

1 Supplementary materials

2 Table of Contents

3

4 Tables.....2

5 *Table 1*2

6 *Table 2*3

7 *Table 3*6

8 Figures10

9 *Supplementary Materials Figure 1*.....10

10 *Supplementary Materials Figure 2*.....11

11 *Supplementary Materials Figure 3*.....12

12 *Supplementary Materials Figure 4*.....13

13 *Supplementary Materials Figure 5*.....14

14 *Supplementary Materials Figure 6*.....15

15 *Supplementary Materials Figure 7*.....16

16 *Supplementary Materials Figure 8*.....17

17 Code18

Tables

Table 1 Summary of the sequencing and analysis characteristics for the Jurkat and HEK293T cells.

	Jurkat		HEK293T	
	SS3X	scLIS-seq	SS3X	scLIS-seq
Sequencing platform	AVITI, PE150	PromethION, R10.4.1	AVITI, PE150	PromethION, R10.4.1
Total yield (M reads)	28.5	30.8*	190.4	52.8*
Average quality score (Q-score)	40.1	16.7	41.8	16
Downstream processing	zUMIs	wf-single-cell; scywalker	zUMIs	wf-single-cell; scywalker

*Reads with Q < 10 were removed during basecalling

24 **Table 2** The i5 and i7 indexes used for SS3X and scLIS-seq library preparation of the jurkat cells. The i5 and i7 indexes
 25 are part of the IDT for Illumina UD Nextera Indexes. For scLIS-seq, only the i5 index is incorporated during library
 26 preparation.

zUMIs_barcode_file	i7_base_sample_sheet (reverse compliment adapter)	i5_base (whitelist scLIS-seq)
TCTCATGATAAACACGTGGA	TCTCATGATA	AACACGTGGA
CGAGGCCAAGGTGTTACCGG	CGAGGCCAAG	GTGTTACCGG
TTCACGAGACAGATTGTTAC	TTCACGAGAC	AGATTGTTAC
GCGTGGATGGTTGACCAATG	GCGTGGATGG	TTGACCAATG
TCCTGGTTGTCTGACCGGCA	TCCTGGTTGT	CTGACCGGCA
TAATTCTGCTTCTCATCAAT	TAATTCTGCT	TCTCATCAAT
CGCACGACTGGGACCAACAG	CGCACGACTG	GGACCAACAG
GAGGTTAGACAATGTATTGC	GAGGTTAGAC	AATGTATTGC
AACCGAGTTCGATCTCTGGA	AACCGAGTTC	GATCTCTGGA
TGTGATAACTCAGGCGCCAT	TGTGATAACT	CAGGCGCCAT
AGTATGCTACTTAATAGACC	AGTATGCTAC	TTAATAGACC
GTAAGTGAAGGGAGTCGCGA	GTAAGTGAAG	GGAGTCGCGA
TCCTCGGACTAACGCCAGAG	TCCTCGGACT	AACGCCAGAG
CTGGAAGTGTGTAATTAAC	CTGGAAGTGT	CGTAATTAAC
GAATATGCGGACGAGACTGA	GAATATGCGG	ACGAGACTGA
GATCGGATAAGTATCGGCCG	GATCGGATAA	GTATCGGCCG
GCTAGACTATAATACGACAT	GCTAGACTAT	AATACGACAT
AGCTACTATAGTTATATGGC	AGCTACTATA	GTTATATGGC
CCACCGGAGTGCCTGCCATG	CCACCGGAGT	GCCTGCCATG
CTTACCGCACTAAGACCTAT	CTTACCGCAC	TAAGACCTAT
TTAGGATATCTATACCATGG	TTAGGATATC	TATACCATGG
TTATACGCGAGCCGTCTGTT	TTATACGCGA	GCCGTCTGTT
CGCTTAGAATCAGAGTGATA	CGCTTAGAAT	CAGAGTGATA
CCGAAGCGCTTGCTAACTAT	CCGAAGCGCT	TGCTAACTAT
CACTATCAACTCAGTTAATG	CACTATCAAC	TCAGTTAATG
TTGCTCTATTGTGACCTTGA	TTGCTCTATT	GTGACCTTGA
TTACAGTTAGACATGCATAT	TTACAGTTAG	ACATGCATAT
CTAAGTACGCAACATACCTA	CTAAGTACGC	AACATACCTA
TAGTTCGGTACCATGTGTAG	TAGTTCGGTA	CCATGTGTAG
CTATTACTACGAGTCTCTCC	CTATTACTAC	GAGTCTCTCC
TAGCATAACCGCTATGCGCA	TAGCATAACC	GCTATGCGCA
ACTCTATTGTATCGCATATG	ACTCTATTGT	ATCGCATATG
TAGTGGAAGCAGTACCTATA	TAGTGGAAGC	AGTACCTATA
CGCCATATCTGACCGGAGAT	CGCCATATCT	GACCGGAGAT
GCTTCATATTCGTTACGCCT	GCTTCATATT	CGTTACGCCT
ACTAGCGCTATTACTTCCTC	ACTAGCGCTA	TTACTTCCTC
GCTCTTAACTCACGTCCACC	GCTCTTAACT	CACGTCCACC
GTGGTATCTGGCTACTATCT	GTGGTATCTG	GCTACTATCT

