Supplementary Materials for

Direct detection of early-stage cancers using circulating tumor DNA

Jillian Phallen[1], Mark Sausen[2], Vilmos Adleff[1], Alessandro Leal[1], Carolyn Hruban[1], James White[1], Valsamo Anagnostou[1], Jacob Fiksel[1], Stephen Cristiano[1], Eniko Papp[1+], Savannah Speir[1], Thomas Reinert[3], Mai-Britt Worm Orntoft[3], Brian D Woodward[4], Derek Murphy[2], Sonya Parpart-Li[2], David Riley[2], Monica Nesselbush[2], Naomi Sengamalay[2], Andrew Georgiadis[2], Qing Kay Li[1], Mogens Rørbæk Madsen[5], Frank Viborg Mortensen[6], Joost Huiskens[7,8], Cornelis Punt[8], Nicole van Grieken[9], Remond Fijneman[10], Gerrit Meijer[10], Hatim Husain[4], Robert B Scharpf[1], Luis Diaz[1+], Siân Jones[2], Sam Angiuoli[2], Torben Ørntoft[3], Hans Jørgen Nielsen[11], Claus Lindbjerg Andersen[3], and Victor E. Velculescu[1*]

*Corresponding author. E-mail: velculescu@jhmi.edu

This PDF file includes:

Other Supplementary Material for this manuscript includes the following:

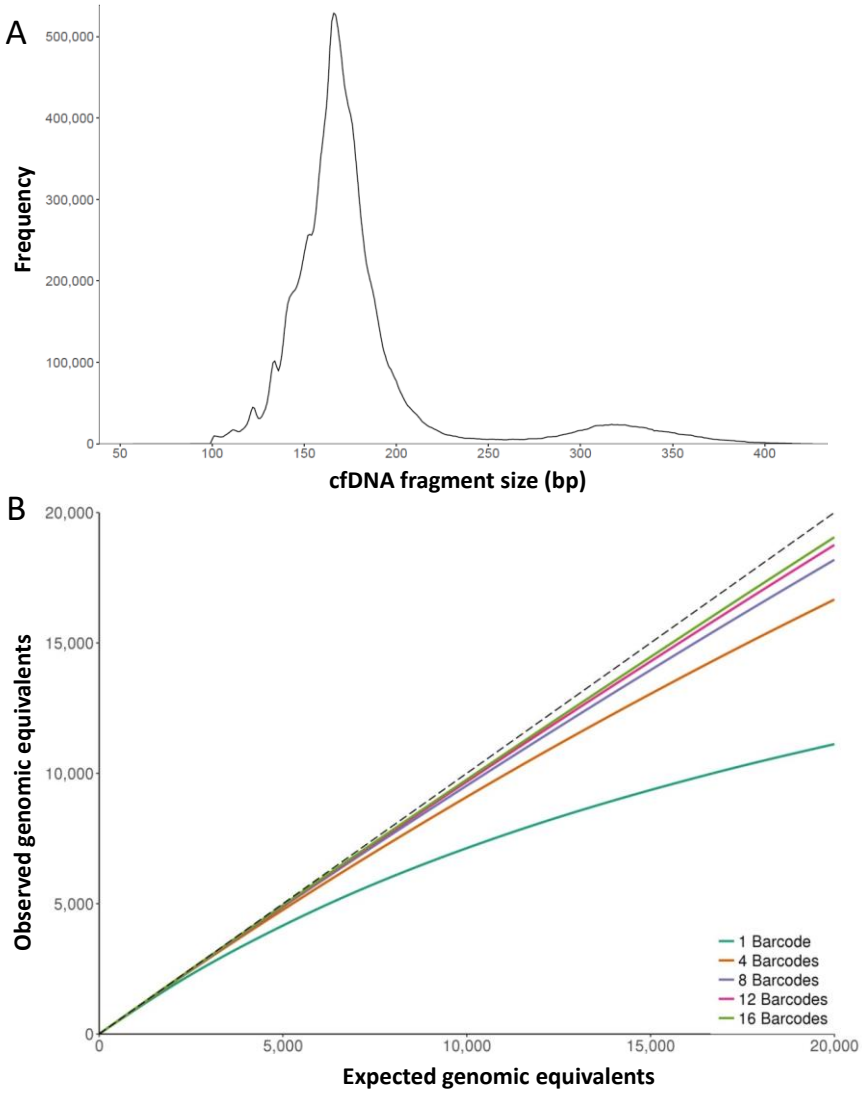

Figure S1. Simulations using limited exogenous barcodes. Monte Carlo simulations were performed to evaluate the effect of varying the quantity of exogenous barcodes on the number of genomic equivalents that can be distinguished through next-generation sequencing as compared to the expected number of genomic equivalents. **(A)** Representative distribution of cfDNA fragment sizes observed from sequencing data of ten colorectal cancer patients. **(B)** Using a sliding number of expected genomic equivalents (F), we sampled F fragment lengths from the distribution in **(A)** with replacement. Sample fragments were then randomly assigned start and end positions relative to an arbitrary base (x). Exogenous barcodes were randomly assigned to each fragment for simulation of 1, 4, 8, 12, and 16 barcodes. The number of observed genomic equivalents was then calculated from the unique combinations of endogenous (start and end position) and exogenous barcodes. These analyses indicate that a limited number of exogenous barcodes with endogenous barcodes improves the number of genome equivalents that can be analyzed at the sequencing depths typically utilized.

