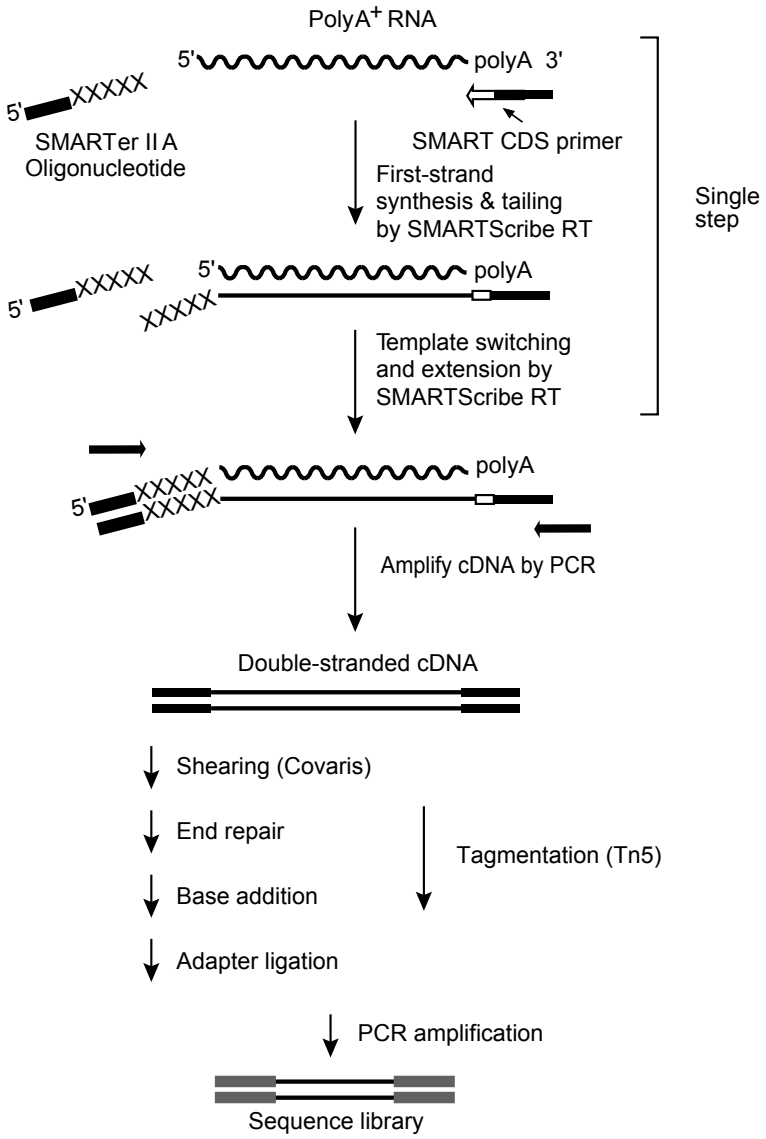
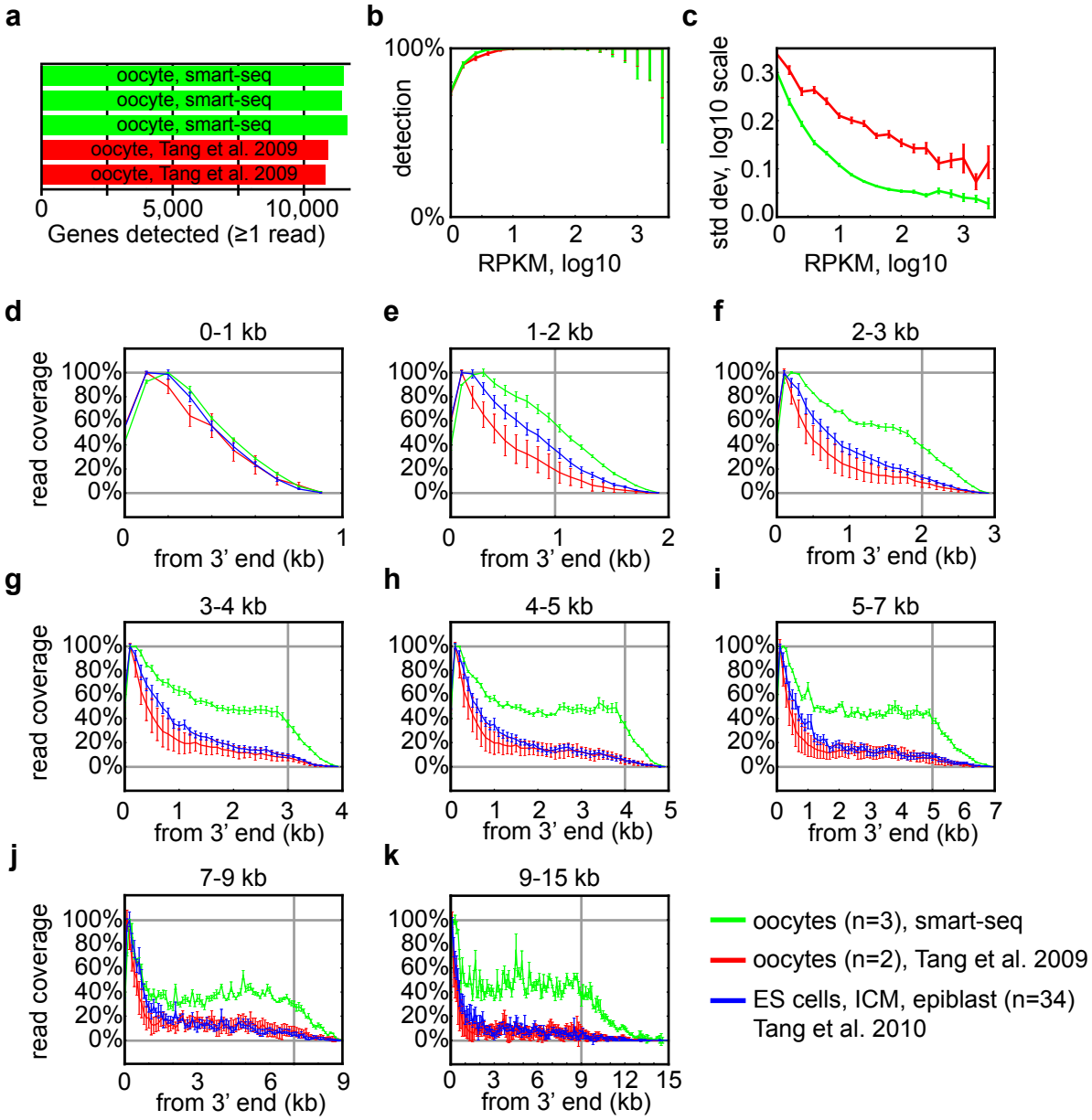
