = 4X, 44169 reads/101945130 bases/212417 exons</b>					
GMAP	-f samse --cross-species -z sense_force	81.93	62.44	52.23	28.47
Minimap2	-ax splice	87.7	90.76	85.07	65.73
deSALT	-d 10 -x clr -s 2	88.39	95.01	95.71	78.98
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.42	95.62	96.29	81.12
<b>Fruit fly PacBio subread, coverage = 10X, 97910 reads/254613734 bases/518798 exons</b>					
GMAP	-f samse --cross-species -z sense_force	82.13	62.94	51.9	27.45
Minimap2	-ax splice	88.15	92.95	93.98	72.28
deSALT	-d 10 -x clr -s 2	88.55	96.4	97.83	83.32
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.56	96.71	98.09	84.55
<b>Fruit fly PacBio subread, coverage = 30X, 277541 reads/763798783 bases/1540635 exons</b>					
GMAP	-f samse --cross-species -z sense_force	82.58	85.25	52.79	27.41
Minimap2	-ax splice	88.3	99	96	74.27
deSALT	-d 10 -x clr -s 2	88.58	96.92	98.34	84.78
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.59	97.16	98.44	85.76
<b>Mouse PacBio subread, coverage = 4X, 46481 reads/65756249 bases/216580 exons</b>					
GMAP	-f samse --cross-species -z sense_force	74.2	52.37	42.76	24.27
Minimap2	-ax splice	84.5	83.3	75.27	52.45
deSALT	-d 10 -x clr -s 2	86.45	90.06	88.57	67.79
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.53	91.19	89.86	71.17
<b>Mouse PacBio subread, coverage = 10X, 102433 reads/163803693 bases/526135 exons</b>					
GMAP	-f samse --cross-species -z sense_force	74.2	52.55	41.02	22.68
Minimap2	-ax splice	85.18	85.98	84.42	58.65
deSALT	-d 10 -x clr -s 2	86.89	92.56	92.89	73.95
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.93	93.16	93.56	76.08
<b>Mouse PacBio subread, coverage = 30X, 289221 reads/491278535 bases/1559863 exons</b>					
GMAP	-f samse --cross-species -z sense_force	74.7	53.32	40.85	22.39

Minimap2	-ax splice	85.49	87.22	87.99	60.97
deSALT	-d 10 -x clr -s 2	86.98	93.5	94.49	76.25
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87	93.9	94.79	77.68
<b>Human PacBio subread, coverage = 4X, 46537 reads/56809800 bases/208419 exons</b>					
GMAP	-f samse --cross-species -z sense_force	71.43	49.52	39.77	22.84
Minimap2	-ax splice	84.29	82.65	73.78	51.12
deSALT	-d 10 -x clr -s 2	86.98	90.05	88.24	68.13
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.12	91.33	89.84	71.52
<b>Human PacBio subread, coverage = 10X, 102533 reads/141326345 bases/504951 exons</b>					
GMAP	-f samse --cross-species -z sense_force	71.47	49.79	37.66	21.19
Minimap2	-ax splice	85.13	85.64	83.57	57.45
deSALT	-d 10 -x clr -s 2	87.57	92.84	93.44	74.29
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.64	93.59	94.32	76.62
<b>Human PacBio subread, coverage = 30X, 289375 reads/423841890 bases/1496976 exons</b>					
GMAP	-f samse --cross-species -z sense_force	71.75	50.28	37.04	20.8
Minimap2	-ax splice	85.48	86.92	87.58	60.13
deSALT	-d 10 -x clr -s 2	87.76	93.85	95.39	76.87
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.8	94.4	95.94	78.75
<b>Simulated ONT 1D datasets</b>					
<b>Fruit fly ONT 1D, coverage = 4X, 44546 reads/101842292 bases/240905 exons</b>					
GMAP	-f samse --cross-species -z sense_force	64.73	20.37	13.58	3.4
Minimap2	-ax splice	89.58	75.05	64.21	33.72
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	95.54	84.89	82.46	47.1
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	95.94	89.32	89.39	58.13
<b>Fruit fly ONT 1D, coverage = 10X, 100127 reads/254608663 bases/589616 exons</b>					
GMAP	-f samse --cross-species -z sense_force	65.95	20.86	13.25	3.25
Minimap2	-ax splice	90.98	77.97	69.81	36.9
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.24	89.12	88	55.5
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.47	92.51	93.29	66.57
<b>Fruit fly ONT 1D, coverage = 30X, 297821 reads/763802581 bases/1764192 exons</b>					



GMAP	-f samse --cross-species -z sense_force	66.01	20.91	13.24	3.2
Minimap2	-ax splice	91.05	78.03	69.96	37.11
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.43	90.94	91.28	61.83
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.56	93.23	94.1	69.45
<b>Mouse ONT 1D, coverage = 4X, 46600 reads/65508866 bases/243889 exons</b>					
GMAP	-f samse --cross-species -z sense_force	53.66	12.83	13.09	2.86
Minimap2	-ax splice	75.32	52.55	39.15	16.76
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	86.68	65.86	57.44	25.43
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.04	74.35	68.35	36.75
<b>Mouse ONT 1D, coverage = 10X, 103647 reads/163775669 bases/595736 exons</b>					
GMAP	-f samse --cross-species -z sense_force	54.88	13.51	12.34	2.66
Minimap2	-ax splice	77.03	55.27	42.83	18.66
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	89.22	74.21	67.75	34.25
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.98	80.11	76.05	45.76
<b>Mouse ONT 1D, coverage = 30X, 309276 reads/491282118 bases/1783381 exons</b>					
GMAP	-f samse --cross-species -z sense_force	35.45	8.14	9.68	2.02
Minimap2	-ax splice	77.13	55.53	42.97	18.87
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	89.75	77.5	72.65	40.99
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	90.16	81.1	77.19	49.07
<b>Human ONT 1D, coverage = 4X, 46644 reads/56517199 bases/233621 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.06	11.52	10.59	2.5
Minimap2	-ax splice	70.68	47.38	33.84	14.5
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	85.16	64.34	55.25	25.24
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	86.64	72.08	65.13	35.27
<b>Human ONT 1D, coverage = 10X, 103532 reads/141296228 bases/569988 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.92	12.13	10.17	2.38
Minimap2	-ax splice	72.45	50.34	37.46	16.32
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	87.73	71.89	65.03	32.99
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.59	77.52	72.78	43.43
<b>Human ONT 1D, coverage = 30X, 309135 reads/423843578 bases/1706339 exons</b>					
GMAP	-f samse --cross-species -z sense_force	50.12	12.05	10.19	2.36
Minimap2	-ax splice	72.6	50.47	37.54	16.4

deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.36	74.52	69.05	38.61
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.88	78.36	73.71	46.23

- a) The table depicts the results of the aligners on the 36 simulated datasets.
- b) Refer to the Supplementary Notes for the command lines of the benchmarked aligners.
- c) Base%: the proportion of bases being correctly aligned to their ground truth positions, i.e., the mapped positions of the bases are within 5 bp of their ground truth positions.
- d) Exon%: the proportion of exons being correctly mapped. An exon in a certain read is considered to be correctly mapped only if its two boundaries are mapped within 5 bp of their ground truth positions in the reference genome.
- e) Read80%: the proportion of Read80% reads. A read is considered to be a Read80% read only if it meets two conditions:  $N_T/N_G > 80\%$  and  $N_T/N_P > 80\%$ , where  $N_G$  is the number of ground truth exons within the read,  $N_P$  is the number of exons predicted by the alignment, and  $N_T$  is the number of true positive exons. Herein, a predicted exon is considered to be a true positive exon only if there is a ground truth exon in the read and the corresponding boundaries of the predicted exon and the ground truth exon are within 5 bp.
- f) Read100%: the proportion of Read100% reads. A read is considered to be a Read100% read only if it meets two conditions:  $N_T/N_G = 100\%$  and  $N_T/N_P = 100\%$ . It is worth noting that a Read100% read indicates that the read has a highly correct full-length alignment.

**Supplementary Table 3. Benchmark results on the simulated reads with short exons <sup>a</sup>**

Aligner	Parameters <sup>b</sup>	Base% <sup>c</sup>	Exons% <sup>d</sup>	Read80% <sup>e</sup>	Read100% <sup>f</sup>
<b>Simulated PacBio ROI datasets</b>					
<b>Fruit fly PacBio ROI, coverage = 4X, 2657 reads/9799450 bases/29172 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.17	95.43	96.39	70.91
Minimap2	-ax splice	99.23	93.47	95.22	54.12
deSALT	-d 10 -x ccs	99.31	96.42	98.27	69.74
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.31	96.46	98.31	69.85
<b>Fruit fly PacBio ROI, coverage = 10X, 6624 reads/25034865 bases/73558 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.88	95.27	96.78	70.73
Minimap2	-ax splice	99.91	93.47	95.79	54.63
deSALT	-d 10 -x ccs	99.97	96.73	98.96	72.95
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.99	97.43	98.96	73.95
<b>Fruit fly PacBio ROI, coverage = 30X, 19879 reads/76182159 bases/222242 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.61	94.62	96.34	69.9
Minimap2	-ax splice	99.68	92.96	95.8	54.1
deSALT	-d 10 -x ccs	99.76	96.29	99.01	72.42
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.77	96.28	99.04	72.5
<b>Mouse PacBio ROI, coverage = 4X, 1369 reads/3348645 bases/18578 exons</b>					
GMAP	-f samse --cross-species -z sense_force	96.43	95.13	95.03	73.05
Minimap2	-ax splice	94.09	88.44	85.24	48.58
deSALT	-d 10 -x ccs	96.04	94.1	92.7	70.27
deSALT + GTF	-d 10 -x ccs -G anno.gtf	96.04	94.45	93.7	70.87
<b>Mouse PacBio ROI, coverage = 10X, 3332 reads/8371168 bases/45615 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.06	94.34	91.42	64.2
Minimap2	-ax splice	97.12	91.36	87.76	51.89
deSALT	-d 10 -x ccs	99.86	97.66	96.82	74.67
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.86	97.88	97.82	74.97
<b>Mouse PacBio ROI, coverage = 30X, 10012 reads/25528211 bases/138591 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.66	93.86	90.86	63.18

Minimap2	-ax splice	96.53	90.93	87.4	51.32
deSALT	-d 10 -x ccs	99.32	97.09	96.47	72.87
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.32	97.89	97.47	73.87
<b>Human PacBio ROI, coverage = 4X, 1244 reads/2645652 bases/16104 exons</b>					
GMAP	-f samse --cross-species -z sense_force	97.35	93.01	93.25	70.34
Minimap2	-ax splice	95.35	91.77	91.88	63.42
deSALT	-d 10 -x ccs	99.08	96.57	97.99	75.56
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.08	96.51	98.15	75.16
<b>Human PacBio ROI, coverage = 10X, 3085 reads/6673839 bases/40650 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.33	93.53	93.45	69.14
Minimap2	-ax splice	96.22	92.49	91.64	63.4
deSALT	-d 10 -x ccs	99.16			99.16
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.76			99.79
<b>Human PacBio ROI, coverage = 30X, 9243 reads/20206849 bases/122929 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.68	93.87	93.48	69.14
Minimap2	-ax splice	96.23	92.45	91.5	63.09
deSALT	-d 10 -x ccs	99.8	97.48	98.05	77.72
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.8	97.5	98.06	78.16
<b>Simulated ONT 2D (1D<sup>2</sup>) datasets</b>					
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 4X, 2669 reads/101867685 bases/229107 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.69	70.17	49.42	15.51
Minimap2	-ax splice	96.01	86.63	85.09	22.22
deSALT	-d 10 -x ont2d -s 2	96.32	92.64	92.13	47.28
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	96.36	93.4	94.01	50.43
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 10X, 6632 reads/25802683 bases/74427 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.05	71.2	50.9	16.3
Minimap2	-ax splice	95.55	86.56	85.57	23.72
deSALT	-d 10 -x ont2d -s 2	95.88	94.18	95.75	56.77
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.9	94.7	96.58	59.86
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 30X, 19914 reads/78707188 bases/225342 exons</b>					

GMAP	-f samse --cross-species -z sense_force	94.6	71.09	50.21	16.26
Minimap2	-ax splice	95.99	86.85	85.87	23.67
deSALT	-d 10 -x ont2d -s 2	96.34	94.92	96.66	60.5
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	96.36	95.44	97.56	62.91
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 4X, 1326 reads/3335501 bases/18089 exons</b>					
GMAP	-f samse --cross-species -z sense_force	88.38	64.7	40.57	12.9
Minimap2	-ax splice	90.03	79.91	65.08	15.08
deSALT	-d 10 -x ont2d -s 2	92.8	88.6	82.5	39.14
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	92.88	89.07	83.33	39.82
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 10X, 3298 reads/8408061 bases/45667 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.22	66.14	40.39	12.49
Minimap2	-ax splice	91.39	81.76	67.4	16.16
deSALT	-d 10 -x ont2d -s 2	94.34	92.01	85.63	47.63
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.38	92.28	86.39	48.85
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 30X, 9934 reads/25658107 bases/139226 exons</b>					
GMAP	-f samse --cross-species -z sense_force	89.4	66.1	41.68	12.86
Minimap2	-ax splice	90.45	80.59	67.3	15.93
deSALT	-d 10 -x ont2d -s 2	93.14	91.02	86.65	48.88
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	93.17	91.25	87.44	49.92
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 4X, 1246 reads/2631404 bases/16424 exons</b>					
GMAP	-f samse --cross-species -z sense_force	89.1	67.68	45.83	17.26
Minimap2	-ax splice	90.25	82.56	71.27	27.61
deSALT	-d 10 -x ont2d -s 2	95.06	90.99	87.24	47.99
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.08	91.5	88.28	48.72
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 10X, 3099 reads/6631932 bases/41196 exons</b>					
GMAP	-f samse --cross-species -z sense_force	87.85	66.83	46.21	16.65
Minimap2	-ax splice	89.13	81.67	71.64	25.69
deSALT	-d 10 -x ont2d -s 2	93.98	91.64	89.64	53.05
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.09	92.11	90.22	54.24
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 30X, 9258 reads/20030911 bases/123352 exons</b>					
GMAP	-f samse --cross-species -z sense_force	88.76	77.41	62.58	23.74
Minimap2	-ax splice	93.5	85.8	77.23	26.07

deSALT	-d 10 -x ont2d -s 2	94.98	92.35	90.13	52.2
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.06	92.94	90.54	54.32
<b>Simulated PacBio subread datasets</b>					
<b>Fruit fly PacBio subread, coverage = 4X, 2425 reads/9649661 bases/26290 exons</b>					
GMAP	-f samse --cross-species -z sense_force	82.91	61.57	31.01	6.85
Minimap2	-ax splice	87.51	85.24	81.81	21.73
deSALT	-d 10 -x clr -s 2	87.86	92.52	91.42	48.95
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.88	93.43	92.49	53.28
<b>Fruit fly PacBio subread, coverage = 10X, 5981 reads/24299878 bases/66138 exons</b>					
GMAP	-f samse --cross-species -z sense_force	84.08	61.94	32.28	6.83
Minimap2	-ax splice	88.61	85.82	83.95	22.34
deSALT	-d 10 -x clr -s 2	89.06	94.6	95.27	58.67
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	89.07	95.04	96.12	61.04
<b>Fruit fly PacBio subread, coverage = 30X, 277541 reads/763798783 bases/1540635 exons</b>					
GMAP	-f samse --cross-species -z sense_force	84.08	61.94	32.28	6.83
Minimap2	-ax splice	88.61	85.82	83.95	22.34
deSALT	-d 10 -x clr -s 2	88.91	95.08	96.43	60.82
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.92	95.5	96.98	63.21
<b>Mouse PacBio subread, coverage = 4X, 1223 reads/3211130 bases/16386 exons</b>					
GMAP	-f samse --cross-species -z sense_force	76.75	54.13	20.2	4.91
Minimap2	-ax splice	83.45	79.3	61.82	13.65
deSALT	-d 10 -x clr -s 2	86.4	88.48	80.95	36.47
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.4	89.13	82.26	38.84
<b>Mouse PacBio subread, coverage = 10X, 2998 reads/8020991 bases/40870 exons</b>					
GMAP	-f samse --cross-species -z sense_force	77.02	54.63	20.28	4.87
Minimap2	-ax splice	83.13	79.02	63.04	12.94
deSALT	-d 10 -x clr -s 2	86.51	90.35	84.96	45.76
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.56	90.73	85.86	46.96
<b>Mouse PacBio subread, coverage = 30X, 9007 reads/24148406 bases/123267 exons</b>					
GMAP	-f samse --cross-species -z sense_force	77.42	54.43	19.77	4.74