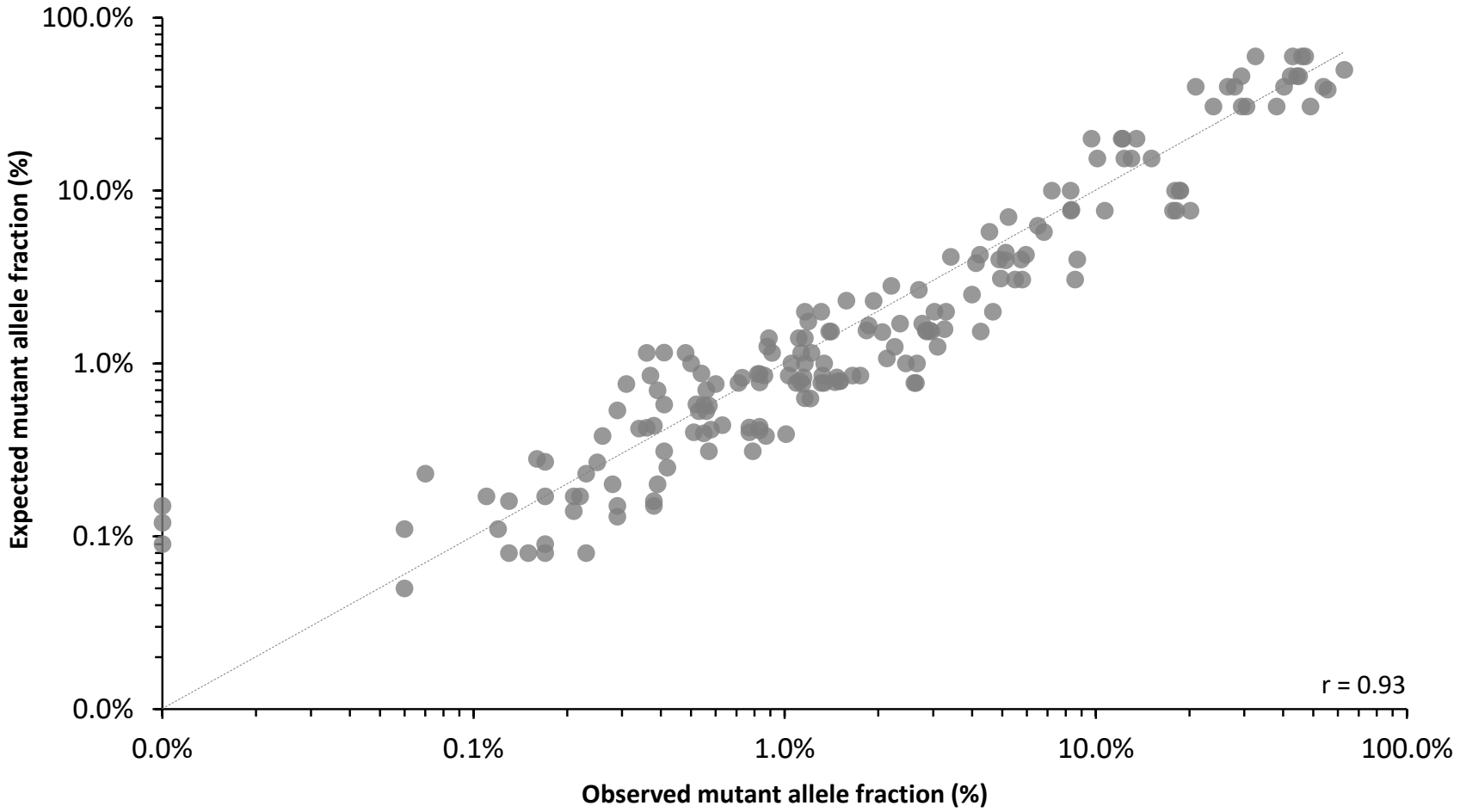

Figure S2. Validation of TEC-Seq approach. Correlation between observed and expected mutant allele fractions from mutant pools of tumor cell line DNA mixed with varying dilutions of genomic DNA (Pearson correlation: r=0.93, 95% CI=0.91–0.95, p<0.0001, r^2=0.87).