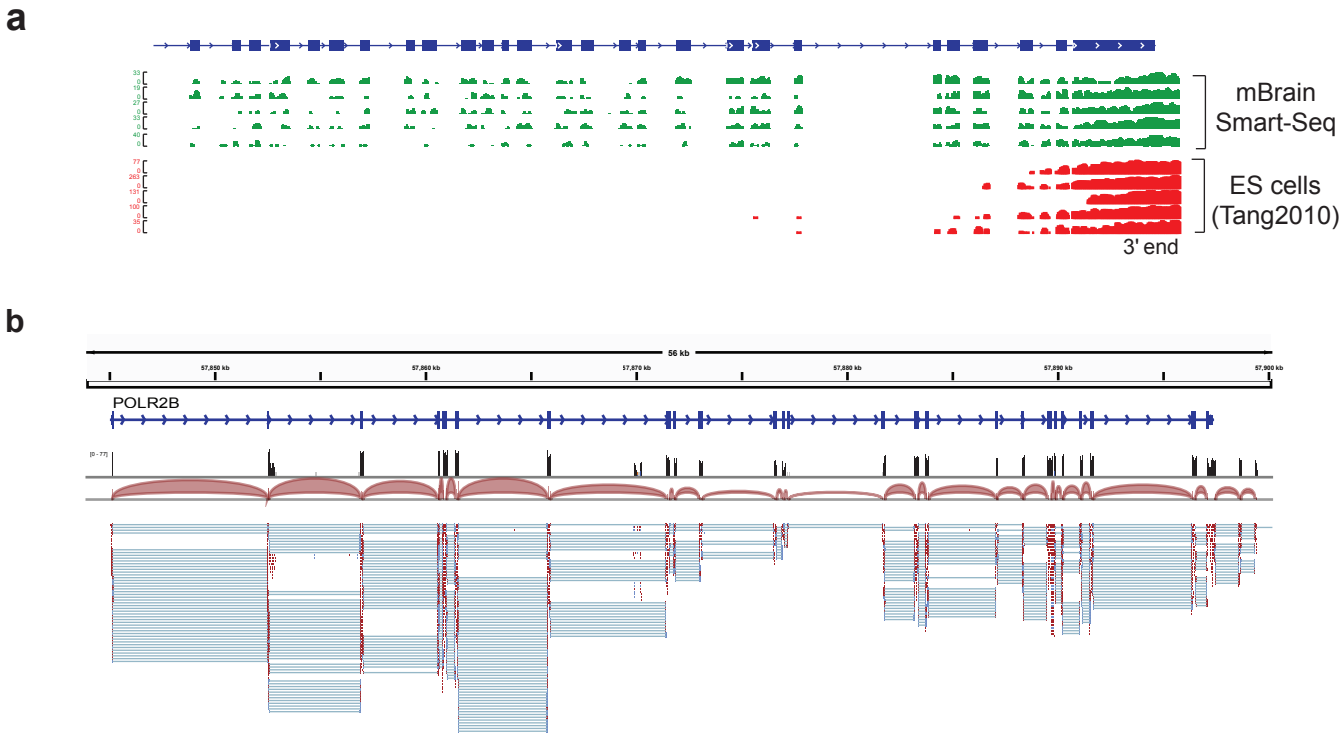