TGACGGCCGTAGTCAACCAT	TGACGGCCGT	AGTCAACCAT
CAGTAATTACCGAGGCGGTA	CAGTAATTAC	CGAGGCGGTA
TACAAGACTTCAGGTGTTCA	TACAAGACTT	CAGGTGTTCA
CTGTGGTGACGACAGACAGG	CTGTGGTGAC	GACAGACAGG
CTCCACTAATTGTACTTGTT	CTCCACTAAT	TGTACTTGTT
ATAGTTAGCACTCTAAGTAG	ATAGTTAGCA	CTCTAAGTAG
ATAGGTCTTAGTCACCACAG	ATAGGTCTTA	GTCACCACAG
TTCTTAACCATCTACATACC	TTCTTAACCA	TCTACATACC
AAGGAAGAGTCACGTTAGGC	AAGGAAGAGT	CACGTTAGGC
GGAAGGAGACTGGTGAGTCT	GGAAGGAGAC	TGGTGAGTCT
TGAACGCGGACTTCGAAGGA	TGAACGCGGA	CTTCGAAGGA
CCTGCAACCTGTAGAGTCAG	CCTGCAACCT	GTAGAGTCAG
TTCATGGTTCGACATTGTCA	TTCATGGTTC	GACATTGTCA
ATCCTCTCAATCCGCAAGGC	ATCCTCTCAA	TCCGCAAGGC
CACTAGACCAACTGCCTTAT	CACTAGACCA	ACTGCCTTAT
ATTATCCACTTACGCACGTA	ATTATCCACT	TACGCACGTA
ATGGCGTGCCCGCTTGAAGT	ATGGCGTGCC	CGCTTGAAGT
TCCAGAGATCCTGCACTTCA	TCCAGAGATC	CTGCACTTCA
ATGTCCAGCACAGCGGACAA	ATGTCCAGCA	CAGCGGACAA
CAACGTTCCGGGATCCGCAT	CAACGTTCCG	GGATCCGCAT
GCGTATTAATTGCGGTGTTG	GCGTATTAAT	TGCGGTGTTG
CCGAATCTGGATGAATCAAG	CCGAATCTGG	ATGAATCAAG
TCTCAATACCGACGTTGCGG	TCTCAATACC	GACGTTGCGG
AAGCATCTTGCATTCAACAA	AAGCATCTTG	CATTCAACAA
TCAGTCTCGTCACGGATTAT	TCAGTCTCGT	CACGGATTAT
TGCAAGATAATTGAGGACGG	TGCAAGATAA	TTGAGGACGG
GTAACAATCTCTGTATAC	GTAACAATCT	CTCTGTATAC
CCATGGTATATCTCGCGGAG	CCATGGTATA	TCTCGCGGAG
TCATACCGTTGGTAACGCAG	TCATACCGTT	GGTAACGCAG
GGCGCCATTGACCGCGCAAT	GGCGCCATTG	ACCGCGCAAT
AGCGAATTAGAGCCGGAACA	AGCGAATTAG	AGCCGGAACA
TTAGACCATGTCCTAGGAAG	TTAGACCATG	TCCTAGGAAG
CACACAGTATTTGAGCCTAA	CACACAGTAT	TTGAGCCTAA
TCTTGTGCGGCCACCTGTGT	TCTTGTGCGC	CCACCTGTGT
TACCGCCTCGCCTCGCAACC	TACCGCCTCG	CCTCGCAACC
CTGTTATATCGTATAGCTGT	CTGTTATATC	GTATAGCTGT
TAACCGGCGAGCTACATTAG	TAACCGGCGA	GCTACATTAG
AAGAGAGTCTTACGAATCTT	AAGAGAGTCT	TACGAATCTT
GTAGGCGAGCTAGGAGCGCA	GTAGGCGAGC	TAGGAGCGCA
AACTTATCCTGTACTGGCGT	AACTTATCCT	GTAAGTGGCGT
ATTATGTCTCAGTTAAGAGC	ATTATGTCTC	AGTTAAGAGC
TATAACAGCTTCGCGTATAA	TATAACAGCT	TCGCGTATAA
CCAATGATACGAGTGTGCCG	CCAATGATAC	GAGTGTGCCG

GAGGCCTATTCTAGTCCGGA	GAGGCCTATT	CTAGTCCGGA
AGCTAAGCGGATTAATACGC	AGCTAAGCGG	ATTAATACGC
CTTCCTAGGACCTAGAGTAT	CTTCCTAGGA	CCTAGAGTAT
CGATCTGTGATAGGAAGACT	CGATCTGTGA	TAGGAAGACT
GTGGACAAGTCCGTGGCCTT	GTGGACAAGT	CCGTGGCCTT
AACAAGTACAGGATATATCC	AACAAGTACA	GGATATATCC
AGATTAAGTGACACCTCTTGG	AGATTAAGTG	CACCTCTTGG
TATCACTCTGAACGTTACAT	TATCACTCTG	AACGTTACAT
AGAATTCGCCCCGGCAAGCTC	AGAATTCGCC	CGGCAAGCTC
CCTGACCACTTCTTGGCTAT	CCTGACCACT	TCTTGGCTAT
AGCTGGAATGACGGAATGCG	AGCTGGAATG	ACGGAATGCG
TGATAACGAGGTTCCGCAGG	TGATAACGAG	GTTCCGCAGG
CATAGTAAGGACCAAGTTAC	CATAGTAAGG	ACCAAGTTAC
ATTGGCTTCTTGGCTCGCAG	ATTGGCTTCT	TGGCTCGCAG
GTACCGATTAACTAACGTT	GTACCGATTA	AACTAACGTT

27

28

29 **Table 3** The i5 and i7 indexes used for SS3X and scLIS-seq library preparation of the HEK293T cells. The i5 and i7
30 indexes are part of the IDT for Illumina UD Nextera Indexes. For scLIS-seq, only the i5 index is incorporated during
31 library preparation.

zUMIs_barcode_file	i7_base_sample_sheet (reverse compliment adapter)	i5_base (whitelist scLIS-seq)
CTAGTGCTCTTGGAAACAGTA	CTAGTGCTCT	TGGAAACAGTA
TCCGTTATGTTTGTAGTGTA	TCCGTTATGT	TTGTAGTGTA
TCCATTGCCGGAATGCACGA	TCCATTGCCG	GAATGCACGA
ATTAACAAGGTGTTGTTTCGT	ATTAACAAGG	TGTTGTTTCGT
ACGCCTTGTTTAAGGAACGT	ACGCCTTGTT	TAAGGAACGT
TCCGCCAATTCTCCGTGCTG	TCCGCCAATT	CTCCGTGCTG
GTGACACGCAGCTTACGGAC	GTGACACGCA	GCTTACGGAC
GCTGACGTTGTGTTTCGCATT	GCTGACGTTG	TGTTTCGCATT
CAATCGGCTGGCTGTAGGAA	CAATCGGCTG	GCTGTAGGAA
TGGCCTCTGTTGAGACAGCG	TGGCCTCTGT	TGAGACAGCG
GTCCTGGATACGGTGGCGAA	GTCCTGGATA	CGGTGGCGAA
GTTGGATGAATTCCAGGTAA	GTTGGATGAA	TTCCAGGTAA
CTGAGGAATACTTAACCACT	CTGAGGAATA	CTTAACCACT
TTGTATCAGGACAGAGGCCA	TTGTATCAGG	ACAGAGGCCA
ATTGCGCGGTTTGGCTCCGC	ATTGCGCGGT	TTGGCTCCGC
TCCGGACTAGGACGAACAAT	TCCGGACTAG	GACGAACAAT
CTTGTACACCAAGCGCGCTT	CTTGTACACC	AAGCGCGCTT
AGCCTATGATCACCACCTAC	AGCCTATGAT	CACCACCTAC
ACGGTCAGGACGGCCTCGTT	ACGGTCAGGA	CGGCCTCGTT
GGACCAGTGGCGCGAGCCTA	GGACCAGTGG	CGCGAGCCTA
CCTCTACATGGATACCTCCT	CCTCTACATG	GATACCTCCT
TTCCTACAGCGGCGAATTCT	TTCCTACAGC	GGCGAATTCT
GTATTCCACCATGTAGACAA	GTATTCCACC	ATGTAGACAA
TTGAATGTTGCGCAATCTAG	TTGAATGTTG	CGCAATCTAG
CGCACGACTGGGACCAACAG	CGCACGACTG	GGACCAACAG
TGCAATGAATCCAGAAGTAA	TGCAATGAAT	CCAGAAGTAA
GAATATGCGGACGAGACTGA	GAATATGCGG	ACGAGACTGA
GGAGAGACTCTTGAAGCAGA	GGAGAGACTC	TTGAAGCAGA
CGCTTAGAATCAGAGTGATA	CGCTTAGAAT	CAGAGTGATA
TTCTCGTGCAATACACAGAG	TTCTCGTGCA	ATACACAGAG
TAGCATAACCGCTATGCGCA	TAGCATAACC	GCTATGCGCA
CAACGACCTAGCGACTCGAT	CAACGACCTA	GCGACTCGAT
TGACGGCCGTAGTCAACCAT	TGACGGCCGT	AGTCAACCAT
TTGAATCCAAGTCACGGTGT	TTGAATCCAA	GTCACGGTGT
AAGGAAGAGTCACGTTAGGC	AAGGAAGAGT	CACGTTAGGC
ATGATTCCGGCTCTGACGTG	ATGATTCCGG	CTCTGACGTG
ATGGCGTGCCCGCTTGAAGT	ATGGCGTGCC	CGCTTGAAGT
TCTTACGCCGATTCATTGCA	TCTTACGCCG	ATTCATTGCA