Minimap2	-ax splice	83.46	79.34	64.32	13.08
deSALT	-d 10 -x clr -s 2	86.18	90.36	85.11	45.81
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.2	90.67	85.69	47.3
<b>Human PacBio subread, coverage = 4X, 1109 reads/2359725 bases/14010 exons</b>					
GMAP	-f samse --cross-species -z sense_force	74.53	55.65	23.08	5.68
Minimap2	-ax splice	82.54	81.52	68.98	22.45
deSALT	-d 10 -x clr -s 2	86.43	90.11	85.75	44.72
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.6	90.66	86.47	45.63
<b>Human PacBio subread, coverage = 10X, 2761 reads/5996015 bases/35333 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.83	56.22	22.6	6.16
Minimap2	-ax splice	84.06	82.85	70.01	23.65
deSALT	-d 10 -x clr -s 2	88.39	92.43	88.7	48.39
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.47	92.94	89.35	50.78
<b>Human PacBio subread, coverage = 30X, 8275 reads/17979517 bases/105647 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.24	56.24	23.08	5.73
Minimap2	-ax splice	83.64	82.73	70.98	22.65
deSALT	-d 10 -x clr -s 2	87.99	93.23	90.21	51.9
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.06	93.4	90.21	52.83
<b>Simulated ONT 1D datasets</b>					
<b>Fruit fly ONT 1D, coverage = 4X, 2829 reads/10366837 bases/31649 exons</b>					
GMAP	-f samse --cross-species -z sense_force	64.43	18.54	1.03	0.18
Minimap2	-ax splice	92.51	72.39	55.11	2.58
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.02	83.14	76.85	15.73
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.3	86.71	85.9	22.62
<b>Fruit fly ONT 1D, coverage = 10X, 7112 reads/26737114 bases/80525 exons</b>					
GMAP	-f samse --cross-species -z sense_force	64.21	18.64	1.03	0.17
Minimap2	-ax splice	92.93	73.51	56.72	2.64
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.38	87.11	83.52	26.07
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.59	89.94	89.57	34.52
<b>Fruit fly ONT 1D, coverage = 30X, 21225 reads/80613021 bases/242158 exons</b>					

GMAP	-f samse --cross-species -z sense_force	64.51	18.77	1.19	0.23
Minimap2	-ax splice	92.94	73.59	57.14	2.95
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.42	88.93	87.59	34.46
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.59	91.57	91.97	43.41
<b>Mouse ONT 1D, coverage = 4X, 1436 reads/3402252 bases/19736 exons</b>					
GMAP	-f samse --cross-species -z sense_force	50.38	12.69	0.49	0.14
Minimap2	-ax splice	77.72	55.96	26.39	1.11
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	87.14	66.07	40.04	3.34
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.59	73.09	53.41	6.62
<b>Mouse ONT 1D, coverage = 10X, 3536 reads/8738830 bases/49776 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.36	12.66	0.51	0.14
Minimap2	-ax splice	77.51	56.37	25.51	0.62
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	87.65	71.71	50.57	8.6
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.68	77.26	61.57	13.24
<b>Mouse ONT 1D, coverage = 30X, 10567 reads/26107819 bases/148826 exons</b>					
GMAP	-f samse --cross-species -z sense_force	29.87	7.28	0.29	0.03
Minimap2	-ax splice	77.74	56.62	25.69	0.68
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.36	74.87	56.04	13.24
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.19	78.9	63.96	18.66
<b>Human ONT 1D, coverage = 4X, 1317 reads/2603348 bases/16963 exons</b>					
GMAP	-f samse --cross-species -z sense_force	47.55	12.77	0.68	0.3
Minimap2	-ax splice	73.25	51.66	22.1	1.52
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	87.69	64.78	39.94	4.78
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.88	73.48	53.38	6.61
<b>Human ONT 1D, coverage = 10X, 3297 reads/6757929 bases/43499 exons</b>					
GMAP	-f samse --cross-species -z sense_force	47.97	13.16	0.55	0.15
Minimap2	-ax splice	72.61	51.49	24.17	1.4
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	87.61	70.77	53.11	10.01
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.63	77.71	62.6	12.68
<b>Human ONT 1D, coverage = 30X, 9917 reads/20588822 bases/132385 exons</b>					
GMAP	-f samse --cross-species -z sense_force	47.94	13.15	0.67	0.17
Minimap2	-ax splice	72.96	52.07	23.59	1.27



deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.29	72.85	55.21	13.21
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	90.03	78.8	63.05	16.17

- a) This table depicts the results of the aligners on all the reads with short exons (< 31 bp), which are extracted from the simulated datasets in advance.
- b) Refer to the Supplementary Notes for the command lines of the benchmarked aligners.
- c) Base%: the proportion of bases being correctly aligned to their ground truth positions, i.e., the mapped positions of the bases are within 5 bp of their ground truth positions.
- d) Exon%: the proportion of the exons being correctly mapped. An exon in a certain read is considered to be correctly mapped only if its two boundaries are mapped within 5 bp of their ground truth positions in the reference genome.
- e) Read80%: the proportion of Read80% reads. A read is considered to be a Read80% read only if it meets two conditions:  $N_T/N_G > 80\%$  and  $N_T/N_p > 80\%$ , where  $N_G$  is the number of ground truth exons within the read,  $N_p$  is the number of exons predicted by the alignment, and  $N_T$  is the number of true positive exons. Herein, a predicted exon is considered to be a true positive exon only if there is a ground truth exon in the read and the corresponding boundaries of the predicted exon and the ground truth exon are within 5 bp.
- f) Read100%: the proportion of Read100% reads. A read is considered to be a Read100% read only if it meets two conditions:  $N_T/N_G = 100\%$  and  $N_T/N_p = 100\%$ . It is worth noting that a Read100% read indicates that the read has a highly correct full-length alignment.

**Supplementary Table 4. Benchmark results on the simulated reads from the transcripts with various numbers of exons <sup>a</sup>**

Aligner	Parameters <sup>b</sup>	Base% <sup>c</sup>	Exons% <sup>d</sup>	Read80% <sup>e</sup>	Read100% <sup>f</sup>
<b>Simulated PacBio ROI datasets</b>					
<b>Fruit fly PacBio ROI, coverage = 4X, &lt; 6 exons events, 29600 reads/40847769 bases/69512 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.68	96.59	98.6	93.17
Minimap2	-ax splice	99.37	97.2	98.3	94.32
deSALT	-d 10 -x ccs	99.62	98.23	99.5	96.2
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.62	98.25	99.49	96.26
<b>Fruit fly PacBio ROI, coverage = 4X, 6-9 exons events, 8731 reads/28454497 bases/63936 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.1	95.8	97.81	84.4
Minimap2	-ax splice	99.13	97.76	99.23	92.11
deSALT	-d 10 -x ccs	99.15	98.29	99.2	94.61
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.16	98.31	99.2	94.62
<b>Fruit fly PacBio ROI, coverage = 4X, &gt; 9 exons events, 6965 reads/33008906 bases/96939 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.02	96.83	97.75	75.95
Minimap2	-ax splice	99.13	98.76	99.45	87.09
deSALT	-d 10 -x ccs	99.14	99.22	99.55	90.09
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.14	99.24	99.81	90.29
<b>Fruit fly PacBio ROI, coverage = 10X, &lt; 6 exons events, 62238 reads/100125992 bases/158496 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.23	96.41	98.07	93.26
Minimap2	-ax splice	98.97	97.55	98.66	95.29
deSALT	-d 10 -x ccs	99.2	98.24	99.31	96.81
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.2	98.23	99.29	96.85
<b>Fruit fly PacBio ROI, coverage = 10X, 6-9 exons events, 21463 reads/70301415 bases/157116 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.01	96.92	98.64	85.36
Minimap2	-ax splice	99.13	98.83	100.14	93.01
deSALT	-d 10 -x ccs	99.15	99.42	99.17	95.7
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.15	99.42	99.19	95.72
<b>Fruit fly PacBio ROI, coverage = 10X, &gt; 9 exons events, 17589 reads/84422121 bases/245861 exon</b>					
GMAP	-f samse --cross-species -z sense_force	99.71	96.57	97.6	75.48

Minimap2	-ax splice	99.82	98.51	99.67	86.74
deSALT	-d 10 -x ccs	99.84	99.02	99.87	90.26
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.84	99.02	99.87	90.41
<b>Fruit fly PacBio ROI, coverage = 30X, &lt; 6 exons events, 171178 reads/296437653 bases/454719 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.39	96.68	98.02	93.26
Minimap2	-ax splice	99.19	98.09	99.08	95.89
deSALT	-d 10 -x ccs	99.36	98.63	99.37	97.18
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.36	98.62	99.34	97.21
<b>Fruit fly PacBio ROI, coverage = 30X, 6-9 exons events, 64162 reads/211010822 bases/469724 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.67	96.49	98.36	84.77
Minimap2	-ax splice	99.72	98.48	99.89	92.58
deSALT	-d 10 -x ccs	99.74	99.13	99.92	95.6
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.74	99.13	99.92	95.6
<b>Fruit fly PacBio ROI, coverage = 30X, &gt; 9 exons events, 53015 reads/256400555 bases/742347 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.92	96.67	97.55	75.45
Minimap2	-ax splice	99.03	98.62	99.84	86.85
deSALT	-d 10 -x ccs	99.93	99.19	99.89	90.8
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.93	99.19	99.92	90.92
<b>Mouse PacBio ROI, coverage = 4X, &lt; 6 exons events, 34144 reads/33327445 bases/77876 exons</b>					
GMAP	-f samse --cross-species -z sense_force	97.75	95.04	97.72	91.76
Minimap2	-ax splice	97.97	93.37	95.87	90.31
deSALT	-d 10 -x ccs	97.69	94.85	97.21	91.93
deSALT + GTF	-d 10 -x ccs -G anno.gtf	97.71	94.95	97.24	92.15
<b>Mouse PacBio ROI, coverage = 4X, 6-9 exons events, 6182 reads/10959544 bases/44384 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.61	97.11	98.06	88.21
Minimap2	-ax splice	98.02	94.98	96.26	87.3
deSALT	-d 10 -x ccs	99.27	97.74	98.79	91.81
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.3	97.79	98.82	91.88
<b>Mouse PacBio ROI, coverage = 4X, &gt; 9 exons events, 6486 reads/21864873 bases/110551 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99	98.67	99.14	84.21
Minimap2	-ax splice	98.78	98	98.21	87.73
deSALT	-d 10 -x ccs	99.26	98.81	98.98	91.4

deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.26	98.8	98.97	91.44
<b>Mouse PacBio ROI, coverage = 10X, &lt; 6 exons events, 71939 reads/81401924 bases/176379 exons</b>					
GMAP	-f samse --cross-species -z sense_force	97.99	93.62	96.03	89.12
Minimap2	-ax splice	98.36	94.94	96.89	92.32
deSALT	-d 10 -x ccs	98.05	95.91	97.45	93.25
deSALT + GTF	-d 10 -x ccs -G anno.gtf	98.08	96.02	97.45	93.48
<b>Mouse PacBio ROI, coverage = 10X, 6-9 exons events, 15326 reads/27583428 bases/110033 exons</b>					
GMAP	-f samse --cross-species -z sense_force	97.99	94.11	95.45	78.21
Minimap2	-ax splice	97.51	94.97	96.22	87.25
deSALT	-d 10 -x ccs	98.7	97.72	98.62	91.8
deSALT + GTF	-d 10 -x ccs -G anno.gtf	98.72	97.72	98.62	91.82
<b>Mouse PacBio ROI, coverage = 10X, &gt; 9 exons events, 16233 reads/55208415 bases/277753 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.91	95.57	96.43	68.52
Minimap2	-ax splice	98.82	98.03	98.16	87.91
deSALT	-d 10 -x ccs	99.29	98.79	99.07	91.33
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.29	98.81	99.08	91.76
<b>Mouse PacBio ROI, coverage = 30X, &lt; 6 exons events, 198377 reads/242311390 bases/506426 exons</b>					
GMAP	-f samse --cross-species -z sense_force	97.74	93.77	95.83	89.41
Minimap2	-ax splice	98.14	95.5	97.3	93.34
deSALT	-d 10 -x ccs	97.76	96.17	97.45	93.87
deSALT + GTF	-d 10 -x ccs -G anno.gtf	97.78	96.3	97.47	94.18
<b>Mouse PacBio ROI, coverage = 30X, 6-9 exons events, 45827 reads/82794202 bases/328948 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.61	94.35	95.85	78.75
Minimap2	-ax splice	98.03	95.12	96.3	87.48
deSALT	-d 10 -x ccs	99.22	97.87	98.76	91.97
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.23	97.87	98.73	92.02
<b>Mouse PacBio ROI, coverage = 30X, &gt; 9 exons events, 48580 reads/166313402 bases/834224 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.01	95.6	96.4	68.59
Minimap2	-ax splice	98.9	98.09	98.19	87.93
deSALT	-d 10 -x ccs	99.4	98.91	99.22	91.73
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.4	98.92	99.22	91.93
<b>Human PacBio ROI, coverage = 4X, &lt; 6 exons events, 34881 reads/28674816 bases/83324 exons</b>					

GMAP	-f samse --cross-species -z sense_force	98.83	94.54	96.88	90.11
Minimap2	-ax splice	98.54	94.65	96.27	91.29
deSALT	-d 10 -x ccs	98.9	96.43	98.09	93.74
deSALT + GTF	-d 10 -x ccs -G anno.gtf	98.93	96.6	98.16	94.05
<b>Human PacBio ROI, coverage = 4X, 6-9 exons events, 6016 reads/9871523 bases/42525 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.99	95.76	97.36	82.35
Minimap2	-ax splice	98.69	97.01	97.94	90.38
deSALT	-d 10 -x ccs	99.11	98.31	98.94	93.67
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.21	98.41	99.02	93.83
<b>Human PacBio ROI, coverage = 4X, &gt; 9 exons events, 5886 reads/18656908 bases/96807 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.95	95.38	97.11	69.42
Minimap2	-ax splice	98.8	97.82	98.76	88.38
deSALT	-d 10 -x ccs	99.69	98.78	99.56	91.44
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.7	98.8	99.56	91.85
<b>Human PacBio ROI, coverage = 10X, &lt; 6 exons events, 73602 reads/69736938 bases/189283 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99	95.03	96.75	90.49
Minimap2	-ax splice	98.88	96.07	97.3	93.25
deSALT	-d 10 -x ccs	99.08	97.18	98.31	94.92
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.09	97.32	98.31	95.26
<b>Human PacBio ROI, coverage = 10X, 6-9 exons events, 15055 reads/25082950 bases/106471 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.84	95.7	97.27	82.42
Minimap2	-ax splice	98.47	96.9	97.72	90.63
deSALT	-d 10 -x ccs	98.95	98.27	98.93	93.82
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.06	98.34	99.02	93.87
<b>Human PacBio ROI, coverage = 10X, &gt; 9 exons events, 14682 reads/46949479 bases/242162 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.95	95.53	96.68	68.73
Minimap2	-ax splice	98.76	98.14	98.89	88.55
deSALT	-d 10 -x ccs	99.58	99.03	99.55	91.94
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.61	99.09	99.59	92.25
<b>Human PacBio ROI, coverage = 30X, &lt; 6 exons events, 203063 reads/207294297 bases/543941 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.99	95.37	96.74	91.01
Minimap2	-ax splice	98.93	96.78	97.92	94.45

deSALT	-d 10 -x ccs	99.09	97.62	98.47	95.75
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.11	97.78	98.47	96.14
<b>Human PacBio ROI, coverage = 30X, 6-9 exons events, 44964 reads/74851113 bases/317959 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.02	95.78	97.26	82.58
Minimap2	-ax splice	98.62	96.84	97.67	90.44
deSALT	-d 10 -x ccs	99.15	98.36	99.01	93.96
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.2	98.39	98.98	94.14
<b>Human PacBio ROI, coverage = 30X, &gt; 9 exons events, 44090 reads/141865473 bases/729517 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.88	95.38	96.56	68.85
Minimap2	-ax splice	98.69	98.01	98.71	88.6
deSALT	-d 10 -x ccs	99.54	98.97	99.48	91.99
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.58	99.03	99.49	92.55
<b>Simulated ONT 2D (1D<sup>2</sup>) datasets</b>					
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 4X, &lt; 6 exons events, 29228 reads/39799847 bases/70477 exons</b>					
GMAP	-f samse --cross-species -z sense_force	93.9	72.86	81.28	55.17
Minimap2	-ax splice	94.19	85.52	87.93	72.23
deSALT	-d 10 -x ont2d -s 2	95.23	90.05	94.99	79.8
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.31	92.06	96.23	83.5
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 4X, 6-9 exons events, 8325 reads/26798031 bases/60669 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.91	75.98	67.1	29.45
Minimap2	-ax splice	96.01	95.39	98.28	76.88
deSALT	-d 10 -x ont2d -s 2	96.18	97.3	99.33	84.11
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	96.22	97.63	99.52	85.85
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 4X, &gt; 9 exons events, 6892 reads/35269807 bases/97961 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.16	73.98	51.94	12.77
Minimap2	-ax splice	95.54	96.06	98.24	67.9
deSALT	-d 10 -x ont2d -s 2	95.66	97.79	99.19	78.63
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.68	98.05	99.22	80.98
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 10X, &lt; 6 exons events, 60612 reads/96611912 bases/158794 exons</b>					
GMAP	-f samse --cross-species -z sense_force	93.65	74.02	81.22	54.95

