# Supplementary Methods

### Cell culture and isolation of total RNA

Seven lung adenocarcinoma cell lines (PC-7, PC-9, H1975, H2228, VMRC-LCD, LC2/ad, and A549) were used in this research. PC-7, PC-9, H1975, and H2228 were cultivated in RPMI medium containing L-glutamine (Wako Pure Chemical Industry) supplemented with 10% FBS, 1x antibiotic-antimycotic (Thermo Fisher), and 1x MEM nonessential amino acids (Sigma Aldrich). VMRC-LCD was cultivated in EMEM medium (Thermo Fisher) supplemented with 10% FBS, 2 mM L-glutamine (Thermo Fisher), 1x antibiotic-antimycotic, and 1x MEM nonessential amino acids. LC2/ad and A549 were cultivated in DMEM (Thermo Fisher) supplemented with 10% FBS, 2 mM L-glutamine, 1x antibiotic-antimycotic, and 1x MEM nonessential amino acids on collagen type I-coated dish (AGC Techno Glass) or cell culture dishes without coating, respectively. Using the RNeasy Mini Kit (QIAGEN), total RNA was isolated from these cells. We performed a quality check of total RNA using the Agilent RNA Nano Kit (Agilent Technologies).

### SMART-Seq using Nextera XT and HiSeq2500

We synthesized the RNA-Seq library from the same FL-cDNA applied for FL-cDNA-Seq following SMART-Seq v4 or the FL-cDNA prepared following the single-cell RNA-Seq protocol by C1 (Fluidigm). Briefly, using the Nextera XT DNA Library Preparation Kit (Illumina), cDNA was fragmented and adapters introduced. The dual index Illumina libraries were obtained by PCR amplification using the Nextera XT Index Kit. Sequencing of 36-bp single ends was performed by HiSeq2500 (Illumina).

### Whole-exome sequencing and SNP calling

The libraries of whole-exome sequencing (WES) were prepared from the purchased DNA of the two individuals using the SureSelect XT kit and SureSelect Human All Exon V5 (Agilent), following the manufacturer's protocol. The 100 pair-end sequencing of these libraries was performed using HiSeq2500. Using WES data for two individuals, SNPs were called by GATK version 3.7 according to the GATK Best Practices for germline short variants[17].

**Sanger sequencing of fusion transcripts**

Using SuperScript II Reverse Transcriptase (Thermo Fisher) and oligo-dT primers, we synthesized first-strand cDNA from total RNA of LC2/ad and PC-9. We amplified the fusion region of CCDC6-RET and WAC-SFMBT2 from cDNA of LC2/ad and ZSCAN22-CHMP2A from PC-9 using Phusion High-Fidelity PCR Master Mix (Thermo Fisher) and the following primers: CCDC6-RET (fwd: GCAGCAAGAGAACAAGGTGC, rev: ACCATCCTAAGTTGCTGGGC), WAC-SFMBT2 (fwd: AGCACAGGTCACAGTAAGGC, rev: AACCTTAGTCACCGACGCAG), ZSCAN22-CHMP2A (fwd: AAGTTGGCTAGTCTCTGCGG and rev: CCCAGCTCATCCAGAACCTG). Sanger sequencing was performed using the 3730xl DNA Analyzer (Thermo Fisher) and forward primers for each primer set.

**Data analysis of conventional RNA-Seq data**

We utilized the data obtained for the TruSeq RNA of LC2/ad from our previous study (DRA001846)[16]. We aligned the RNA-Seq and SMART-Seq reads to the longest coding isoforms of each RefGene using BWA-MEM with default parameters. For expression quantification, we counted the number of reads mapped to each isoform and calculated the tpm (transcript per million) as the expression level.

To detect the transcript isoform, we aligned RNA-Seq reads of LC2/ad to hg38 using tophat2.1.0 and

assembled them with cufflinks 2.2.1 according to the RefSeq annotation guide or not.

## Supplementary Reference

Carninci, P., Sandelin, A., Lenhard, B., et al. 2006, Genome-wide analysis of mammalian promoter architecture and evolution. Nat. Genet., 38, 626–35.

**Supplementary Table 1.** Statistics for FL-cDNA-Seq of lung cancer cell lines (2D sequencing)

| Cell line | Flow cell | # of total reads | # of pass 2d reads | Average length of pass 2d reads (bp) | Human genome | |
|---|---|---|---|---|---|---|
| | | | | | # of mapped reads | Mapping rate |
| H1975 | R9 | 251,062 | 65,006 | 1,017 | 59,837 | 92% |
| PC-9 | R9 | 135,717 | 35,578 | 1,053 | 32,910 | 93% |
| PC-7 | R9 | 142,933 | 44,345 | 910 | 35,378 | 80% |
| H2228 | R9 | 240,567 | 58,134 | 1,144 | 53,926 | 93% |
| VMRC-LCD | R9 | 84,733 | 19,313 | 918 | 43,678 | 94% |
| VMRC-LCD | R9 | 104,863 | 27,359 | | | |
| LC2/ad | R9 | 169,401 | 49,593 | 908 | 45,381 | 92% |
| LC2/ad | R9.4 | 233,331 | 136,753 | 917 | 665,212 | 92% |
| LC2/ad | R9.4 | 252,951 | 142,245 | | | |
| LC2/ad | R9.4 | 311,128 | 188,805 | | | |
| LC2/ad | R9.4 | 410,826 | 255,947 | | | |

**Supplementary Table S2**. Statistics for FL-cDNA-Seq of A549 ($1D^2$ sequencing)

| Flow cell | Read type | # of reads | Average length (bp) | All RefSeq | | |
|---|---|---|---|---|---|---|
| | | | | # of mapped reads | Mapping rate | Average identity |
| R9.5 | 1d | 994,111 | 926 | 619,458 | 62% | 87% |
| | 1dsq pass | 931 | 1,712 | 643 | 69% | 95% |

**Supplementary Table S3**. Statistics for alignment of FL-cDNA-Seq reads to RefSeq

| Cell line | Mapper | All RefSeq | | Longest isoforms | |
|---|---|---|---|---|---|
| | | # of mapped reads | Mapping rate | # of mapped reads | Mapping rate |
| LC2/ad (R9.4) | BWA | 690,528 | 95% | 649,666 | 90% |
| | LAST | 532,956 | 74% | 633,427 | 88% |

