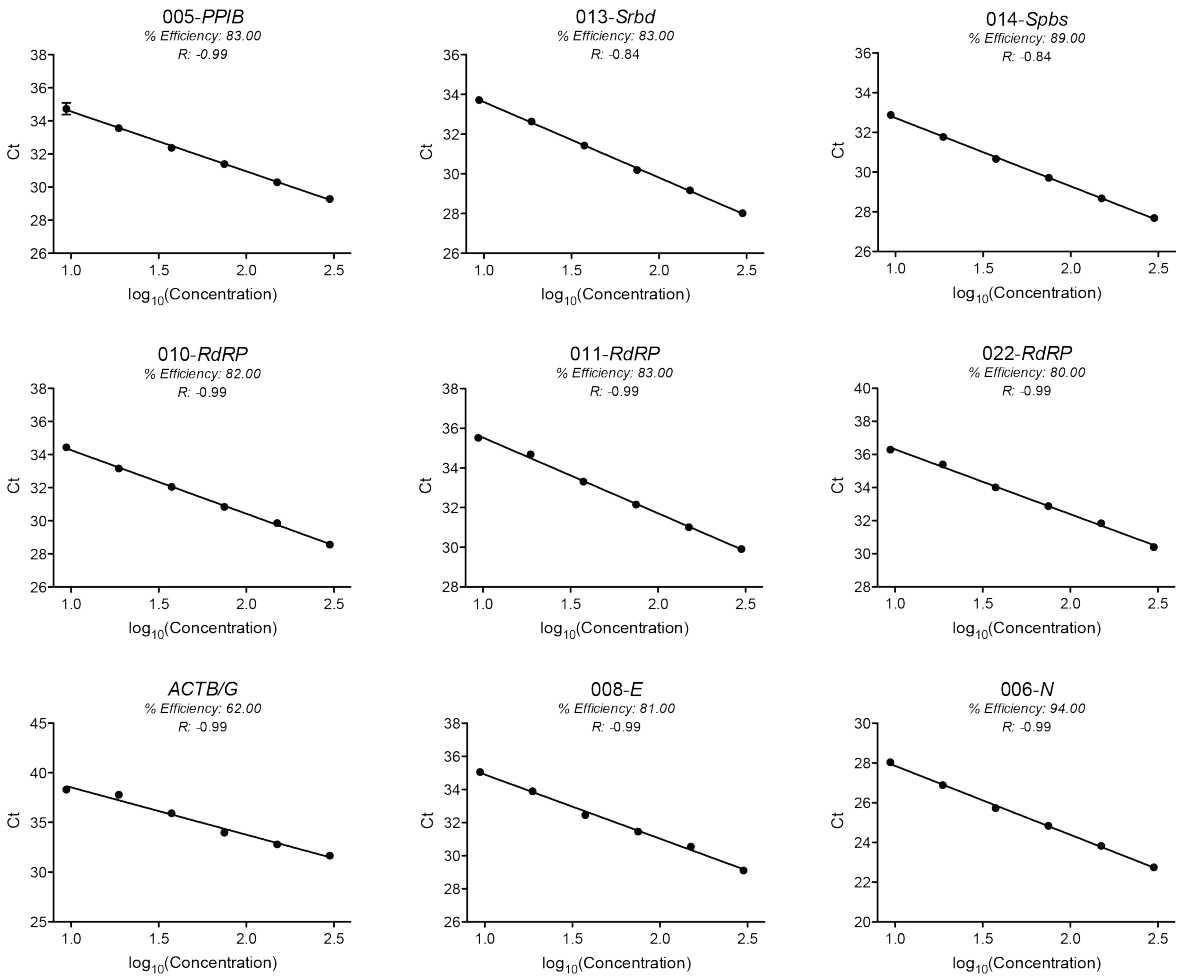
A Multiplexed, Next Generation Sequencing Platform for High-Throughput Detection of SARS-CoV-2

Marie-Ming Aynaud[1**], J. Javier Hernandez[1,3**], Seda Barutcu[1**], Ulrich Braunschweig[3,4], Kin Chan[1], Joel D. Pearson[1], Daniel Trcka[1], Suzanna L. Prosser[1], Jaeyoun Kim[1], Miriam Barrios-Rodiles[1], Mark Jen[1], Siyuan Song[4,5], Jess Shen[1], Christine Bruce[2], Bryn Hazlett[2], Susan Poutanen[2], Liliana Attisano[4,5], Rod Bremner[1], Benjamin J. Blencowe[3,4], Tony Mazzulli[2], Hong Han[3,4], Laurence Pelletier[1,3*], Jeffrey L. Wrana[1,3*].