Minimap2	-ax splice	94.45	89.27	94.17	77.36
deSALT	-d 10 -x ont2d -s 2	94.94	93.05	97.09	84.4
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.99	94.27	97.71	87
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 10X, 6-9 exons events, 20799 reads/67295749 bases/151677 exons</b>					
GMAP	-f samse --cross-species -z sense_force	95.1	75.84	66.7	29.9
Minimap2	-ax splice	96.05	95.33	98.34	76.65
deSALT	-d 10 -x ont2d -s 2	96.24	97.95	99.76	87.16
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	96.27	98.11	99.77	88.03
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 10X, &gt; 9 exons events, 17569 reads/90693397 bases/250238 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.87	74.79	53.67	13.32
Minimap2	-ax splice	96.08	96.36	98.83	68.21
deSALT	-d 10 -x ont2d -s 2	96.26	98.46	99.63	82.36
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	96.28	98.65	99.72	84.13
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 30X, &lt; 6 exons events, 172988 reads/284891590 bases/457490 exons</b>					
GMAP	-f samse --cross-species -z sense_force	93.89	74.26	80.99	54.83
Minimap2	-ax splice	94.67	89.78	94.19	77.88
deSALT	-d 10 -x ont2d -s 2	95.21	94.45	97.68	87
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.24	95.09	97.89	88.35
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 30X, 6-9 exons events, 62337 reads/203862649 bases/455175 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.66	75.8	66.71	30
Minimap2	-ax splice	95.69	95.27	98.17	77.32
deSALT	-d 10 -x ont2d -s 2	95.89	97.93	99.44	88.68
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.9	98.08	99.54	89.09
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 30X, &gt; 9 exons events, 53012 reads/275060750 bases/758300 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.87	74.5	53.02	13.36
Minimap2	-ax splice	96.17	96.43	98.82	68.84
deSALT	-d 10 -x ont2d -s 2	96.33	98.6	99.89	83.42
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	96.34	98.76	99.92	84.78
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 4X, &lt; 6 exons events, 34108 reads/33184744 bases/79605 exons</b>					
GMAP	-f samse --cross-species -z sense_force	89.85	64.92	74.43	48.65
Minimap2	-ax splice	90.58	73.75	77.37	59.24
deSALT	-d 10 -x ont2d -s 2	92.59	83.27	89.16	71.22

deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	92.75	85.75	90.64	75.33
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 4X, 6-9 exons events, 6186 reads/10762016 bases/44279 exons</b>					
GMAP	-f samse --cross-species -z sense_force	89.2	67.6	56.19	23.7
Minimap2	-ax splice	91.05	85.64	87.75	58.62
deSALT	-d 10 -x ont2d -s 2	93.47	92.37	94.86	73.86
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	93.56	92.98	95.28	76.27
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 4X, &gt; 9 exons events, 6297 reads/21637571 bases/108257 exons</b>					
GMAP	-f samse --cross-species -z sense_force	92.19	71.55	44.66	10.1
Minimap2	-ax splice	94.42	94.89	95.74	60.79
deSALT	-d 10 -x ont2d -s 2	95.34	97.18	97.92	74.48
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.41	97.6	98.35	77.39
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 10X, &lt; 6 exons events, 71315 reads/81729449 bases/179515 exons</b>					
GMAP	-f samse --cross-species -z sense_force	89.92	66.47	75.33	49.35
Minimap2	-ax splice	91.33	79.22	86.43	66.36
deSALT	-d 10 -x ont2d -s 2	92.82	87.65	93.06	77.56
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	92.9	89.12	93.77	80.39
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 10X, 6-9 exons events, 15705 reads/27421329 bases/112575 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.17	68.71	57.19	24.43
Minimap2	-ax splice	92.08	86.5	88.67	59.91
deSALT	-d 10 -x ont2d -s 2	94.44	93.56	95.46	77.48
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.49	93.84	95.7	78.39
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 10X, &gt; 9 exons events, 15859 reads/54611806 bases/273794 exons</b>					
GMAP	-f samse --cross-species -z sense_force	92.56	71.78	45.03	10.71
Minimap2	-ax splice	94.73	95.01	95.94	61.59
deSALT	-d 10 -x ont2d -s 2	95.83	97.86	98.47	78.75
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.85	98.03	98.58	80.4
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 30X, &lt; 6 exons events, 204708 reads/242324237 bases/521418 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.12	66.84	74.97	49.55
Minimap2	-ax splice	91.56	80	86.86	66.91
deSALT	-d 10 -x ont2d -s 2	92.85	88.66	93.38	79.34
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	92.91	89.62	93.8	81.15
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 30X, 6-9 exons events, 46551 reads/82067115 bases/333544 exons</b>					



GMAP	-f samse --cross-species -z sense_force	90.85	68.82	56.81	24.85
Minimap2	-ax splice	92.54	86.55	88.81	60.26
deSALT	-d 10 -x ont2d -s 2	95.05	93.93	95.9	78.17
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.08	94.14	96.04	79.09
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 30X, &gt; 9 exons events, 48154 reads/166937719 bases/833114 exons</b>					
GMAP	-f samse --cross-species -z sense_force	92.12	71.35	44.75	10.42
Minimap2	-ax splice	94.38	94.75	95.61	61.45
deSALT	-d 10 -x ont2d -s 2	95.39	97.62	98.4	78.73
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.4	97.77	98.5	80.3
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 4X, &lt; 6 exons events, 34832 reads/28480470 bases/85272 exons</b>					
GMAP	-f samse --cross-species -z sense_force	89.73	65.78	73.14	48.58
Minimap2	-ax splice	89.62	73.95	75.41	57.82
deSALT	-d 10 -x ont2d -s 2	93.08	84.52	88.9	72.05
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	93.26	86.98	90.64	76.07
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 4X, 6-9 exons events, 6042 reads/9765619 bases/42767 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.44	69.62	57.12	24.96
Minimap2	-ax splice	93.35	88.32	90.75	59.98
deSALT	-d 10 -x ont2d -s 2	95.64	94.29	96.69	74.5
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.83	95.14	97.67	77.31
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 4X, &gt; 9 exons events, 5720 reads/18358704 bases/94986 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.81	71.59	45.1	9.48
Minimap2	-ax splice	94.26	94.75	96.24	63.18
deSALT	-d 10 -x ont2d -s 2	95.67	97.24	98.64	75.84
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.65	97.48	98.74	78.37
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 10X, &lt; 6 exons events, 73054 reads/69878562 bases/193650 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.78	67.28	73.9	49.47
Minimap2	-ax splice	91.69	79.78	85.76	65.67
deSALT	-d 10 -x ont2d -s 2	94.32	89.3	93.71	78.92
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.42	90.93	94.67	82.07
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 10X, 6-9 exons events, 15404 reads/24970788 bases/109140 exons</b>					
GMAP	-f samse --cross-species -z sense_force	89.96	68.63	56.58	24.38
Minimap2	-ax splice	92.03	87.32	89.94	59.13

deSALT	-d 10 -x ont2d -s 2	94.44	94.04	96.41	76.69
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.53	94.5	96.75	78.26
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 10X, &gt; 9 exons events, 14462 reads/46434095 bases/240227 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.51	71.05	44.27	10.28
Minimap2	-ax splice	94.19	94.56	96.18	61.19
deSALT	-d 10 -x ont2d -s 2	95.67	97.45	98.88	76.75
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.7	97.7	99	78.88
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 30X, &lt; 6 exons events, 209725 reads/206910570 bases/562271 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.72	76.31	80.9	60.34
Minimap2	-ax splice	91.73	80.34	85.8	66.08
deSALT	-d 10 -x ont2d -s 2	94.29	90.16	93.91	80.6
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.38	91.27	94.59	82.66
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 30X, 6-9 exons events, 45842 reads/75221616 bases/324667 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.55	77.96	70.31	35.68
Minimap2	-ax splice	92.58	87.8	90.44	59.93
deSALT	-d 10 -x ont2d -s 2	95.01	94.68	96.92	77.87
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.07	95.05	97.31	79.07
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 30X, &gt; 9 exons events, 43830 reads/141765970 bases/728962 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.53	81.64	65.6	18.49
Minimap2	-ax splice	94.15	94.64	96.2	61.94
deSALT	-d 10 -x ont2d -s 2	95.57	97.4	98.63	76.84
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.6	97.68	98.81	79.21
<b>Simulated PacBio subread datasets</b>					
<b>Fruit fly PacBio subread, coverage = 4X, &lt; 6 exons events, 30278 reads/42410655 bases/69852 exons</b>					
GMAP	-f samse --cross-species -z sense_force	77.3	54.04	56.02	36.73
Minimap2	-ax splice	85.67	80.26	78.86	63.38
deSALT	-d 10 -x clr -s 2	87.06	88.58	93.84	77.42
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.12	89.82	94.61	79.65
<b>Fruit fly PacBio subread, coverage = 4X, 6-9 exons events, 7707 reads/26875546 bases/56191 exons</b>					
GMAP	-f samse --cross-species -z sense_force	85.99	66.81	51.81	15.67

Minimap2	-ax splice	89.52	95.21	98.31	74.35
deSALT	-d 10 -x clr -s 2	89.72	97.55	99.68	83.95
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	89.74	97.89	99.92	85.62
<b>Fruit fly PacBio subread, coverage = 4X, &gt; 9 exons events, 6184 reads/32658929 bases/86374 exons</b>					
GMAP	-f samse --cross-species -z sense_force	84.61	66.39	34.22	4.03
Minimap2	-ax splice	88.83	96.34	98.93	66.51
deSALT	-d 10 -x clr -s 2	89.02	98.57	99.92	80.4
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	89.03	98.84	100	82.7
<b>Fruit fly PacBio subread, coverage = 10X, &lt; 6 exons events, 63048 reads/104050616 bases/159211 exons</b>					
GMAP	-f samse --cross-species -z sense_force	78.16	56.02	56.77	36.86
Minimap2	-ax splice	87.3	87.06	91.54	73.09
deSALT	-d 10 -x clr -s 2	87.98	92.35	96.82	82.58
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88	92.99	97.2	83.85
<b>Fruit fly PacBio subread, coverage = 10X, 6-9 exons events, 19163 reads/66676969 bases/139521 exons</b>					
GMAP	-f samse --cross-species -z sense_force	85.12	66.6	51.14	15.66
Minimap2	-ax splice	88.82	95.05	98.19	74.38
deSALT	-d 10 -x clr -s 2	89.01	98	99.62	86.73
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	89.02	98.18	99.73	87.58
<b>Fruit fly PacBio subread, coverage = 10X, &gt; 9 exons events, 15699 reads/83886149 bases/220066 exons</b>					
GMAP	-f samse --cross-species -z sense_force	84.67	65.63	33.26	4.05
Minimap2	-ax splice	88.68	95.89	98.6	66.44
deSALT	-d 10 -x clr -s 2	88.88	98.31	99.68	82.14
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.88	98.45	99.67	83.69
<b>Fruit fly PacBio subread, coverage = 30X, &lt; 6 exons events, 172125 reads/306614209 bases/452111 exons</b>					
GMAP	-f samse --cross-species -z sense_force	79.26	58.07	58.55	37.94
Minimap2	-ax splice	87.76	89.15	94.56	76.3
deSALT	-d 10 -x clr -s 2	88.18	93.68	97.53	84.8
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.18	94.08	97.66	85.57
<b>Fruit fly PacBio subread, coverage = 30X, 6-9 exons events, 57839 reads/201638578 bases/421444 exons</b>					
GMAP	-f samse --cross-species -z sense_force	85.16	66.46	51.19	15.43
Minimap2	-ax splice	88.83	94.95	98.15	74.32
deSALT	-d 10 -x clr -s 2	89.03	98.05	99.66	86.92

deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	89.03	98.22	99.72	87.77
<b>Fruit fly PacBio subread, coverage = 30X, &gt; 9 exons events, 47577 reads/255545996 bases/667078 exons</b>					
GMAP	-f samse --cross-species -z sense_force	84.52	65.91	33.88	3.87
Minimap2	-ax splice	88.52	95.95	98.59	66.85
deSALT	-d 10 -x clr -s 2	88.72	98.39	99.69	82.09
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.72	98.58	99.68	83.98
<b>Mouse PacBio subread, coverage = 4X, &lt; 6 exons events, 35080 reads/34447409 bases/78452 exons</b>					
GMAP	-f samse --cross-species -z sense_force	68.44	37.59	46.23	29.75
Minimap2	-ax splice	82.79	68.74	69.99	51.76
deSALT	-d 10 -x clr -s 2	85.21	80.67	86.16	66.92
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	85.29	82.76	87.69	70.13
<b>Mouse PacBio subread, coverage = 4X, 6-9 exons events, 5724 reads/10741011 bases/41117 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.18	50.73	36.25	11.2
Minimap2	-ax splice	84.52	84.86	87.46	54
deSALT	-d 10 -x clr -s 2	86.97	91.68	94.13	70.18
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.08	92.61	94.95	73.55
<b>Mouse PacBio subread, coverage = 4X, &gt; 9 exons events, 5677 reads/20567829 bases/97011 exons</b>					
GMAP	-f samse --cross-species -z sense_force	83.33	65.01	27.85	3.61
Minimap2	-ax splice	87.34	94.4	95.61	55.15
deSALT	-d 10 -x clr -s 2	88.26	96.96	97.89	70.74
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.31	97.42	98.15	75.15
<b>Mouse PacBio subread, coverage = 10X, &lt; 6 exons events, 73842 reads/85200651 bases/179030 exons</b>					
GMAP	-f samse --cross-species -z sense_force	68.37	37.28	44.81	28.69
Minimap2	-ax splice	83.92	74.98	81.56	59.95
deSALT	-d 10 -x clr -s 2	85.78	85.59	91.32	73.65
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	85.83	86.86	92.17	75.89
<b>Mouse PacBio subread, coverage = 10X, 6-9 exons events, 14180 reads/51689732 bases/243516 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.24	50.13	35.25	11.48
Minimap2	-ax splice	85.14	85.32	87.88	54.86
deSALT	-d 10 -x clr -s 2	87.68	93.1	95.41	74.48
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.73	93.47	95.75	76.02
<b>Mouse PacBio subread, coverage = 10X, &gt; 9 exons events, 17084 reads/56231420 bases/296347 exons</b>					