**Supplementary Table S4.** Statistics for alignment of Illumina sequencing

| Cell line | Library prep. method | Read length | Single/Pair | # of total reads | Longest isoforms | | Genome | | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | # of mapped reads | Mapping rate | # of mapped reads | Mapping rate | # of novel isoforms |
| LC-2/ad | TruSeq | 100 bp | paired-end | 120,312,052 | 83,123,354 | 69% | 96,422,580 | 80% | 27,690 |
| | SMART-Seq | 36 bp | single-end | 2,662,312 | 1,811,231 | 68% | N/A | N/A | N/A |

**Supplementary Table S5.** Statistics for direct RNA-Seq reads

| Cell line | Flow cell | # of runs | # of 1d reads | Average length | All RefSeq | | | Longest isoform | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | # of mapped reads | Mapping rate | Average identity | # of mapped reads | Mapping rate |
| LC2/ad | R9.5 | 1 | 556,195 | 926 | 279,251 | 50% | 83% | 373,320 | 67% |

**Supplementary Table S7.** Statistics for genome alignment by BLAT

| Cell line | Flow cell | # of pass 2d reads | # of mapped reads | Mapping rate |
|---|---|---|---|---|
| H1975 | R9 | 65,006 | 57,411 | 88% |
| PC-9 | R9 | 35,578 | 31,573 | 89% |
| PC-7 | R9 | 44,345 | 34,559 | 78% |
| H2228 | R9 | 58,134 | 50,967 | 88% |
| VMRC-LCD | R9 | 46,672 | 39,673 | 85% |
| LC2/ad | R9 | 49,593 | 44,316 | 89% |
| LC2/ad | R9.4 | 723,750 | 679,983 | 94% |

**Supplementary Table S10.** Statistics for FL-cDNA-Seq of tissues (2D sequencing)

| Tissue | Male | | | | Female | | | |
|---|---|---|---|---|---|---|---|---|
| | RIN score | Average length (bp) | # of raw reads | # of 2D reads | RIN score | Average length (bp) | # of raw reads | # of 2D reads |
| Liver | 7.5 | 383 | 1,877,492 | 950,526 | 8.6 | 575 | 2,408,040 | 1,704,811 |
| Kidney | 6.1 | 300 | 2,169,687 | 1,038,648 | 8.1 | 455 | 1,219,741 | 622,673 |
| Lung | 7.1 | 436 | 2,196,944 | 1,102,784 | 8.4 | 671 | 1,982,029 | 1,314,034 |
| Skeletal muscle | 6.7 | 1,377 | 967,999 | 714,931 | 6.8 | 524 | 1,140,199 | 756,317 |
| Pancreas | 8.6 | 429 | 1,136,345 | 563,300 | 6.3 | 496 | 750,299 | 436,896 |
| Heart | 6.4 | 348 | 1,596,992 | 776,182 | 8.6 | 523 | 1,235,122 | 781,718 |
| Colon | 5.6 | 300 | 809,755 | 434,140 | 6.9 | 395 | 2,030,195 | 1,215,465 |

**Supplementary Table S11.** Statistics for alignment of FL-cDNA-Seq reads of tissues to genome

| Tissue | Male | | Female | |
|---|---|---|---|---|
| | # of mapped reads | Mapping rate | # of mapped reads | Mapping rate |
| Liver | 387,961 | 41% | 1,006,366 | 59% |
| Kidney | 363,702 | 35% | 327,385 | 53% |
| Lung | 489,070 | 44% | 797,848 | 61% |
| Skeletal muscle | 260,451 | 36% | 499,246 | 66% |
| Pancreas | 291,912 | 52% | 260,465 | 60% |
| Heart | 295,878 | 38% | 507,065 | 65% |
| Colon | 120,296 | 28% | 507,887 | 42% |

**Supplementary Table S12.** Statistics for whole exome sequencing using Illumina

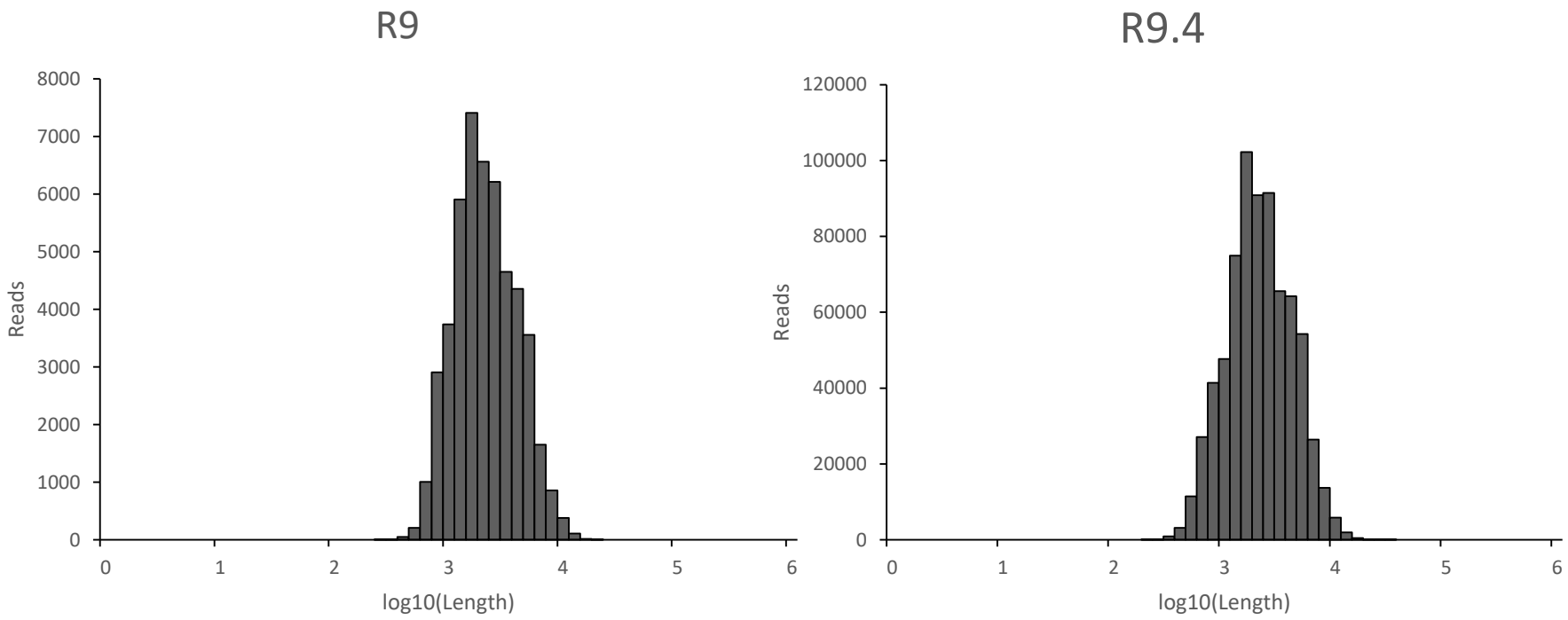
| Sample gender | Read length | Single/Pair | # of total reads | Genome | | # of SNPs | # of hetero SNPs |
|---|---|---|---|---|---|---|---|
| | | | | # of mapped reads | Mapping rate | | |
| Male | 100 | paired-end | 217,421,430 | 202,431,918 | 93% | 1,128,434 | 437,646 |
| Female | 100 | paired-end | 206,732,004 | 194,230,944 | 94% | 1,128,724 | 395,486 |