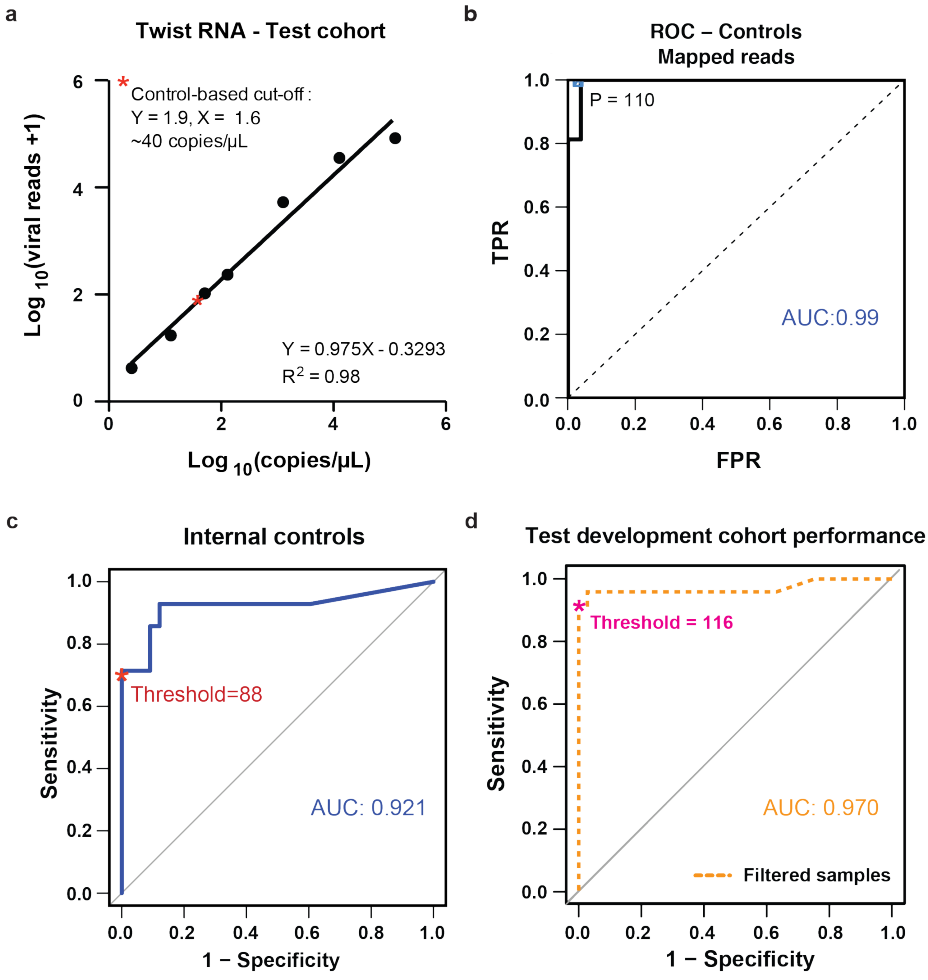
Extended data Figures



Extended data Fig. 1

Extended data Fig. 1: Efficiency of multiplex primers. Standard curve of Ct values (Y-axis) and $\log_{10}$(Concentration) (X-axis) of 6 limited dilutions of SARS-CoV-2^{high} sample (LTRI-18) for 9 pairs of primers (see Supplementary Table S1). Each condition was tested

in duplicate. Means are plotted for each point. The percent efficiency and the correlation (r) are calculated for each pair of primers after linear regression.



