From Background to Signal: Simultaneously derived Copy Numbers from Methylation-Dependent Sequencing cell-free DNA

Daan M Hazelaar[1], Ruben G Boers[2], Joachim B Boers[2], Vanja de Weerd[1], Maurice PHM Jansen[1], Henk MW Verheul[1], Cornelis Verhoef[3], Joost Gribnau[2], John WM Martens[1], Stavros Makrodimtris[1,4,*], Saskia M Wilting[1,*]

[1]Erasmus MC, Medical Oncology, Erasmus MC Cancer Institute, University Hospital Rotterdam, The Netherlands.

[2]Erasmus MC, Developmental Biology, Erasmus MC Cancer Institute, University Hospital Rotterdam, The Netherlands.

[3]Erasmus MC, Oncological and Gastrointestinal Surgery, Erasmus MC Cancer Institute, University Hospital Rotterdam, The Netherlands.

[4]Delft University of Technology, Pattern Recognition and Bioinformatics, Technical University Delft, the Netherlands.

[*] contributed equally.

Abstract

Cell-free DNA (cfDNA) analysis offers a powerful, non-invasive approach to cancer diagnostics and monitoring by revealing tumor-specific genomic and epigenetic alterations. Here, we demonstrate the versatility of MeD-seq, a methylation-dependent sequencing assay, for comprehensive cfDNA analysis, including methylation profiling, chromosomal copy number (CN) alterations, and tumor fraction (TF) estimation. MeD-seq-derived CN profiles and TF estimates from 38 colorectal cancer with liver metastases (CRLM) and 5 ovarian cancer patients were highly comparable to shallow whole-genome sequencing (sWGS) validating our approach. These findings establish MeD-seq as a robust and cost-effective platform for detecting cancer-specific signals directly from plasma without prior tissue-based information. Future work should expand its application to other cancer types, solidifying MeD-seq as a versatile tool for minimally-invasive cfDNA profiling.

Introduction

Over the past decade, liquid biopsies i.e. the utilization of bodily fluids, has surged in cancer research. While not yet widely integrated into standard diagnostic or disease monitoring protocols, liquid biopsies hold significant potential for early detection and tracking of disease progression and treatment efficacy (Connal et al., 2023; Penny et al., 2024). Specifically, circulating cell-free DNA (cfDNA) fragments in the blood stream originating from dying cells provide a promising means to obtain crucial information on their cells of origin (Wan et al., 2017).

To distinguish between tumor-derived cfDNA (ctDNA) and cfDNA from healthy cells , one can focus on a number of general oncogenic hallmarks of ctDNA such as tumor-specific mutations (André et al., 2024), chromosomal alterations (Adalsteinsson et al., 2017), altered DNA methylation profiles (Y. Liu et al., 2024; Zhou et al., 2022), and so-called fragmentomics features (Mathios et al., 2021; Moldovan et al., 2024; Mouliere et al., 2018). Previous studies have highlighted the potential of cfDNA analysis in cancer research, demonstrating its utility for clinical predictions, cancer subtyping, and identifying biological signatures across diverse cancer types (Cristiano et al., 2019; Doebley et al., 2022). However, the limited amount of cfDNA in a blood sample restricts the range and number of assays one can perform on a single sample. Moreover, these assays are often costly, making it unfeasible to perform all of them even if sufficient material were available.

To overcome these limitations, bioinformatics efforts are being made to allow multimodal read-outs from individual assays with the goal to conserve patient material, reduce costs, and improve ctDNA detection (Katsman et al., 2022; J. Liu et al., 2024; Y. Liu et al., 2024; Nguyen et al., 2023; Thierry, 2023; Wang et al., 2023). Inspired by these initiatives, we here investigate whether we can extract additional, relevant information from the data generated by our previously described Methylation Dependent Sequencing (MeD-seq) assay for ctDNA methylation profiling (Deger et al., 2021).

MeD-seq uses a methylation-dependent restriction enzyme (LpnPI) to digest isolated, methylated (cf)DNA into 32bp fragments, followed by size selection and sequencing. The

sequenced fragments that contain an LpnPI recognition site at the expected position ("methylated reads") are then mapped to the human genome to associate each of them with a CpG site and construct a methylation profile (Boers et al., 2023). This approach offers some key advantages over bisulfite sequencing of cfDNA: a) the broad coverage of CpG sites (>50% genome wide), b) the lack of need for DNA conversion, c) its low DNA input requirement (5-10ng), d) its low cost (only 20-30M reads needed), and e) the short hands-on time (Boers et al., 2018; Deger et al., 2021).