**Supplementary Table S14.** Statistics for single cell analysis (MinION)

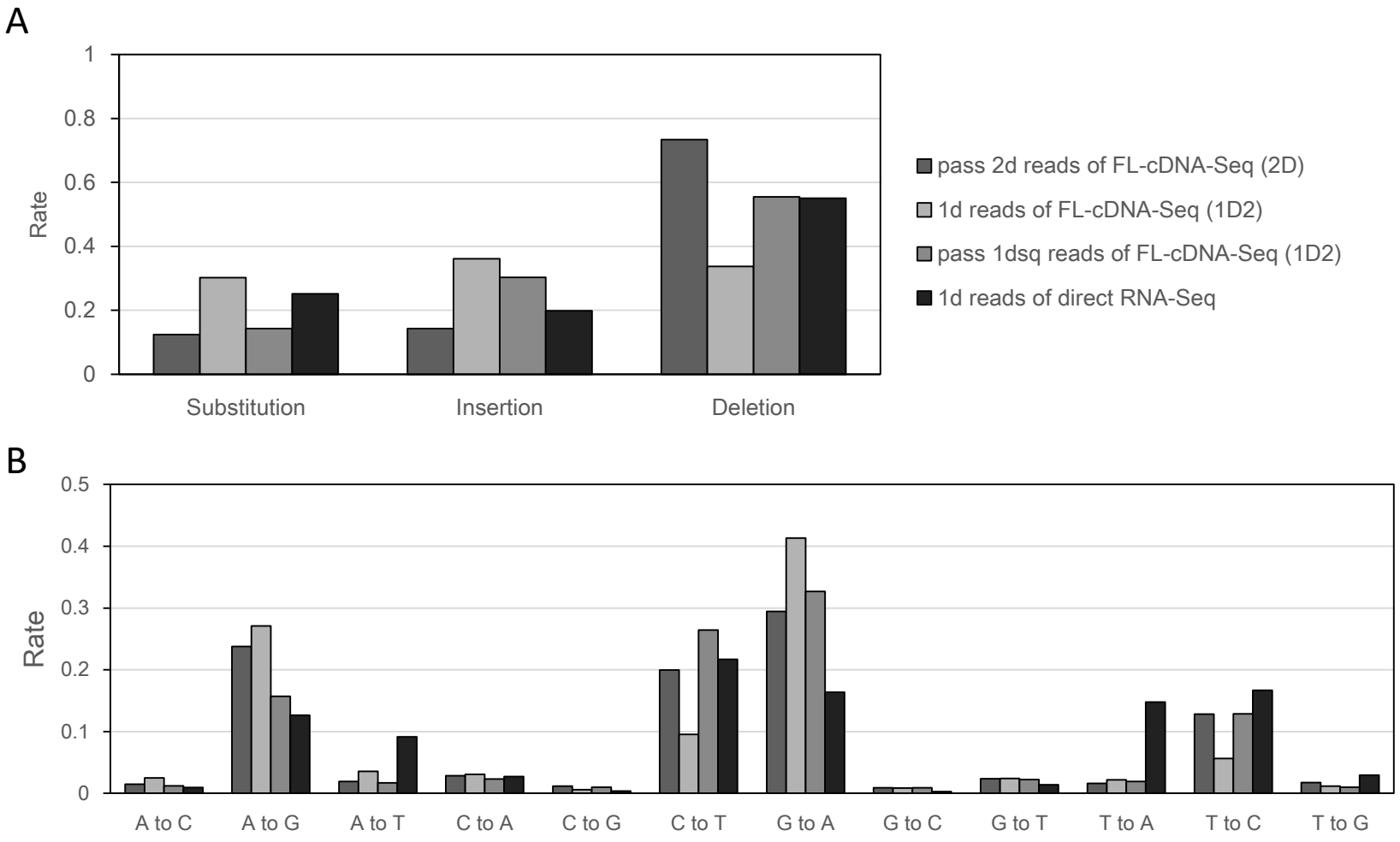
| Sample name | # of total reads | Longest isoforms | | All RefSeq | |
|---|---|---|---|---|---|
| | | # of mapped reads | Mapping rate | # of mapped reads | Mapping rate |
| bulk | 6,522 | 4,808 | 74% | 2,872 | 44% |
| sc1 | 1,488 | 728 | 49% | 443 | 30% |
| sc2 | 469 | 199 | 42% | 123 | 26% |
| sc3 | 1,187 | 570 | 48% | 345 | 29% |
| sc4 | 3,126 | 1,606 | 51% | 1,127 | 36% |
| sc5 | 884 | 425 | 48% | 281 | 32% |
| sc6 | 584 | 100 | 17% | 59 | 10% |
| sc7 | 2,918 | 1,372 | 47% | 1,000 | 34% |
| sc8 | 1,247 | 608 | 49% | 476 | 38% |
| sc9 | 3,426 | 2,147 | 63% | 1,615 | 47% |
| doublet | 2,966 | 1,650 | 56% | 1,261 | 43% |