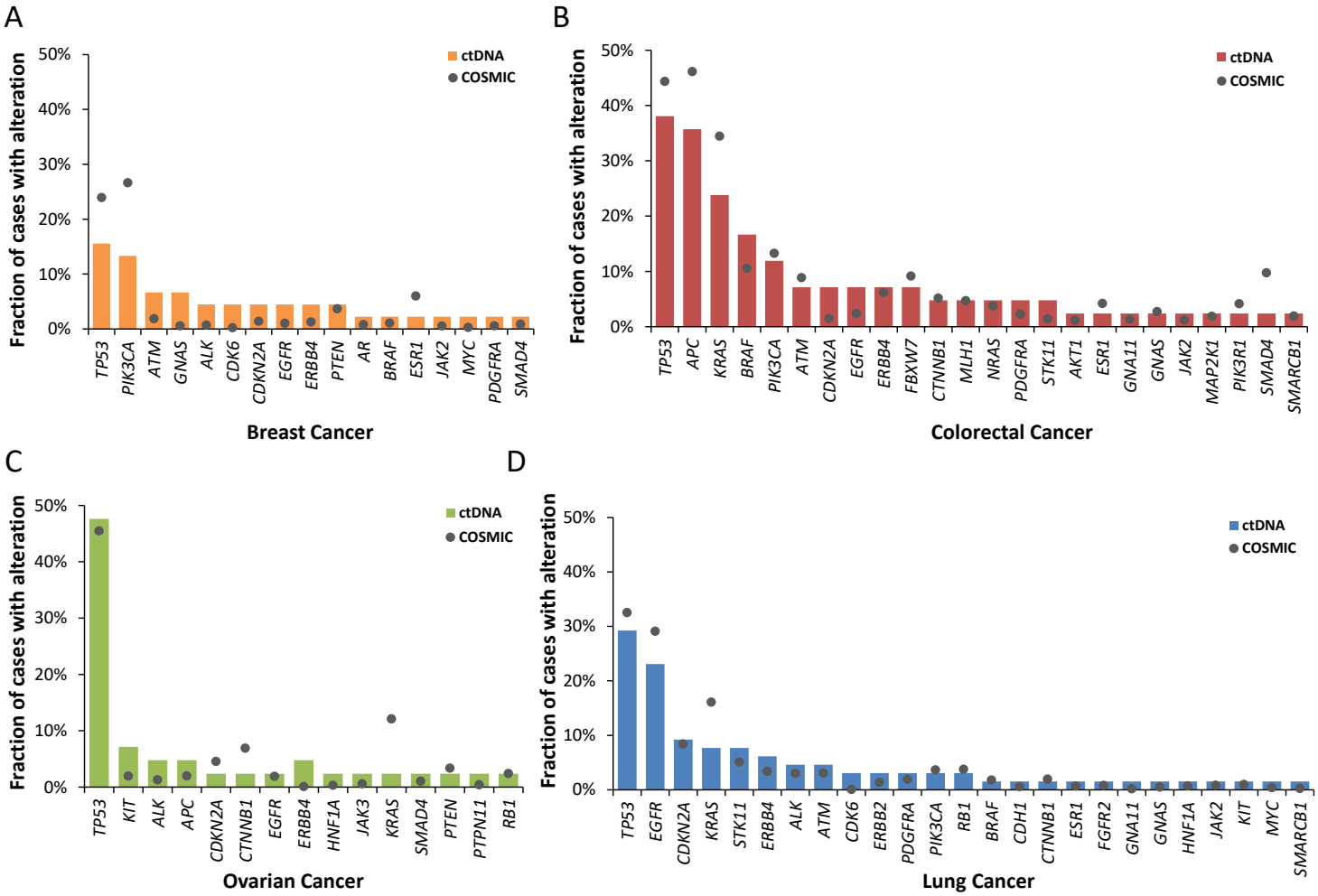

Figure S3. Mutation frequencies in cancer genes. Bar charts depict the fraction of patients with an alteration in a cancer driver gene observed in the plasma using TEC-Seq for breast **(A)**, colorectal **(B)**, ovarian **(C)**, and lung **(D)** cancer cohorts. The fraction of cancer cases reported in the COSMIC database with an alteration in the same genes is shown in the overlaid dot plot. The fraction of patients in our study and in the COSMIC database with an alteration in the genes of interest was similar for 75 out of 81 genes analyzed ($P>0.05$ for 75 of 81 genes, Fisher's exact test).

