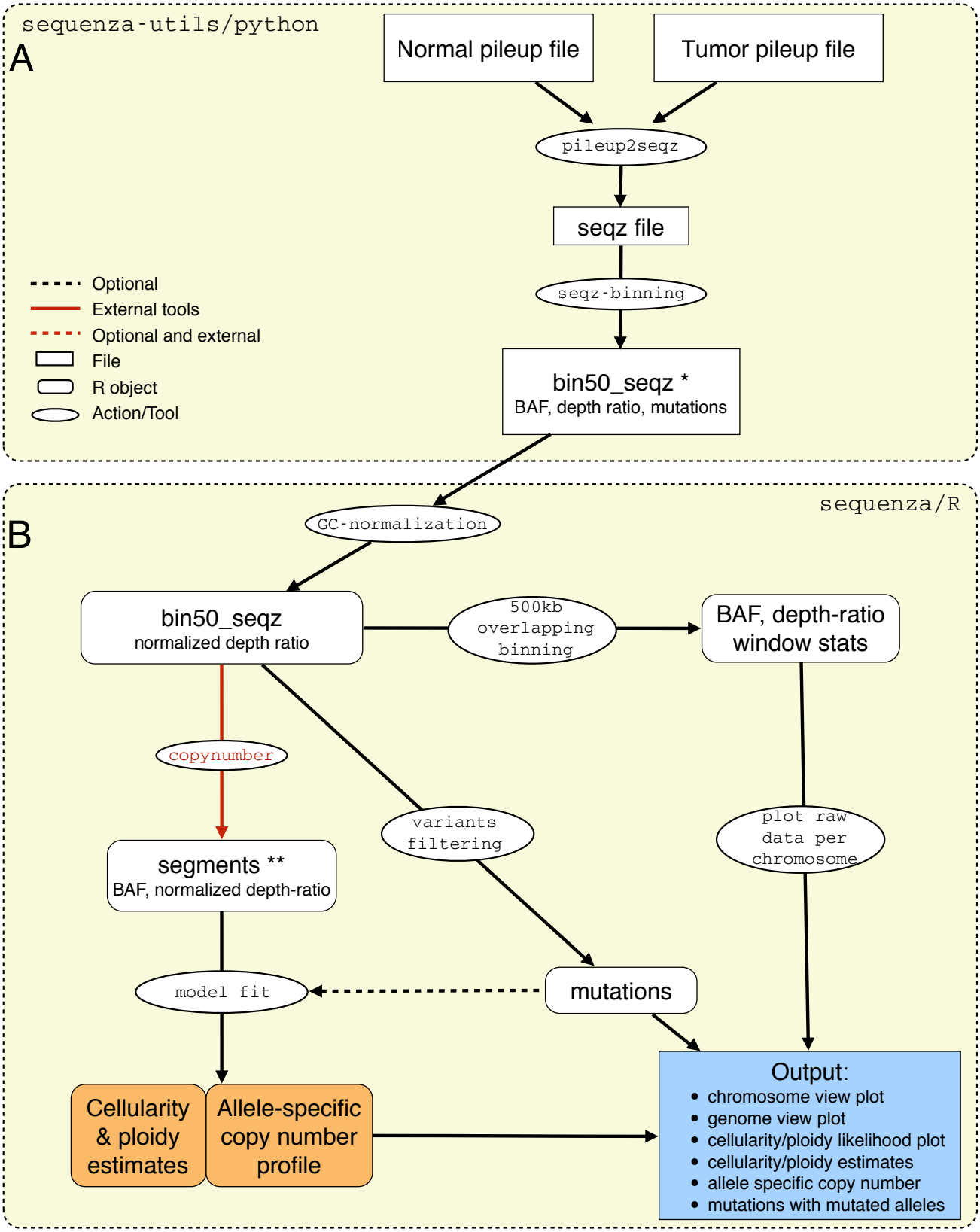

* A "seqz" file can also be obtained from the output of VarScan2 "somatic" and "copynumber" programs, using the R function "VarScan2seqz".
** Alternatively, breakpoint coordinates generated from other algorithms can be specified.