**Supplementary Table S15.** Statistics for single cell RNA-Seq (Illumina)

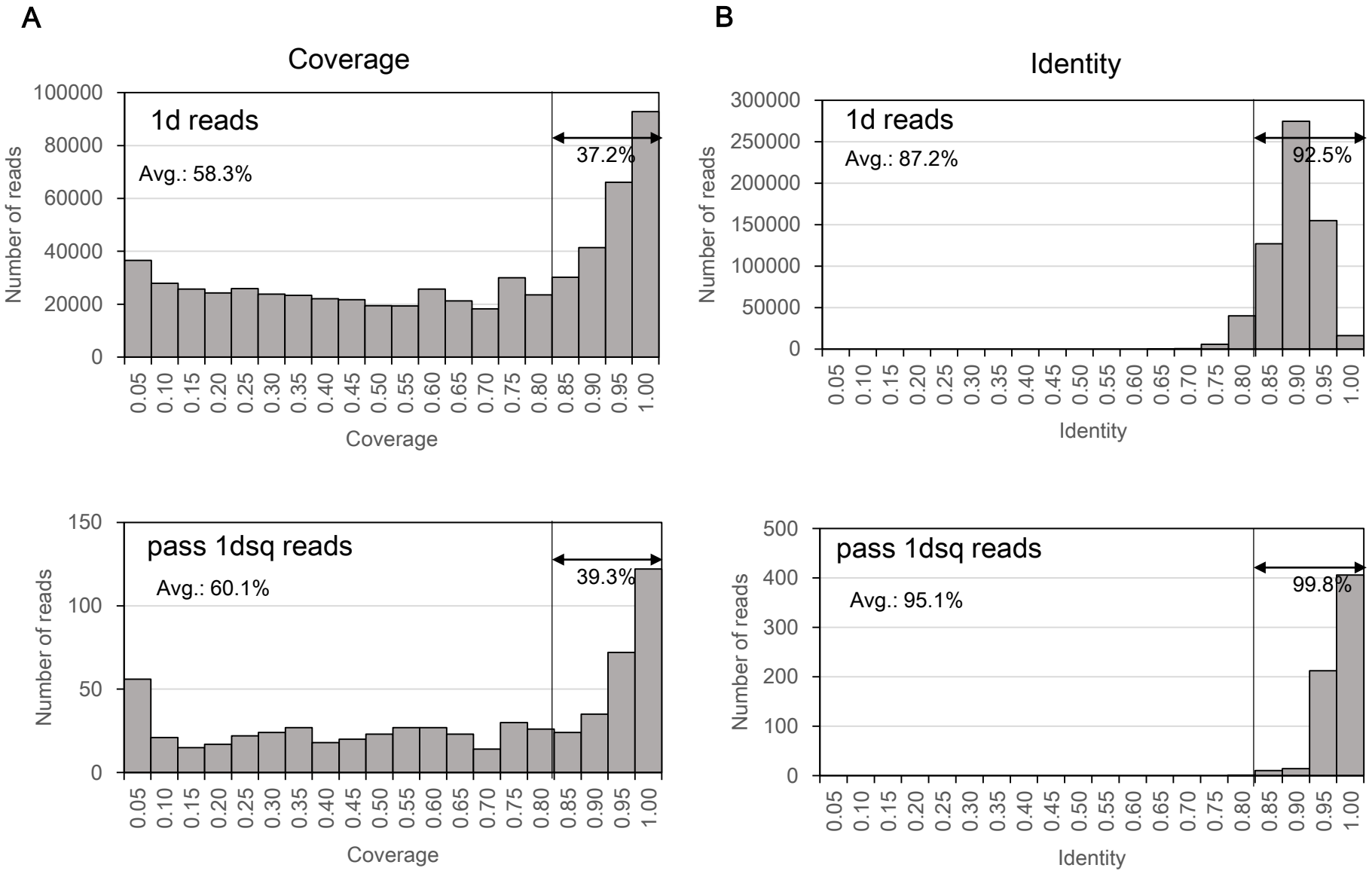
| Sample name | # of total reads | Longest isoforms | |
|---|---|---|---|
| | | # of mapped reads | Mapping rate |
| bulk | 3,980,593 | 2,039,135 | 51% |
| sc1 | 1,490,135 | 423,866 | 28% |
| sc2 | 1,306,780 | 323,265 | 25% |
| sc3 | 1,685,942 | 453,687 | 27% |
| sc4 | 1,589,393 | 468,952 | 30% |
| sc5 | 616,877 | 157,873 | 26% |
| sc6 | 932,335 | 58,323 | 6% |
| doublet | 1,353,973 | 478,578 | 35% |
| sc7 | 1,467,764 | 351,158 | 24% |
| sc8 | 1,556,215 | 547,531 | 35% |
| sc9 | 1,679,026 | 737,134 | 44% |
| Total | 17,659,033 | 6,039,502 | 34% |