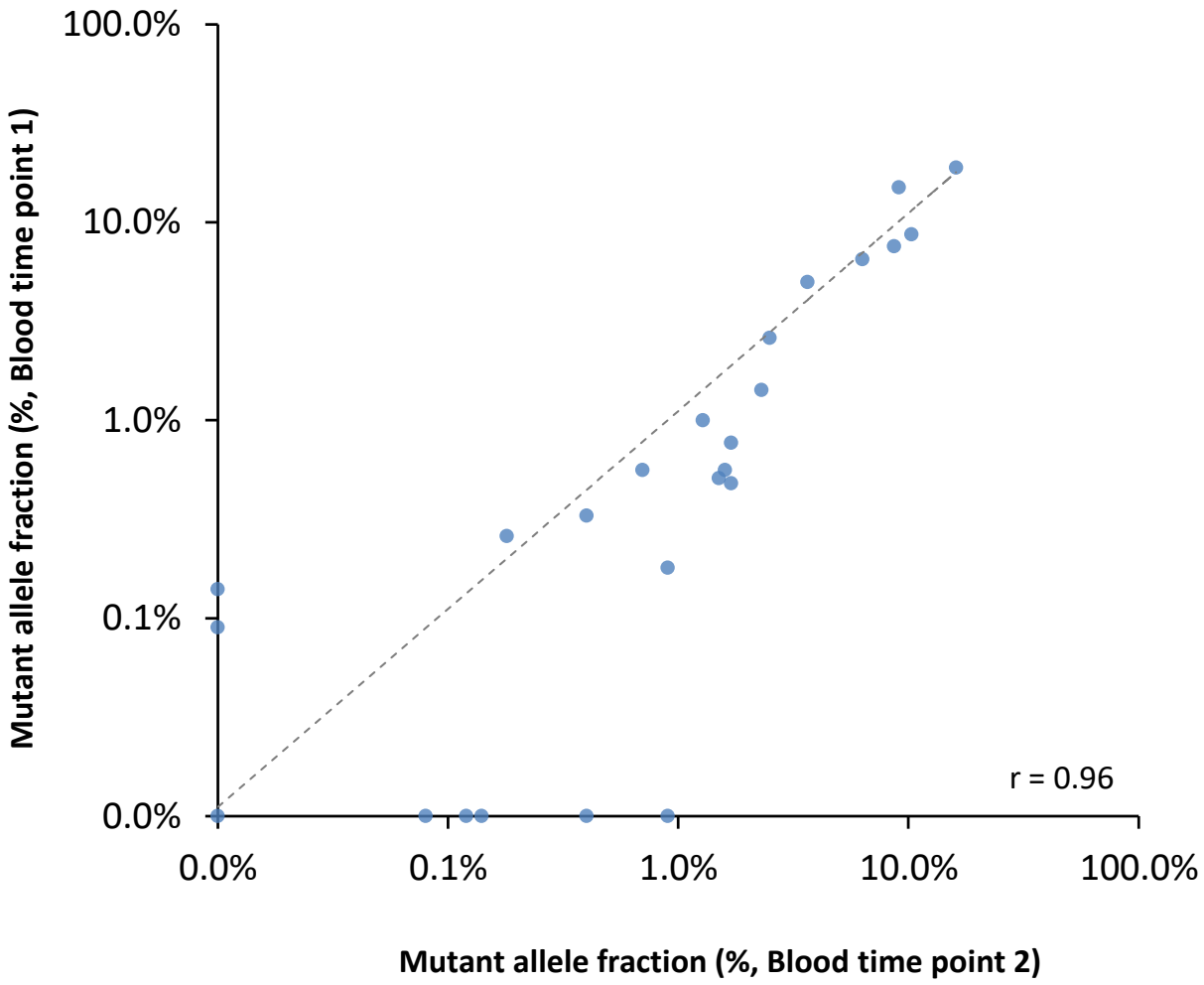

Figure S4. ctDNA mutant allele fractions in serial blood draws. Mutant allele fractions for alterations identified in two serial blood draws from six patients are indicated for each time point (Pearson r=0.96, 95% CI=0.92 – 0.98, p<0.0001, r^2=0.93).