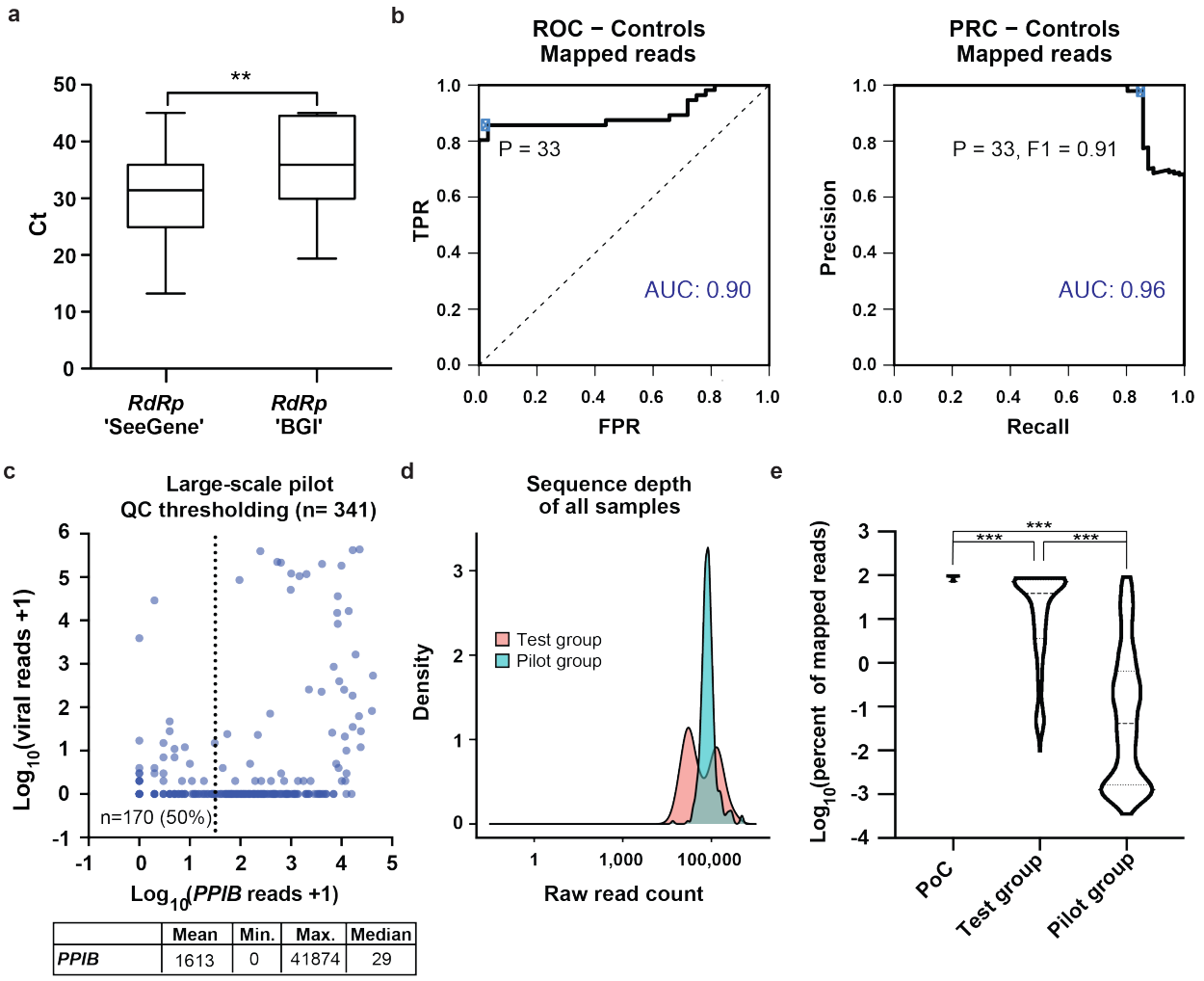
Cut-off selection

Optimal cut-off: 88 (PROC01)	*Value*	*Lower Limit*	*Upper Limit*
Sensitivity	*0.714*	*0.419*	*0.916*
Specificity	*1*	*0.894*	*1*

Extended data Fig. 2

Extended data Fig. 2: Using embedded controls as a training set for a control-based PR and ROC classifier. a, Total viral read counts are plotted against estimated viral copies (copies/µL) obtained using synthetic Twist SARS-CoV-2 RNA with statistics indicated. The cutoff defined by PROC analysis (see panel **c**) is marked with a red