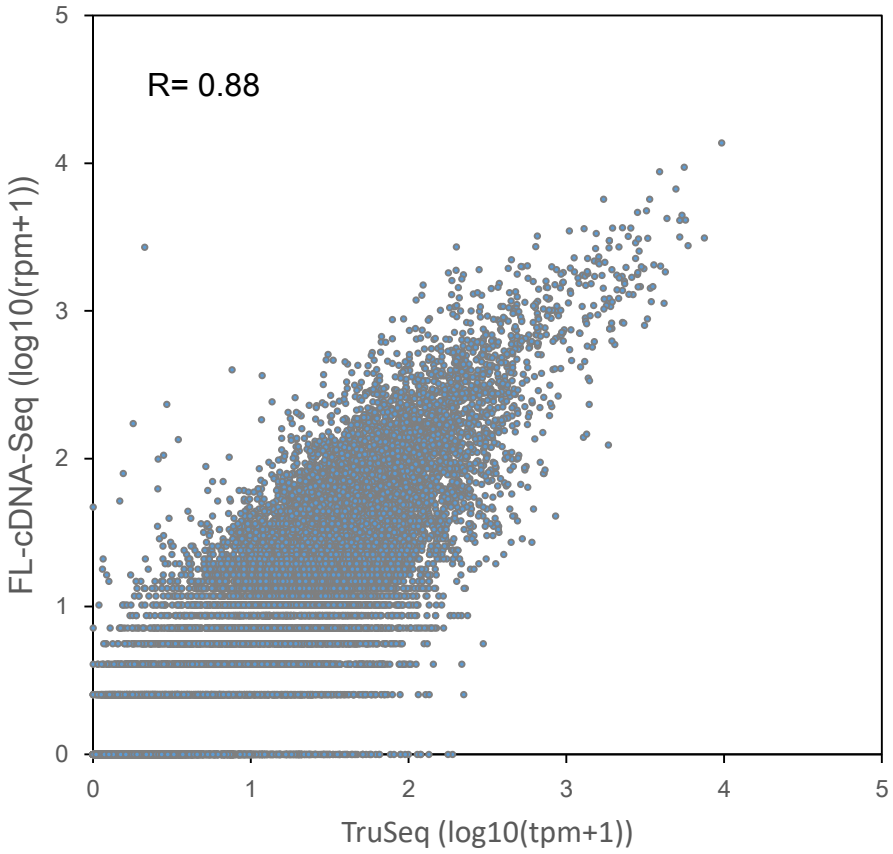

**Supplementary Figure S1**
Length distribution of FL-cDNA-Seq reads of LC2/ad. Length distribution of cDNA-Seq reads sequenced by R9 (left) and R9.4 (right).



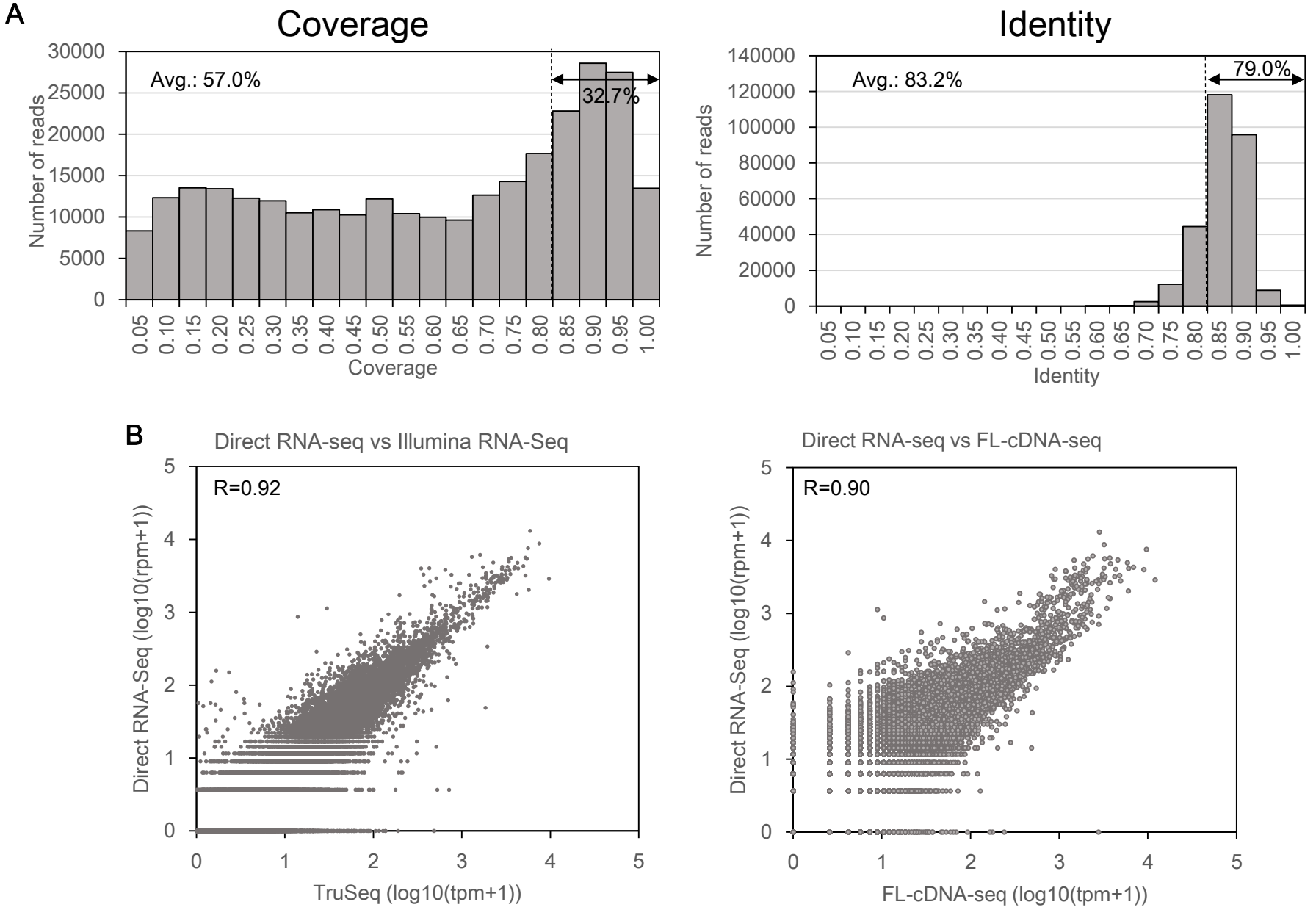

**Supplementary Figure S2**
Error pattern of MinION sequencing. (A) Error pattern of four kinds of MinION RNA-Seq reads. (B) Substitution pattern of both.




**Supplementary Figure S3**
Coverage and identity of 1D$^2$ sequencing. (A) Coverage of 1d reads (top) and pass 1dsq reads (bottom). (B) Identity of 1d reads (top) and pass 1dsq reads (bottom).