Figure S1: Detailed schema of sequenza workflow. The top panel (**A**) describes the necessary steps to obtain the input file from the sequencing data, using the provided sequenza-utils command line programs. The bottom panel (**B**) describes the steps implemented in the Sequenza R package workflow programs (*sequenza.extract*, *sequenza.fit* and *sequenza.results*) to analyse the input .seqz file.

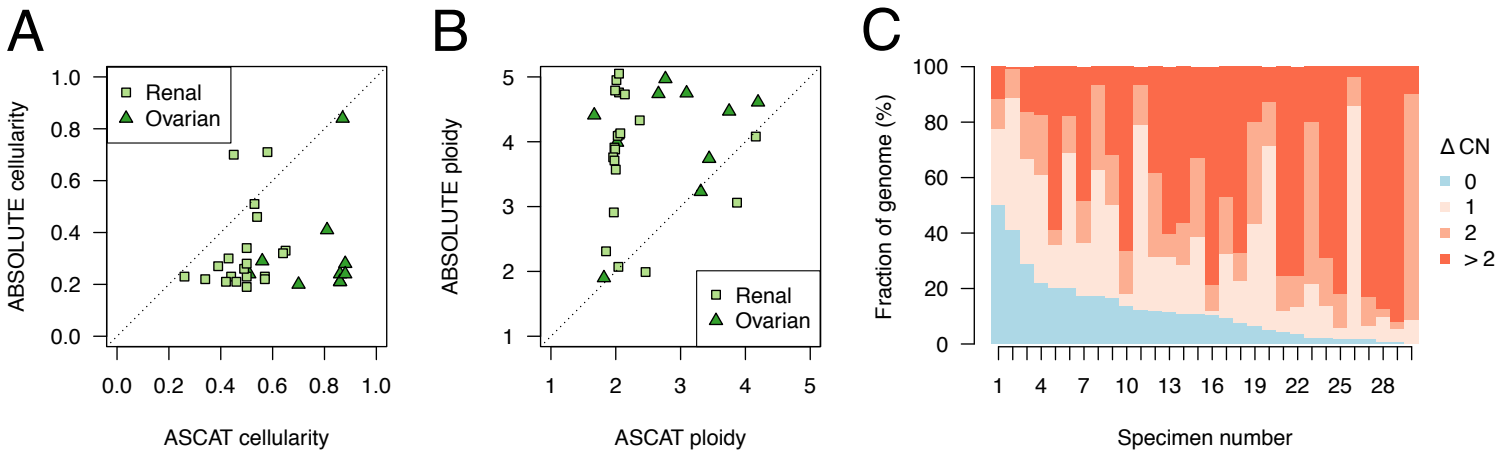

Figure S2: ABSOLUTE analysis of TCGA tumors. Comparison of cellularity and ploidy estimates and copy number profiles derived from exome sequence with ABSOLUTE to those derived from SNP array / ASCAT. See caption of figure 2 A-C for details.