GMAP	-f samse --cross-species -z sense_force	83.26	64.81	27.14	2.74
Minimap2	-ax splice	87.27	94.34	95.83	55.73
deSALT	-d 10 -x clr -s 2	88.31	97.45	98.52	74.97
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.32	97.66	98.62	77.15
<b>Mouse PacBio subread, coverage = 30X, &lt; 6 exons events, 202805 reads/253239643 bases/510067 exons</b>					
GMAP	-f samse --cross-species -z sense_force	69.05	38.03	44.8	28.92
Minimap2	-ax splice	84.43	77.88	86.28	63.48
deSALT	-d 10 -x clr -s 2	85.9	87.74	93.4	76.81
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	85.93	88.54	93.79	78.16
<b>Mouse PacBio subread, coverage = 30X, 6-9 exons events, 42947 reads/156407534 bases/737207 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.42	50.47	35.51	11.03
Minimap2	-ax splice	85.04	85.32	88.07	54.52
deSALT	-d 10 -x clr -s 2	87.55	93.31	95.58	74.96
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.57	93.55	95.67	76.06
<b>Mouse PacBio subread, coverage = 30X, &gt; 9 exons events, 51149 reads/169009353 bases/889311 exons</b>					
GMAP	-f samse --cross-species -z sense_force	83.48	65.11	27.6	3.09
Minimap2	-ax splice	87.44	94.49	95.94	55.62
deSALT	-d 10 -x clr -s 2	88.43	97.58	98.53	74.89
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.44	97.77	98.62	77.06
<b>Human PacBio subread, coverage = 4X, &lt; 6 exons events, 35848 reads/29729832 bases/84394 exons</b>					
GMAP	-f samse --cross-species -z sense_force	64.19	34.61	42.4	27.72
Minimap2	-ax splice	82.41	68.55	67.9	49.68
deSALT	-d 10 -x clr -s 2	86.22	81.56	85.51	67.07
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.41	83.93	87.4	70.55
<b>Human PacBio subread, coverage = 4X, 6-9 exons events, 5522 reads/9604060 bases/39360 exons</b>					
GMAP	-f samse --cross-species -z sense_force	73.23	49.35	34.66	10.41
Minimap2	-ax splice	84.69	87.27	90.22	55.12
deSALT	-d 10 -x clr -s 2	86.88	93.59	96.32	71.86
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.99	94.41	97.23	74.81
<b>Human PacBio subread, coverage = 4X, &gt; 9 exons events, 5167 reads/17475908 bases/84665 exons</b>					
GMAP	-f samse --cross-species -z sense_force	82.76	64.46	26.96	2.32
Minimap2	-ax splice	87.27	94.56	97.08	56.82

deSALT	-d 10 -x clr -s 2	88.34	96.87	98.55	71.51
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.41	97.28	98.86	74.72
<b>Human PacBio subread, coverage = 10X, &lt; 6 exons events, 75600 reads/73059595 bases/192890 exons</b>					
GMAP	-f samse --cross-species -z sense_force	63.84	34.39	40.36	26.37
Minimap2	-ax splice	83.69	75.08	80.05	57.98
deSALT	-d 10 -x clr -s 2	86.87	87.01	91.81	74.16
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.95	88.44	92.9	76.61
<b>Human PacBio subread, coverage = 10X, 6-9 exons events, 13999 reads/24325764 bases/99703 exons</b>					
GMAP	-f samse --cross-species -z sense_force	73.59	48.48	33.8	10.38
Minimap2	-ax splice	85.48	87.17	90.41	55.2
deSALT	-d 10 -x clr -s 2	87.89	94.52	97.3	75.15
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.97	95.06	97.78	77.13
<b>Human PacBio subread, coverage = 10X, &gt; 9 exons events, 12934 reads/43940986 bases/212358 exons</b>					
GMAP	-f samse --cross-species -z sense_force	82.97	64.4	26.09	2.61
Minimap2	-ax splice	87.32	94.52	96.76	56.77
deSALT	-d 10 -x clr -s 2	88.58	97.34	98.79	74.12
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.61	97.58	98.87	76.16
<b>Human PacBio subread, coverage = 30X, &lt; 6 exons events, 208020 reads/217338313 bases/552141 exons</b>					
GMAP	-f samse --cross-species -z sense_force	64.47	34.56	39.63	26.19
Minimap2	-ax splice	84.51	78.07	85.3	61.76
deSALT	-d 10 -x clr -s 2	87.34	89.44	94.37	77.77
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.39	90.45	95.06	79.57
<b>Human PacBio subread, coverage = 30X, 6-9 exons events, 42098 reads/73146856 bases/300035 exons</b>					
GMAP	-f samse --cross-species -z sense_force	73.9	49.04	33.94	10.93
Minimap2	-ax splice	85.59	87.31	90.4	55.44
deSALT	-d 10 -x clr -s 2	88.06	94.75	97.38	75.64
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.1	95.16	97.72	77.26
<b>Human PacBio subread, coverage = 30X, &gt; 9 exons events, 39257 reads/133356721 bases/644800 exons</b>					
GMAP	-f samse --cross-species -z sense_force	82.45	64.31	26.67	2.77
Minimap2	-ax splice	87	94.31	96.63	56.49
deSALT	-d 10 -x clr -s 2	88.27	97.2	98.67	73.42
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.28	97.43	98.71	75.98

**Simulated ONT 1D datasets**

**Fruit fly ONT 1D, coverage = 4X, < 6 exons events, 28467 reads/39435293 bases/72362 exons**

GMAP	-f samse --cross-species -z sense_force	61.47	17.82	19.83	5.22
Minimap2	-ax splice	82.28	53.68	54.23	31.67
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	93.55	69.03	76.24	44.05
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.16	77.71	84.96	56.03

**Fruit fly ONT 1D, coverage = 4X, 6-9 exons events, 8741 reads/26204640 bases/63690 exons**

GMAP	-f samse --cross-species -z sense_force	70.89	23.48	4.4	0.31
Minimap2	-ax splice	93.02	79.65	79.76	41.47
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.63	89.4	91.99	55.15
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.97	92.84	96.68	65.08

**Fruit fly ONT 1D, coverage = 4X, > 9 exons events, 7338 reads/36202359 bases/104853 exons**

GMAP	-f samse --cross-species -z sense_force	63.83	20.25	0.26	0
Minimap2	-ax splice	95.02	86.99	84.4	32.41
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.94	93.1	95.2	49.33
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	97.14	95.2	97.87	58

**Fruit fly ONT 1D, coverage = 10X, < 6 exons events, 59321 reads/93320200 bases/158771 exons**

GMAP	-f samse --cross-species -z sense_force	64.41	19.4	20.71	5.38
Minimap2	-ax splice	85	59.02	60.68	36.11
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	94.62	75.88	82.23	51.75
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.98	83.74	89.86	65.15

**Fruit fly ONT 1D, coverage = 10X, 6-9 exons events, 21936 reads/67655362 bases/160155 exons**

GMAP	-f samse --cross-species -z sense_force	71.28	23.68	4.22	0.27
Minimap2	-ax splice	93.8	81.19	81.54	42.85
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	97.48	92.91	95.85	65.21
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	97.65	94.9	98.04	72.22

**Fruit fly ONT 1D, coverage = 10X, > 9 exons events, 18870 reads/93633101 bases/270690 exons**

GMAP	-f samse --cross-species -z sense_force	63.65	20.04	0.3	0.01
Minimap2	-ax splice	94.9	87.18	84.89	32.46
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.96	94.65	96.99	56.02

deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	97.11	96.24	98.55	64.49
<b>Fruit fly ONT 1D, coverage = 30X, &lt; 6 exons events, 175750 reads/278301970 bases/471337 exons</b>					
GMAP	-f samse --cross-species -z sense_force	64.44	19.42	20.76	5.3
Minimap2	-ax splice	85.31	59.25	60.82	36.3
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	95.24	80.73	87.29	59.52
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	95.41	85.68	91.19	68.67
<b>Fruit fly ONT 1D, coverage = 30X, 6-9 exons events, 65335 reads/202340209 bases/477630 exons</b>					
GMAP	-f samse --cross-species -z sense_force	70.93	23.65	4.21	0.32
Minimap2	-ax splice	93.5	81.02	81.35	43.2
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	97.21	94.11	96.87	70.58
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	97.31	95.25	98.04	74.71
<b>Fruit fly ONT 1D, coverage = 30X, &gt; 9 exons events, 56736 reads/283160402 bases/815225 exons</b>					
GMAP	-f samse --cross-species -z sense_force	64.05	20.17	0.33	0.01
Minimap2	-ax splice	94.95	87.13	85.18	32.59
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	97.03	94.98	97.17	58.9
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	97.15	96.4	98.56	65.83
<b>Mouse ONT 1D, coverage = 4X, &lt; 6 exons events, 33267 reads/33137109 bases/81062 exons</b>					
GMAP	-f samse --cross-species -z sense_force	52.82	8.51	18.01	3.96
Minimap2	-ax splice	69.84	26.5	34.42	16.32
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	83.21	43.42	53.48	25.35
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	84.46	54.55	63.54	36.66
<b>Mouse ONT 1D, coverage = 4X, 6-9 exons events, 6683 reads/10709673 bases/47920 exons</b>					
GMAP	-f samse --cross-species -z sense_force	51.12	11.36	1.42	0.19
Minimap2	-ax splice	72.41	45.86	40.55	18.49
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	86.7	65.57	60.27	27.97
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.18	74.28	73.87	39.29
<b>Mouse ONT 1D, coverage = 4X, &gt; 9 exons events, 6650 reads/21662084 bases/114907 exons</b>					
GMAP	-f samse --cross-species -z sense_force	56.21	16.48	0.17	0
Minimap2	-ax splice	85.14	73.72	61.41	17.25
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	91.98	81.82	74.42	23.28
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	93.46	88.35	86.87	34.65
<b>Mouse ONT 1D, coverage = 10X, &lt; 6 exons events, 70045 reads/79985115 bases/180952 exons</b>					



GMAP	-f samse --cross-species -z sense_force	54.83	9.25	17.82	3.9
Minimap2	-ax splice	72.29	29.92	38.64	18.88
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	86.21	54.05	62.89	33.5
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	86.84	63.1	71.61	45.91
<b>Mouse ONT 1D, coverage = 10X, 6-9 exons events, 16518 reads/27559134 bases/118437 exons</b>					
GMAP	-f samse --cross-species -z sense_force	52.25	11.96	1.7	0.19
Minimap2	-ax splice	73.04	46.47	41.17	18.67
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.78	73.11	72.29	37.87
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.49	77.96	79.19	47.14
<b>Mouse ONT 1D, coverage = 10X, &gt; 9 exons events, 17084 reads/56231420 bases/296347 exons</b>					
GMAP	-f samse --cross-species -z sense_force	56.24	16.74	0.18	0
Minimap2	-ax splice	85.74	74.27	61.59	17.75
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	93.72	86.96	83.28	33.79
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.68	91.34	91.24	43.81
<b>Mouse ONT 1D, coverage = 30X, &lt; 6 exons events, 208841 reads/239854618 bases/540647 exons</b>					
GMAP	-f samse --cross-species -z sense_force	39.27	6.32	14.06	2.97
Minimap2	-ax splice	72.13	29.99	38.68	19.04
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	86.71	59.48	68.31	40.35
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	86.99	65.24	73.12	49.5
<b>Mouse ONT 1D, coverage = 30X, 6-9 exons events, 49286 reads/82418147 bases/353423 exons</b>					
GMAP	-f samse --cross-species -z sense_force	31.22	7.16	1.07	0.09
Minimap2	-ax splice	73.56	46.85	41.52	19.05
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	89.71	76.74	77.04	45.45
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.96	78.92	79.9	50.42
<b>Mouse ONT 1D, coverage = 30X, &gt; 9 exons events, 51149 reads/169009353 bases/889311 exons</b>					
GMAP	-f samse --cross-species -z sense_force	32.11	9.64	0.1	0
Minimap2	-ax splice	85.96	74.51	61.87	18.03
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	94.08	88.75	86.14	39.28
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.77	91.62	91.19	46.02
<b>Human ONT 1D, coverage = 4X, &lt; 6 exons events, 34028 reads/28318016 bases/87076 exons</b>					
GMAP	-f samse --cross-species -z sense_force	46.46	7.29	14.23	3.39
Minimap2	-ax splice	63	23.28	28.98	13.89

deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	80.4	43.58	50.62	24.75
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	81.84	53.06	59.82	34.9
<b>Human ONT 1D, coverage = 4X, 6-9 exons events, 6578 reads/9907975 bases/46628 exons</b>					
GMAP	-f samse --cross-species -z sense_force	47.59	10.02	1.32	0.21
Minimap2	-ax splice	69.67	42.53	36.8	16.3
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	86.66	66.16	61.25	28.05
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.12	74.01	73.72	37.46
<b>Human ONT 1D, coverage = 4X, &gt; 9 exons events, 6038 reads/18291208 bases/99917 exons</b>					
GMAP	-f samse --cross-species -z sense_force	53.87	15.9	0.17	0
Minimap2	-ax splice	83.11	70.64	58	15.98
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	91.71	81.58	74.81	24.93
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	93.26	87.76	85.76	34.96
<b>Human ONT 1D, coverage = 10X, &lt; 6 exons events, 71640 reads/68392014 bases/194695 exons</b>					
GMAP	-f samse --cross-species -z sense_force	47.8	7.95	14.32	3.41
Minimap2	-ax splice	65.07	25.9	32.42	16.1
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	83.7	53.08	59.85	32.28
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	84.49	61.44	68.11	44.11
<b>Human ONT 1D, coverage = 10X, 6-9 exons events, 16291 reads/25117618 bases/115369 exons</b>					
GMAP	-f samse --cross-species -z sense_force	48.9	10.7	1.5	0.12
Minimap2	-ax splice	70.85	43.98	39.06	16.89
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.52	72.73	71.59	36.63
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.13	76.61	77.32	43.42
<b>Human ONT 1D, coverage = 10X, &gt; 9 exons events, 15601 reads/47786596 bases/259924 exons</b>					
GMAP	-f samse --cross-species -z sense_force	53.49	15.89	0.19	0
Minimap2	-ax splice	83.86	71.47	58.92	16.7
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	93.08	85.61	81.92	32.45
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.17	89.97	89.49	40.32
<b>Human ONT 1D, coverage = 30X, &lt; 6 exons events, 213756 reads/204186862 bases/582406 exons</b>					
GMAP	-f samse --cross-species -z sense_force	48.17	7.84	14.37	3.38
Minimap2	-ax splice	65.45	25.95	32.57	16.12
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	84.79	57.84	64.66	38.58
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	85.13	63.28	69.44	47.31

<b>Human ONT 1D, coverage = 30X, 6-9 exons events, 48744 reads/75971120 bases/345208 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.09	10.68	1.51	0.15
Minimap2	-ax splice	70.63	43.96	38.9	17.23
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.73	75.23	75.15	41.17
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	88.97	77.33	78.07	45.83
<b>Human ONT 1D, coverage = 30X, &gt; 9 exons events, 46635 reads/143685596 bases/778725 exons</b>					
GMAP	-f samse --cross-species -z sense_force	53.43	15.81	0.14	0
Minimap2	-ax splice	83.81	71.7	58.94	16.79
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	93.24	86.68	82.79	36.09
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.16	90.1	88.73	41.73

- a) This table depicts the results of the aligners on the reads from the transcripts with various numbers of exons. For each of the simulated datasets, the reads from the transcripts with < 6 exons, 6–9 exons, and >9 exons are respectively divided into three groups, and the benchmark is implemented for each of the groups.
- b) Refer to the Supplementary Notes for the command lines of the benchmarked aligners.
- c) Base%: the proportion of bases being correctly aligned to their ground truth positions, i.e., the mapped positions of the bases are within 5 bp of their ground truth positions.
- d) Exon%: the proportion of the exons being correctly mapped. An exon in a certain read is considered to be correctly mapped only if its two boundaries are mapped within 5 bp of their ground truth positions in the reference genome.
- e) Read80%: the proportion of Read80% reads. A read is considered to be a Read80% read only if it meets two conditions:  $N_T/N_G > 80\%$  and  $N_T/N_P > 80\%$ , where  $N_G$  is the number of ground truth exons within the read,  $N_P$  is the number of exons predicted by the alignment, and  $N_T$  is the number of true positive exons. Herein, a predicted exon is considered to be a true positive exon only if there is a ground truth exon in the read and the corresponding boundaries of the predicted exon and the ground truth exon are within 5 bp.
- f) Read100%: the proportion of Read100% reads. A read is considered to be a Read100% read only if it meets two conditions:  $N_T/N_G = 100\%$  and  $N_T/N_P = 100\%$ . It is worth noting that a Read100% read indicates that the read has a highly correct full-length alignment.