The original MeD-seq computational pipeline focused on methylated reads and ignored reads that do not contain an LpnPI recognition site —so called background reads— as these are not directly informative for the methylation status of the DNA (Boers et al., 2018). In practice, this means only a median of 37% of all sequencing reads in cfDNA are used (Deger et al., 2021). We hypothesize that these background reads represent the sample's cfDNA fragment distribution before digestion and may thus provide information on the chromosomal coverage of the sample. Here, we aim to utilize background reads from MeD-seq data to construct sample-specific chromosomal Copy Number (CN) profiles and estimate Tumor Fraction (TF).

Materials and Methods

Patient and control samples

We used 132 blood samples from five consenting patients with metastatic ovarian cancer and 113 consenting patients with colorectal liver metastases (CRLM; 43 female, 70 male). Of these 132 samples, 113 were collected preoperatively (T0) and 19 postoperatively (T3). The clinical studies from which the used samples were obtained were both approved by the Medical Ethics Committee of the Erasmus University Medical Center (MEC 15-289, and MEC 17-238). In addition, 10 ml of blood was obtained from 51 consenting anonymous healthy blood donors (HBDs; 22 females and 29 males) via the Dutch National blood bank (Sanquin).

All blood samples were collected in CellSave tubes and plasma was isolated from the obtained blood within 96 h after blood draw by 2 sequential centrifugation steps at room temperature (10 min at 1711 g followed by 10 min at 12,000 g) and stored at −80 °C (van Dessel et al., 2017). cfDNA from plasma was extracted by using the QiaAmp kit (Qiagen) according to manufacturer's instructions to isolate cell free DNA from plasma. cfDNA concentration was determined using the Quanti-IT dsDNA high-sensitivity Assay (Invitrogen) and stored at -30 °C until further handling.

We performed sWGS and MeD-seq on cfDNA isolated from the same blood sample for five ovarian cancer patients and 38 samples from 19 CRLM patients (6 female, 13 male). For the remaining 94 CRLM patients, only MeD-seq analysis was conducted on the pre-operative sample. For the HBDs, sWGS was performed on cfDNA from 20 individuals (10 female, 10 male), while MeD-seq analysis was performed on a different set of 31 healthy individuals (12 female, 19 male).

MeD-seq and sWGS analyses of cfDNA

MeD-Seq was performed as previously described (Boers et al., 2018; Deger et al., 2021). In short, 10ng of cfDNA was digested with LpnPI (New England Biolabs, Ipswich, MA) yielding 32 bp fragments around the fully methylated recognition site containing a CpG. Samples were prepped for sequencing using the ThruPLEX DNA-seq 96D kit and the ThruPLEX DNA-Seq HV kit (Rubicon Genomics, Takara Bio Europe, Saint-Germain-en-Laye, France), for

CRLM and ovarian cancer samples respectively, and purified on a Pippin HT system with 3% agarose gel cassettes (Sage Science, Beverly, MA). Libraries were multiplexed and sequenced until a total of ~20M reads on either an Illumina HiSeq 2500 for 50 bp single-end reads (CRLM samples) or an Illumina NEXTseq2000 (ovarian cancer) for 50 bp paired-end reads, according to the manufacturer's instructions (Illumina, San Diego, CA).

For sWGS, indexed sequencing libraries were prepared using the ThruPLEX DNA-Seq HV kit with an input of 5 ng of cfDNA per sample. Libraries were sequenced to a median 1.0x depth of coverage on an Illumina NEXTseq2000, generating at least 10M 150 bp paired-end reads.

Quality control, pre-processing and read alignment

For quality control of sWGS and MeDseq FASTQ and BAM files, we utilized a combination of tools. FastQC (v0.12.1) and MultiQC (v1.21) were used to assess the quality of the raw sequencing data. Samtools (v1.20) stats, coverage, flagstat, and idxstats were employed to evaluate various quality metrics of the aligned BAM files, including alignment statistics, coverage, flag information, and index statistics, respectively. We used Fastp (v0.23.4) for adapter removal and deduplication, and BWA-MEM (v0.7.18) with soft-clipping enabled to align reads to the hg38 reference genome. Next, duplicate reads were removed using Sambamba (v1.0.1) markdup, as well as reads that mapped to blacklisted regions (ENCODE v2) and reads with mapping quality below 5.