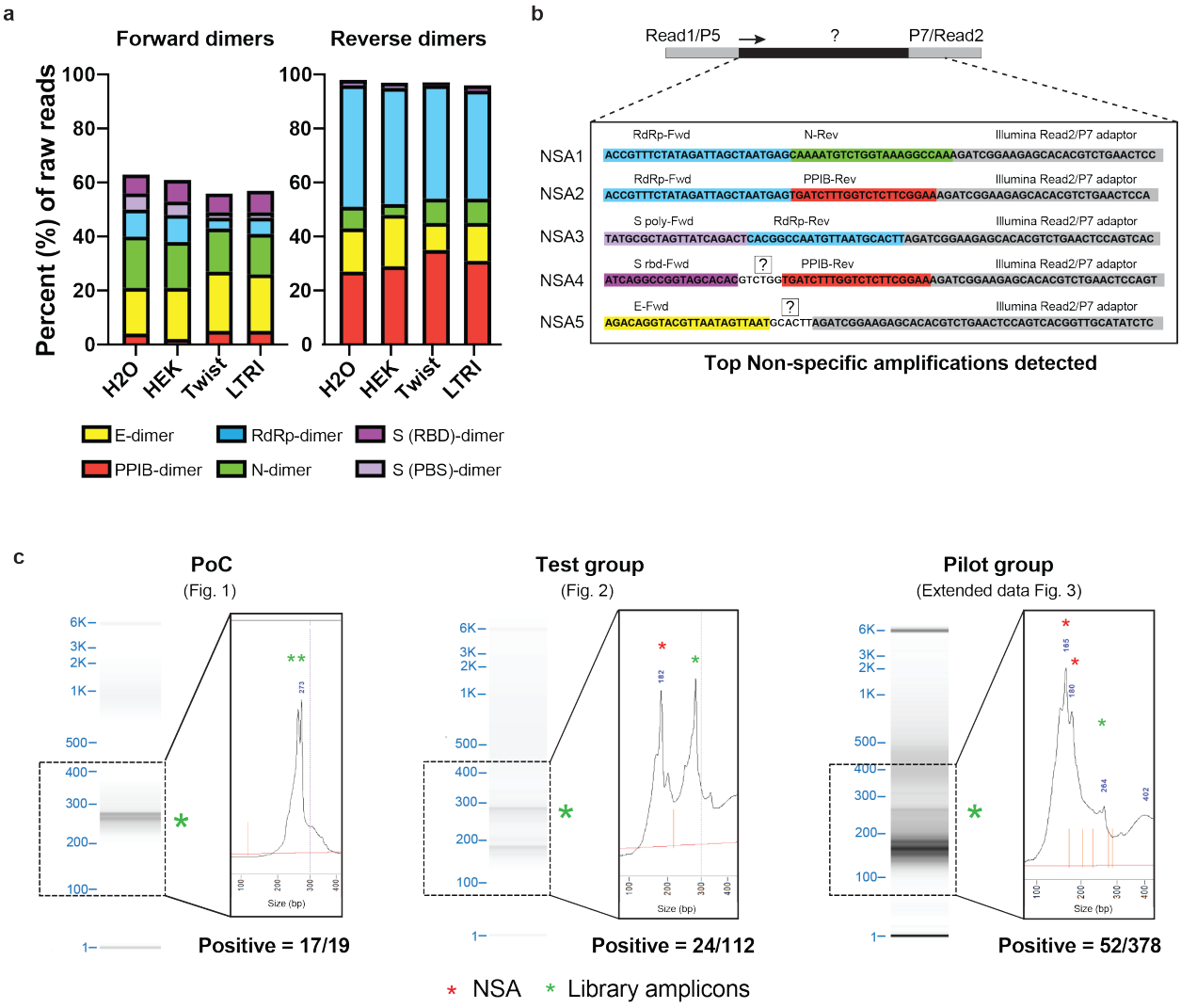
asterisk. **b**, Thresholding sample quality. coPR analysis on control samples: ROC of control samples for accurate detection of mapped reads are plotted. The optimal precision and recall read cut-off associated (P = 110) with the highest F1 (0.97) score, and AUC (area under the curve) is indicated on the ROC plot. **c**, Threshold for classification of positives in the test cohort. Total viral reads of negative (H2O and HEK293T) and positive (Twist dilutions) samples are used to calculate optimum cut-off by PROC and the defined threshold (P = 88) is plotted on the ROC curve. Values of sensitivity, and specificity at this cut-off are indicated (below). **d**, Performance of C19-SPAR-Seq. ROC analysis on patient samples that passed RNA-QC threshold was performed using clinical diagnostic results (Seegene Allplex qRT-PCR assay, **Supplementary Table 3**) and total viral reads for patient samples (n = 112). AUC is indicated on the graph.