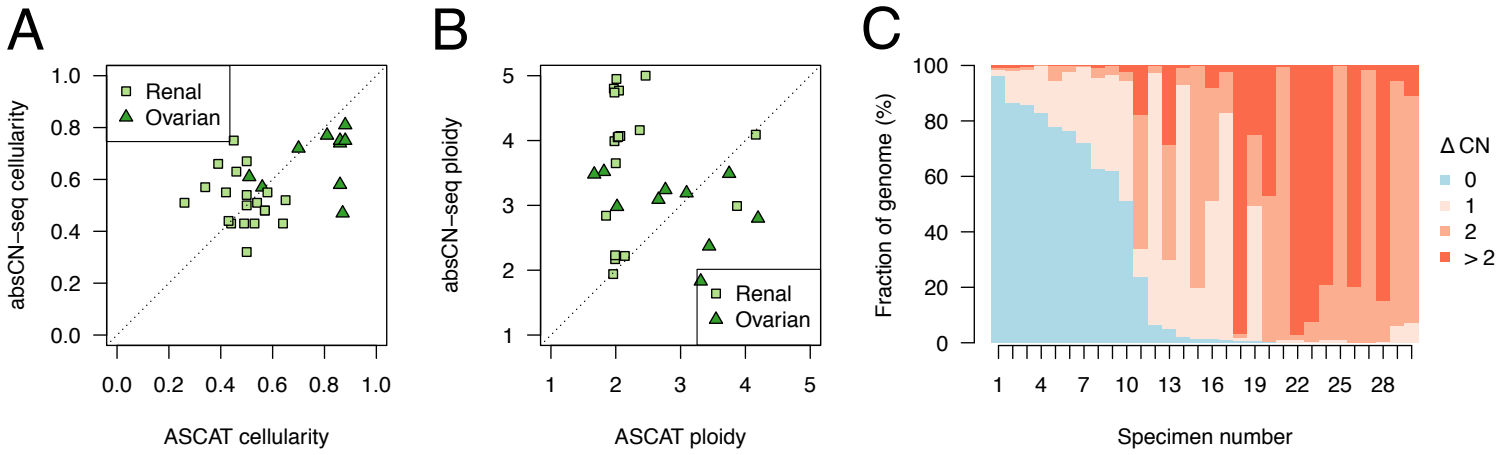

Figure S3: absCN-seq analysis of TCGA tumors. Comparison of cellularity and ploidy estimates and copy number profiles derived from exome sequence with absCN-seq to those derived from SNP array / ASCAT. See caption of figure 2 A-C for details.

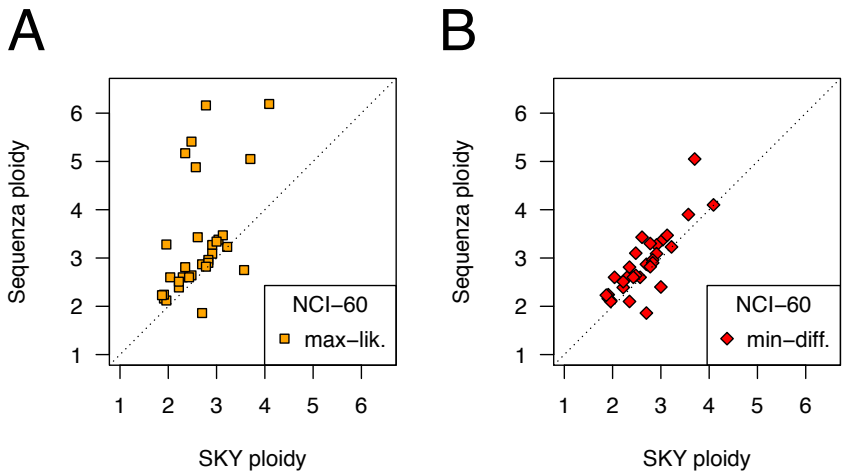

Figure S4: Sequenza analysis of 31 NCI-60 cell lines. Sequenza ploidy estimates are compared with the SKY-derived chromosome counts. (A) Maximum likelihood estimated ploidy estimates yield a RMSE of 1.20. (B) Manual inspection of the results identified 8 cell lines with non-optimal solutions; these were replaced with an alternative solution in agreement with the SKY ploidy, reducing the RMSE to 0.44**.**