**Supplementary Table 5. Benchmark results on the simulated reads from the genes with one/multiple isoforms <sup>a</sup>**

Aligner	parameters <sup>b</sup>	Base% <sup>c</sup>	Exons% <sup>d</sup>	Read80% <sup>e</sup>	Read100% <sup>f</sup>
<b>Simulated PacBio ROI datasets</b>					
<b>Fruit fly PacBio ROI, coverage = 4X, reads from the genes with single isoform, 14301 reads/15998364 bases/33963 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.9	95.98	98.6	92.51
Minimap2	-ax splice	98.88	97.48	98.61	95.12
deSALT	-d 10 -x ccs	99	98.13	99.33	96.36
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99	98.13	99.29	96.43
<b>Fruit fly PacBio ROI, coverage = 4X, reads from the genes with multiple isoforms, 30995 reads/86312808 bases/196424 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.76	96.56	98.18	87.14
Minimap2	-ax splice	99.67	98.1	98.81	91.7
deSALT	-d 10 -x ccs	99.78	98.76	99.65	94.3
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.78	98.78	99.67	94.38
<b>Fruit fly PacBio ROI, coverage = 10X, reads from the genes with single isoform, 31515 reads/39591445 bases/80112 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.9	96.21	98.43	92.84
Minimap2	-ax splice	98.89	97.91	99.04	95.89
deSALT	-d 10 -x ccs	99.04	98.41	99.37	96.94
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.04	97.28/99.52	99.31	96.98
<b>Fruit fly PacBio ROI, coverage = 10X, reads from the genes with multiple isoforms, 69775 reads/215258083 bases/481361 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.77	96.69	97.97	86.54
Minimap2	-ax splice	99.7	98.4	99.2	92.16
deSALT	-d 10 -x ccs	99.79	98.99	99.69	94.76
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.79	99	99.7	94.82
<b>Fruit fly PacBio ROI, coverage = 30X, reads from the genes with single isoform, 88943 reads/118375102 bases/234312 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.95	96.28	98.32	92.75
Minimap2	-ax splice	98.95	98.24	99.4	96.41
deSALT	-d 10 -x ccs	99.05	98.53	99.32	97.13
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.04	98.49	99.25	97.19
<b>Fruit fly PacBio ROI, coverage = 30X, reads from the genes with multiple isoforms, 199412 reads/645473928 bases/1432478 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.77	96.68	97.87	86.02

Minimap2	-ax splice	99.74	98.47	99.4	92.19
deSALT	-d 10 -x ccs	99.81	99.1	99.74	95
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.81	99.1	99.75	95.03
<b>Mouse PacBio ROI, coverage = 4X, reads from the genes with single isoform, 12435 reads/12982861 bases/27141 exons</b>					
GMAP	-f samse --cross-species -z sense_force	96.28	93.78	97.36	91.71
Minimap2	-ax splice	97.4	93.63	96.24	91.5
deSALT	-d 10 -x ccs	96.2	93.2	96.41	91.57
deSALT + GTF	-d 10 -x ccs -G anno.gtf	96.25	93.5	96.54	92.01
<b>Mouse PacBio ROI, coverage = 4X, reads from the genes with multiple isoforms, 34377 reads/53169001 bases/205670 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.8	97.61	98.18	89.72
Minimap2	-ax splice	98.46	96.17	96.25	88.85
deSALT	-d 10 -x ccs	99.03	97.82	98.12	91.94
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.03	97.82	98.11	92.02
<b>Mouse PacBio ROI, coverage = 10X, reads from the genes with single isoform, 27414 reads/32112338 bases/63714 exons</b>					
GMAP	-f samse --cross-species -z sense_force	96.28	91.85	96.4	89.13
Minimap2	-ax splice	97.66	94.63	97.05	93.01
deSALT	-d 10 -x ccs	96.36	93.63	96.4	92.29
deSALT + GTF	-d 10 -x ccs -G anno.gtf	96.42	93.91	96.48	92.73
<b>Mouse PacBio ROI, coverage = 10X, reads from the genes with multiple isoforms, 76084 reads/132081429 bases/500451 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.79	95.03	95.86	82.52
Minimap2	-ax splice	98.55	96.7	96.97	90.11
deSALT	-d 10 -x ccs	99.11	98.2	98.4	92.9
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.12	98.21	98.38	93.05
<b>Mouse PacBio ROI, coverage = 30X, reads from the genes with single isoform, 77424 reads/95970680 bases/185808 exons</b>					
GMAP	-f samse --cross-species -z sense_force	96.32	92.12	96.43	89.29
Minimap2	-ax splice	97.65	94.87	97.26	93.59
deSALT	-d 10 -x ccs	96.34	93.87	96.46	92.63
deSALT + GTF	-d 10 -x ccs -G anno.gtf	96.39	94.19	96.56	93.22
<b>Mouse PacBio ROI, coverage = 30X, reads from the genes with multiple isoforms, 215360 reads/395448314 bases/1483790 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.8	95.14	95.75	82.49
Minimap2	-ax splice	98.56	96.95	97.3	90.78
deSALT	-d 10 -x ccs	99.1	98.38	98.49	93.43

deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.1	98.39	98.46	93.56
<b>Human PacBio ROI, coverage = 4X, reads from the genes with single isoform, 12012 reads/10139594 bases/22474 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.54	93.64	97.5	92.38
Minimap2	-ax splice	98.79	95.19	97.07	94.03
deSALT	-d 10 -x ccs	98.96	96.32	98.14	95.55
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.07	97.05	98.43	96.33
<b>Human PacBio ROI, coverage = 4X, reads from the genes with multiple isoforms, 34881 reads/28674816 bases/83324 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.97	95.31	96.78	84.48
Minimap2	-ax splice	98.62	96.62	96.7	89.69
deSALT	-d 10 -x ccs	99.24	97.98	98.47	92.71
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.26	98	98.45	92.85
<b>Human PacBio ROI, coverage = 10X, reads from the genes with single isoform, 26452 reads/25006370 bases/52122 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.59	93.84	97.46	92.32
Minimap2	-ax splice	98.93	95.69	97.61	94.8
deSALT	-d 10 -x ccs	98.97	96.65	98.29	96.05
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.1	97.48	98.63	96.93
<b>Human PacBio ROI, coverage = 10X, reads from the genes with multiple isoforms, 76887 reads/116762997 bases/485794 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.03	95.56	96.59	84.13
Minimap2	-ax splice	98.73	97.32	97.58	91.31
deSALT	-d 10 -x ccs	99.28	98.4	98.67	93.75
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.29	98.41	98.58	93.84
<b>Human PacBio ROI, coverage = 30X, reads from the genes with single isoform, 74648 reads/74627529 bases/151031 exons</b>					
GMAP	-f samse --cross-species -z sense_force	98.7	94.11	97.49	92.77
Minimap2	-ax splice	98.97	96.14	98.03	95.47
deSALT	-d 10 -x ccs	99.11	97.06	98.46	96.4
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.15	97.72	98.72	97.38
<b>Human PacBio ROI, coverage = 30X, reads from the genes with multiple isoforms, 217469 reads/349383354 bases/1440386 exons</b>					
GMAP	-f samse --cross-species -z sense_force	99.02	95.6	96.56	84.17
Minimap2	-ax splice	98.76	97.48	97.99	92.08
deSALT	-d 10 -x ccs	99.28	98.53	98.79	94.39
deSALT + GTF	-d 10 -x ccs -G anno.gtf	99.31	98.56	98.7	94.58

**Simulated ONT 2D (1D<sup>2</sup>) datasets**

<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 4X, reads from the genes with single isoform, 14295 reads/15799802 bases/34025 exons</b>					
GMAP	-f samse --cross-species -z sense_force	92.81	72.32	83.39	55.54
Minimap2	-ax splice	93.25	86.09	87.17	72.47
deSALT	-d 10 -x ont2d -s 2	94.46	89.32	94.35	78.44
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.57	92.21	95.87	83.93
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 4X, reads from the genes with multiple isoforms, 30150 reads/86067883 bases/195082 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.52	74.49	69.66	38.2
Minimap2	-ax splice	95.48	93.78	93.51	72.41
deSALT	-d 10 -x ont2d -s 2	95.84	96.32	97.45	81.37
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.88	96.77	97.99	83.36
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 10X, reads from the genes with single isoform, 31482 reads/39446943 bases/80732 exons</b>					
GMAP	-f samse --cross-species -z sense_force	93.25	73.1	84.47	55.37
Minimap2	-ax splice	94.14	89.61	94.84	78.23
deSALT	-d 10 -x ont2d -s 2	94.78	92.69	97.18	83.68
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.86	94.37	97.75	87.53
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 10X, reads from the genes with multiple isoforms, 67498 reads/215154115 bases/479977 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.69	75.15	68.05	36.2
Minimap2	-ax splice	95.69	94.83	96.35	74.35
deSALT	-d 10 -x ont2d -s 2	95.94	97.48	98.53	85.05
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.96	97.75	98.84	86.32
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 30X, reads from the genes with single isoform, 91597 reads/118351500 bases/239087 exons</b>					
GMAP	-f samse --cross-species -z sense_force	93.24	73.41	83.99	55.02
Minimap2	-ax splice	94.19	89.97	94.72	78.36
deSALT	-d 10 -x ont2d -s 2	94.88	94.51	97.78	87.25
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.91	95.17	97.8	88.84
<b>Fruit fly ONT 2D (1D<sup>2</sup>), coverage = 30X, reads from the genes with multiple isoforms, 196740 reads/645463489 bases/1431878 exons</b>					
GMAP	-f samse --cross-species -z sense_force	94.67	75.02	67.53	35.7
Minimap2	-ax splice	95.72	95.02	96.46	75.05
deSALT	-d 10 -x ont2d -s 2	95.96	97.74	98.79	86.45
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	95.98	97.97	99	87.39

<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 4X, reads from the genes with single isoform, 12412 reads/12823106 bases/27050 exons</b>					
GMAP	-f samse --cross-species -z sense_force	88.16	63.35	79.66	50.76
Minimap2	-ax splice	90.44	77.27	80.98	63.37
deSALT	-d 10 -x ont2d -s 2	90.55	80.03	88.62	69.1
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	90.85	84.87	90.85	76.38
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 4X, reads from the genes with multiple isoforms, 34179 reads/52761225 bases/205091 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.08	69.2	63.74	36.27
Minimap2	-ax splice	92.28	87.01	81.33	57.91
deSALT	-d 10 -x ont2d -s 2	94.39	93.01	92	73.07
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.47	93.68	92.83	75.5
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 10X, reads from the genes with single isoform, 27346 reads/31974136 bases/63843 exons</b>					
GMAP	-f samse --cross-species -z sense_force	88.8	64.56	81.71	51.2
Minimap2	-ax splice	91.12	80.88	88.63	69.04
deSALT	-d 10 -x ont2d -s 2	91.36	85.62	92.66	76.46
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	91.51	88.3	93.56	81.36
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 10X, reads from the genes with multiple isoforms, 75533 reads/131788448 bases/502041 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.34	70.11	62.89	35.39
Minimap2	-ax splice	92.95	89.25	88.1	63.05
deSALT	-d 10 -x ont2d -s 2	94.76	94.8	94.84	78.19
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.79	95.14	95.26	79.62
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 30X, reads from the genes with single isoform, 79605 reads/95941227 bases/189014 exons</b>					
GMAP	-f samse --cross-species -z sense_force	88.89	65.11	81.37	51.56
Minimap2	-ax splice	91.4	81.73	89.14	69.64
deSALT	-d 10 -x ont2d -s 2	91.41	87.49	93.41	79.52
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	91.53	88.92	93.73	82.2
<b>Mouse ONT 2D (1D<sup>2</sup>), coverage = 30X, reads from the genes with multiple isoforms, 219808 reads/395387844 bases/1499062 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.42	70	62.18	35.02
Minimap2	-ax splice	93	89.43	88.36	63.32
deSALT	-d 10 -x ont2d -s 2	94.73	94.96	95	78.89
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.75	95.24	95.33	80.15
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 4X, reads from the genes with single isoform, 11992 reads/9984792 bases/22360 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.59	63.81	81.8	54.48



Minimap2	-ax splice	90.88	74.24	78.68	63.31
deSALT	-d 10 -x ont2d -s 2	93.29	80.05	89.15	71.68
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	93.7	86.61	92.19	79.69
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 4X, reads from the genes with multiple isoforms, 34602 reads/46620001 bases/200665 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.72	69.57	62.7	35.95
Minimap2	-ax splice	91.96	86.83	80.4	57.18
deSALT	-d 10 -x ont2d -s 2	94.59	93.12	91.79	73.23
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.64	93.73	92.67	75.41
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 10X, reads from the genes with single isoform, 26423 reads/24855256 bases/52295 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.23	65.39	83.43	54.93
Minimap2	-ax splice	92.29	79.25	88.14	69.95
deSALT	-d 10 -x ont2d -s 2	94.1	86.32	94.27	78.79
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.35	90.59	95.69	85.04
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 10X, reads from the genes with multiple isoforms, 76497 reads/116428189 bases/490722 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.8	69.63	61.51	35.12
Minimap2	-ax splice	92.64	88.75	87.76	62.03
deSALT	-d 10 -x ont2d -s 2	94.93	94.67	95.04	78.1
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.97	95.07	95.56	79.68
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 30X, reads from the genes with single isoform, 76874 reads/74590314 bases/154293 exons</b>					
GMAP	-f samse --cross-species -z sense_force	91.18	76.08	87.58	67.12
Minimap2	-ax splice	92.43	79.84	88.39	70.21
deSALT	-d 10 -x ont2d -s 2	94.31	88.97	94.94	82.3
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.52	91.56	95.87	86.07
<b>Human ONT 2D (1D<sup>2</sup>), coverage = 30X, reads from the genes with multiple isoforms, 222523 reads/349307842 bases/1461607 exons</b>					
GMAP	-f samse --cross-species -z sense_force	90.91	79.36	73.4	44.68
Minimap2	-ax splice	92.74	89.18	87.91	62.57
deSALT	-d 10 -x ont2d -s 2	94.96	94.9	95.11	78.71
deSALT + GTF	-d 10 -x ont2d -s 2 -G anno.gtf	94.99	95.28	95.55	80.06

**Simulated PacBio subread datasets**

**Fruit fly PacBio subread, coverage = 4X, reads from the genes with single isoform, 14281 reads/15851135 bases/32392 exons**