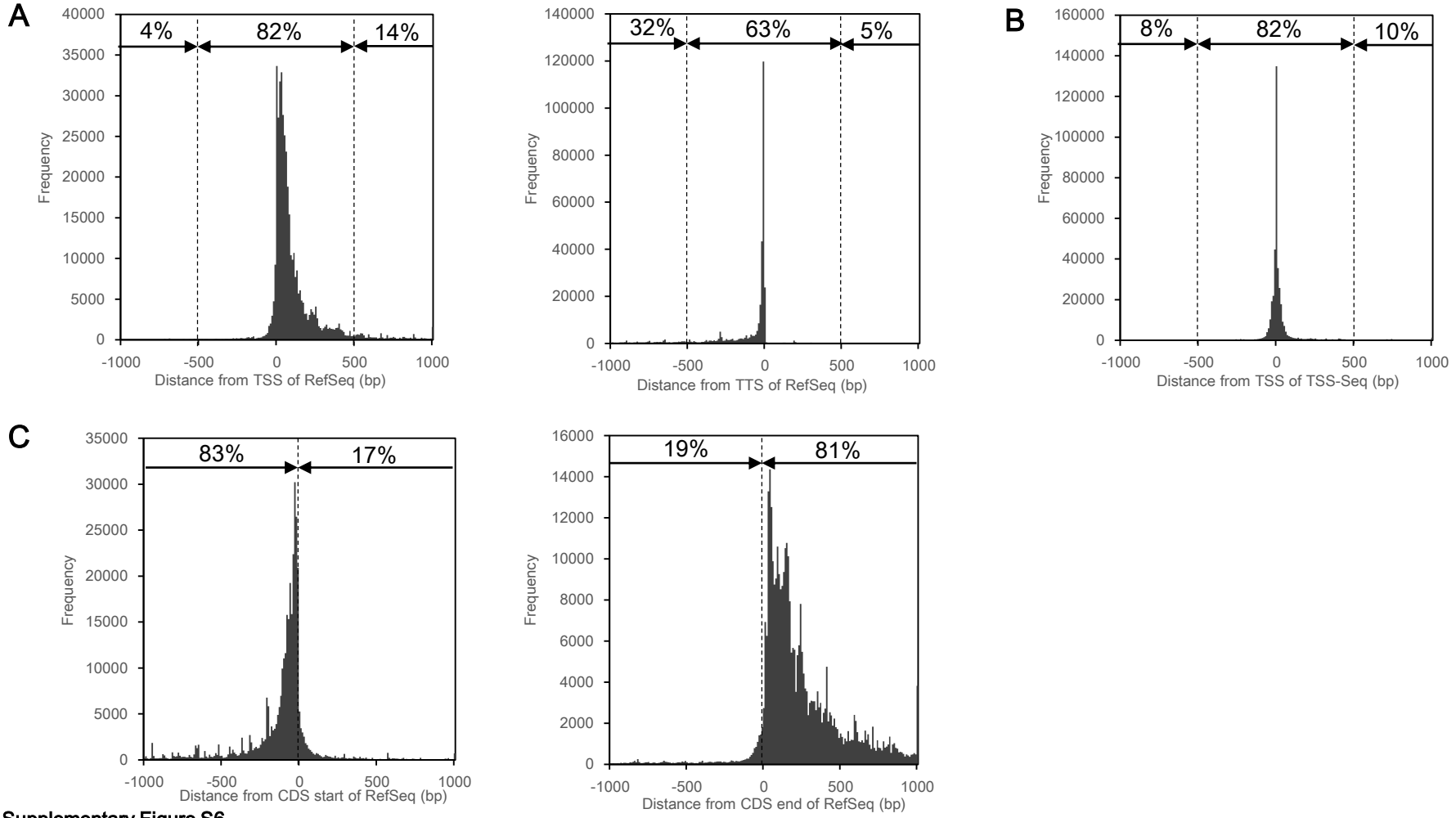



**Supplementary Figure S4**
Correlation of the expression level between FL-cDNA-Seq (R9.4) and Illumina TruSeq of LC2/ad using BWA-processed data.




**Supplementary Figure S5**

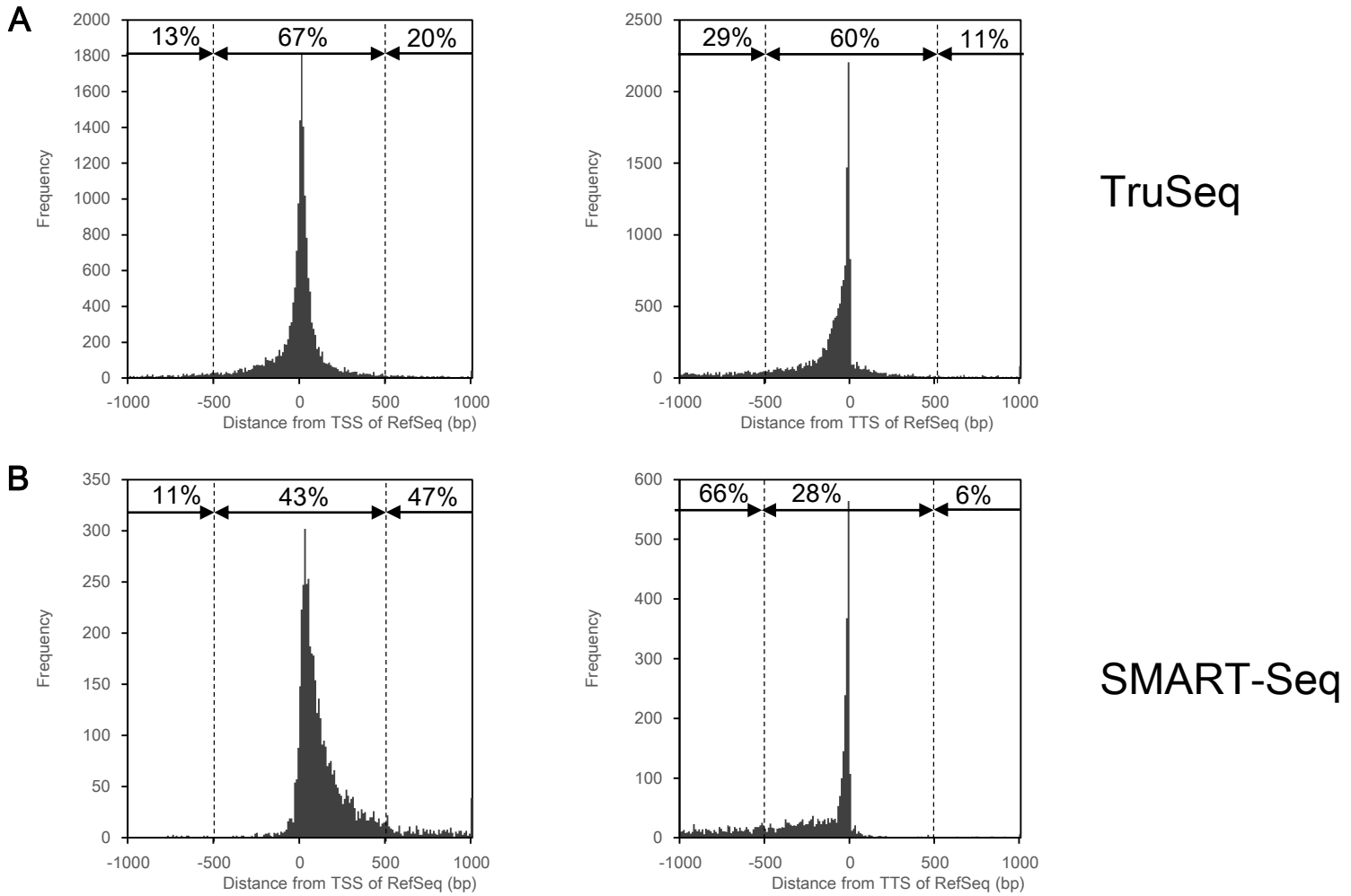
Evaluation of direct RNA-Seq. (A) Distribution of the coverage (left) and identity (right) of 1d reads of LC2ad, aligned by LAST. The percentages of reads with a coverage or identity greater than 0.8 is shown on the graphs. (B) Comparison of direct RNA-Seq and other methods. The gene expression of FL-cDNA-Seq of LC2/ad was compared to that of TruSeq RNA (left) and FL-cDNA-Seq (right). Pearson correlation coefficients are shown on the graph.