Supplementary Figure 1. The Smart-Seq protocol
Protocol for cDNA amplification and library preparation from few or single cells. Cells are first lysed in reverse transcription compatible buffer and the reaction is initiated with oligo-dT containing primer. First strand synthesis is completed with the addtion of a few untemplated C nucleotides followed by template switching and the incorporation of SMARTer IIA oligonucleotide. Full-length cDNAs are amplified using PCR to obtain a few nanograms of DNA. Fragmentation and adapter introduction was performed using either acoustic shearing or transposase tagmentation based procedures to generate sequencing libraries.

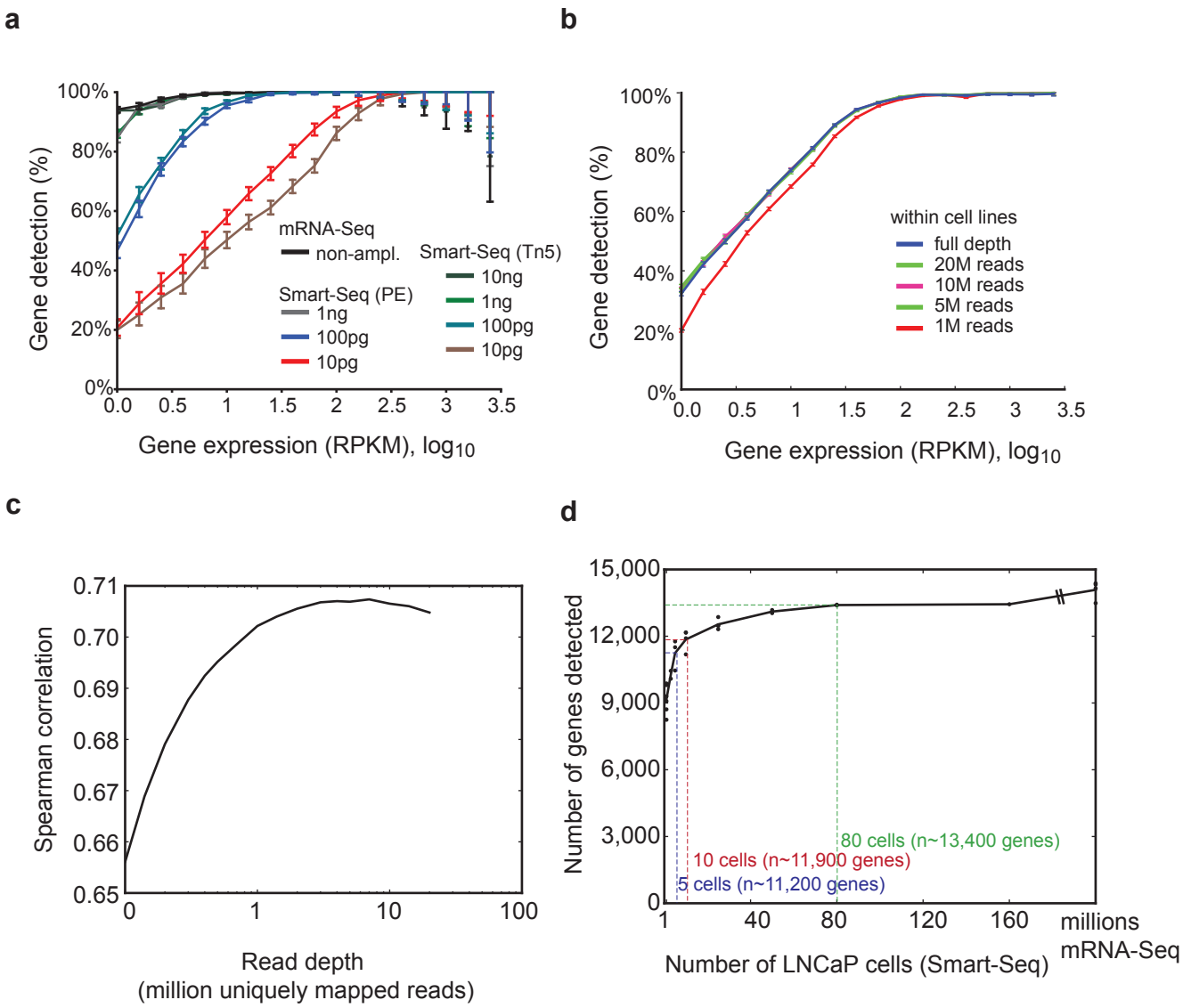


Supplementary Figure 2. Comparison of Smart-Seq and Tang et al. data (2009, 2010)