Extended data Fig. 3

Extended data Fig. 3: Quality metrics assignment for the pilot cohort. a, Comparison of Ct (*RdRP*) values in 'SeeGene' *versus* 'BGI' tests of the positive archival samples. **b**, coPR analysis on control samples. ROC and PRC of control samples are plotted and the optimal precision and recall cut-off (P = 33) associated with the highest F1 score (0.91)

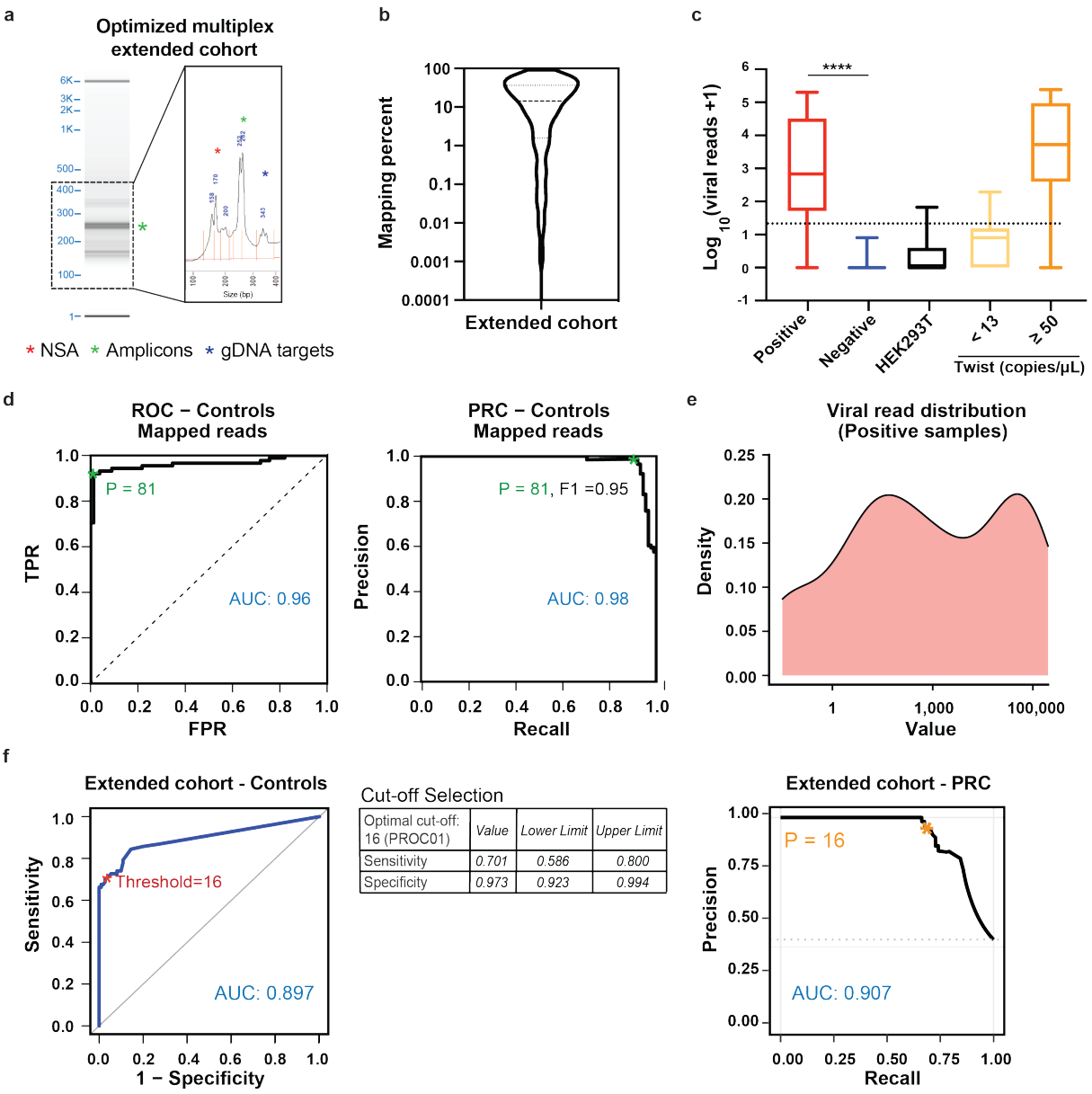
was calculated, as indicated in the PRC plot. **c**, coPR thresholding of the pilot cohort. Plot of total viral reads +1 (Y-axis) *versus PPIB* reads +1 (X-axis) of 341 patient samples in a pilot cohort (see Methods) is shown with the threshold (*PPIB* read counts > 33) to filter low-input samples marked. 170/341 (50%) samples were inconclusive (upper panel). Mean, minimum, maximum, and median values of *PPIB* and total viral read counts are indicated in the table (lower panel). **d**, Sequencing depth of test development and pilot cohort. Distribution density of raw read counts for the test development (pink) and pilot (turquoise) cohorts are shown. **e**, Read mapping percentages. Comparison of overall read mapping percentages between the PoC (**Fig. 1**), test (**Fig. 2**) and pilot cohort (n = 341). One way ANOVA - Tukey's multiple comparison test (****: $p < 0.0001$).