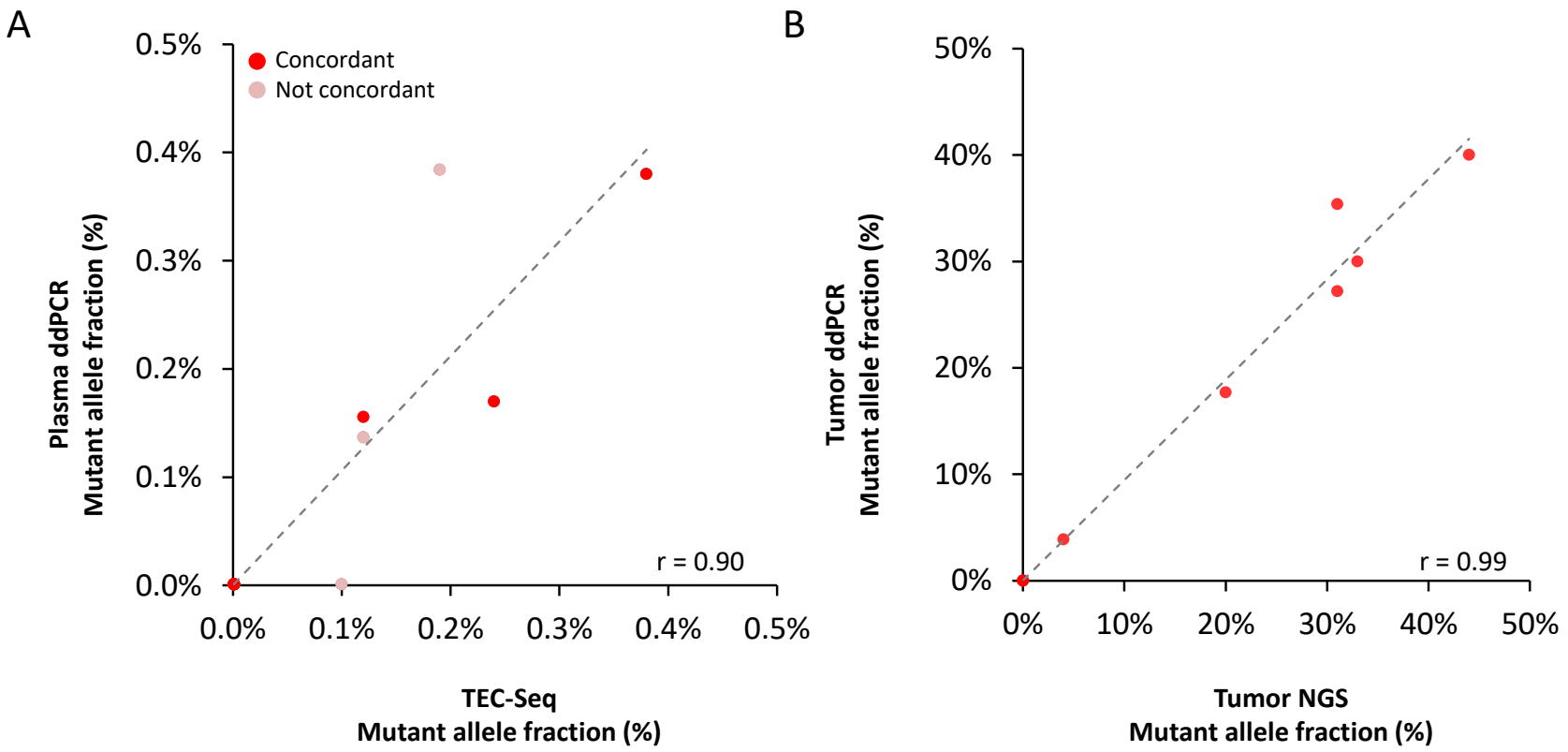

Figure S5. Comparison of ctDNA mutant allele fractions measured by TEC-Seq and ddPCR. Correlation of independent detection of alterations in cfDNA using ddPCR and TEC-Seq (**A.** Pearson r=0.90, 95% CI=0.72–0.96, p<0.0001, r^2=0.81) and in tumor tissue using ddPCR and conventional NGS (**B.** Pearson r=0.99, 95% CI=0.95–1.00, p<0.0001, r^2=0.98). Nine alterations in panel **A** were not detected by either plasma ddPCR or TEC-Seq. Alterations analyzed in the plasma **(A)** that were confirmed to be concordant with alterations in the matched tumor are indicated in bright red whereas alterations not concordant with those in the tumor appear in light red.