TCAGTCTCGTCACGGATTAT	TCAGTCTCGT	CACGGATTAT
ACTTACTTCATCTGGAATTA	ACTTACTTCA	TCTGGAATTA
CACACAGTATTTGAGCCTAA	CACACAGTAT	TTGAGCCTAA
TCGTTGCTGCGTCTCGTGAA	TCGTTGCTGC	GTCTCGTGAA
ATTATGTCTCAGTTAAGAGC	ATTATGTCTC	AGTTAAGAGC
AGGTATGGCGGTGGCGAGAC	AGGTATGGCG	GTGGCGAGAC
AACAAGTACAGGATATATCC	AACAAGTACA	GGATATATCC
TCTATTCAGTTCTCAGTACA	TCTATTCAGT	TCTCAGTACA
ATTGGCTTCTTGGCTCGCAG	ATTGGCTTCT	TGGCTCGCAG
TATGAACTTGAATCCTTAGG	TATGAACTTG	AATCCTTAGG
GATCAAGGCACCTTGTTAAT	GATCAAGGCA	CCTTGTTAAT
GGTCTATTAACCACGACACG	GGTCTATTAA	CCACGACACG
CGGTTACGGCAAGACTATAG	CGGTTACGGC	AAGACTATAG
TGATGGCTACCCAACAACAT	TGATGGCTAC	CCAACAACAT
TTCTACATACCTAACTGTAA	TTCTACATAC	CTAACTGTAA
AGGACAGGCCCATTCAGCT	AGGACAGGCC	CATTCAGCT
ACAGTGTATGGAACATACGG	ACAGTGTATG	GAACATACGG
CTAATAACCGAACCGCATCG	CTAATAACCG	AACCGCATCG
TATGTAGTCACGCACTAATG	TATGTAGTCA	CGCACTAATG
CCAGGCACCAACGCTAATTA	CCAGGCACCA	ACGCTAATTA
CAGTGGCACTTAACAATAGG	CAGTGGCACT	TAACAATAGG
AATCCAATTGCATGTAGAGG	AATCCAATTG	CATGTAGAGG
AACGCACGAGGCCTCGGATA	AACGCACGAG	GCCTCGGATA
CACAGCGGTCAATCCTATTG	CACAGCGGTC	ATTCCTATTG
TGGCGCGAACAACCTGATACT	TGGCGCGAAC	AACTGATACT
GTGACGGAGCTGGCGGTCCA	GTGACGGAGC	TGGCGGTCCA
ACACAGGTGGTGAGCGTTGT	ACACAGGTGG	TGAGCGTTGT
TATGACAATCTTGCTTGAT	TATGACAATC	TTGCTTGAT
GGCAAGCCAGCAAGCATCCG	GGCAAGCCAG	CAAGCATCCG
CCTTCTAACAGATAAGCTCT	CCTTCTAACA	GATAAGCTCT
GGAGCGTGTAATCCGTAAGT	GGAGCGTGTA	ATCCGTAAGT
ATCCAGGTATCAACGTCAGC	ATCCAGGTAT	CAACGTCAGC
CCTCCGTCCACACATCGGTG	CCTCCGTCCA	CACATCGGTG
CCGCTCCGTTTACGGCGAAG	CCGCTAACAC	GAGTTGTAAT
GAGGTTAGACAATGTATTGC	GAGGTTAGAC	AATGTATTGC
CTATGAAGGACTTATACCTG	CTATGAAGGA	CTTATACCTG
GATCGGATAAGTATCGGCCG	GATCGGATAA	GTATCGGCCG
CCGCTCCGTTTACGGCGAAG	CCGCTCCGTT	TACGGCGAAG
CCGAAGCGCTTGCTAACTAT	CCGAAGCGCT	TGCTAACTAT
GCTAGGAAGTTCTCGGACGA	GCTAGGAAGT	TCTCGGACGA
ACTCTATTGTATCGCATATG	ACTCTATTGT	ATCGCATATG
CGGTCGGCATCTAGGCAAGG	CGGTCGGCAT	CTAGGCAAGG
CAGTAATTACCGAGGCGGTA	CAGTAATTAC	CGAGGCGGTA