GMAP	-f samse --cross-species -z sense_force	73.19	53.19	56.82	36.44
Minimap2	-ax splice	85.65	80.93	79.25	62.91
deSALT	-d 10 -x clr -s 2	87.17	87.62	93.45	75.34
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.29	89.41	94.34	78.67
<b>Fruit fly PacBio subread, coverage = 4X, reads from the genes with multiple isoforms, 29888 reads/86093995 bases/180025 exons</b>					
GMAP	-f samse --cross-species -z sense_force	83.55	64.11	50.04	24.67
Minimap2	-ax splice	88.08	92.53	87.85	67.08
deSALT	-d 10 -x clr -s 2	88.61	96.35	96.79	80.71
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.63	96.74	97.23	82.29
<b>Fruit fly PacBio subread, coverage = 10X, reads from the genes with single isoform, 31433 reads/39457740 bases/76495 exons</b>					
GMAP	-f samse --cross-species -z sense_force	73.32	54.06	55.63	35.48
Minimap2	-ax splice	86.61	86.36	91.24	71.71
deSALT	-d 10 -x clr -s 2	87.58	91.43	96.47	80.89
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.6	92.22	96.84	82.54
<b>Fruit fly PacBio subread, coverage = 10X, reads from the genes with multiple isoforms, 66477 reads/215155994 bases/442303 exons</b>					
GMAP	-f samse --cross-species -z sense_force	83.74	64.48	50.13	23.65
Minimap2	-ax splice	88.44	94.09	95.27	72.54
deSALT	-d 10 -x clr -s 2	88.72	97.26	98.47	84.47
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.73	97.48	98.68	85.5
<b>Fruit fly PacBio subread, coverage = 30X, reads from the genes with single isoform, 88646 reads/118338677 bases/223878 exons</b>					
GMAP	-f samse --cross-species -z sense_force	73.94	55.27	56.78	36
Minimap2	-ax splice	87.03	88.6	95.02	75.72
deSALT	-d 10 -x clr -s 2	87.68	92.96	97.38	83.84
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.68	93.26	97.34	84.53
<b>Fruit fly PacBio subread, coverage = 30X, reads from the genes with multiple isoforms, 188895 reads/645460106 bases/1316757 exons</b>					
GMAP	-f samse --cross-species -z sense_force	84.16	65.2	50.91	23.38
Minimap2	-ax splice	88.53	94.55	96.46	73.59
deSALT	-d 10 -x clr -s 2	88.75	97.59	98.8	85.22
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	88.75	97.83	98.95	86.33
<b>Mouse PacBio subread, coverage = 4X, reads from the genes with single isoform, 12407 reads/12875901 bases/25870 exons</b>					
GMAP	-f samse --cross-species -z sense_force	68.9	43.9	51.44	32.43
Minimap2	-ax splice	82.84	72.44	74.45	56.44

deSALT	-d 10 -x clr -s 2	83.27	76.92	85.62	64.28
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	83.5	80.67	87.85	69.36
<b>Mouse PacBio subread, coverage = 4X, reads from the genes with multiple isoforms, 34074 reads/52880348 bases/190710 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.49	53.51	39.6	21.3
Minimap2	-ax splice	84.9	84.77	75.57	50.99
deSALT	-d 10 -x clr -s 2	87.23	91.84	89.65	69.07
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.26	92.62	90.59	71.82
<b>Mouse PacBio subread, coverage = 10X, reads from the genes with single isoform, 27315 reads/31993814 bases/60557 exons</b>					
GMAP	-f samse --cross-species -z sense_force	68.87	44.16	50.78	31.73
Minimap2	-ax splice	83.59	77.42	84.42	63.76
deSALT	-d 10 -x clr -s 2	84.11	82.9	90.82	72.04
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	84.22	85.38	92.16	75.97
<b>Mouse PacBio subread, coverage = 10X, reads from the genes with multiple isoforms, 75118 reads/131809879 bases/465578 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.49	53.64	37.47	19.39
Minimap2	-ax splice	85.56	87.09	84.42	56.79
deSALT	-d 10 -x clr -s 2	87.57	93.81	93.65	74.65
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.58	94.18	94.07	76.12
<b>Mouse PacBio subread, coverage = 30X, reads from the genes with single isoform, 77032 reads/95918127 bases/176769 exons</b>					
GMAP	-f samse --cross-species -z sense_force	69.45	44.82	50.49	32.01
Minimap2	-ax splice	84.15	79.81	88.74	67
deSALT	-d 10 -x clr -s 2	84.3	85.61	93.17	76
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	84.37	86.74	93.45	77.91
<b>Mouse PacBio subread, coverage = 30X, reads from the genes with multiple isoforms, 212189 reads/395360408 bases/1383094 exons</b>					
GMAP	-f samse --cross-species -z sense_force	75.98	54.41	37.35	18.9
Minimap2	-ax splice	85.82	88.17	87.71	58.78
deSALT	-d 10 -x clr -s 2	87.63	94.51	94.97	76.34
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.64	94.82	95.27	77.6
<b>Human PacBio subread, coverage = 4X, reads from the genes with single isoform, 11985 reads/10043761 bases/21582 exons</b>					
GMAP	-f samse --cross-species -z sense_force	66.55	39.39	48.94	33.08
Minimap2	-ax splice	83.2	68.72	72.33	55.94
deSALT	-d 10 -x clr -s 2	85.81	76.5	85.95	66.25
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	86.2	82.11	88.99	72.66

<b>Human PacBio subread, coverage = 4X, reads from the genes having multiple isoforms, 34552 reads/46766039 bases/186837 exons</b>					
GMAP	-f samse --cross-species -z sense_force	72.48	50.69	36.59	19.29
Minimap2	-ax splice	84.52	84.26	74.29	49.44
deSALT	-d 10 -x clr -s 2	87.23	91.61	89.03	68.78
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.32	92.39	90.13	71.12
<b>Human PacBio subread, coverage = 10X, reads from the genes with single isoform, 26391 reads/24880108 bases/50077 exons</b>					
GMAP	-f samse --cross-species -z sense_force	66.51	39.33	48.04	31.75
Minimap2	-ax splice	84.5	74.4	82.88	63.16
deSALT	-d 10 -x clr -s 2	86.85	83.49	92.34	74.29
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.04	86.82	93.88	78.59
<b>Human PacBio subread, coverage = 10X, reads from the genes with multiple isoforms, 76142 reads/116446237 bases/454874 exons</b>					
GMAP	-f samse --cross-species -z sense_force	72.53	50.94	34.07	17.53
Minimap2	-ax splice	85.26	86.88	83.81	55.47
deSALT	-d 10 -x clr -s 2	87.73	93.87	93.82	74.29
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.77	94.33	94.48	75.94
<b>Human PacBio subread, coverage = 30X, reads from the genes with single isoform, 74399 reads/74561310 bases/145347 exons</b>					
GMAP	-f samse --cross-species -z sense_force	66.68	39.48	47.43	31.89
Minimap2	-ax splice	85.04	77.14	87.65	67.14
deSALT	-d 10 -x clr -s 2	87.23	86.91	95.07	78.91
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.33	88.89	95.85	81.72
<b>Human PacBio subread, coverage = 30X, reads from the genes with multiple isoforms, 214976 reads/349280580 bases/1351629 exons</b>					
GMAP	-f samse --cross-species -z sense_force	72.84	51.44	33.45	16.96
Minimap2	-ax splice	85.57	87.97	87.55	57.7
deSALT	-d 10 -x clr -s 2	87.87	94.6	95.5	76.16
deSALT + GTF	-d 10 -x clr -s 2 -G anno.gtf	87.89	94.99	95.97	77.72
<b>Simulated ONT 1D datasets</b>					
<b>Fruit fly ONT 1D, coverage = 4X, reads from the genes with single isoform, 14293 reads/15779696 bases/35183 exons</b>					
GMAP	-f samse --cross-species -z sense_force	55	16.79	21.24	5.37
Minimap2	-ax splice	77.7	53.85	52.75	29.2
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	91.41	64.14	74.17	37.98

deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	92.5	74.83	83.75	52.65
<b>Fruit fly ONT 1D, coverage = 4X, reads from the genes with multiple isoforms, 30253 reads/86062596 bases/205722 exons</b>					
GMAP	-f samse --cross-species -z sense_force	66.52	20.99	9.96	2.47
Minimap2	-ax splice	91.75	78.67	69.62	35.85
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.3	88.44	86.37	51.41
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.57	91.8	92.05	60.72
<b>Fruit fly ONT 1D, coverage = 10X, reads from the genes with multiple isoforms, 31603 reads/39450304 bases/83839 exons</b>					
GMAP	-f samse --cross-species -z sense_force	56.83	17.69	22.26	5.5
Minimap2	-ax splice	80.44	58.74	59.48	33.02
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	93.49	72.62	81	45.71
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.14	83.44	90.37	64.02
<b>Fruit fly ONT 1D, coverage = 10X, reads from the genes with multiple isoforms, 68524 reads/215158359 bases/505777 exons</b>					
GMAP	-f samse --cross-species -z sense_force	67.63	21.38	9.1	2.21
Minimap2	-ax splice	92.91	81.16	74.58	38.69
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.75	91.86	91.23	60.02
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.9	94.02	94.63	67.75
<b>Fruit fly ONT 1D, coverage = 30X, reads from the genes with single isoform, 94542 reads/118339414 bases/251194 exons</b>					
GMAP	-f samse --cross-species -z sense_force	56.91	17.82	22.36	5.35
Minimap2	-ax splice	80.52	58.79	59.47	33.02
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	94.16	79.38	87.37	55.79
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	94.44	85.64	91.65	68.08
<b>Fruit fly ONT 1D, coverage = 30X, reads from the genes with multiple isoforms, 203279 reads/645463167 bases/1512998 exons</b>					
GMAP	-f samse --cross-species -z sense_force	67.68	21.43	8.99	2.2
Minimap2	-ax splice	92.98	81.22	74.84	39.01
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	96.84	92.86	93.09	64.64
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	96.94	94.49	95.24	70.09
<b>Mouse ONT 1D, coverage = 4X, reads from the genes with single isoform, 12411 reads/12792340 bases/27910 exons</b>					
GMAP	-f samse --cross-species -z sense_force	53.16	10.43	25.52	4.99
Minimap2	-ax splice	71.45	36.85	41.33	17.86
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	80.42	40.42	56.36	22.19
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	82.18	53.6	64.66	35.3
<b>Mouse ONT 1D, coverage = 4X, reads from the genes with multiple isoforms, 34189 reads/52716526 bases/215979 exons</b>					

GMAP	-f samse --cross-species -z sense_force	53.78	13.14	8.58	2.09
Minimap2	-ax splice	76.26	54.58	38.36	16.37
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.2	69.15	57.84	26.61
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.47	77.03	69.69	37.28
<b>Mouse ONT 1D, coverage = 10X, reads from the genes with single isoform, 27474 reads/31977856 bases/66240 exons</b>					
GMAP	-f samse --cross-species -z sense_force	54.18	10.91	25.96	5.25
Minimap2	-ax splice	73.68	40.19	46.08	20.29
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	84.28	50.1	63.62	26.97
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	85.52	65.68	74.86	45.52
<b>Mouse ONT 1D, coverage = 10X, reads from the genes with multiple isoforms, 76173 reads/131797813 bases/529496 exons</b>					
GMAP	-f samse --cross-species -z sense_force	55.05	13.84	7.43	1.73
Minimap2	-ax splice	77.85	57.16	41.66	18.08
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	90.42	77.23	69.24	36.87
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	91.06	81.91	76.48	45.85
<b>Mouse ONT 1D, coverage = 30X, reads from the genes with single isoform, 82162 reads/95919067 bases/198286 exons</b>					
GMAP	-f samse --cross-species -z sense_force	52.78	10.46	24.45	4.84
Minimap2	-ax splice	73.62	40.46	46	20.46
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	85.43	59.81	70.44	36.03
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	85.99	69.95	76.83	51.43
<b>Mouse ONT 1D, coverage = 30X, reads from the genes with multiple isoforms, 227114 reads/395363051 bases/1585095 exons</b>					
GMAP	-f samse --cross-species -z sense_force	31.25	7.85	4.34	1
Minimap2	-ax splice	77.98	57.42	41.87	18.3
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	90.8	79.71	73.45	42.78
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	91.18	82.5	77.32	48.21
<b>Human ONT 1D, coverage = 4X, reads from the genes with single isoform, 11997 reads/9944834 bases/23000 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.37	8.23	22.53	4.86
Minimap2	-ax splice	67.38	27	35.75	15.57
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	79.56	33.02	53.16	20.81
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	81.77	46.52	61.08	33.38
<b>Human ONT 1D, coverage = 4X, reads from the genes with multiple isoforms, 34647 reads/46572365 bases/210621 exons</b>					
GMAP	-f samse --cross-species -z sense_force	48.99	11.88	6.46	1.68
Minimap2	-ax splice	71.38	49.6	33.18	14.13

deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	86.35	67.76	55.98	26.78
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	87.68	74.87	66.54	35.92
<b>Human ONT 1D, coverage = 10X, reads from the genes with single isoform, 26530 reads/24859119 bases/53998 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.99	8.82	23.85	5.18
Minimap2	-ax splice	69.61	30.46	40.37	18.11
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	84.04	41.56	60.32	25.42
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	85.63	59.81	71.78	44.71
<b>Human ONT 1D, coverage = 10X, reads from the genes with multiple isoforms, 77002 reads/116437109 bases/515990 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.9	12.47	5.46	1.41
Minimap2	-ax splice	73.06	52.42	36.46	15.7
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.52	75.06	66.65	35.6
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.22	79.38	73.13	42.99
<b>Human ONT 1D, coverage = 30X, reads from the genes with single isoform, 79328 reads/74561914 bases/161631 exons</b>					
GMAP	-f samse --cross-species -z sense_force	51	9.01	24.12	5.08
Minimap2	-ax splice	69.62	29.94	40.1	17.9
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	85.5	51.65	66.98	33.7
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	86.26	64.53	73.98	49.99
<b>Human ONT 1D, coverage = 30X, reads from the genes with multiple isoforms, 229807 reads/349281664 bases/1544708 exons</b>					
GMAP	-f samse --cross-species -z sense_force	49.93	12.37	5.39	1.42
Minimap2	-ax splice	73.24	52.62	36.66	15.88
deSALT	-d 10 -x ont1d -s 2 -k 14 -s 2	88.98	76.91	69.76	40.31
deSALT + GTF	-d 10 -x ont1d -s 2 -k 14 -s 2 -G anno.gtf	89.44	79.81	73.61	44.93

- a) This table depicts the results of the aligners on the reads from the genes with single and multiple isoforms. For each of the simulated datasets, the reads from the genes with single and multiple isoforms are respectively divided into two groups, and the benchmark is implemented for each of the groups.
- b) Refer to the Supplementary Notes for the command lines of the benchmarked aligners.
- c) Base%: the proportion of bases being correctly aligned to their ground truth positions, i.e., the mapped positions of the bases are within 5 bp of their ground truth positions.
- d) Exon%: the proportion of the exons being correctly mapped. An exon in a certain read is considered to be correctly mapped only if its two boundaries are mapped within 5 bp of their ground truth positions in the reference genome.
- e) Read80%: the proportion of Read80% reads. A read is considered to be a Read80% read only if it meets two conditions:  $N_T/N_G > 80\%$  and  $N_T/N_P > 80\%$ , where  $N_G$  is the number of ground truth exons within the read,  $N_P$  is the number of exons predicted by the alignment, and  $N_T$  is the number of true positive exons. Herein, a predicted exon is considered to be a true positive exon only if there is a ground truth exon in the read and the corresponding boundaries of the predicted exon and the ground truth exon are within 5 bp.

- f) Read100%: the proportion of Read100% reads. A read is considered to be a Read100% read only if it meets two conditions:  $N_T/N_G = 100\%$  and  $N_T/N_P = 100\%$ . It is worth noting that a Read100% read indicates that the read has a highly correct full-length alignment.