Construction of CNV profiles using ichorCNA

To evaluate chromosomal alterations, we employed ichorCNA, an established CNV calling method for sWGS data of cfDNA (Adalsteinsson et al., 2017). For MeD-seq data, methylated reads, which have an LpnPI recognition site 16bp from either end were excluded, as they are methylation-dependent, and only the remaining background reads were kept for copy number analysis. For sWGS data, cfDNA fragments up to 150 bp were selected in silico. This was done to mimic the MeD-seq library preparation protocol, which includes a purification step, enriching for fragments ranging from 148 to192bp (including sequencing adapters). This process is known to also enrich for tumor-derived DNA fragments (Mouliere et al., 2018).

For both assays, the genome was binned into intervals of 1Mbp, and the number of reads in each bin was counted. Bins with a copy number gain or loss are expected to have more or fewer reads, respectively, compared to a diploid genome. ichorCNA employs a set of cfDNA profiles from healthy individuals (panel of normals) to estimate the expected read counts in each bin. To avoid potential biases due to the use of a different assay, we constructed assay-specific panels of normals using either MeD-seq background reads (n=31) or sWGS data (n=20). Within each sample, GC content and mappability bias were corrected for each bin's read count as described in (Adalsteinsson et al., 2017). Subsequently, the log ratio of the corrected read count in each bin to the corresponding median corrected read count in the appropriate panel of normals was calculated. These log ratios were then used as input for the ichorCNA Hidden Markov Model, which returns the most likely copy number state of each bin as well as an estimate of the tumor fraction.

We used default parameters for the ichorCNA model (v0.3.2) except chromosome 19 was excluded from the training phase, and the model was initialized using two prior normal fractions of 0.5 and 0.6 and the result with the highest likelihood was selected. Additionally, for MeD-seq samples, we adapted the ichorCNA source code to include sex as a parameter as determining the sex from the binned data showed to be inconsistent.

Evaluation TF estimates and CN-profiles

To assess systematic bias in the ichorCNA data and model we employed Bland-Altman (BA) analysis (Altman & Bland, 1983; Giavarina, 2015). In short, BA analysis is typically used to compare two continuous measurements, a and b, of the same quantity by plotting the mean of a and b on the x-axis, and the difference between a and b on the y-axis. Additionally, we calculated summary statistics, including the mean difference between the measurements and the 95% confidence intervals for both the data points and the mean difference, using bootstrapping (n=10,000). Additionally, we used Spearman correlation to compare TF estimates and ichorCNA logR ratios based on MeD-seq and sWGS data.

Agreement between CN-profiles

The ichorCNA model assigns a discrete state to each bin (ranging from 1-5; with 1 being hemizygous loss and 5 high-level amplification), representing the copy number state of that

bin. We quantified the agreement between the copy number states estimated from MeD-seq and sWGS data from the same sample using Cohen's Kappa.

$$k \equiv \frac{p_0 - p_e}{1 - p_e}$$

With k as the value for kappa, bounded between -1 (perfect disagreement) and 1 (perfect agreement), p_0 as the observed agreement and p_e as the probability of a by change agreement.

Association between noise and observed agreement in CN-profiles

Low sequencing depth leads to noisy coverage profiles which can have a detrimental effect on the detection of CNVs. To quantify noise within a CN-profile, we calculated the Median Absolute Deviation (MAD) of each CN-profile by:

$$MAD = median(|X_{klm} - median(X_{lm})|)$$

With, X as the log ratio of bin k in CN segment l in sample m.

To determine if differences in noise levels between MeD-Seq and WGS-derived CN profiles could be explained by differences in coverage, we down-sampled the BAM files of five high-coverage ovarian cancer sWGS samples to 75%, 50%, 25%, 10%, and 5% of the original read count and compared the MAD scores of these down-sampled samples and their MeD-seq counterparts.

Implementation details

We developed a freely available data processing pipeline, called emc-medseqcn, that starts from the output files of the in-house scripts described by Boers et al for methylation profiling (Boers et al., 2018). To ensure our pipeline is portable, scalable, and can be applied reproducibly we used the nf-core template of Nextflow (Di Tommaso et al., 2017; Ewels et al., 2020). All components of the pipeline are provided as Docker and Singularity containers through the Biocontainers initiative (da Veiga Leprevost et al., 2017). The pipeline includes

multiple components for quality control, alignment, processing, and fitting the ichorCNA model for both MeD-seq and sWGS sequencing data. It produces a structured output to support subsequent downstream analyses and is available at https://github.com/daanhazelaar/nf-core-medseqcn. All R scripts used for differential methylation analysis and all subsequent analysis is available at https://github.com/daanhazelaar/analysis_medseqcn