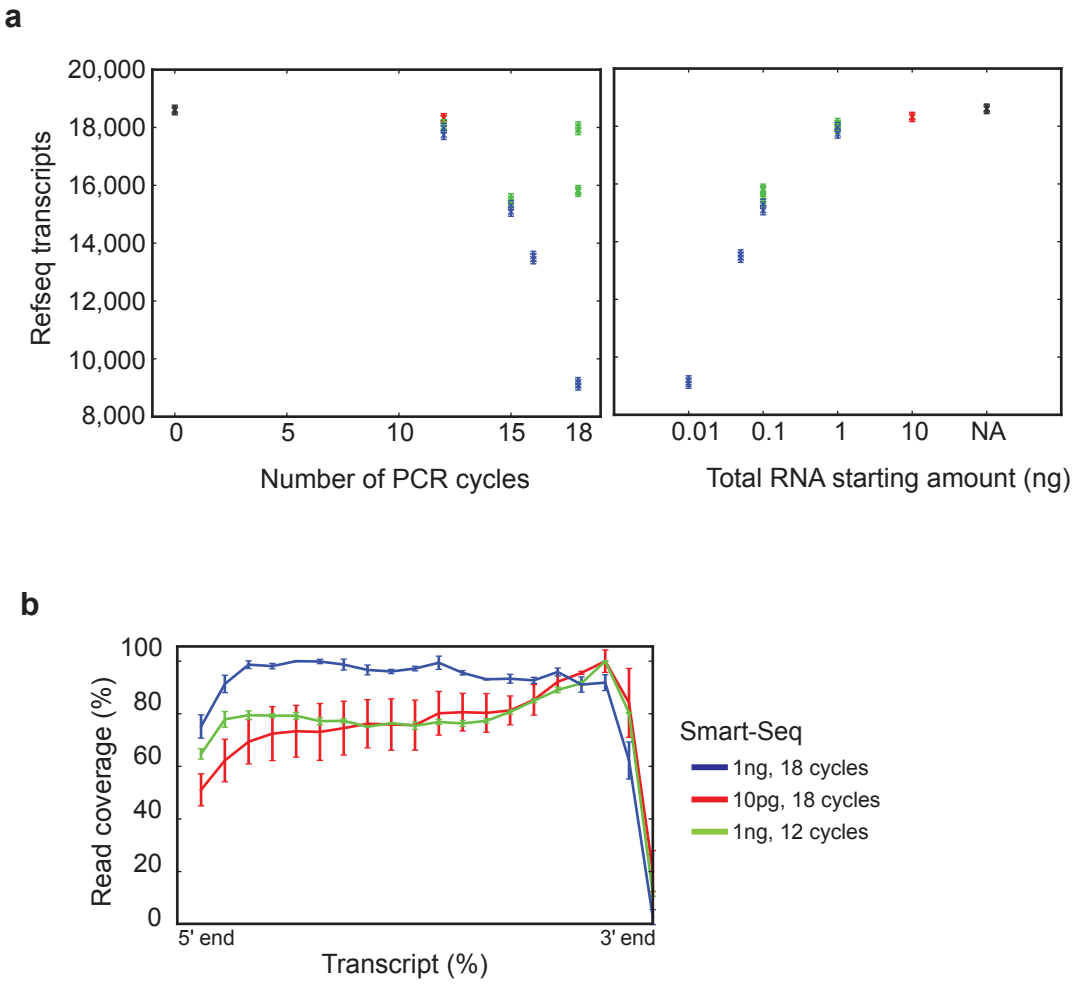
Comparisons of sensitivity between mouse oocytes generated with Smart-Seq (green) and from Tang et al. 2009 (red), using 5 million reads per cell. (**a**) The number of genes detected in each oocyte (result not sensitive to read cutoff for detection). (**b**) The percentage of genes detected in pairs of oocytes, as a function of the gene expression level (max over biologial replicates). (**c**) Variation in expression level estimation between pairs of oocytes as a function of the mean expression. (**d-k**) Read coverage across transcripts for Smart-Seq oocytes, oocytes from Tang et al. 2009 and cells from Tang et al. 2010. Transcripts has been separated according to their lengths as indicated over the plots. The gray vertical line indicate the point from where the coverage is expected to drop based on the span of transcript lengths included in each figure.



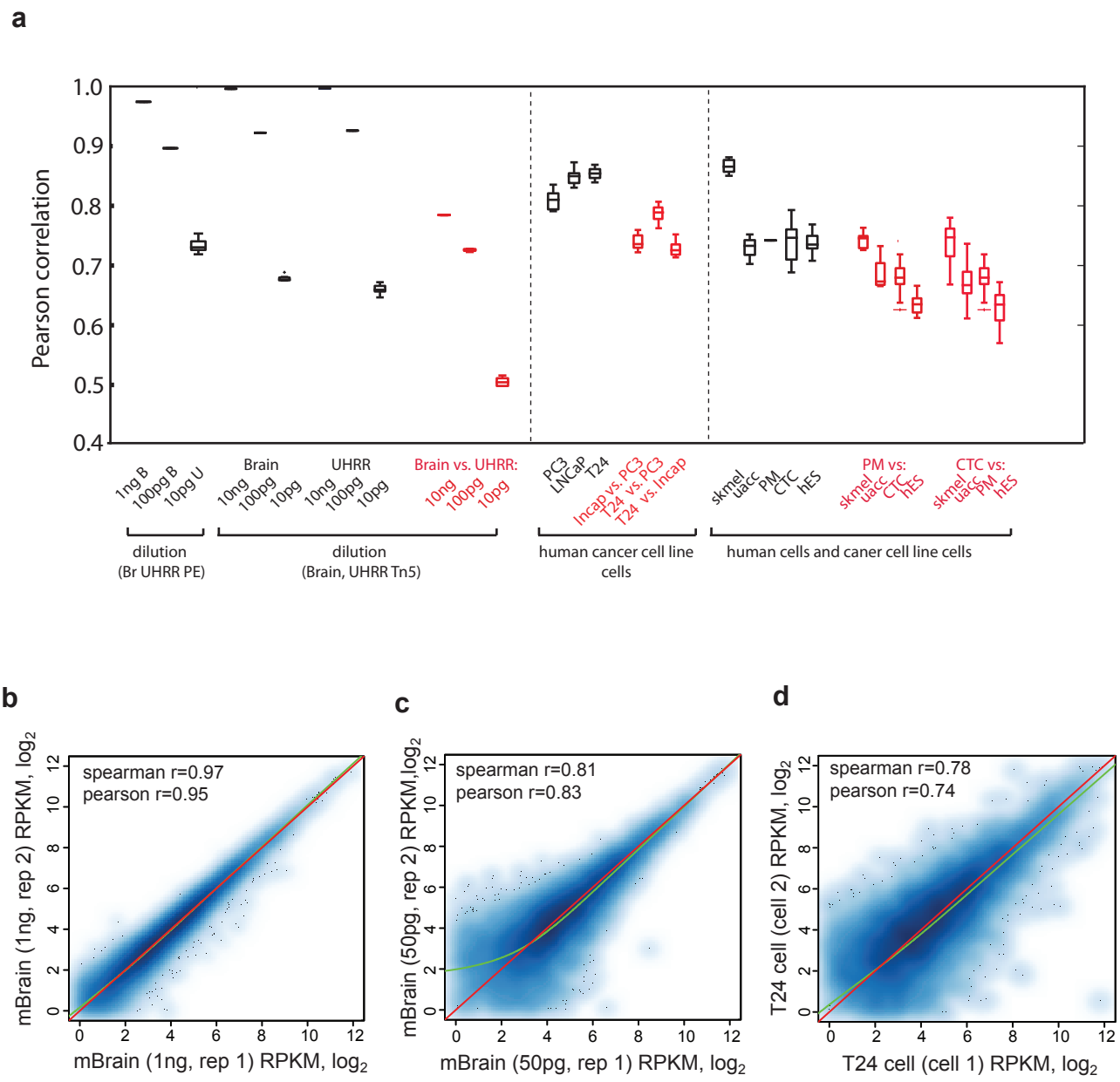
Supplementary Figure 3. Full-transcript coverage and reconstruction
(**a**) Smart-Seq reads mapping to a 14-kb region of the RNA polymerase II alpha gene locus containing 27 of the 28 exons. The read coverage is displayed for Smart-Seq data from diluted amounts of mouse brain RNA (green) and for previously published single-cell data from mouse embryonic stem (ES) cells (red) from Tang et al. 2010. (**b**) Example of a fully reconstructed transcript (human POLR2B) from Smart-Seq reads from an individual T24 cancer cell line cell. Reads were visualized using the Integrated Genome Viewer (IGV, Broad Institute).