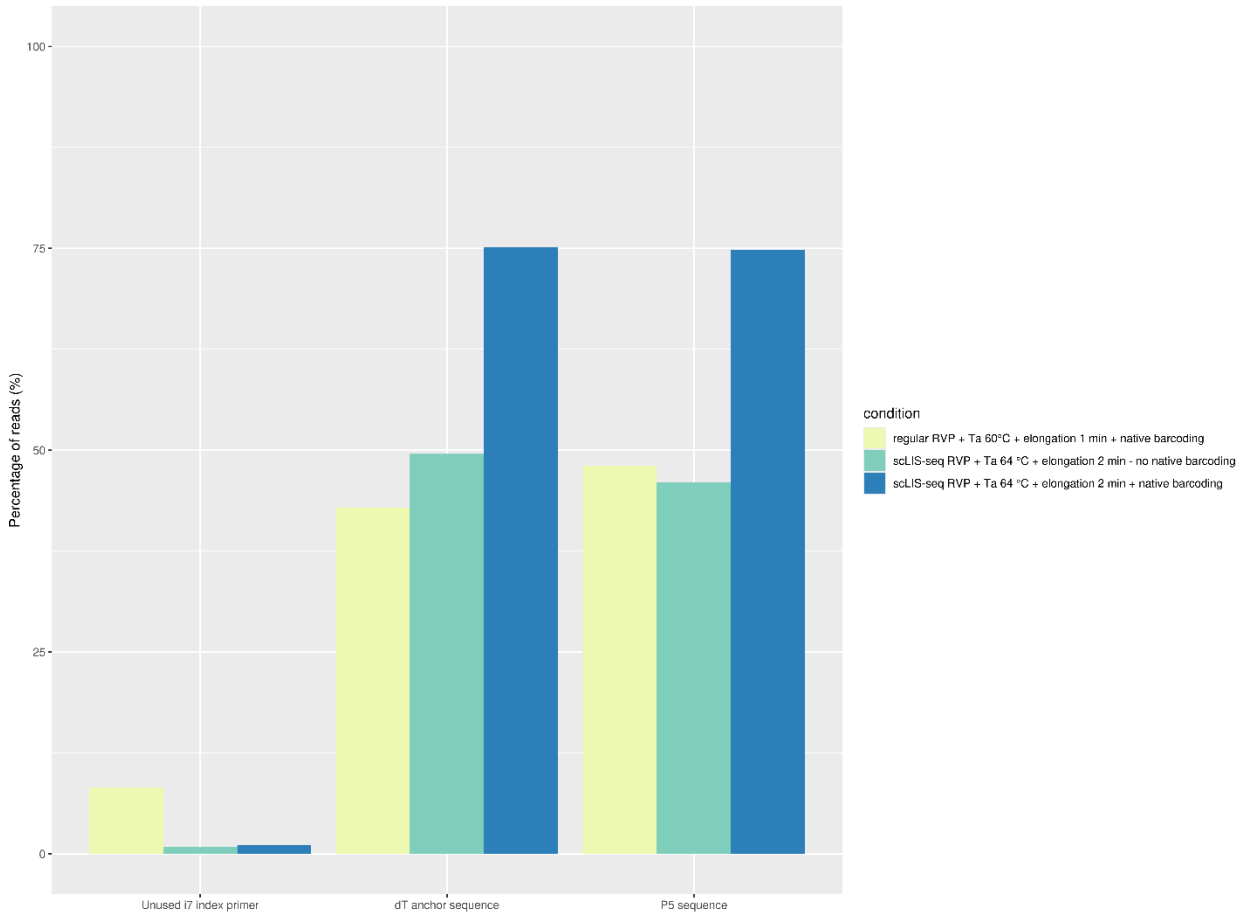
CACGGCTAGTGGTGTACAAG	CACGGCTAGT	GGTGTACAAG
GGAAGGAGACTGGTGAGTCT	GGAAGGAGAC	TGGTGAGTCT
TTAGGCTCAATCGAATGGAA	TTAGGCTCAA	TCGAATGGAA
TCCAGAGATCCTGCACTTCA	TCCAGAGATC	CTGCACTTCA
AGCTGATGTCTCATGTCCTG	AGCTGATGTC	TCATGTCCTG
TGCAAGATAATTGAGGACGG	TGCAAGATAA	TTGAGGACGG
CGCATTCCGTATTAGTGAG	CGCATTCCGT	ATTAGTGAG
TCTTGTCGGCCACCTGTGT	TCTTGTCGGC	CCACCTGTGT
CGACAAGGATAAGGCCACCT	CGACAAGGAT	AAGGCCACCT
TATAACAGCTTCGCGTATAA	TATAACAGCT	TCGCGTATAA
TCCAGCCTGCGCCAGATCCA	TCCAGCCTGC	GCCAGATCCA
AGATTAAGTGACCTCTTGG	AGATTAAGTG	CACCTCTTGG
CAAGACGTCCAGTATACGGA	CAAGACGTCC	AGTATACGGA
GTACCGATTAACTAACGTT	GTACCGATTA	AACTAACGTT
AGTGGTTAAGGGTGGAATAC	AGTGGTTAAG	GGTGGAATAC
AACATCGCGCTCCACTTGT	AACATCGCGC	GTCCACTTGT
ATGGCGCCTGTCCACGGCCT	ATGGCGCCTG	TCCACGGCCT
GAAGCGGCACCGGCTCTACT	GAAGCGGCAC	CGGCTCTACT
TAATTAGCGTTGGTTAAGAA	TAATTAGCGT	TGGTTAAGAA
TATTGCGCTCCTGTGTTAGG	TATTGCGCTC	CTGTGTTAGG
ATACTTGTTCTAACCGCCGA	ATACTTGTTT	TAACCGCCGA
CCACCAGGCACTTATGGAAT	CCACCAGGCA	CTTATGGAAT
AGGAGGTATCCCAGTTGGCA	AGGAGGTATC	CCAGTTGGCA
GTGCAGACAGTACCATCCGT	GTGCAGACAG	TACCATCCGT
TGGTGTTATGTCCAATTCTA	TGGTGTTATG	TCCAATTCTA
AGATCCATTAATCTCTACCA	AGATCCATTA	ATCTCTACCA
ACAACACTGAGTTATCACA	ACAACACTG	AGTTATCACA
GCATAAGCTTTTCCGTCGCA	GCATAAGCTT	TTCCGTCGCA
TAGAGAATACTATAGATTCTG	TAGAGAATAC	TATAGATTCTG
AACACTGTTAGCAAGTCTCA	AACACTGTTA	GCAAGTCTCA
CGATGCGGTTCTGCTTGGT	CGATGCGGTT	CCTGCTTGGT
GTATTGACGTGGAATTCCAA	GTATTGACGT	GGAATTCCAA
CATATGCGATCCTTGAACGG	CATATGCGAT	CCTTGAACGG
CATTAGTGCGTGACTACATA	CATTAGTGCG	TGACTACATA
ATTGTTTCGTCTCTCTAGATT	ATTGTTTCGT	TCTCTAGATT
TTGGAATTCCACGTCAATAC	TTGGAATTCC	ACGTCAATAC
AATTCTTGGATTGTCAACTT	AATTCTTGG	TTGTCAACTT
TCATAGATTGGATTAAGGTG	TCATAGATTG	GATTAAGGTG
CACCTTAATCTTAGGATAGA	CACCTTAATC	TTAGGATAGA
TAATTCTGCTTCTCATCAAT	TAATTCTGCT	TCTCATCAAT
AAGATACACGTGTGCTAACA	AAGATACACG	TGTGCTAACA
CTGGAAGTGTGTAATTAAC	CTGGAAGTGT	CGTAATTAAC
ATTGAGGTCCCGATTATAT	ATTGAGGTCC	CGGATTATAT