**Supplementary Table 6. Alignment speed on simulated human datasets with various numbers of threads <sup>a</sup>**

Aligner	Parameters	1	4	8	16	1	4	8	16	1	4	8	16
		thread	threads	threads	threads	thread	threads	threads	threads	thread	threads	threads	threads
		<b>Human PacBio ROI 4x dataset</b> (46783 reads, 57203247 bases)				<b>Human PacBio ROI 10x dataset</b> (103339 reads, 141769367 bases)				<b>Human PacBio ROI 30x dataset</b> (292117 reads, 424010833 bases)			
GMAP	-f samse --cross-species -z sense_force -t thread	3016	798	421	411	7944	1980	1036	946	24370	6112	3529	2628
Minimap2	-ax splice -t thread	977	280	159	128	2391	654	347	276	6698	1885	975	740
deSALT	-d 10 -t thread	942	253	161	128	2126	555	308	236	5881	1541	826	597
deSALT+GTF	-d 10 -G anno.gtf -t thread	660	216	134	109	1745	473	266	200	4682	1332	696	465
		<b>Human ONT 2D (1D<sup>2</sup>) 4x dataset</b> (46594 reads, 56604793 bases)				<b>Human ONT 2D (1D<sup>2</sup>) 10x dataset</b> (102920 reads, 141283445 bases)				<b>Human ONT 2D (1D<sup>2</sup>) 30x dataset</b> (299397 reads, 423898156 bases)			
GMAP	-f samse --cross-species -z sense_force -t thread	18288	4336	2194	1262	43528	11051	5787	3566	130273	34327	17543	10178
Minimap2	-ax splice -t thread	977	296	172	116	2389	694	370	236	6811	2055	1048	625
deSALT	-d 10 -t thread	1075	349	193	137	2555	701	400	243	6636	1982	1012	593
deSALT+GTF	-d 10 -G anno.gtf -t thread	951	265	181	133	2122	595	347	238	6082	1686	935	574
		<b>Human PacBio subread 4x dataset</b> (46537 reads, 56809800 bases)				<b>Human PacBio subread 10x dataset</b> (102533 reads, 141326345 bases)				<b>Human PacBio subread 30x dataset</b> (289375 reads, 423841890 bases)			
GMAP	-f samse --cross-species -z sense_force -t thread	30398	7427	3969	3136	84699	20885	11068	9558	245062	63679	31066	23901
Minimap2	-ax splice -t thread	921	251	149	134	2355	579	322	274	6644	1687	903	777
deSALT	-d 10 -t thread	1088	339	173	161	2528	647	349	280	7078	1777	952	746
deSALT+GTF	-d 10 -G anno.gtf -t thread	852	289	159	122	2030	592	324	232	5860	1676	885	600
		<b>Human ONT 1D 4x dataset</b> (46644 reads, 56517199 bases)				<b>Human ONT 1D 10x dataset</b> (103532 reads, 141296288 bases)				<b>Human ONT 1D 30x dataset</b> (309135 reads, 423843578 bases)			
GMAP	-f samse --cross-species -z sense_force -t thread	66923	17327	9810	1262	229494	58722	41764	3566	688482	140132	91460	10178
Minimap2	-ax splice -t thread	863	255	147	124	2243	592	321	273	6329	1675	891	737
deSALT	-d 10 -t thread	1181	350	220	194	3025	746	418	345	8155	2115	1183	974
deSALT+GTF	-d 10 -G anno.gtf -t thread	927	276	172	135	2392	616	340	262	6484	1766	971	692

- a) The benchmark of speed contains the 12 human datasets in four kinds of sequencing error models (PacBio ROI, PacBio subreads, ONT2D, and ONT1D) and three kinds of sequencing depths (4x, 10x, 30x). (Refer to datasets 1–12 in Supplementary Table 1)

**Supplementary Table 7. Benchmark results on the real sequencing datasets**

Aligner	Parameters	#BaseA <sup>c</sup>	#BaseGA <sup>d</sup>	#ExonP <sup>e</sup>	#ExonGO <sup>f</sup>	#ExonGA <sup>g</sup>	#ReadGA <sup>h</sup>	Running Time (in seconds) <sup>i</sup>
<b>real HUMAN ONT dataset, 15152101 reads, 14134831170 bases</b>								
GMAP	-f samse --cross-species -z sense_force <sup>a</sup>	9948549314	8402753847	39905636	37625378	17132563	3107592	1856119
Minimap2	-ax splice -uf -k14 <sup>b</sup>	12068583337	10297615525	50372792	49621879	25201774	4746517	13201
deSALT	-d 10 -k 15	12035091373	10555134761	53626292	52661753	31887344	6892999	10865
deSALT + GTF	-d 10 -k 15 -G anno.gtf	12082199478	10603380632	54383709	53339284	32374563	6975312	10003
<b>real mouse PacBio dataset 2269795 reads, 3213849871 bases</b>								
GMAP	-f samse --cross-species -z sense_force	2256745964	1926549648	8889112	8341684	5464835	791874	489082
Minimap2	-ax splice	2749969434	2451570407	10792312	10755064	8103391	1396677	4160
deSALT	-d 10	2769323628	2475226314	11274040	11149070	8389832	1463815	3385
deSALT + GTF	-d 10 -G anno.gtf	2768404864	2474675193	11255344	11137885	8375354	1463945	3347

- a) The parameter was recommended by the PacBio team ([https://github.com/Magdoll/cDNA\\_Cupcake/wiki/Best-practice-for-aligning-Iso-Seq-to-reference-genome:-minimap2,-GMAP,-STAR,-BLAT](https://github.com/Magdoll/cDNA_Cupcake/wiki/Best-practice-for-aligning-Iso-Seq-to-reference-genome:-minimap2,-GMAP,-STAR,-BLAT)).
- b) The parameter of Minimap2 for the NA12878 dataset was used by referring to <https://github.com/nanopore-wgs-consortium/NA12878/blob/master/RNA.md>.
- c) #BaseA: the number of bases being aligned.
- d) #BaseGA: the number of bases aligned to the positions within annotated exons.
- e) #ExonP: the number of exons predicted by the alignments (also termed “predicted exons”). Herein, the predicted exons in various reads are independently considered.
- f) #ExonGO: the number of predicted exons being overlapped by annotated exons (also termed “overlapped exons”). Herein, a predicted exon is considered to be overlapped by annotated exons only if there is at least one annotated exon and at least 10 bp overlapping between the predicted exon and the annotated exon.
- g) #ExonGA: the number of predicted exons being exactly matched by annotated exons (also termed “exactly matched exons”). Herein, a predicted exon is considered to be exactly matched by annotated exons only if there is an annotated exon and the positions of the two boundaries of the predicted exon are within 5 bp of the corresponding boundaries of the annotated exon.
- h) #ReadGA: the number of ReadGA reads. A read is considered to be a ReadGA read only if each of the intron boundaries implied by its alignment is within 5 bp of an annotated exon. Herein, a ReadGA read indicates that the read could have a correct full-length alignment.
- i) The alignment time of aligners. All the aligners were run in 24 threads for the HUMAN ONT dataset and in 32 threads for the MOUSE PacBio dataset.

**Supplementary Table 8. Numbers of exactly matched exons of various lengths on the real sequencing datasets <sup>a</sup>**

Aligner	ExonGA(20)	ExonGA(30)	ExonGA(40)	ExonGA(50)	ExonGA(60)
<b>real HUMAN ONT dataset, 15152101 reads, 14134831170 bases</b>					
GMAP	9332	208819	447638	893442	1604262
Minimap2	1775	138108	357615	891225	1859905
deSALT	7534	242548	602534	1367186	2725536
deSALT + GTF	7658	243037	603863	1368751	2727465
<b>real mouse PacBio dataset 2269795 reads, 3213849871 bases</b>					
GMAP	4111	45191	137456	227049	365372
Minimap2	625	30058	129886	256000	463676
deSALT	5068	57585	178308	312060	531633
deSALT + GTF	4930	58157	178617	312166	531212

a) This table describes the various ExonGA(x) (x<60) statistics of the aligners, which reflect their ability to handle relatively small exons.

**Supplementary Table 9. Numbers of in silico transcript sequences from various sets of genes <sup>a</sup>**

<b>Species</b>	<b># of transcripts from genes having single isoforms</b>	<b># of transcripts from genes having multiple isoforms</b>	<b># of transcripts from genes having short (&lt;31 bp) exons</b>	<b># of all the generated transcripts</b>
Fruit fly	3009	3065	2880	8954
Mouse	3072	4080	2783	9935
Human	2998	3603	3346	9947

a) This table describes the number of *in silico* transcript sequences being extracted from various sets of genes and used to generate simulated datasets.

## Supplementary Notes

### 1. Alignment skeleton generation (first-pass alignment)

#### 1.1 Genome indexing

The reference genome is organized and indexed using the RdBG-index approach, which was initially developed in deBGA [4] for short DNA-seq read alignment. Specifically, a de Bruijn graph of the reference genome (RDBG) is constructed with a given  $k$ -mer size ( $k$  parameter of the tool), and the RdBG is further organized by a hash table-based data structure to index the vertices ( $k$ -mers) of the graph as well as their unitig IDs and offsets.

In practice, no significant difference is observed in the results of the long RNA-seq read alignment with various  $k$  parameter settings ( $k=20$  to  $28$ , data not shown). This is mainly due to the fact that the seed length used in deSALT is much shorter than  $k$  bp, so that tuning the  $k$  parameter has little effect on the generation of alignment skeletons. In this situation, deSALT uses the setting  $k=22$  in all the benchmarks. Moreover, users can change the setting on  $k$  if necessary.

#### 1.2 Match-block generation

For a given read, deSALT extracts  $l$ -mers (default value:  $l=15$ ) at every  $m$  bp (default value:  $m=5$ ) as seeds and finds exact matches between the seeds and the unitigs of RdBG via the RdBG-index. Since  $l < k$ , a seed possibly has matches to multiple unitigs. deSALT extends each of the matches in both the upstream and downstream direction to generate a maximal exact match between the read and the corresponding unitig (termed a U-MEM). Moreover, for a certain seed, only its longest U-MEM is saved for further processing and others (if applicable) are filtered out. Furthermore, some seeds could have identical U-MEMs; such U-MEMs would be merged to remove redundant ones. After the processing, a set of non-redundant U-MEMs is produced, and each of them is termed as a specific 5-tuple,  $M_i = (M_i^{RS}, M_i^{RE}, M_i^U, M_i^{US}, M_i^{UE})$ ,  $i = 1, \dots, |M|$ , where  $M_i^{RS}$  and  $M_i^{RE}$  respectively indicate the start and end positions of the match on the read,  $M_i^U$  indicates the unitig ID of the match,  $M_i^{US}$  and  $M_i^{UE}$  respectively indicate the start and end positions of the match on unitig  $M_i^U$ , and  $|M|$  indicates the total number of U-MEMs.

deSALT further merges co-linear U-MEMs on same unitigs as super U-MEMs (SU-MEMs). This is implemented by the following rule: for two U-MEMs  $M_1$  and  $M_2$  on the same unitig  $ux$ ,  $M_1 = (M_1^{RS}, M_1^{RE}, M_{ux}^U, M_1^{US}, M_1^{UE})$  and  $M_2 = (M_2^{RS}, M_2^{RE}, M_{ux}^U, M_2^{US}, M_2^{UE})$ , if  $|(M_2^{RS} - M_1^{RS}) - (M_2^{US} - M_1^{US})| < T_{SU}$ ,  $M_1^R < M_2^R$  and  $M_1^O < M_2^O$ , the two U-MEMs can be merged, where  $T_{SU}$  is a threshold (default value:  $T_{SU} = 20$ ). An SU-MEM,  $SM_1 = (M_1^{RS}, M_2^{RE}, M_{ux}^U, M_1^{US}, M_2^{UE})$ , is generated after the merging. It is worth noting that a SU-MEM is defined by a similar 5-tuple, i.e.,  $SM_1 = (M_i^{RS}, M_i^{RE}, M_i^U, M_i^{US}, M_i^{UE})$ . This definition is compatible with U-MEMs and easy to use for the merging operations between a U-MEM and an SU-MEM or two SU-MEMs.

In practice, deSALT separately processes the unitigs with U-MEMs. For each of the unitigs, deSALT uses a greedy approach to iteratively check the U-MEMs from upstream to downstream and merges the U-MEMs meeting the conditions to produce SU-MEMs. It is also worth noting that there are possibly a proportion of U-MEMs (if not all) that cannot be merged and remain as individual U-MEMs. The SU-MEMs and individual U-MEMs are then mapped to the reference genome through the genomic positions of the unitigs (which are recorded by the RdBG-index in advance). It is worth noting that a unitig could have multiple copies in the reference genome, so that a SU-MEM/U-MEM could be split into multiple copies and mapped to all the corresponding genomic positions. Each of the mapped SU-MEMs/U-MEMs is termed a match block (MB) and defined as a 4-tuple  $MB_i = (MB_i^{RS}, MB_i^{RE}, MB_i^{GS}, MB_i^{GE})$ ,  $i = 1, \dots, |MB|$ , where  $MB_i^{RS}$  and  $MB_i^{RE}$  respectively indicate the start and end positions of the match on the read,  $MB_i^{GS}$  and  $MB_i^{GE}$  respectively indicate the start and end positions of the match on the reference genome, and  $|MB|$  indicates the total number of MBs.

#### 1.3 SDP-based alignment skeleton generation

deSALT builds a direct acyclic graph (DAG) consisting of MBs. Each MB is a vertex, and there is an edge  $MB_i \rightarrow MB_j$  if the placement of the two MBs,  $MB_i$  and  $MB_j$ , meets the following condition:

$$MB_i^{RE} - \varepsilon < MB_j^{RS} \text{ and } MB_i^{GE} - \varepsilon < MB_j^{GS} < MB_i^{GE} + \delta,$$

where  $\delta$  is a user-defined parameter depicting the maximum allowed intron length (default value:  $\delta = 200,000$  bp), and  $\varepsilon$  is the maximum allowed overlap length between two MBs (default value:  $\varepsilon = 5$  bp).  $\varepsilon$  is helpful for handling some small repeats around the boundaries of exons well. Such repeats could make some neighboring MBs have small overlaps, and  $\varepsilon$  helps to correctly link such MBs and prevent false negatives.

A weight  $s(MB_i \rightarrow MB_j) = w(MB_i \rightarrow MB_j) - p(MB_i \rightarrow MB_j)$  is assigned to each of the edges.

Here,  $w(MB_j)$  is a weight score for depicting the match length of  $MB_j$ :

$$w(MB_i \rightarrow MB_j) = \min(MB_j^{RE} - MB_j^{RS}, MB_j^{RE} - MB_i^{RE}, MB_j^{GE} - MB_i^{GE})$$

The last two elements in the brackets are applicable when two MBs have overlapping bases.

$p(MB_i \rightarrow MB_j)$  is the intron length-based penalty score for preventing an ill-defined intron:

$$p(MB_i \rightarrow MB_j) = \begin{cases} \alpha |MB_j^{GS} - MB_i^{GE}| / w(MB_i \rightarrow MB_j), & MB_j^{GS} - MB_i^{GE} \geq 0 \\ \min(\alpha |MB_j^{GS} - MB_i^{GE}| / w(MB_i \rightarrow MB_j), \beta) & MB_j^{GS} - MB_i^{GE} < 0 \end{cases}$$

where  $\alpha$  is a constant empirically set as 2 to scale the penalty score, and  $\beta = \alpha T_{SU} / |MB_j^{RE} - MB_j^{RS}|$  is a constant penalty for an edge.

deSALT performs an SDP on the MBs by the following recursion equation to generate the alignment skeletons:

$$S_{AS}(MB_j) = \max\{S_{AS}(MB_i) + w(MB_i \rightarrow MB_j)\}, i \in Precursor\{MB_j\},$$

where  $Precursor\{MB_j\}$  is the precursor set of  $MB_j$  in which each of its elements  $MB_i$  has an edge  $MB_i \rightarrow MB_j$ .

Moreover, we define a vertex  $MB_j$  as an isolated vertex if  $Precursor\{MB_j\} = \emptyset$ , and set  $S_{AS}(MB_j) = MB_j^{GE} - MB_j^{GS}$ .

With the recursion equation, deSALT assigns a score to all the vertices and also records the selected precursors for each of the vertices. The end vertex with the highest score is then selected, and deSALT builds the best alignment skeleton from it with backtracking operations. Here, end vertices indicate the vertices without successors.