Results

MeD-seq background reads are enriched around LpnPI recognition sites

To assess whether MeD-seq background reads provide information on chromosomal alterations, sequencing metrics were evaluated to confirm sufficient coverage (mean: 10.5%, min: 2.5%, max: 19.2%) and read mapping quality (mean: 42.5, min: 37.8, max: 53.1) (Figure 1A, Supplemental Figure 1). Next, we analyzed the impact of LpnPI digestion on the genomic distribution of methylated and background reads, revealing increased LpnPI recognition site frequencies at +12 bp and -12 bp from the start and end of sequencing reads for both read types (Supplemental Figure 2). This pattern, absent in sWGS reads, suggests that MeD-seq background reads are partially influenced by LpnPI digestion, introducing a bias that needs to be corrected before downstream analysis.

Bias correction removes GC-bias and LpnPI Recognition site bias

We observed a strong correlation between the per bin read counts and both GC content (Spearman $\rho = 0.96$, $p<0.001$) and LpnPI recognition site counts (Spearman $\rho = 0.92$, $p<0.001$), for MeD-seq background reads, which could negatively impact the identification of CNs (Figure 1B, Figure Supplemental Figure 3). Additionally, GC content and LpnPI recognition site counts were highly correlated as well ($\rho = 0.95$, $p < 0.001$) (Supplemental Figure 4). To address these biases, we applied LOESS regression, and used a separate panel of MeD-seq normals to generate normalized read count ratios (logR). This approach effectively reduced both GC bias ($\rho = 0.11$, $p<0.001$) and LpnPI recognition site bias ($\rho = 0.22$, $p<0.001$) in the logR from MeD-seq (Figure 1B, Supplemental Figure 3).

Next, we compared MeD-seq logR values to sWGS across the same genomic bins and samples. We observed strong correlation ($\rho=0.75$, $p < 0.001$) and high agreement (mean relative difference: 0.08%, 95% CI: [-0.04%, 0.19%]), which supports the applicability of MeD-seq logR values for ichorCNA model fitting (Figure 1C).

Figure 1, Sequencing statistics, LpnpI bias removal, and strong logR correlation in MeD-Seq data

A) Scatter and box plots compare sequencing coverage (left) and read mapping quality (right) per sample (points) for MeD-seq background reads and sWGS data (colors) separate for cancer types.

B) Scatter plots show for each genomic bin (points), normalized read counts (before correction, left) and logR (after correction, right) versus LpnPI site counts (x-axis) for MeD-seq and sWGS (colors). Blue lines indicate second-order regression, with the shaded area indicating the standard error of the estimate. With Spearman correlation and corresponding p-value in top left.

C) Left heatmaps compare logR from MeD-seq and sWGS for bins without and with CNAs, including linear regression and Spearman correlation. Right heatmaps show Bland-Altman plots for logR differences, split by CNA presence, with regression lines, confidence intervals, and Spearman correlation.

Copy number profiles and ctDNA load estimates from MeD-seq closely resemble those from sWGS

We fitted ichorCNA using both MeD-seq logR values and sWGS logR values of the same 43 samples. A representative example output for the same sample is show in Figure 2A. We detected a non-zero TF (range: [15%-43%] in the same 21 samples, and we observed a strong correlation between TF-estimates (ρ=0.97, $p < 0.001$) (Figure 2B). Additionally, we performed BA-analysis which showed high agreement between TF-estimates (mean relative difference: -0.43%, 95%CI: [-1.5%, 0.71%]) (Figure 2C).

As an independent validation of the estimated TF, mutation-based variant allele frequencies were available for 19 CRLM samples from a targeted colon-specific NGS panel. A weak to moderate correlation between these mutation-based Variant Allele Frequencies (VAFs) for both MeD-seq (ρ = 0.25, p = 0.3) and sWGS-based TF estimates (ρ = 0.44, p = 0.06) was observed (supplemental figure 5). In two samples, no mutations were detected using the NGS panel, whereas a non-zero TF was detected using ichorCNA both in MeD-seq and sWGS data, indicating these samples contained mutations that were not covered by the NGS panel. Copy-number based TF estimates were generally higher than VAFs, likely due to the MeD-seq experimental protocol selecting shorter fragments, and the in silico size-selection for sWGS, both of which enrich for ctDNA fragments (Mouliere et al., 2018).