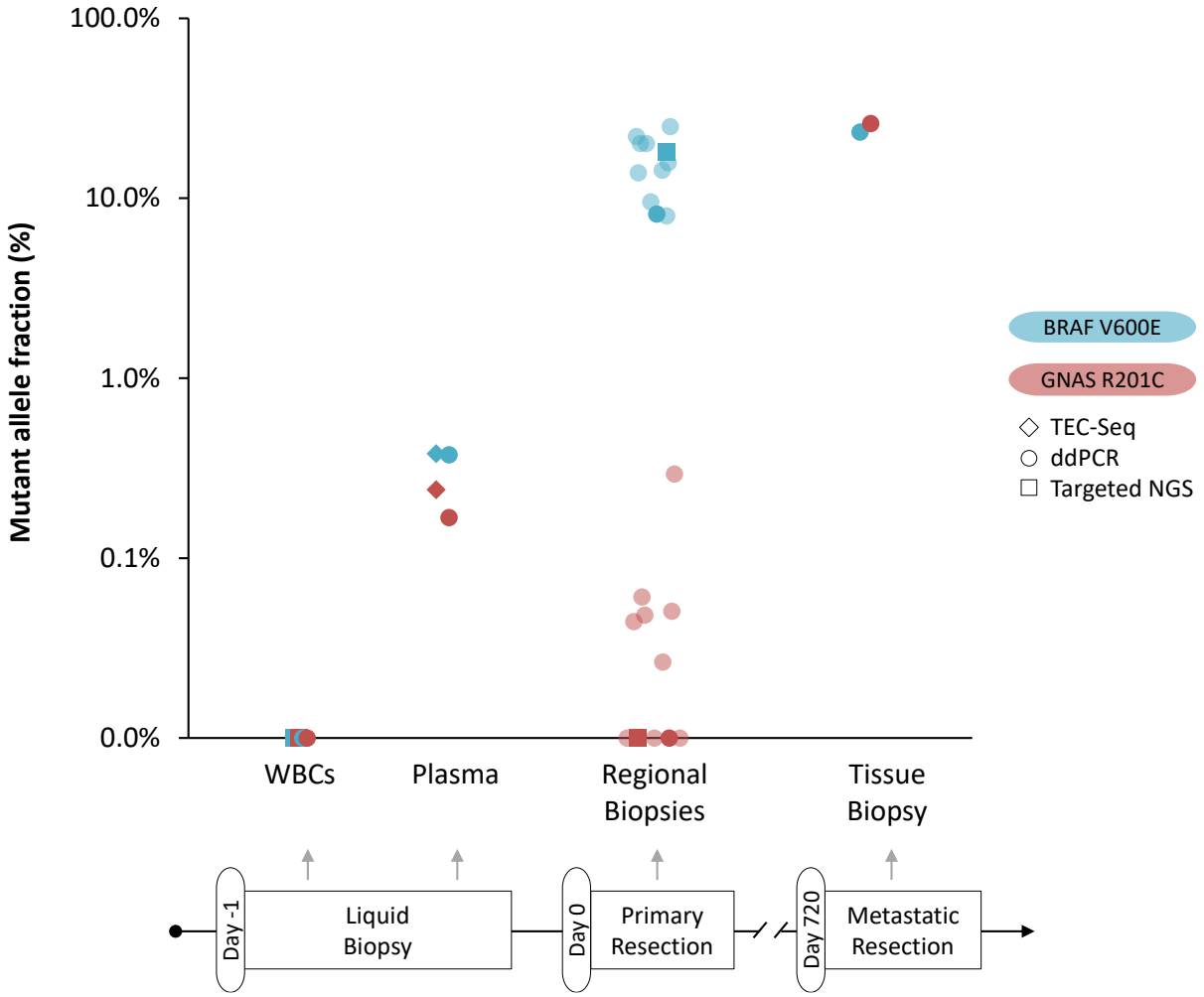

Figure S6. ctDNA and tumor heterogeneity. Analysis of two alterations, BRAF V600E (blue) and GNAS R201C (red), identified in a stage II CRC patient by three independent methods: TEC-Seq (diamonds), ddPCR (circles), and targeted NGS (squares). A liquid biopsy obtained one day before primary resection of the tumor yielded white blood cells and plasma for analysis of germline DNA and cfDNA, respectively. Both alterations were assessed in the white blood cells by targeted NGS and ddCPR, and in the plasma with TEC-Seq and ddPCR. Tissue from the primary resection was cored to obtain multiple biopsies, each analyzed separately by ddPCR for both alterations. Two analyses of one biopsy performed using targeted NGS and ddPCR are shown in darker shades compared to biopsies assessed by ddPCR alone. Tissue from a metastatic lesion was analyzed with ddPCR for both alterations. These analyses indicate that alterations identified in the plasma using TEC-Seq may represent heterogeneous changes that are present in only a portion of the primary tumor and/or occult lesions.