TTATACGCGAGCCGTCTGTT	TTATACGCGA	GCCGTCTGTT
CACCTCTCTTCTCGACTCCT	CACCTCTCTT	CTCGACTCCT
CTATTACTACGAGTCTCTCC	CTATTACTAC	GAGTCTCTCC
TGTGTTAGTATACTCTGTTA	TGTGTTAGTA	TACTCTGTTA
GTGGTATCTGGCTACTATCT	GTGGTATCTG	GCTACTATCT
TGGATCTGGCCAATGCTGAA	TGGATCTGGC	CAATGCTGAA
TTCTTAACCATCTACATACC	TTCTTAACCA	TCTACATACC
AATGACTGGTTCTCACGCGT	AATGACTGGT	TCTCACGCGT
ATTATCCACTTACGCACGTA	ATTATCCACT	TACGCACGTA
CGAACGCACCATGACAGAAC	CGAACGCACC	ATGACAGAAC
AAGCATCTTGCAATTCAACAA	AAGCATCTTG	CATTCAACAA
GCCTAACGTGAGCTCTCAAG	GCCTAACGTG	AGCTCTCAAG
TTAGACCATGTCCTAGGAAG	TTAGACCATG	TCCTAGGAAG
TTACGCACCTGCTCTCGTTG	TTACGCACCT	GCTCTCGTTG
AACTTATCCTGTACTGGCGT	AACTTATCCT	GTACTGGCGT
TACTCTGCGCCCTTACTATG	TACTCTGCGC	CCTTACTATG
GTGGACAAGTCCGTGGCCTT	GTGGACAAGT	CCGTGGCCTT
CTTCAACCACTTATGCGCCT	CTTCAACCAC	TTATGCGCCT
CATAGTAAGGACCAAGTTAC	CATAGTAAGG	ACCAAGTTAC
TCAACCACGACAAGTTCATA	TCAACCACGA	CAAGTTCATA

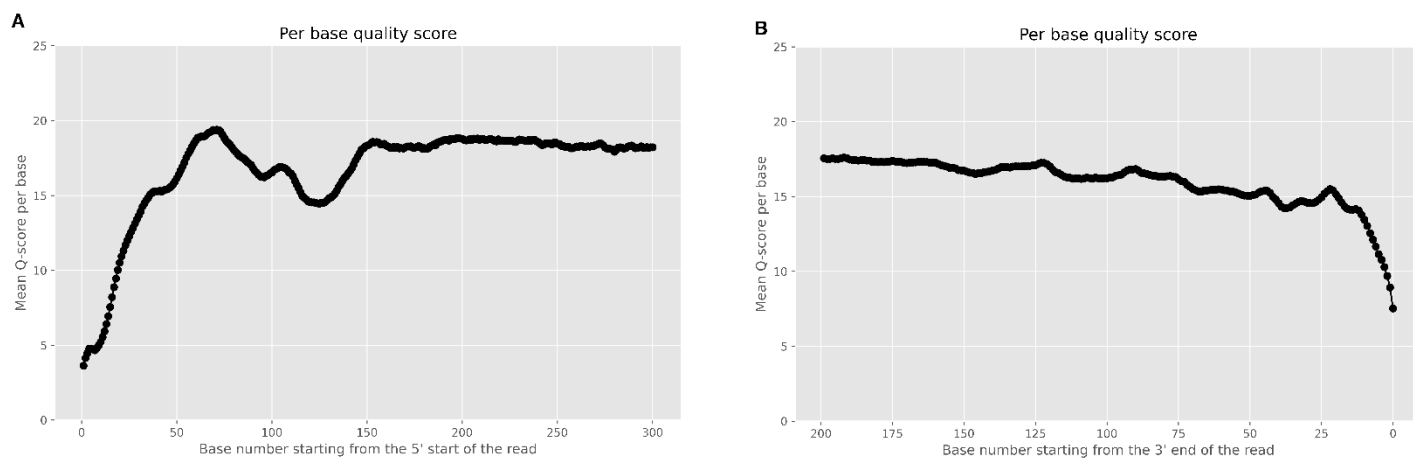
32

33

Figures



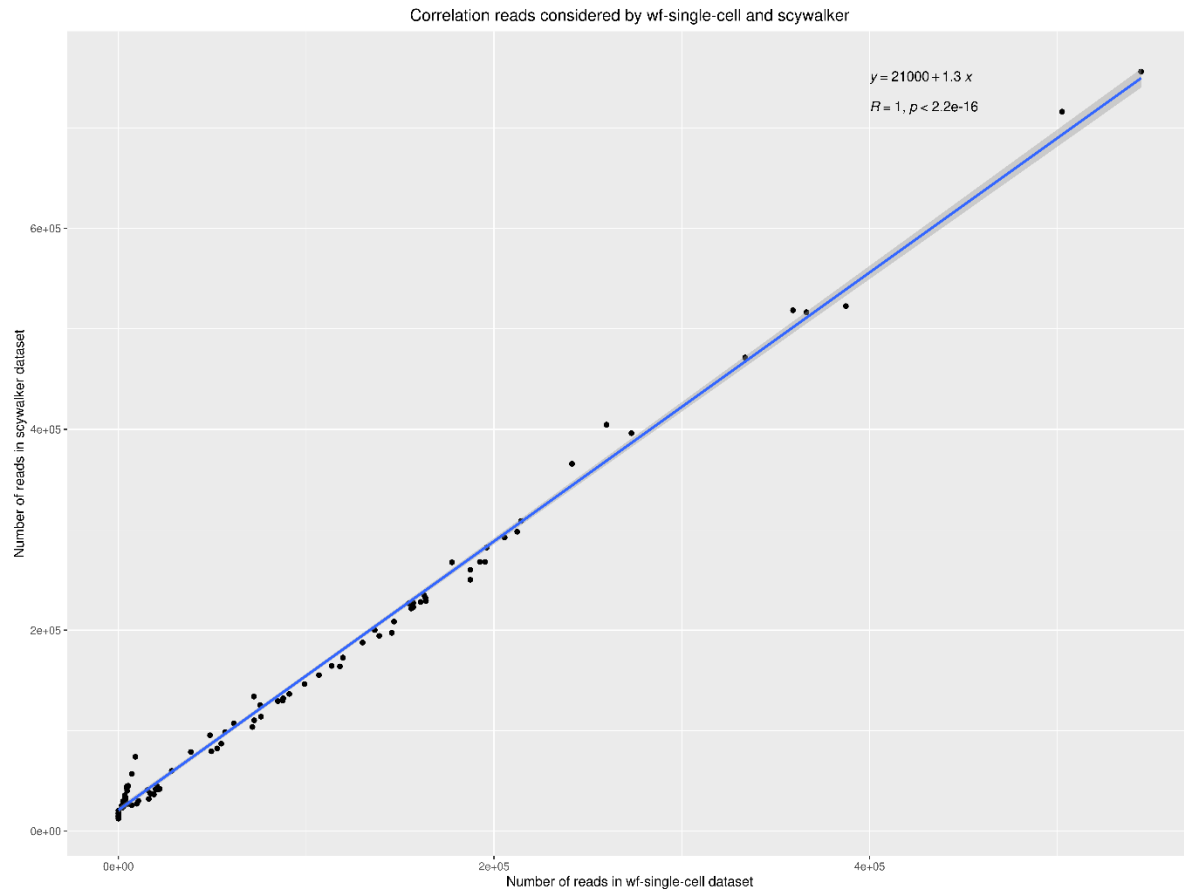
Supplementary Materials Figure 1 Summary of the optimizations performed to reduce the percentage of reads originating from the unused i7 index primer, increase the proportion of reads having detectable oligodT anchor sequences indicative for the end of the transcript and increase the proportion of reads having detectable P5 sequences, indicative for the 5' end of the transcript. Regular RVP: 5'-GCATCAGCAGCATACGA-3', Ta: annealing temperature. The yellow condition was performed on single cell equivalents (10 pg) of priorly extracted RNA from the Loucy cell line. The other conditions were performed on the cDNA of Jurkat cells. The green condition was performed on a R10.4.1. Flongle flow cell, using the LSK-SQK114 kit.