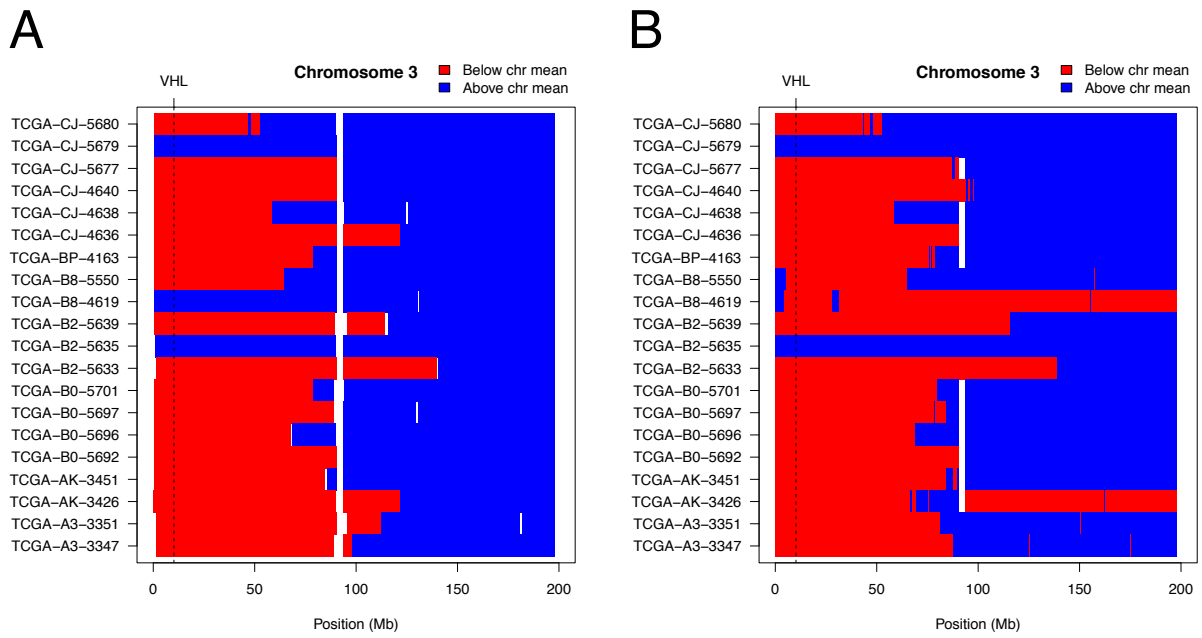

Figure S5:Occurrence of 3p loss in the KIRC dataset. Gains and losses in Chromosome 3 were analyzed in the 20 KIRC samples, using copy number estimates from (A) Sequenza using exome sequencing data, or (B) ASCAT using SNP-array data. The location of the VHL gene, which is frequently lost in renal clear cell carcinoma, is indicated. The colors indicate the copy number value relative to the entire chromosome: red indicates values smaller then the chromosomal mean; blue indicates values larger than the chromosomal mean.