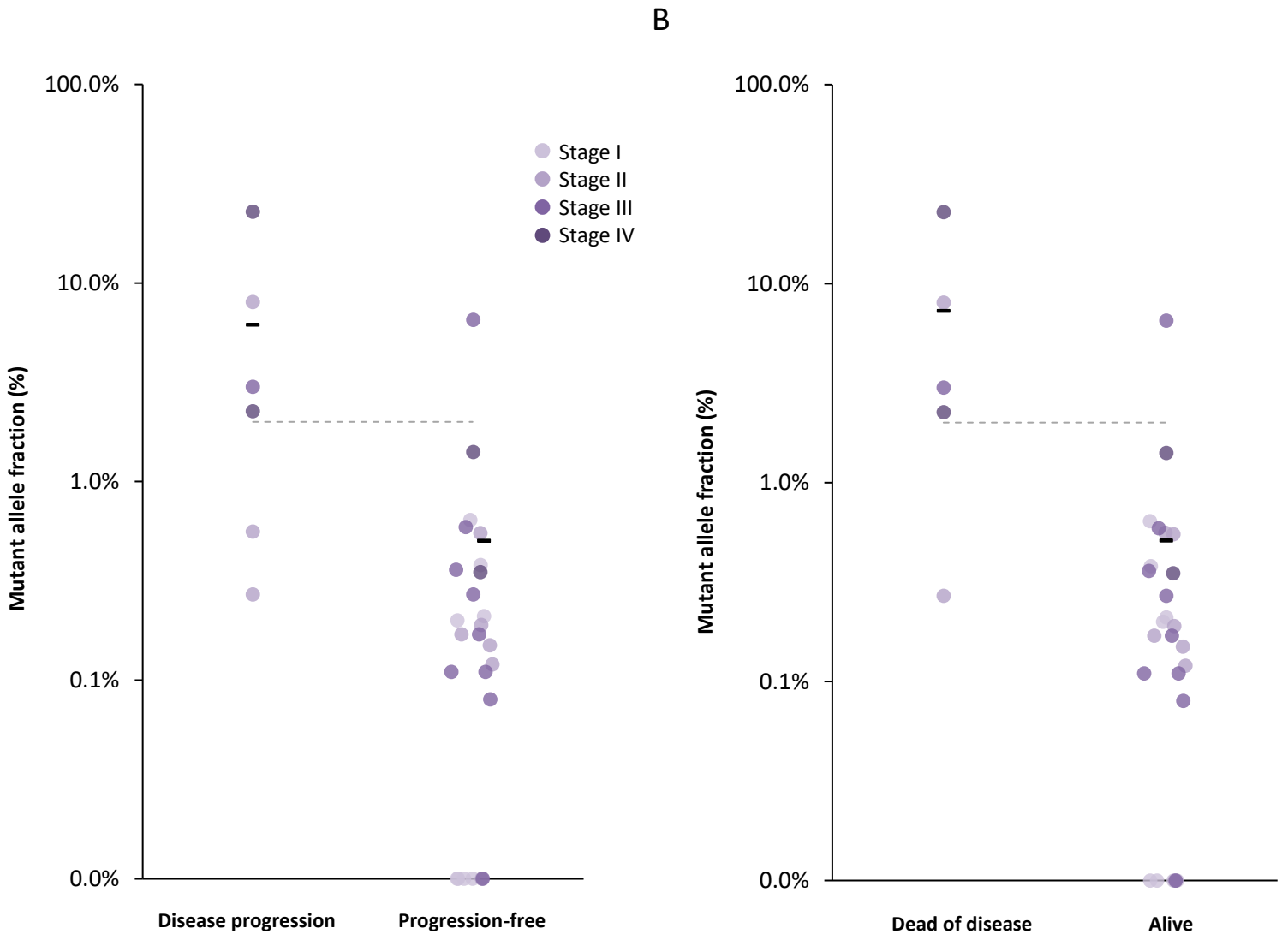

Figure S7. Pre-operative ctDNA mutant allele fractions in colorectal cancer patients. Mutant allele fractions of 31 CRC patients with stage I – IV disease organized based on progression-free survival status (**A**, P=0.0026, unpaired t test), and overall survival status (**B**, P=0.0006, unpaired t test). The dotted line represents a mutant allele fraction threshold of 2%.