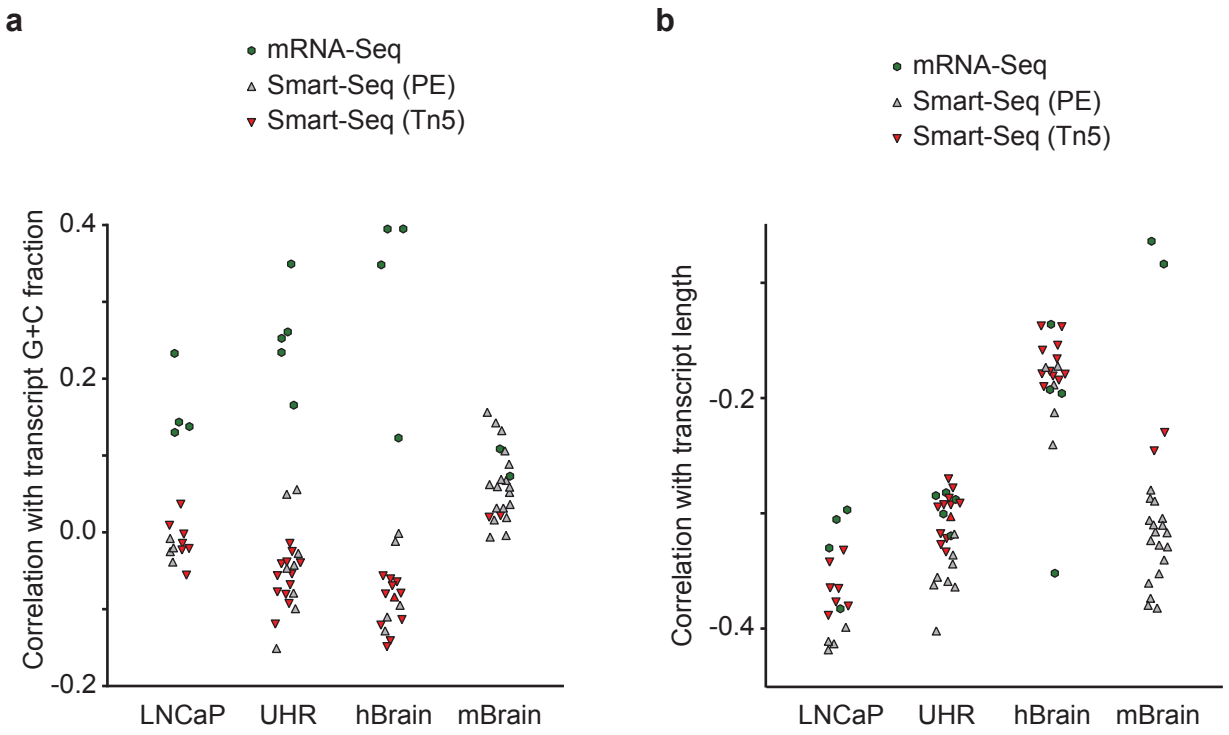
Supplementary Figure 4. Sensitivity and variation in Smart-Seq data
(**a**) Gene sensitivity presented as the mean percentage of genes detected in replicates, binned according to expression levels. We performed all pair-wise comparisons within groups of replicates and report the mean and 95% confidence interval. Comparing sequence libraries constructed using either transposase "tagmentation" (Tn5) or shearing followed by ligation (PE) from indicated low amounts of starting materials. (**b**) Gene detection sensitivity for cancer cells as in (Fig. 2b) but using different sequence depths. (**c**) Spearman correlation (rho) in gene expression levels between different cells of the same origin, as a function of the sequence depth. (**d**) The number of detected genes in individual cells or when combining data from two or more cells as indicated on the x-axis. For comparison, the last data point show the number of genes detected in standard mRNA-Seq from millions of cells.