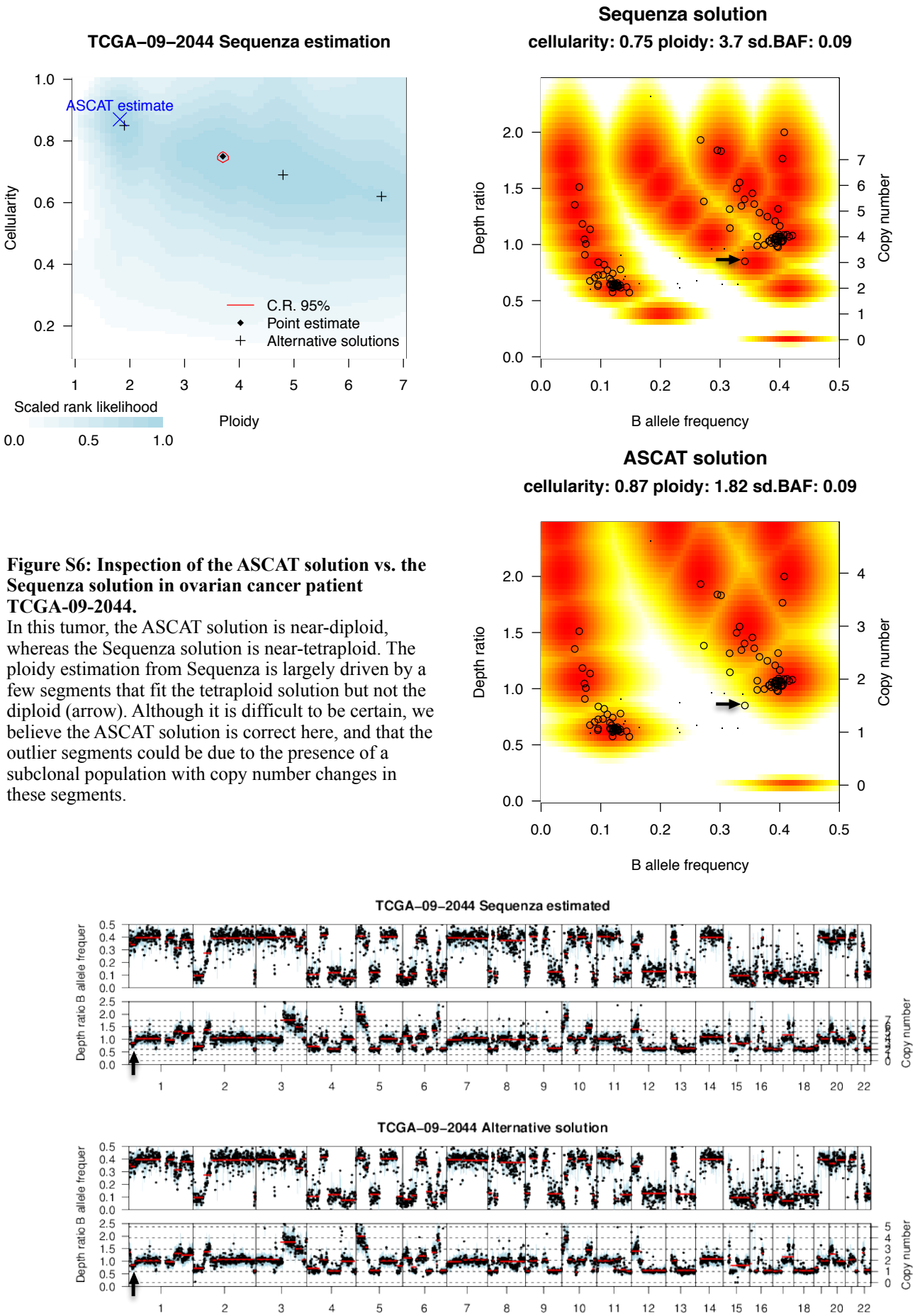

Figure S6: Inspection of the ASCAT solution vs. the Sequenza solution in ovarian cancer patient TCGA-09-2044.
In this tumor, the ASCAT solution is near-diploid, whereas the Sequenza solution is near-tetraploid. The ploidy estimation from Sequenza is largely driven by a few segments that fit the tetraploid solution but not the diploid (arrow). Although it is difficult to be certain, we believe the ASCAT solution is correct here, and that the outlier segments could be due to the presence of a subclonal population with copy number changes in these segments.