Extended data Fig. 4

Extended data Fig. 4: Non-specific amplification (NSA) in pilot cohort. a, Analysis of NSAs in the pilot cohort. NSAs contaminating the C19-SPAR-Seq library were quantified and percentage of reads mapping to the indicated forward and reverse primers are plotted. **b**, Schematic examples and sequences of the top 5 NSAs are shown. **c**,

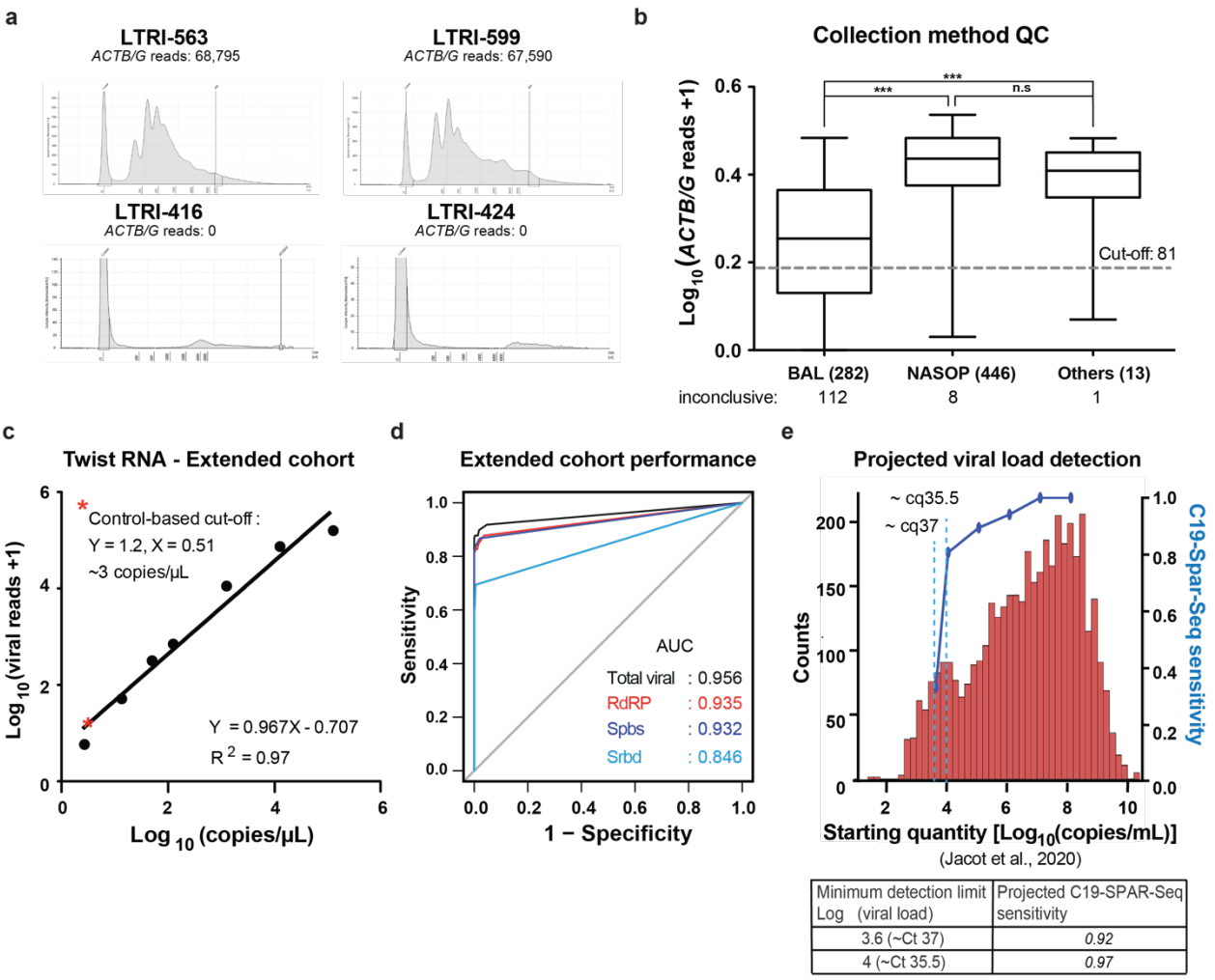
Comparison of fragment analyzer profile of the PoC, test development, and pilot cohort libraries after 0.8X SPRI bead purification. Fragment separation (DNA gel) and blow up view of the product abundance (electropherogram) are shown. Expected library amplicons (green stars) and non-specific amplicons (red stars).