Moreover, deSALT could output multiple alignment skeletons with similar scores. Specifically, deSALT also builds alignment skeletons for the end vertices meeting the following condition:  $S_{AS}(MB_h) / S_{AS}(MB_{best}) > T_{AS}$ , where  $MB_{best}$  is the end vertex with the highest score, and  $T_{AS}$  is a user-defined threshold (default value:  $T_{AS} = 0.9$ ). For an  $MB_h$ , if the built skeleton has no shared vertex with the best skeleton, deSALT saves it as an alternative skeleton; otherwise, it would be discarded.

deSALT stores the best and the alternative skeletons in a temporary file for further processing. Each of the skeletons is recorded as a permutation of MBs:

$$AS_i = MB_{i_1} \rightarrow MB_{i_2} \rightarrow \dots \rightarrow MB_{i_{|AS_i|}}, i = 1, \dots, |AS|,$$

where each  $MB_{i_j}, j = 1, \dots, |AS_i|$  indicates an MB in the skeleton,  $|AS_i|$  indicates the total number of MBs the skeleton has, and  $|AS|$  indicates the total number of generated alignment skeletons.

## 2. Exon inference

### 2.1 Draft exon generation

deSALT integrates the alignment skeletons in an iterative approach to produce a set of draft exons. Herein, the exon set is formulated as  $EX = \{EX_i, i = 1, \dots, |EX|\}$ , where  $|EX|$  is the total number of exons, and each  $EX_i$  depicts an exon as a 3-tuple:  $EX_i = (EX_i^{GS}, EX_i^{GE}, EX_i^C)$ , where  $EX_i^{GS}$  and  $EX_i^{GE}$  are respectively the start and end positions of the exon on the reference genome, and  $EX_i^C$  depicts the number of alignment skeletons covered on the exons.

In practice, deSALT first extracts all the MBs of all the alignment skeletons and then sorts the MBs by their starting

positions on the reference genome, i.e.,  $MB_{i,j}^{GS}$ ,  $i = 1, \dots, |AS|$ ,  $j = 1, \dots, |AS_i|$ . The initial exon set is constructed by the mapped MBs as follows:

$$EX = \{EX_p, p = 1, \dots, \sum_{t=1}^{|AS|} |AS_t|\}, \text{ each } EX_p = (MB_{x,y}^{GS}, MB_{x,y}^{GE}, 1),$$

where the subscript  $x$  and  $y$  defines a specific MB, and all  $EX_p$  are sorted by their upstream position.

deSALT then iteratively merges the  $EX_p$  set to update  $EX$  from upstream to downstream. For each  $EX_p$  ( $p > 1$ ), deSALT checks if it meets the following condition:

$$EX_p^{GS} - \mu < EX_q^{GE} < EX_p^{GE},$$

where  $EX_q$  is the exon closest to  $EX_p$  upstream,  $EX_q^{GE}$  is the end position of  $EX_q$ , and  $\mu$  is the user-defined minimum intron length (default value:  $\mu = 20$ ). If it does, deSALT merges the two exons to generate an updated exon  $EX_q = (EX_q^{GS}, EX_p^{GE}, EX_p^C + EX_q^C)$ ; otherwise  $EX_p$  and  $EX_q$  would be not merged.

An updated exon set is generated after processing all the  $EX_p$ s, and deSALT filters the redundant and ill-defined exons. An exon  $EX_p$  would be filtered out if it meets one of the two following conditions:

- 1)  $EX_p$  is completely within another  $EX_q$ , i.e.,  $EX_p^{GS} > EX_q^{GS}$  and  $EX_p^{GE} < EX_q^{GE}$ ;
- 2)  $EX_p^C = 1$  and  $EX_p^{GE} - EX_p^{GS} < l + 5$ , where  $l$  is the seed length used in the alignment generation step.

## 2.2 Exon refinement

There could be some small differences between the boundaries of the draft exons ( $EX_p^{GS}$ s and  $EX_p^{GE}$ s) and the real splicing sites, mainly due to the effect of sequencing errors. deSALT predicts real splicing sites around the boundaries using a local sequence-based scoring system to refine the draft exons.

For a draft exon  $EX_p$ , deSALT selects two local regions,  $EA_p = [EX_p^{GS} - 10, EX_p^{GS} + 10]$  and  $ED_p = [EX_p^{GE} - 10, EX_p^{GE} + 10]$ , and uses two scoring matrixes—Acceptor Scoring Matrix and Donor Scoring Matrix (see below), to respectively score the positions in  $EA_p$  and  $ED_p$ . For a given position, the corresponding matrix is used as a mask to simultaneously score the position itself, upstream 5 bp (-5 to -1 bp), and downstream 4 bp (+1 to +4 bp) according to their nucleotides and to sum them up as the score of the position. This sum score can be seen as the possibility of the base permutation at the real acceptor or donor splicing sites. deSALT chooses the positions with the highest sum scores in  $EA_p$  and  $ED_p$  as the predicted acceptor and donor sites, respectively, and updates the exon  $EX_p$ .

After the prediction, a more accurate exon set is produced to implement the second-pass read alignment.

**Acceptor Scoring Matrix**

position nucleotide	-5	-4	-3	-2	-1	0	1	2	3	4
A	-201	115	61	2548	-364	704	-142	-166	-33	12
C	431	176	838	-364	-364	-222	-170	-92	201	485
G	-121	199	-191	-364	2548	346	-117	973	157	-91
T	350	-107	-169	-364	-364	-271	1030	-137	47	9

**Donor Scoring Matrix**

position nucleotide	-5	-4	-3	-2	-1	0	1	2	3	4
A	-92	132	-97	922	-107	-344	-344	355	939	222
C	168	252	273	-208	-288	-344	-344	-236	-177	-206
G	-114	-110	-167	-132	1104	2406	-344	543	-160	1145
T	103	-138	143	-202	-225	-344	2406	-223	-207	185



### 3. Refined alignment (second-pass alignment)

#### 3.1 Construction of a local spliced reference sequence (LSRS)

deSALT splits a read  $R$  into a series of unaligned parts according to its alignment skeleton  $AS_R$  (assuming  $R$  has a unique skeleton), in which each part is represented as one of the substrings of  $R$ :

$$R[1, MB_{R_1}^{RS}], R[MB_{R_1}^{RS}, MB_{R_2}^{RS}], \dots, R[MB_{R_j}^{RS}, MB_{R_k}^{RS}], \dots, R[MB_{R_{|AS_R|}}^{RE}, |R|],$$

where  $MB_{R_j}$  and  $MB_{R_k}$  are two neighboring MBs, and  $|R|$  is the length of  $R$ .

For each  $R[MB_{R_j}^{RS}, MB_{R_k}^{RS}]$ , deSALT recognizes the exon meeting the following two conditions as the spanning exon: 1) the exon is between  $MB_{R_j}^{GS}$  and  $MB_{R_k}^{GS}$ ; 2) there is at least one short match between  $R[MB_{R_j}^{RS}, MB_{R_k}^{RS}]$  and the exon. Specifically, assuming that the exons containing  $MB_{R_j}$  and  $MB_{R_k}$  are respectively  $EX_{R_j}(EX_{R_j}^{GS}, EX_{R_j}^{GE}, EX_{R_j}^C)$  and  $EX_{R_k}(EX_{R_k}^{GS}, EX_{R_k}^{GE}, EX_{R_k}^C)$ , deSALT selects the two local genomic regions  $(MB_{R_j}^{GS}, EX_{R_j}^{GE})$  and  $(EX_{R_k}^{GS}, MB_{R_k}^{GS})$  as the first and last spanning exons for  $R[MB_{R_j}^{RS}, MB_{R_k}^{RS}]$ . Furthermore, for each of the exons within  $EX_{R_j}^{GE}$  and  $EX_{R_k}^{GS}$ , if there is at least one  $l$ -mer match between it and  $R[MB_{R_j}^{RS}, MB_{R_k}^{RS}]$ , the exon is also recognized as a spanning exon. The  $l$ -mer matches are quickly retrieved by a hash table-based index which indexes all the  $l$ -mers of read  $R$ . Here,  $l$  is a user-defined parameter (default value:  $l=8$ ). deSALT then stitches all the spanning exons by their genomic positions to compose a sequence as the LSRS for  $R[MB_{R_j}^{RS}, MB_{R_k}^{RS}]$ .

For  $R[1, MB_{R_1}^{RS}]$ , deSALT first recognizes the local genomic region  $(EX_{R_1}^{GS}, MB_{R_1}^{GS})$  as a spanning exon. Furthermore, deSALT checks five inferred exons upstream  $EX_{R_1}$  and selects the one(s) with at least one  $l$ -mer match to  $R[1, MB_{R_1}^{RS}]$  as the spanning exon(s) as well. The LSRS is then composed with all the spanning exons.

For  $R[MB_{R_{|AS_R|}}^{RE}, |R|]$ , the LSRS is composed in a similar way to the LSRS of  $R[1, MB_{R_1}^{RS}]$ , the difference being that deSALT checks five inferred exons downstream  $EX_{R_{|AS_R|}}$ .

deSALT splits a read with multiple alignment skeletons multiple times with the various skeletons and separately produces a set of LSRSs for each of the skeletons.

#### 3.2 Alignment of the read part against the LSRS

deSALT employs a fast SIMD-based global alignment implementation [5] to align read parts against their LSRSs. For each of the alignments, the CIGAR string is investigated. If there are large deletions, deSALT considers that there are exons with alternative splicing sites, i.e., the bases of the read part could be from some parts of the exons only. In this situation, deSALT updates the LSRS by removing the bases corresponding to the large deletion(s) and realigns the read part.

For a certain read, deSALT integrates the alignment of all its read parts into a full-length read alignment if the read has a unique alignment skeleton. If more than one alignment skeleton is applicable, multiple full-length read alignments will be produced for a read. Full-length read alignments are scored by summing the scores of all the alignments of its read parts. For reads with multiple alignments, deSALT outputs the alignment with the highest score as the primary alignment and the other alignments as secondary alignments.

## 4. Availability of datasets

### 4.1 The availability of the human ONT datasets

<http://s3.amazonaws.com/nanopore-human-wgs/rna/fastq/NA12878-cDNA-1D.pass.dedup.fastq>

### 4.2 The availability of the mouse PacBio dataset (Accession Number: SRR6238555)

<ftp://ftp.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByRun/sra/SRR/SRR623/SRR6238555/SRR6238555.sra>

### 4.3 The availability of the simulated datasets

[https://drive.google.com/drive/folders/16RpDYkdTCwOHmvoWNeHnUp7nxrNt\\_Q7T?usp=sharing](https://drive.google.com/drive/folders/16RpDYkdTCwOHmvoWNeHnUp7nxrNt_Q7T?usp=sharing)

These datasets are respectively generated by PBSim with the following command lines:

#### 1) PacBio ROI-like datasets:

```
pbsim Transcripome_File
  --data-type CCS \
  --model_gc model_gc_clr \
  --length-mean 6000 \
  --length-min 100
  --difference-ratio 75:5:20 \
  --accuracy-mean 0.98 \
  --accuracy-min 0.8 \
  --depth 4/10/30
```

#### 2) PacBio subread-like datasets:

```
pbsim Transcripome_File
  --data-type CLR \
  --model_gc model_gc_clr \
  --length-mean 7800 \
  --length-min 100
  --difference-ratio 1:12:2 \
  --accuracy-mean 0.85 \
  --accuracy-min 0.75 \
  --depth 4/10/30
```

#### 3) ONT 2D (1D<sup>2</sup>)-like datasets:

```
pbsim Transcripome_File
  --model_gc model_gc_clr \
  --length-mean 7800 \
  --length-min 100
  --difference-ratio 33:36:31 \
  --accuracy-mean 0.87 \
  --accuracy-min 0.8 \
  --depth 4/10/30
```

#### 4) ONT1D-like datasets:

```
pbsim Transcripome_File
```

```

--model_gc model_gc_clr \
--length-mean 7800 \
--length-min 100
--difference-ratio 48:15:37 \
--accuracy-mean 0.75 \
--accuracy-min 0.7 \
--depth 4/10/30

```

## 5. Availability of references

### 5.1 The availability of the reference genomes

<http://hgdownload.soe.ucsc.edu/goldenPath/hg38/bigZips/>, human GRCh38

<http://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/>, mouse, GRCm38

<http://hgdownload.soe.ucsc.edu/goldenPath/dm6/bigZips/>, fruit fly, DM6

### 5.2 The availability of the gene annotations

[ftp://ftp.ensembl.org/pub/release-94/gtf/homo\\_sapiens/](ftp://ftp.ensembl.org/pub/release-94/gtf/homo_sapiens/), human, version 94

[ftp://ftp.ensembl.org/pub/release-94/gtf/mus\\_musculus](ftp://ftp.ensembl.org/pub/release-94/gtf/mus_musculus/), mouse, version 94

[ftp://ftp.ensembl.org/pub/release-94/gtf/drosophila\\_melanogaster](ftp://ftp.ensembl.org/pub/release-94/gtf/drosophila_melanogaster/), fruit fly, version 94

## 6. Implementation of benchmarking

All the benchmarks were implemented on a server with Intel Xeon E4280 CPU at 2.0GHZ and 1 Terabytes RAM, running Linux Ubuntu 16.04. We referred to a tutorial proposed by the PacBio team to configure the parameters of GMAP and Minimap2 to align the PacBio reads. The tutorial is available at

[https://github.com/Magdoll/cDNA\\_Cupcake/wiki/Best-practice-for-aligning-Iso-Seq-to-reference-genome:-minimap2,-deSALT,-GMAP,-STAR,-BLAT](https://github.com/Magdoll/cDNA_Cupcake/wiki/Best-practice-for-aligning-Iso-Seq-to-reference-genome:-minimap2,-deSALT,-GMAP,-STAR,-BLAT)

The parameters of GMAP and Minimap2 were configured to align ONT reads by referring to a previous study:

<https://github.com/nanopore-wgs-consortium/NA12878/blob/master/RNA.md>

The alignment results on simulated and real datasets were evaluated by two in-house scripts, respectively, which are available at

<https://github.com/hitbc/deSALT>

The command lines and versions of the aligners used in the benchmarking were as follows.

### 6.1 Command lines of deSALT (version 1.1)

Genome indexing:

```
deSALT index Genome_File Genome_index_dir (the index model refer to deBGA)
```

Read alignment (default settings)

```
deSALT aln -o out.sam Genome_Index_Dir Fastq_File ref.fa
```

Read alignment (important parameters)

```
deSALT aln -k seeding-kmer -a local_hash_kmer -s seed_step -c min_chain_score
-gtf GTF_path -B batch_size -x read_type -f temp-file-perfix -o out.sam
Genome_Index_Dir Fastq_File Genome_File
```

## 6.2 Command lines of Minimap2 (version 2.14-r883)

Genome indexing:

```
minimap2 -ax splice -d ref.mmi ref.fa
```

Read alignment

```
minimap2 -ax splice -t 24 ref.mmi Fastq_File > aln.sam
```

## 6.3 Command lines of GMAP (version 2017-10-12)

Genome indexing:

```
gmap_build -d <genome> ref.fa
```

Read alignment

```
gmap -D Genome_Index_Dir -d Genome_Database -f samse -n 0 -t 24 --cross-species  
--max-intronlength 200000 -z sense_force Fastq_File > out.sam 2> out.sam.log
```

## References

- [1] Robinson JT et al. Integrative Genomics Viewer. *Nat Biotechnology* (2011), 29(1): 24–26.
- [2] Weirather JL et al. Comprehensive comparison of Pacific Biosciences and Oxford Nanopore Technologies and their applications to transcriptome analysis. *F1000Res* (2017), 6: 100.
- [3] Carneiro MO et al. Pacific biosciences sequencing technology for genotyping and variation discovery in human data. *BMC Genomics* (2012), 13:375.
- [4] Liu B, et al. deBGA: read alignment with de Bruijn graph-based seed and extension. *Bioinformatics* (2016), 32(21): 3224-3232.
- [5] Suzuki H and Masahiro K. Introducing difference recurrence relations for faster semi-global alignment of long sequences. *BMC Bioinformatics* (2018), 19(1): 45.