44

45 **Supplementary Materials Figure 2** Illustration of the drop in per base quality at the beginning and end of the ONT
 46 reads. The data from the Jurkat cell line is shown.

47

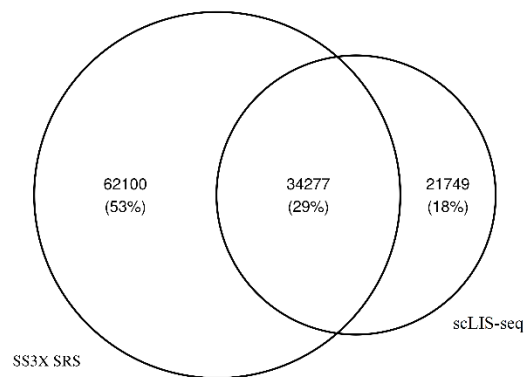


48

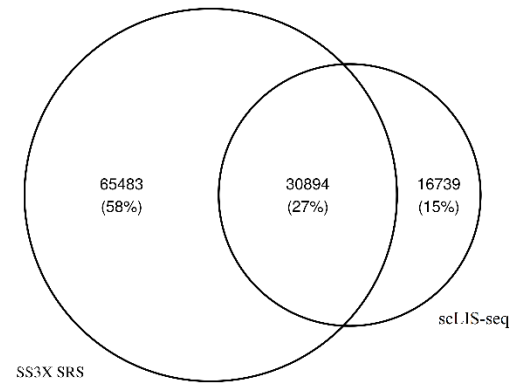
49 **Supplementary Materials Figure 3** Correlation of the number of reads considered by the adapted wf-single-cell
 50 workflow vs. the reads considered by the scywalker workflow for the HEK293T cells.