Extended data Fig. 5

Extended data Fig. 5: Suppressing non-specific amplicons and quality metrics assignment for the extended cohort. a, Fragment analyzer profile of the extended cohort library using an optimized multiplex primer set targeting *ACTB/G*, *Spoly*, *Srbd*, and *RdRP*. Fragment separation (DNA gel) and blow up view of the product abundance

(electropherogram) is shown. **b**, Mapping percentage of the extended cohort. **c**, Overall distribution of total viral reads in the indicated positive samples (n = 98, red), negative samples (n = 444, blue), HEK293T (n = 21, black), synthetic SARS-CoV-2-RNA (< 13.2 copies/µL, n = 6, yellow), and synthetic SARS-CoV-2-RNA (>= 50 copies/µL, n = 30, orange) are plotted. Unpaired *t*-test of negative *versus* positive samples (****: $p < 0.0001$). **d,** coPR thresholding of sample quality and classification in the extended cohort. coPR analysis on control samples for sample quality yielded an optimal precision and recall read cut-off (P = 81) as indicated. **e**, Distribution of $\log_{10}$ total reads +1 of the positive (n = 98) samples. **f**, Threshold for classification of the extended cohort. ROC on control samples (HEK293T and synthetic SARS-CoV-2 RNA control) was assessed to identify an optimal cut-off (P = 16) for classifying patient samples. Performance on the controls is summarized.



Minimum detection limit Log (viral load)	Projected C19-SPAR-Seq sensitivity
3.6 (~Ct 37)	*0.92*
4 (~Ct 35.5)	*0.97*

Extended data Fig. 6

Extended data Fig. 6: C19-SPAR-Seq performance. a, RNA profile of BALs. RNA purified from ten BALs above and 10 below the QC threshold was profiled and two representative traces of each group are shown. *ACTB/G* reads are indicated for each sample. **b,** *ACTB/G* reads according to collection type. *ACTB/G* reads are plotted for each

collection type as a box and whisker (median ± 95% confidence interval, and the maximum and minimum values). The number of samples filtered by coPR (ACTB/G reads < 81) are indicated for each group. 1way ANOVA - Tukey's multiple comparison test (****: $p < 0.0001$, ns: non significative) **c,** Standard curve of total viral reads plotted against synthetic SARS-CoV-2 RNA concentrations obtained from C19-SPAR-Seq analysis of the extended cohort. **d,** ROC curve analysis was performed for each of the indicated viral amplicons and the AUC is shown. **e**, Projection of our C19-SPAR-Seq sensitivity onto the viral load data of ~4,000 patients from Jacot *et al.*, 2020 study[19]. Minimum detection limit and C19-SPAR-Seq sensitivity values are indicated in the table below.