Supplementary Figure 5. Transcript detection and coverage with varying PCR cycles (**a**) The number of Refseq transcripts detected as a function of PCR cycles (left) and starting amount of total RNA (right). A RPKM threshold of 0.1 was used for detection. Colors separate independent dilution series. The error bars show standard error. (**b**) Mean read coverage over transcripts. We compared read coverage for Smart-Seq data generated from 1ng or 10pg of total RNA using either 12 or 18 cycles of PCR. No reproducible trends of PCR cycles on read coverage were observed. Errors bars represent standard deviation.

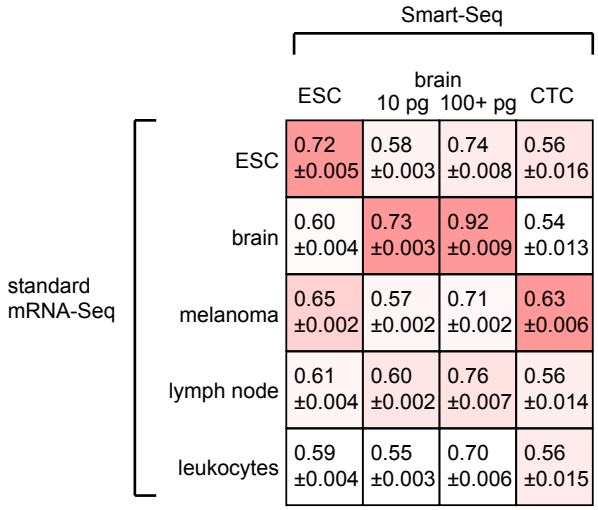


Supplementary Figure 6. Correlations in gene expression levels between technical and biological replicates. (**a**) Correlations over gene expression levels in dilution- or single cell replicates were shown as boxplots (black). Analyses of different types of cells or diluted RNA of different origins were shown in red. Correlations were computed from log2 transformed RPKM gene expression levels. (**b-d**) Representative scatter plots of gene expression levels from libraries generated from independent technical replicates of 1ng total RNA (b), or 50pg total RNA (c), or libraries generated from independent T24 cancer cell line cells (d). These scatter plots are representative of gene expression levels variations found and the results from all pairwise comparisons are summarized in Fig. 2c,d.

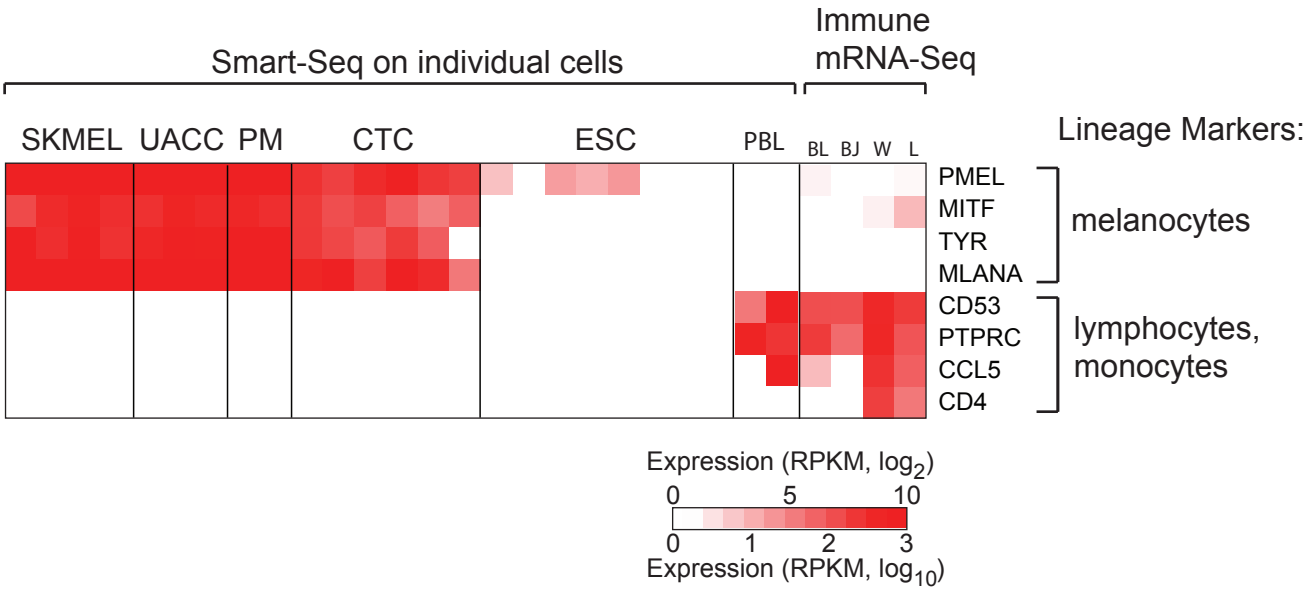


Supplementary Figure 7. Analyses of GC and length biases in Smart-Seq and mRNA-Seq data

(**a**) Pearson correlation coefficients between log transformed expression values and the transcript GC content for pre-amplified (Smart-Seq) and standard mRNA-Seq data. (**b**) Pearson correlation coefficients between log transformed expression values and log transcript length for pre-amplified (Smart-Seq) and standard mRNA-Seq. As the true biological GC and length biases are not known, there is no obvious best correlation. We found no systematic difference between Smart-Seq data generated with shearing and ligation (PE) or Tn5-mediated "tagmentation" (Tn5) or standard mRNA-Seq samples, but a tendency towards lower GC content correlations for Smart-Seq data.