Figure 2. TF Estimates and CN Profiles from MeD-seq Are highly Comparable to sWGS
A) CN profiles from MeD-seq (top) and sWGS (bottom) from patient CRLM_3_T0, showing logR values (y-axis) and predicted CN states (colors) for each genomic bin (x-axis).
B) Scatterplot of tumor fraction (TF) estimates from MeD-seq (x-axis) and sWGS (y-axis) with Spearman correlation and corresponding p-value.
C) Bland-Altman plot showing mean TF estimates (x-axis) versus differences between MeD-seq and sWGS TF estimates (y-axis), with bootstrapped error (solid lines) and confidence intervals.
D) Bar plot of Cohen's kappa values (y-axis) for CN-profile agreement between MeD-seq and sWGS per patient (38 CRLM, 5 ovary).
E) Median Absolute Deviation (MAD) in logR values (y-axis) versus sequencing coverage (x-axis) for MeD-seq and sWGS. Down-sampled sWGS data from five high-coverage ovary samples are shown with connected symbols and MAD values.

Next, we assessed the inter-rater agreement of CN-profiles by comparing the integer CN-state assigned to each bin. This analysis showed an average Cohen's kappa of 0.92 (range: 0.74-1.00) for samples with non-zero TF and perfect agreement for zero TF samples (Figure 2D). We quantified noise levels in CN-profiles by calculating the Median Absolute Deviation (MAD) of logR values (Supplemental Figure 6).

The five ovarian cancer samples showed significantly higher noise in MeD-seq CN-profiles compared to sWGS (p=0.008, Mann-Whitney U test), which we hypothesized was due to higher sequencing coverage in sWGS. To test this, we down-sampled the ovarian sWGS samples to match the MeD-seq coverage, after which similar noise levels were observed (mean MAD, MeD-seq: 0.096, sWGS: 0.092). This provides further evidence that the observed noise differences stem from sequencing coverage and indicates that the quality of MeD-seq CN-profiles is highly similar to those based on sWGS. (Figure 2E).

Lastly, we generated frequency plots to compare the distribution of observed CN alterations based on MeD-seq and sWGS in ovarian and CRLM patients (supplemental figure 7, 8). Aggregated CRLM profiles were highly similar between MeD-seq and sWGS data and revealed common copy number gains in chromosomes 7, 8q, 13, and 20, as well as losses in chromosomes 8p and 18, consistent with chromosomal alterations frequently observed in metastatic colorectal cancer (Mendelaar et al., 2021).

Discussion

In this study we evaluated simultaneous methylation- and chromosomal CN profiling, and TF estimation from cfDNA using data generated by our previously developed MeD-seq assay (Boers et al., 2018; Deger et al., 2021). We showed that is possible to extract CN-profiles and TF-estimates from MeD-seq that are highly comparable to those obtained from sWGS. Moreover, we combined CN and TF information with methylation profiling, which enables the identification of CRLM-specific Differentially Methylated Regions (DMRs), making MeD-seq an affordable and efficient approach to multimodal cfDNA analyses.

We leveraged background reads of MeD-seq data to derive CN profiles and TF estimates, using the ichorCNA model, without altering the experimental protocol of the assay. This reduces the need for additional biological material and adds relevant information to readily performed MeD-seq experiments. Even though MeD-seq data contains a bias due to LpnpI digestion, this bias can consistently be reduced alongside GC bias, resulting in high quality CN-profiles that can be constructed using a sample coverage as low as 2.5% from a total of 30 million reads sequenced per sample. The resulting CN-profiles give insight on a per sample basis, enabling genomic profiling and may aid clinical decision making for CRLM patients (Eikenboom et al., 2023; Mendelaar et al., 2021; Smid et al., 2022; Wullaert et al., 2023).

The analysis of cfDNA holds immense potential for cancer diagnostics and disease monitoring, offering clinical and research benefits. However, challenges remain regarding cost, handling time, sensitivity, and reliability, which must be addressed to integrate cfDNA analysis into routine clinical practice. In this study, we provided a approach to extract additional genomic features from the MeD-seq assay for cfDNA. MeD-seq shows to be a versatile tool, combining methylation profiling, CN profiling, and TF estimation into a single assay. Expanding MeD-seq datasets to include a wider range of cancer types will further validate its utility. Such advancements could establish MeD-seq as a cost-effective, accessible, and versatile platform for non-invasive cancer diagnostics and cfDNA biomarker discovery.