**Supplementary Figure S6**

Evaluation of alignment of FL-cDNA-Seq reads of LC2/ad (R9.4) to genome. (A) The distance between the ends of mapped reads and TSS (Left) or TTS (Right) of RefSeq. (B) The distance between the ends of mapped reads and TSS cluster detected by TSS-Seq of LC2/ad[16]. (C) The distance between the ends of mapped reads and CDS start (Left) or CDS end (Right) of RefSeq.

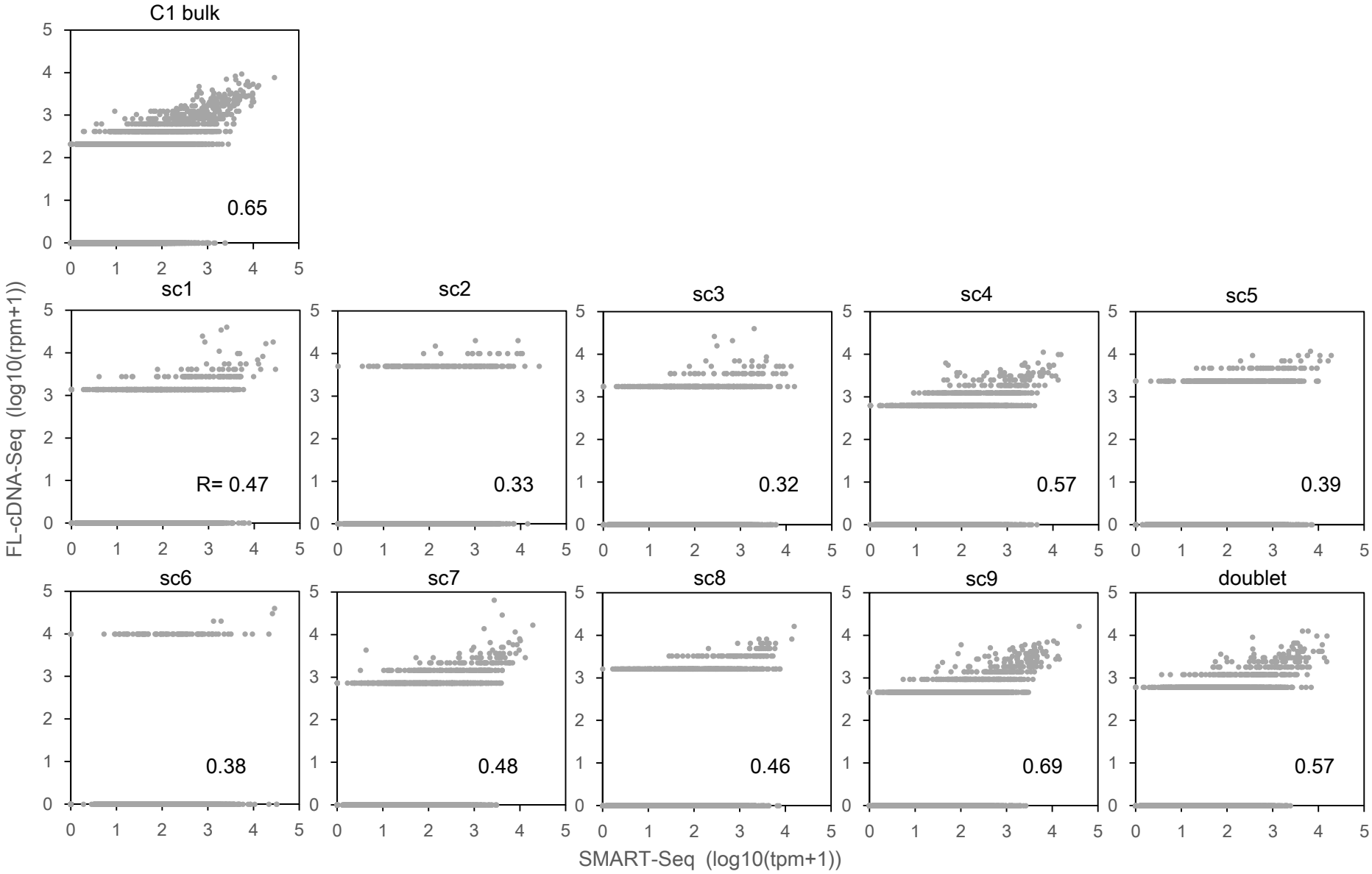
In comparison of FL-cDNA-Seq reads and TSS and TTS of RefSeq, 82% and 63% of the TSS and TTS detected by MinION were located within ± 500 bp from those of the RefSeq transcripts, respectively. Although TTS detected by FL-cDNA-Seq distributed sharply at TTS of RefSeq, TSS detected by FL-cDNA-Seq showed broad distributions around TSS. This should reflect the characteristic feature of the broad TSS as previously reported (Carninci et al. 2006). Fourteen and thirty-two percent of the TSS and TTS detected by MinION were located more than 500 bp downstream of TSS and -500 bp upstream of TTS of RefSeq, respectively. These fraction may represent the truncated form of cDNAs, which occurred in the process of cDNA synthesis or sequencing. To further evaluate rate of the possible cDNA truncation, TSS detected by MinION was compared with TSS detected by our previous TSS-Seq data in LC2/ad. TSS detected by MinION almost evenly located in upstream and downstream of the TSSs of the TSS-Seq, suggesting that the frequency of the cDNA truncation is not high, if any. To evaluate the coverage of full-length of CDS by MinION reads, we also compared TSS and TTS detected by MinION with CDS start and end of RefSeq. Eighty-three and eighty-one percent of the TSS and TTS, respectively, detected by MinION were located upstream of CDS start and downstream of CDS end of RefSeq. Full-length of CDS was covered by 67% of mapped MinION reads. Judging from these observations, we concluded that the FL-cDNA-Seq should be able to detect full-length of CDS with at reasonably high efficiency.