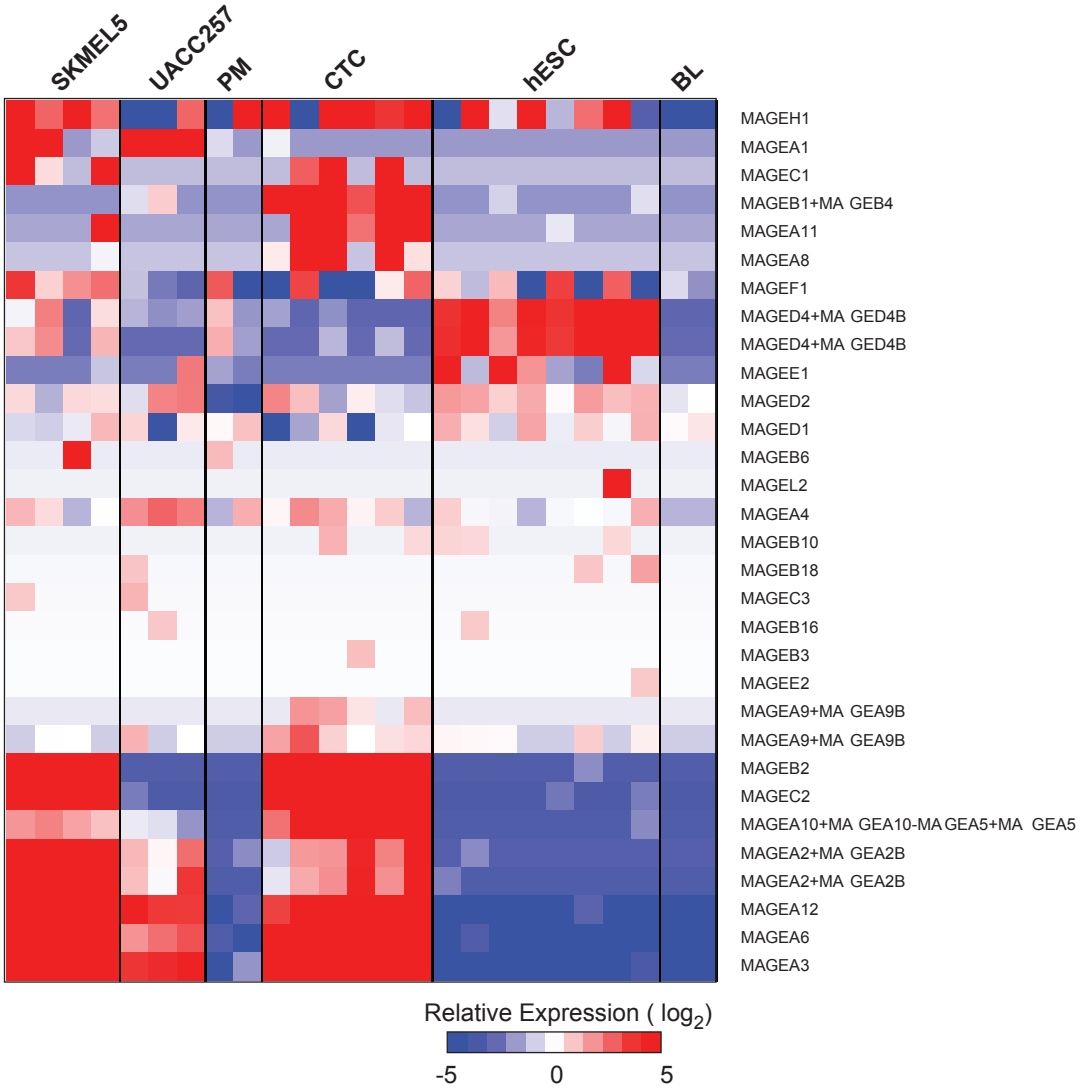


Supplementary Figure 8. Correlation analyses between Smart-Seq and mRNA-Seq gene expression levels.
Individual human ES cells (ESC, n=8), diluted brain RNA at 10pg and 100pg or more, and individual CTCs (n=6), compared with mRNA-Seq data from five different cell types or tissues. Mean spearman correlation and the standard error of the mean is shown for each comparison. Within each Smart-Seq sample (i.e. columns) the matrix cells are colored according to a linear white to red scale, with white at lowest mean correlation red at maximum mean correlation. Note that individual ES cells and CTCs cells are significantly higher correlated to mRNA-Seq data from ES cells and melanoma respectively. Identical Brain RNA prepared with Smart-Seq and mRNA-Seq correlate higher with increasing starting amounts in Smart-Seq. ESC RNA-Seq data was downloaded from GEO (GSM672836).