References

Adalsteinsson, V. A., Ha, G., Freeman, S. S., Choudhury, A. D., Stover, D. G., Parsons, H. A., Gydush, G., Reed, S. C., Rotem, D., Rhoades, J., Loginov, D., Livitz, D., Rosebrock, D., Leshchiner, I., Kim, J., Stewart, C., Rosenberg, M., Francis, J. M., Zhang, C.-Z., … Meyerson, M. (2017). Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nature Communications*, *8*(1), 1324. https://doi.org/10.1038/s41467-017-00965-y

Altman, D. G., & Bland, J. M. (1983). Measurement in Medicine: The Analysis of Method Comparison Studies. *Journal of the Royal Statistical Society. Series D (The Statistician)*, *32*(3), 307–317. https://doi.org/10.2307/2987937

André, F., Solovieff, N., Su, F., Bardia, A., Neven, P., Yap, Y. S., Tripathy, D., Lu, Y.-S., Slamon, D., Chia, S., Joshi, M., Chakravartty, A., Lteif, A., Taran, T., & Arteaga, C. L. (2024). Acquired gene alterations in patients treated with ribociclib plus endocrine therapy or endocrine therapy alone using baseline and end-of-treatment circulating tumor DNA samples in the MONALEESA-2, -3, and -7 trials. *Annals of Oncology*. https://doi.org/10.1016/j.annonc.2024.09.010

Boers, R., Boers, J., Hoon, B. de, Kockx, C., Ozgur, Z., Molijn, A., IJcken, W. van, Laven, J., & Gribnau, J. (2018). Genome-wide DNA methylation profiling using the methylation-dependent restriction enzyme LpnPI. *Genome Research*, *28*(1), 88–99. https://doi.org/10.1101/gr.222885.117

Boers, R., Boers, J., Tan, B., van Leeuwen, M. E., Wassenaar, E., Sanchez, E. G., Sleddens, E., Tenhagen, Y., Mulugeta, E., Laven, J., Creyghton, M., Baarends, W., van IJcken, W. F. J., & Gribnau, J. (2023). Retrospective analysis of enhancer activity and transcriptome history. *Nature Biotechnology*, *41*(11), Article 11. https://doi.org/10.1038/s41587-023-01683-1

Connal, S., Cameron, J. M., Sala, A., Brennan, P. M., Palmer, D. S., Palmer, J. D., Perlow, H., & Baker, M. J. (2023). Liquid biopsies: The future of cancer early detection. *Journal of Translational Medicine*, *21*(1), 118. https://doi.org/10.1186/s12967-023-03960-8

Cristiano, S., Leal, A., Phallen, J., Fiksel, J., Adleff, V., Bruhm, D. C., Jensen, S. Ø., Medina, J. E., Hruban, C., White, J. R., Palsgrove, D. N., Niknafs, N., Anagnostou, V., Forde, P., Naidoo, J., Marrone, K., Brahmer, J., Woodward, B. D., Husain, H., … Velculescu, V. E. (2019). Genome-wide cell-free DNA fragmentation in patients with cancer. *Nature*, *570*(7761), Article 7761. https://doi.org/10.1038/s41586-019-1272-6

da Veiga Leprevost, F., Grüning, B. A., Alves Aflitos, S., Röst, H. L., Uszkoreit, J., Barsnes, H., Vaudel, M., Moreno, P., Gatto, L., Weber, J., Bai, M., Jimenez, R. C., Sachsenberg, T., Pfeuffer, J., Vera Alvarez, R., Griss, J., Nesvizhskii, A. I., & Perez-Riverol, Y. (2017). BioContainers: An open-source and community-driven framework for software standardization. *Bioinformatics*, *33*(16), 2580–2582. https://doi.org/10.1093/bioinformatics/btx192

Deger, T., Boers, R. G., de Weerd, V., Angus, L., van der Put, M. M. J., Boers, J. B., Azmani, Z., van IJcken, W. F. J., Grünhagen, D. J., van Dessel, L. F., Lolkema, M. P. J. K., Verhoef, C., Sleijfer, S., Martens, J. W. M., Gribnau, J., & Wilting, S. M. (2021). High-throughput and affordable genome-wide methylation profiling of circulating cell-free DNA by methylated DNA sequencing (MeD-seq) of LpnPI digested fragments. *Clinical Epigenetics*, *13*(1), 196. https://doi.org/10.1186/s13148-021-01177-4

Di Tommaso, P., Chatzou, M., Floden, E. W., Barja, P. P., Palumbo, E., & Notredame, C. (2017). Nextflow enables reproducible computational workflows. *Nature Biotechnology*, *35*(4), 316–319. https://doi.org/10.1038/nbt.3820