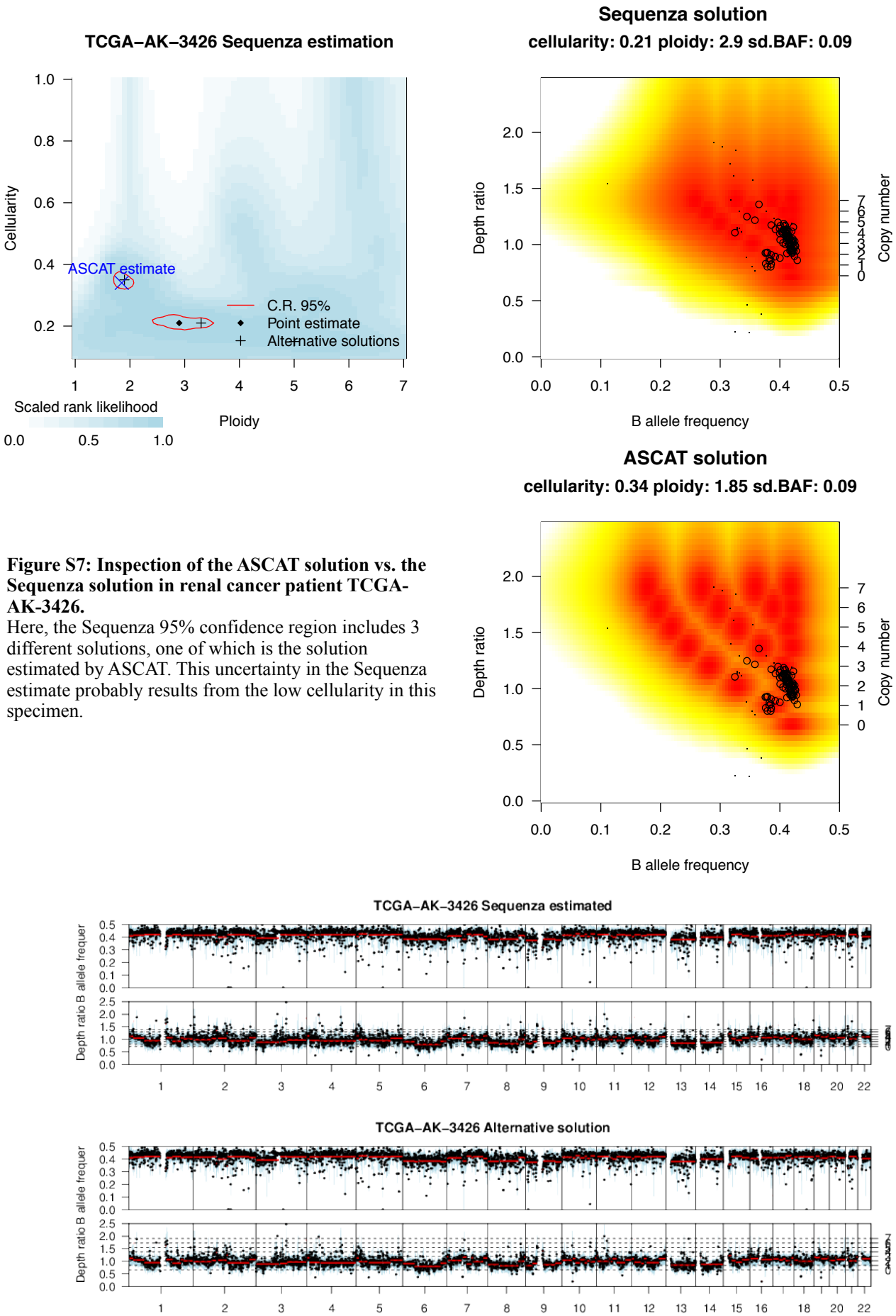

Figure S7: Inspection of the ASCAT solution vs. the Sequenza solution in renal cancer patient TCGA-AK-3426.
Here, the Sequenza 95% confidence region includes 3 different solutions, one of which is the solution estimated by ASCAT. This uncertainty in the Sequenza estimate probably results from the low cellularity in this specimen.