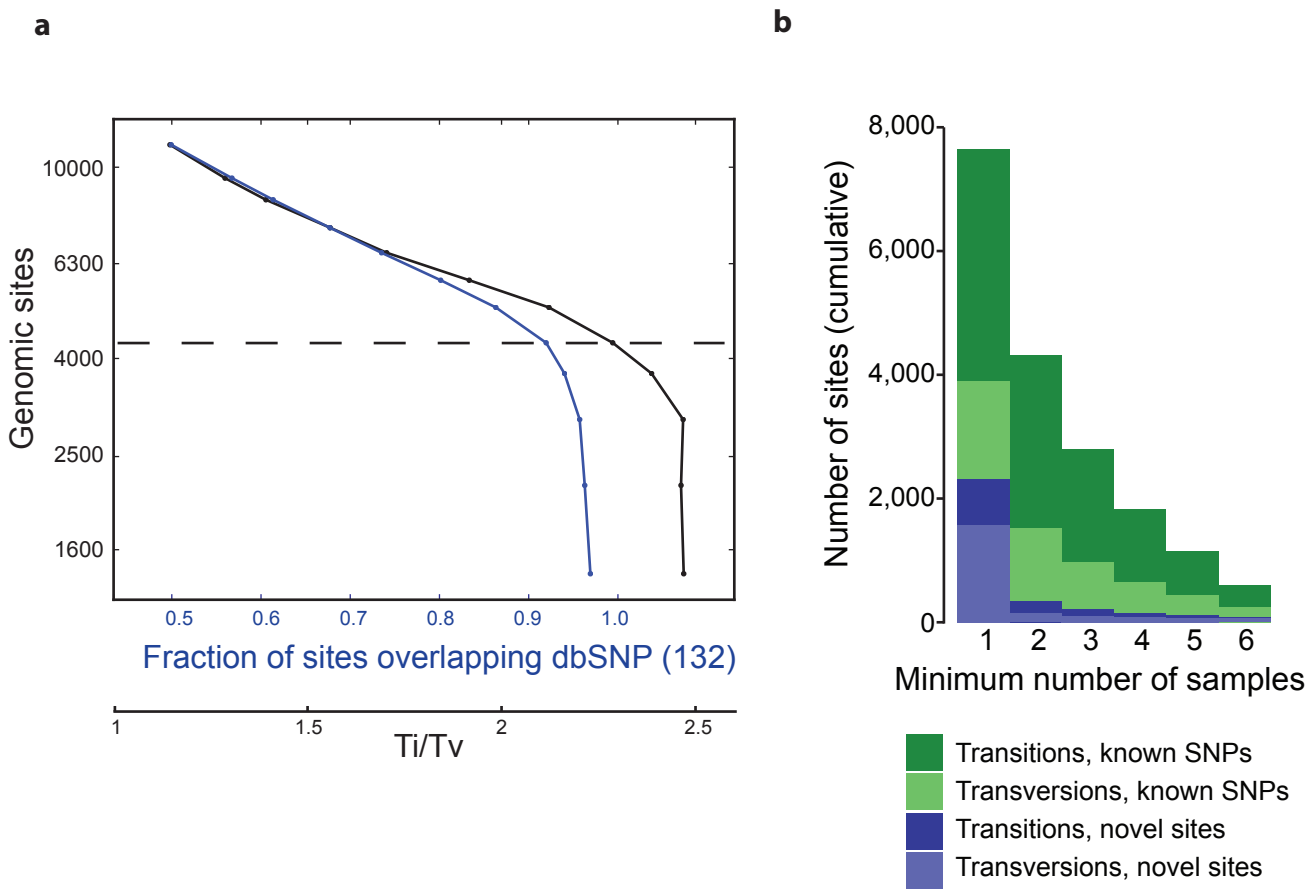


Supplementary Figure 9. Single-cell cDNAs from peripheral blood lymphocytes

Two individual peripheral blood lymphocytes (PBLs) were prepared using Smart-Seq and shallow sequenced on a MiSeq. These two cells (PBL) were compared with Smart-Seq transcriptomes from melanoma cancer cell lines (SKMEL5, UACC257), primary melanocytes (PM), circulating tumor cells (CTC), human embryonic stem cells (ESC) and mRNA-Seq transcriptome data from burkitts lymphoma cell lines (BL41, BJAB), white blood cells (W) and lymphnodes (L). Gene expression levels (RPKM, log2) for marker genes of melanocyte and hematopoietic lineages are shown in a heatmap. Despite a shallow sequence depth, we detected markers for hematopoietic lineage in both PBLs, in contrast to melanocyte lineage markers.



Supplementary Figure 10. Expression of all melanoma-associated antigens (MAGEs) Heatmap showing the expression levels of all MAGE genes in CTC, PM, melanoma cell lines (SKMEL5 and UACC257), and in human ESC and Burkitts lymphoma cell lines BL41 and BJAB (BL).



Supplementary Figure 11. Detection of SNPs and mutations in CTCs
(**a**) Genomic sites with support for allelic difference between single-cell CTCs and reference sequence were sorted according to the QUAL values (Phred scaled probability of polymorphism). For bins of genomic sites, we analyzed the fraction of sites that were already present in dbSNP (build 132) and the ratio of transitions to transversions (Ti/Tv). In genomic DNA re-sequencing studies, one expects to find ~90% of individual differences to be present in dbSNP and a Ti/Tv ratio of ~2.1 (genome-wide) or ~2.8 (exons). We found that sites with QUAL values of 500 or higher (dashed line) had 92% overlap with SNPdb and a Ti/Tv ratio of 2.3, suggesting that genotype calls above that QUAL threshold are reliable. At this threshold we found 4,312 genomic sites using transcriptome data from the six CTCs and requiring that the allele be supported in two or more CTCs. (**b**) The number of sites indentified requiring support in an increasing number of CTCs. The barplot shows the number of genomic sites with known or novel transitions and transversions. We found a large difference between sites supported in one or two cells, likely reflecting a large amount of false positives in calls from a single cell. Requiring support in two or more cells led to more consistent and slowly declining number of SNP or mutation calls that probably reflect power. The QUAL threshold used is 500 as indicated with the dashed line in (a).