Doebley, A.-L., Ko, M., Liao, H., Cruikshank, A. E., Santos, K., Kikawa, C., Hiatt, J. B., Patton, R. D., De Sarkar, N., Collier, K. A., Hoge, A. C. H., Chen, K., Zimmer, A., Weber, Z. T., Adil, M., Reichel, J. B., Polak, P., Adalsteinsson, V. A., Nelson, P. S., … Ha, G. (2022). A framework for clinical cancer subtyping from nucleosome profiling of cell-free DNA. *Nature Communications*, *13*(1), 7475. https://doi.org/10.1038/s41467-022-35076-w

Eikenboom, E. L., Wilting, S. M., Deger, T., Srebniak, M. I., Van Veghel-Plandsoen, M., Boers, R. G., Boers, J. B., van IJcken, W. F. J., Gribnau, J. H., Atmodimedjo, P., Dubbink, H. J., Martens, J. W. M., Spaander, M. C. W., & Wagner, A. (2023). Liquid Biopsies for Colorectal Cancer and Advanced Adenoma Screening and Surveillance: What to Measure? *Cancers*, *15*(18), Article 18. https://doi.org/10.3390/cancers15184607

Ewels, P. A., Peltzer, A., Fillinger, S., Patel, H., Alneberg, J., Wilm, A., Garcia, M. U., Di Tommaso, P., & Nahnsen, S. (2020). The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology*, *38*(3), 276–278. https://doi.org/10.1038/s41587-020-0439-x

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/BM.2015.015

Katsman, E., Orlanski, S., Martignano, F., Fox-Fisher, I., Shemer, R., Dor, Y., Zick, A., Eden, A., Petrini, I., Conticello, S. G., & Berman, B. P. (2022). Detecting cell-of-origin and cancer-specific methylation

features of cell-free DNA from Nanopore sequencing. *Genome Biology*, *23*(1), 158. https://doi.org/10.1186/s13059-022-02710-1

Liu, J., Dai, L., Wang, Q., Li, C., Liu, Z., Gong, T., Xu, H., Jia, Z., Sun, W., Wang, X., Lu, M., Shang, T., Zhao, N., Cai, J., Li, Z., Chen, H., Su, J., & Liu, Z. (2024). Multimodal analysis of cfDNA methylomes for early detecting esophageal squamous cell carcinoma and precancerous lesions. *Nature Communications*, *15*(1), 3700. https://doi.org/10.1038/s41467-024-47886-1

Liu, Y., Reed, S. C., Lo, C., Choudhury, A. D., Parsons, H. A., Stover, D. G., Ha, G., Gydush, G., Rhoades, J., Rotem, D., Freeman, S., Katz, D. W., Bandaru, R., Zheng, H., Fu, H., Adalsteinsson, V. A., & Kellis, M. (2024). FinaleMe: Predicting DNA methylation by the fragmentation patterns of plasma cell-free DNA. *Nature Communications*, *15*(1), 2790. https://doi.org/10.1038/s41467-024-47196-6

Mathios, D., Johansen, J. S., Cristiano, S., Medina, J. E., Phallen, J., Larsen, K. R., Bruhm, D. C., Niknafs, N., Ferreira, L., Adleff, V., Chiao, J. Y., Leal, A., Noe, M., White, J. R., Arun, A. S., Hruban, C., Annapragada, A. V., Jensen, S. Ø., Ørntoft, M.-B. W., … Velculescu, V. E. (2021). Detection and characterization of lung cancer using cell-free DNA fragmentomes. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-021-24994-w

Mendelaar, P. A. J., Smid, M., van Riet, J., Angus, L., Labots, M., Steeghs, N., Hendriks, M. P., Cirkel, G. A., van Rooijen, J. M., Ten Tije, A. J., Lolkema, M. P., Cuppen, E., Sleijfer, S., Martens, J. W. M., & Wilting, S. M. (2021). Whole genome sequencing of metastatic colorectal cancer reveals prior treatment effects and specific metastasis features. *Nature Communications*, *12*(1), Article 1. https://doi.org/10.1038/s41467-020-20887-6

Moldovan, N., Van Der Pol, Y., Van Den Ende, T., Boers, D., Verkuijlen, S., Creemers, A., Ramaker, J., Vu, T., Bootsma, S., Lenos, K. J., Vermeulen, L., Fransen, M. F., Pegtel, M., Bahce, I., Van Laarhoven, H., & Mouliere, F. (2024). Multi-modal cell-free DNA genomic and fragmentomic patterns enhance cancer survival and recurrence analysis. *Cell Reports Medicine*, *5*(1), 101349. https://doi.org/10.1016/j.xcrm.2023.101349