51

A



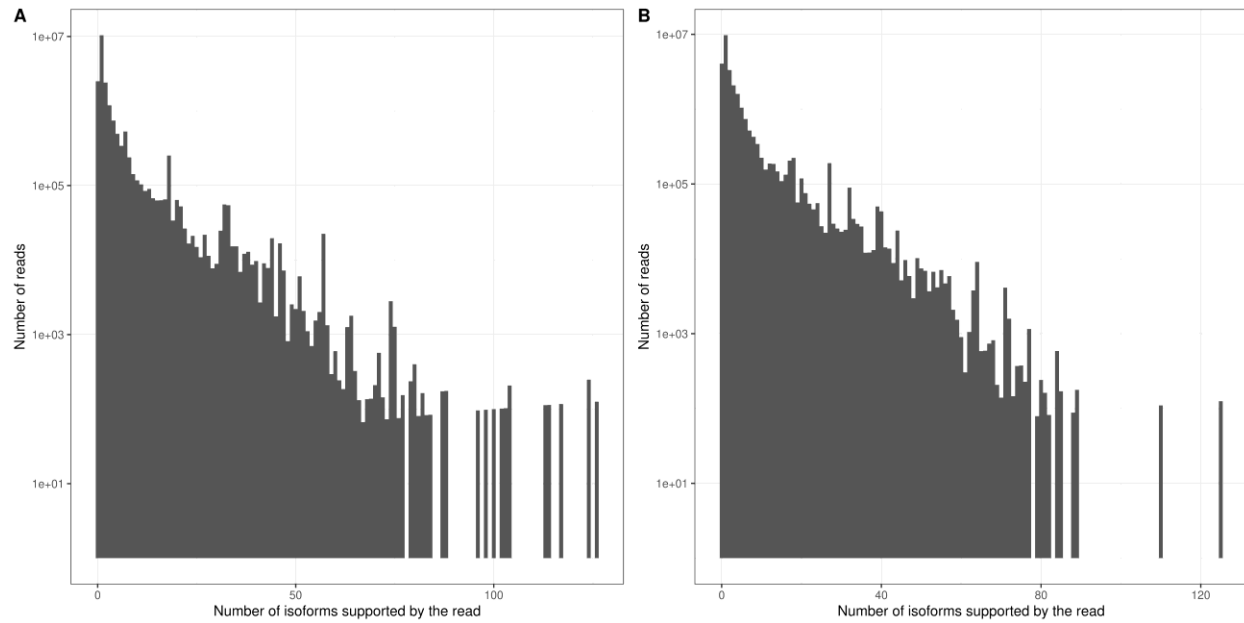
B



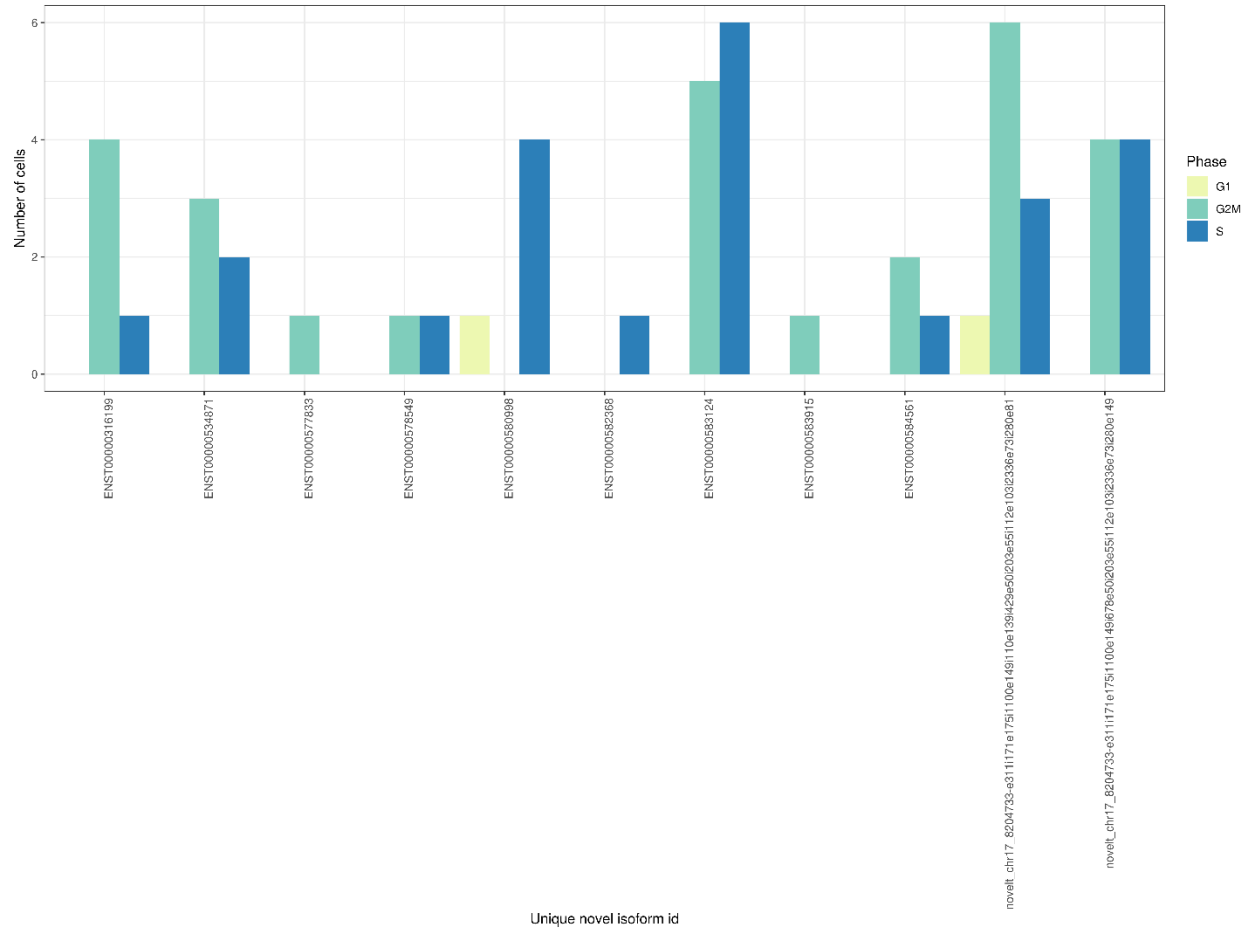
52

53 **Supplementary Materials Figure 4** Evaluation of the overlap in chromosome-CB-UMI combinations between SS3X
 54 and scLIS-seq, as analyzed using wf-single-cell. **A.** Overlapping combination for Jurkat cell 'GTATAGCTGT'. **B.**
 55 Overlapping combinations for the same cell but only considering UMI-sequences in which every base has at least a Q-
 56 score of 30.

57



Supplementary Materials Figure 5 Histograms illustrating the number of reads with valid CB and UMI in relation to their respective compatible isoforms. **A.** Jurkat cell line; **B.** HEK293T cell line.



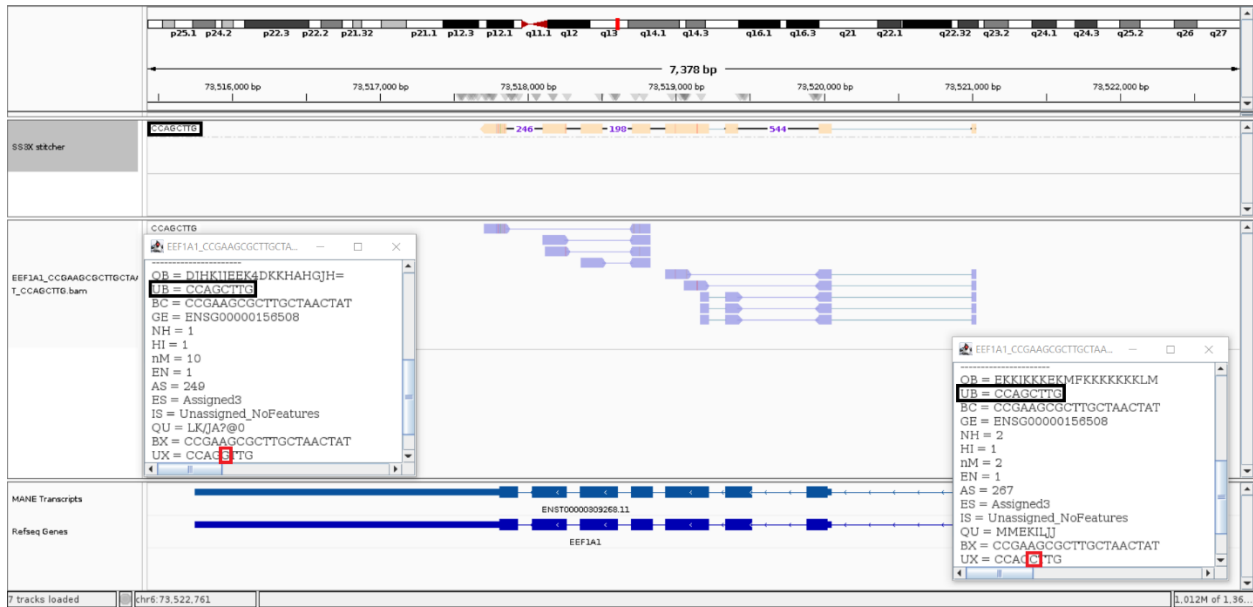
62

63 **Supplementary Materials Figure 6** Cell cycle dependent expression of isoforms annotated as novel by scywalker

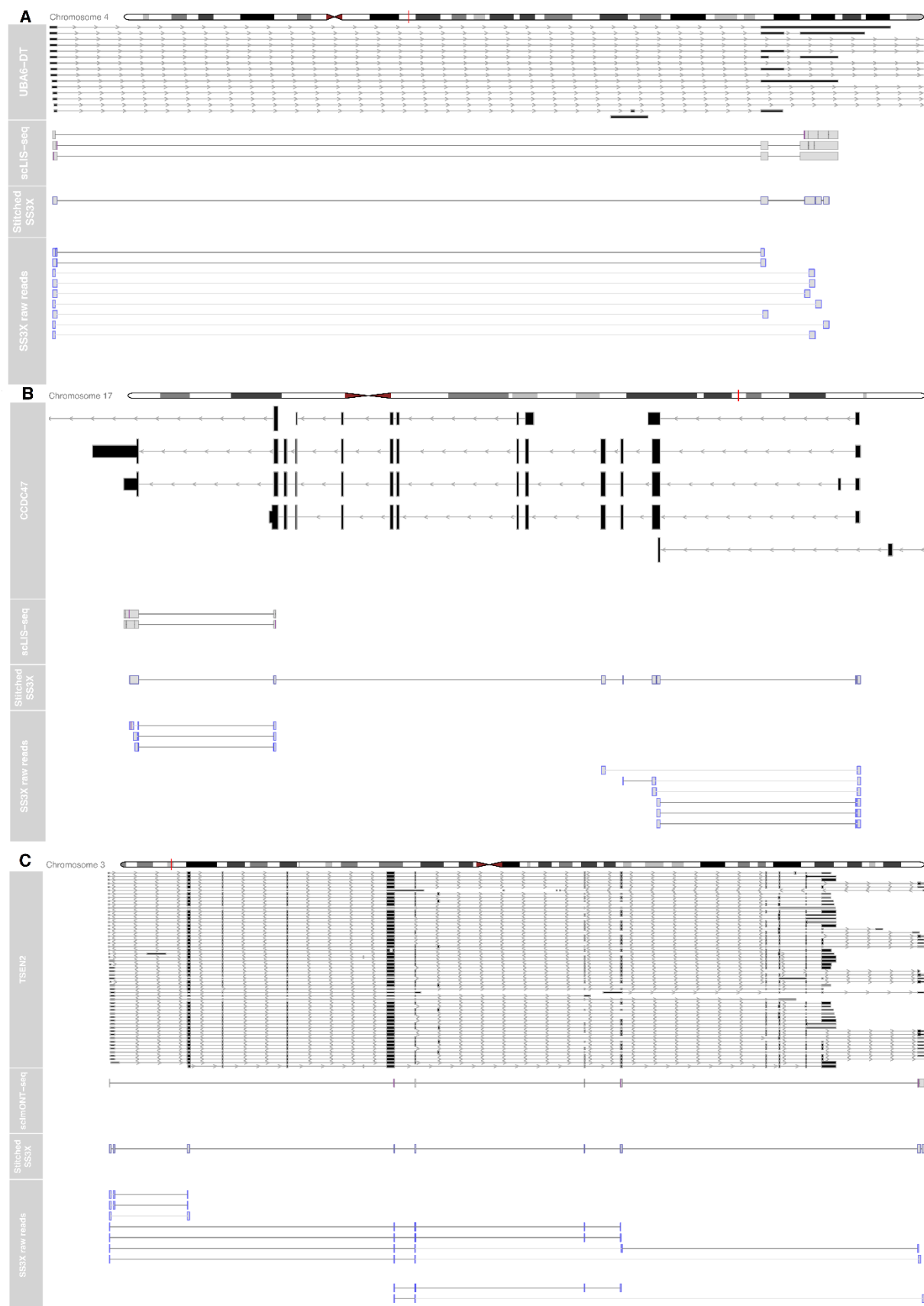
64 for the AURKB gene.