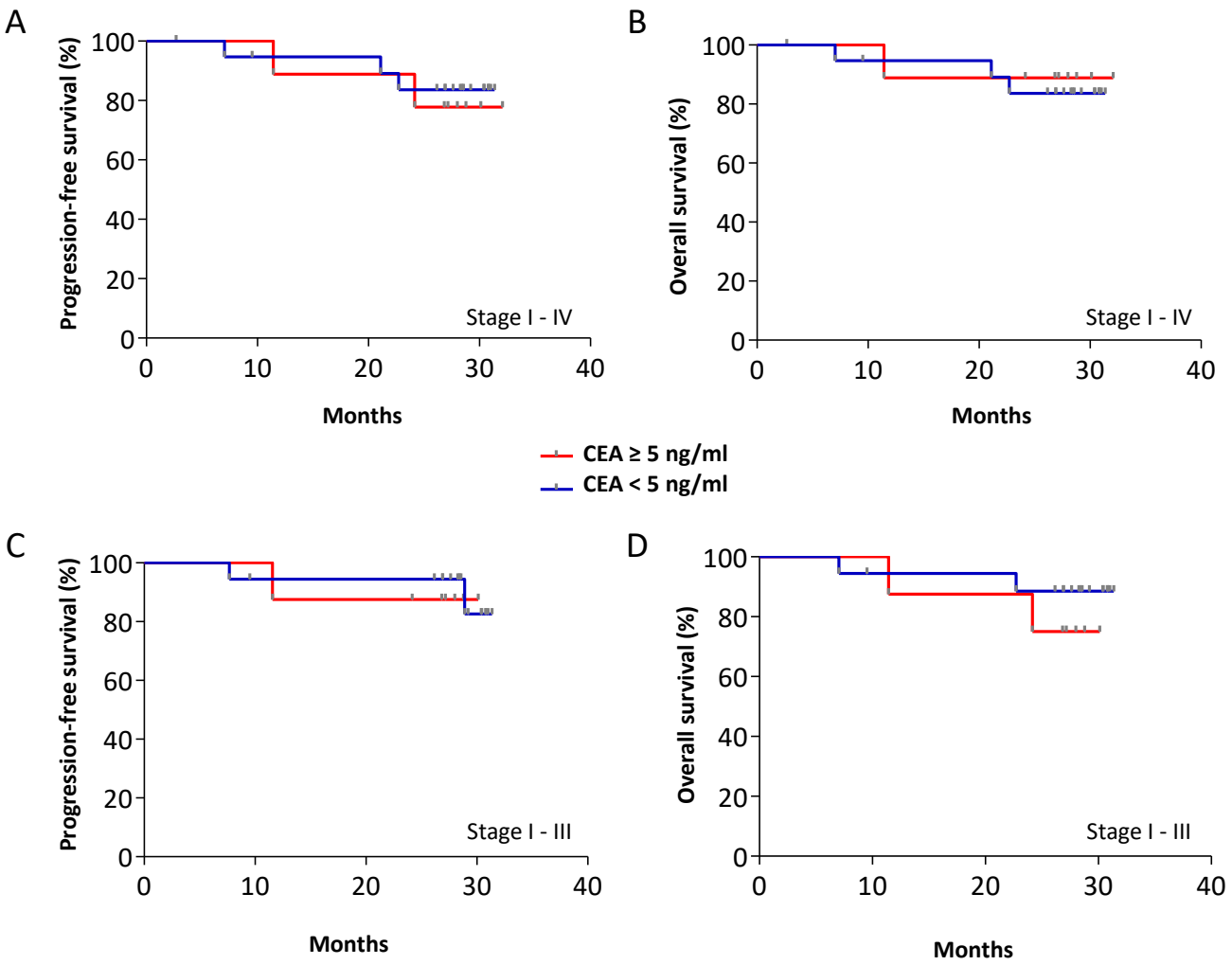

Figure S8. Pre-operative CEA in colorectal cancer patients. Kaplan-Meier curves depict progression-free survival (**A**, p = 0.7533, Log-rank test) and overall survival (**B**, p = 0.7329, Log-rank test) of 31 CRC patients, stage I – IV, stratified based on a CEA threshold of 5 ng/ml. Kaplan-Meier analyses of the 27 patients with stage I – III disease for progression-free survival (**C**, Log-rank test p = 0.4282) and overall survival (D, Log-rank test p = 7345) were performed using the same threshold in order to examine the association of CEA with outcome in patients without stage IV disease. Similar results were obtained using CEA thresholds of 2.5 ng/ml and 3 ng/ml.

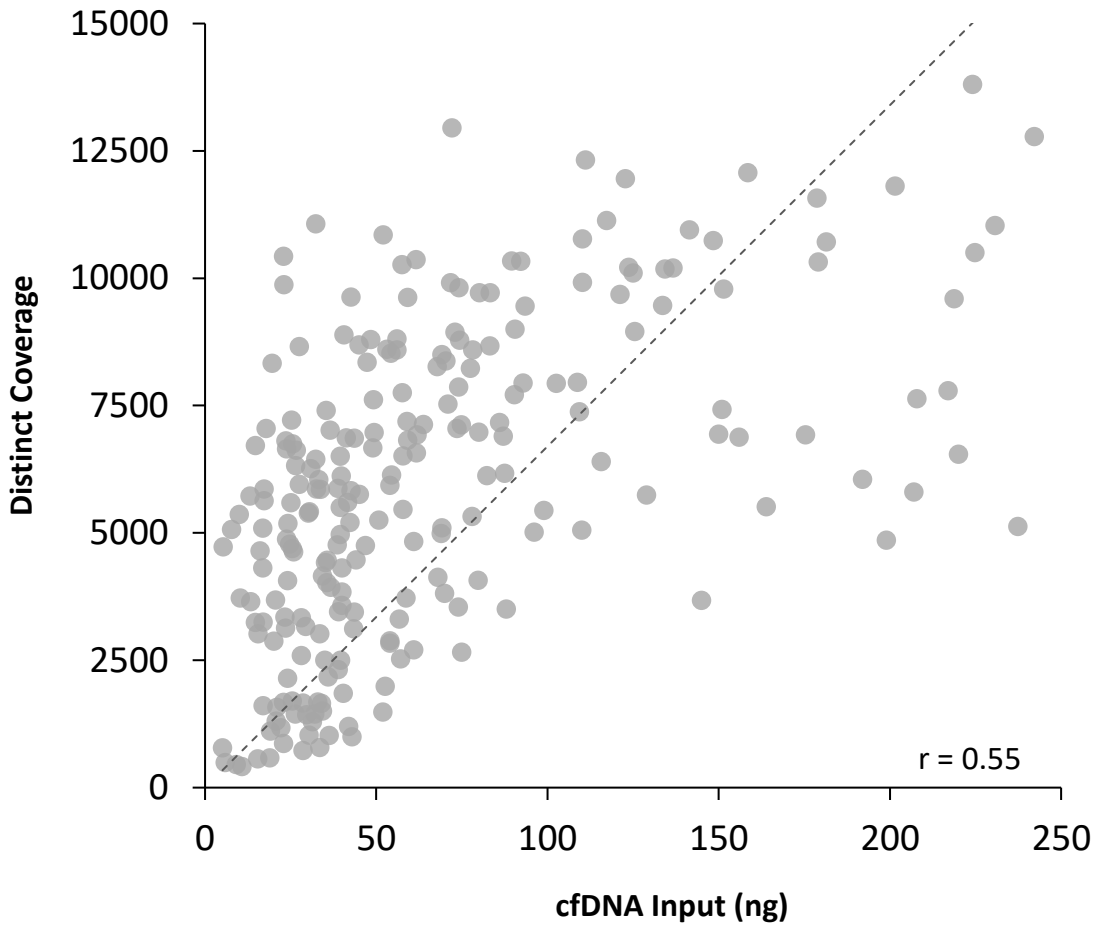

Figure S9. Conversion efficiency of cfDNA. Correlation of library input cfDNA with distinct sequencing coverage in cases with cfDNA <250 ng (n=230) (Pearson correlation: r=0.55, 95% CI=0.46–0.64, p<0.0001).