**Supplementary Figure S7**

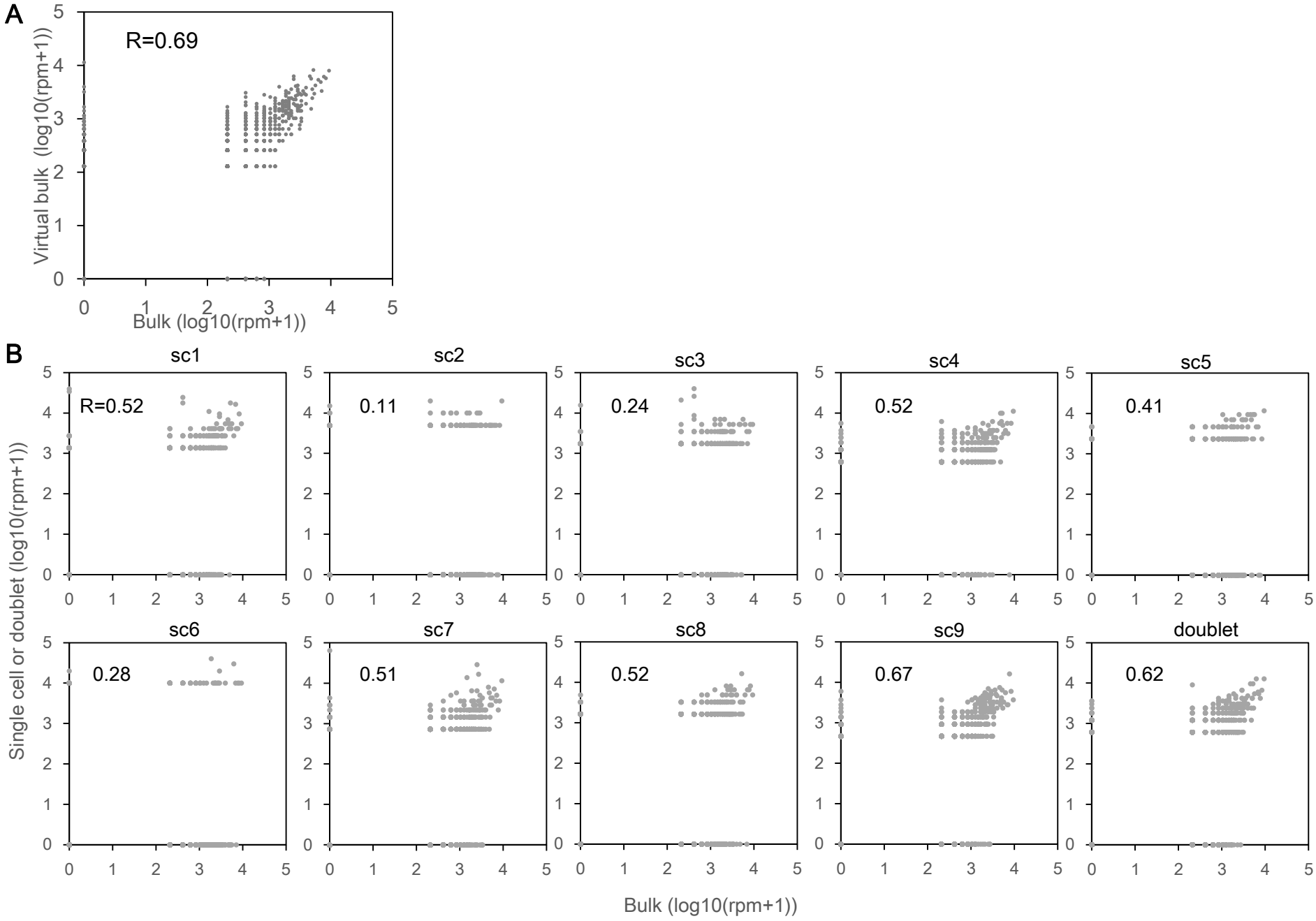
(A) The distance between the ends of isoform detected by TruSeq and TSS (Left) or TTS (Right) of RefSeq. (B) The distance between the ends of isoform detected by SMART-Seq and TSS (Left) or TTS (Right) of RefSeq. For assembling of Illumina short reads to transcript isoforms, we used cufflinks2.2.1 without guide of transcript models.

To evaluate the difference between the MinION and the Illumina sequencing, Illumina short reads were assembled to transcript isoforms, using cufflinks for the TruSeq and SMART-Seq reads of LC2/ad. Using the TSS and TTS as of the assembled isoforms assembled, we compared their positions with those of RefSeq. We found the TSS and the TTS of Illumina isoforms were frequently located more than 500 bp downstream of the TSS and more than -500 bp upstream of the TTS of RefSeq. Since these fractions were lower than those of the MinION reads, we believe that the overall performance of FL-cDNA-Seq is superior than previous method. Also note that the truncated cDNA forms showed similar distribution patterns both at the TSS and the TTS between Illumina reads and the MinION reads, they are independent of the sequencing platforms, but derived from the RNA degradation at starting the starting material.




**Supplementary Figure S8**

Correlation between FL-cDNA-Seq and SMART-Seq of single cells. The Pearson correlation coefficient was calculated using the detected genes for both datasets.




**Supplementary Figure S9**

Comparison of single cell data and bulk data for FL-cDNA-Seq. (A) Comparison of bulk and virtual bulk of nine single cells. (B) Comparison of bulk and nine single cells or one doublet cell. The Pearson correlation coefficient was calculated using detected genes for both datasets.