65

66



Supplementary Materials Figure 7 Illustration of false UMI-stitching of the SS3X Jurkat data. **Top:** reconstructed transcript as generated by *stitcher.py*. **Middle:** Corresponding short-reads with identical CB, UB, and GE, which contributed to the reconstructed transcript. Due to UMI-correction in zUMIs based on Hamming distance, the two probably unrelated short-read groups with different 5' alignment start positions (blue), are assigned with an identical UB-tag. Their uncorrected UMI-sequence is different (red boxes). These two short-reads groups were assumed as one during the reconstruction of the transcript, and can in that way result in the false reconstructions.



Supplementary Materials Figure 8 Graphical comparison of the identical transcripts, with identical UMI and CB, obtained using scLIS-seq, SS3X UMI-stitching reconstructions, and the SS3X short-reads used for stitching. In each subpanel, from top to bottom: UCSC derived reference gene isoforms; scLIS-seq alignment; reconstructed transcript based on UMI-stitching of the SS3X short-reads; SS3X short reads. **A.** Comparison for the *UBA6-DT* gene, illustrating successful stitching of the short-reads and good overlap with the scLIS-seq. **B.** Comparison for the *CCDC47* gene, illustrating false stitching between the cluster of the three top alignments and the bottom read cluster. **C.** Comparison for *TSEN2* gene, illustrating false stitching between the cluster of the two bottom alignments and the first alignment cluster.

Code

For wf-single-cell

In the `~/nextflow/assets/epi2me-labs/wf-single-cell/` directory:

- The following changes were performed to `kit_configs.csv`:

5prime v1 16 8 UDI.txt.gz

*# in which UDI.txt.gz contains the i5 indices used, precedent by the sequence
'CTACAC' to make them 16 nucleotides long. The UDI.txt.gz file is placed in
~/nextflow/assets/epi2me-labs/wf-single-cell/data/ directory.*

- The following changes were performed to `~/nextflow/assets/epi2me-labs/wf-single-cell/bin/workflow_glue/sc_util.py`:

```
'5prime': {  
    'adapter1': 'AATGATACGGCGACCACCGAGAT',  
    'adapter2': 'TCGTATGCTGCTGATGC'  
}
```

*#with adapter 1 partial P5 adapter sequence and adapter 2 scLIS-seq reverse primer
sequence*

- The following changes were performed to `~/nextflow/assets/epi2me-labs/wf-single-cell/bin/workflow_glue/ adapter_scan_vsearch.py`

line 135: `--minseqlength 16`

```

102         line 136: --minwordmatches 10
103
104     • The following changes were performed to ~/nextflow/assets/epi2me-labs/wf-single-
105       cell/bin/workflow_glue/ extract_barcode.py
106
107       # From line 272 on:
108
109       #Removed:
110
111         barcode = query_alignment[bc_start_pos: bc_start_pos + barcode_length]
112
113         umi = query_alignment[
114
115         bc_start_pos + barcode_length: bc_start_pos + barcode_length + umi_length]
116
117       #Added:
118
119         barcode_marker = "CCACCGAGAT"
120
121         umi_marker = "ATTGCGCAATG"
122
123
124
125         barcode_start = p_alignment.traceback.query.find(barcode_marker) +
126         len(barcode_marker)
127
128         umi_start = p_alignment.traceback.query.find(umi_marker) + len(umi_marker)
129
130
131
132         barcode = p_alignment.traceback.query[barcode_start: barcode_start +
133         barcode_length]
134
135         umi = p_alignment.traceback.query[umi_start: umi_start + umi_length]
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200

```

```

123      #Line 363-370: adjust
124      probe_seq = "{a1}{bc}{a2}(9){umi}".format(
125          a1=adapter1_probe_seq,
126          bc="N" * args.barcode_length,
127          a2="TCGTCGGCAGCGTCAGATGTGTATA",
128          tso="AGAGACAGATTGCGCAATG",
129          umi="N" * args.barcode_length
130      )
131      For scywalker
132      In cg_sc_barcode.tcl:
133
134      set umi_start [find_with_mismatch $seq "ATTGCGCAATG"]
135      if {$umi_start != -1} {
136          set umi_start [expr {$umi_start + 11}]
137          set umi [string range $seq $umi_start [expr {$umi_start + $umisize - 1}]]
138      } else {
139          set umi ""
140
141

```

```
142 For profiling TSO-mediated strand-invasion in SQANTI3

143 In SQANTI3_5.3.6/src/classification_classifiers.py, lines 30-37 were replaced by

144     if trec.strand == "+":

145         pos_TTS = trec.exonStarts[0]

146         seq_downTTS = str(genome_dict[trec.chrom].seq[pos_TTS-nPolyA:pos_TTS]).upper()

147     else: # id on - strand

148         pos_TTS = trec.exonEnds[-1]

149         seq_downTTS =

150 str(genome_dict[trec.chrom].seq[pos_TTS:pos_TTS+nPolyA].reverse_complement()).upper()

151     sequences = ['ATGGG', 'TAGGG', 'AAGGG', 'TTGGG'] # or for 10X: change to 'ATGGG'

152     percA = sum(seq_downTTS.count(seq) for seq in sequences) / nPolyA * 100

153 # For compatibility with the rest of the sqanti3_qc.py code, the name 'percA' was kept as

154 variable.

155
```