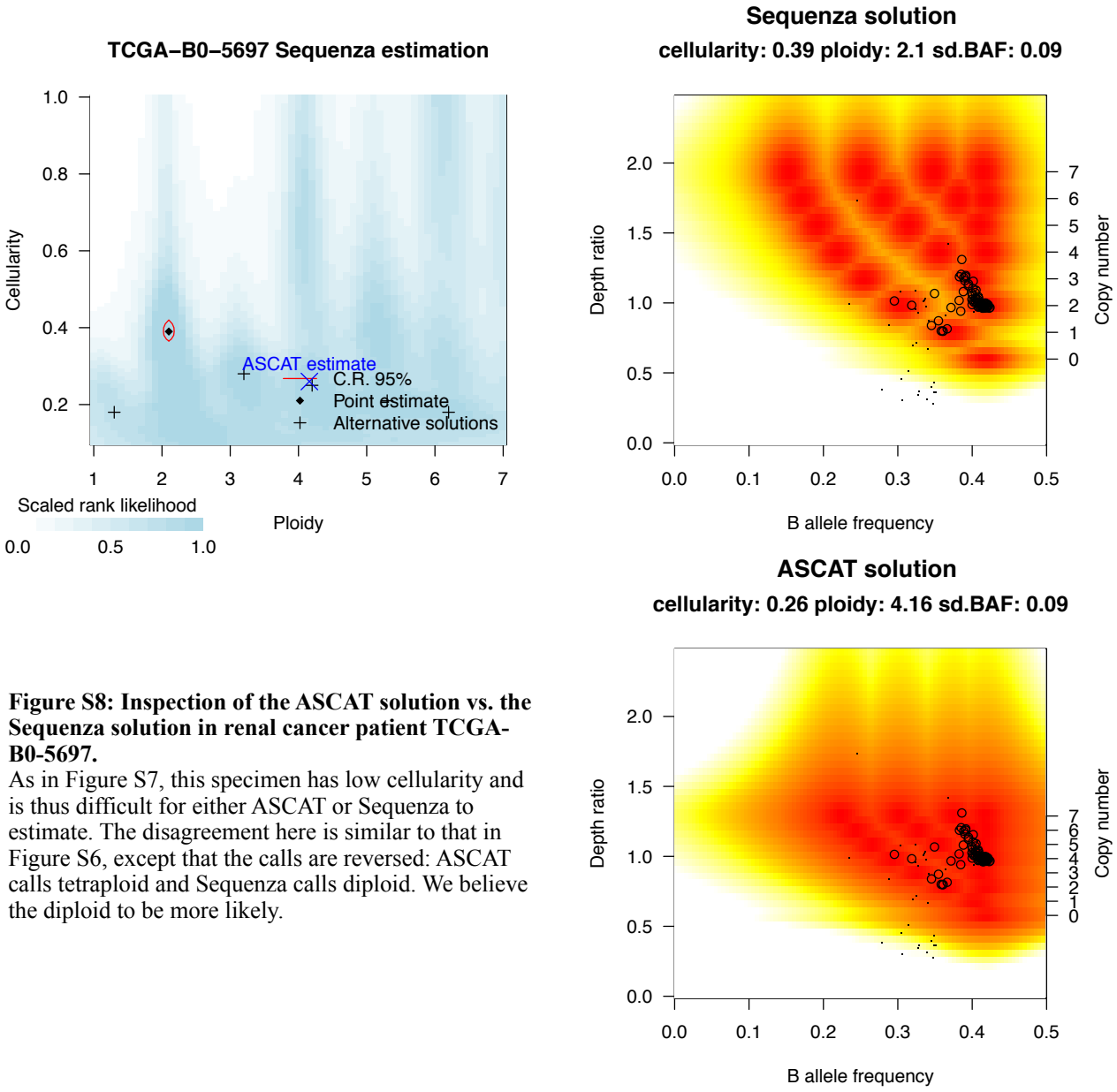

Figure S8: Inspection of the ASCAT solution vs. the Sequenza solution in renal cancer patient TCGA-B0-5697.
As in Figure S7, this specimen has low cellularity and is thus difficult for either ASCAT or Sequenza to estimate. The disagreement here is similar to that in Figure S6, except that the calls are reversed: ASCAT calls tetraploid and Sequenza calls diploid. We believe the diploid to be more likely.