Mouliere, F., Chandrananda, D., Piskorz, A. M., Moore, E. K., Morris, J., Ahlborn, L. B., Mair, R., Goranova, T., Marass, F., Heider, K., Wan, J. C. M., Supernat, A., Hudecova, I., Gounaris, I., Ros, S., Jimenez-Linan, M., Garcia-Corbacho, J., Patel, K., Østrup, O., … Rosenfeld, N. (2018). Enhanced detection of circulating

tumor DNA by fragment size analysis. *Science Translational Medicine, 10*(466), eaat4921. https://doi.org/10.1126/scitranslmed.aat4921

Nguyen, V. T. C., Nguyen, T. H., Doan, N. N. T., Pham, T. M. Q., Nguyen, G. T. H., Nguyen, T. D., Tran, T. T. T., Vo, D. L., Phan, T. H., Jasmine, T. X., Nguyen, V. C., Nguyen, H. T., Nguyen, T. V., Nguyen, T. H. H., Huynh, L. A. K., Tran, T. H., Dang, Q. T., Doan, T. N., Tran, A. M., … Tran, L. S. (2023). Multimodal analysis of methylomics and fragmentomics in plasma cell-free DNA for multi-cancer early detection and localization. *eLife*, *12*, RP89083. https://doi.org/10.7554/eLife.89083

Penny, L., Main, S. C., De Michino, S. D., & Bratman, S. V. (2024). Chromatin- and nucleosome-associated features in liquid biopsy: Implications for cancer biomarker discovery. *Biochemistry and Cell Biology*, *102*(4), 291–298. https://doi.org/10.1139/bcb-2024-0004

Smid, M., Wilting, S. M., & Martens, J. W. M. (2022). Lost by Transcription: Fork Failures, Elevated Expression, and Clinical Consequences Related to Deletions in Metastatic Colorectal Cancer. *International Journal of Molecular Sciences*, *23*(9), Article 9. https://doi.org/10.3390/ijms23095080

Thierry, A. R. (2023). Circulating DNA fragmentomics and cancer screening. *Cell Genomics*, *3*(1). https://doi.org/10.1016/j.xgen.2022.100242

van Dessel, L. F., Beije, N., Helmijr, J. C. A., Vitale, S. R., Kraan, J., Look, M. P., de Wit, R., Sleijfer, S., Jansen, M. P. H. M., Martens, J. W. M., & Lolkema, M. P. (2017). Application of circulating tumor DNA in prospective clinical oncology trials – standardization of preanalytical conditions. *Molecular Oncology*, *11*(3), 295–304. https://doi.org/10.1002/1878-0261.12037

Wan, J. C. M., Massie, C., Garcia-Corbacho, J., Mouliere, F., Brenton, J. D., Caldas, C., Pacey, S., Baird, R., & Rosenfeld, N. (2017). Liquid biopsies come of age: Towards implementation of circulating tumour DNA. *Nature Reviews. Cancer*, *17*(4), 223–238. https://doi.org/10.1038/nrc.2017.7

Wang, Y., Douville, C., Cohen, J. D., Mattox, A., Curtis, S., Silliman, N., Popoli, M., Ptak, J., Dobbyn, L., Nehme, N., Dudley, J. C., Summers, M., Zhang, M., Ho-Pham, L. T., Tran, B. N. H., Tran, T. S., Nguyen, T. V., Bettegowda, C., Papadopoulos, N., … Vogelstein, B. (2023). Detection of rare mutations, copy number alterations, and methylation in the same template DNA molecules. *Proceedings of the National Academy of Sciences*, *120*(15), e2220704120. https://doi.org/10.1073/pnas.2220704120

Wullaert, L., van Rees, J. M., Martens, J. W. M., Verheul, H. M. W., Grünhagen, D. J., Wilting, S. M., & Verhoef, C. (2023). Circulating Tumour DNA as Biomarker for Colorectal Liver Metastases: A Systematic Review and Meta-Analysis. *Cells*, *12*(21), 2520. https://doi.org/10.3390/cells12212520

Zhou, Q., Kang, G., Jiang, P., Qiao, R., Lam, W. K. J., Yu, S. C. Y., Ma, M.-J. L., Ji, L., Cheng, S. H., Gai, W., Peng, W., Shang, H., Chan, R. W. Y., Chan, S. L., Wong, G. L. H., Hiraki, L. T., Volpi, S., Wong, V. W. S., Wong, J., … Lo, Y. M. D. (2022). Epigenetic analysis of cell-free DNA by fragmentomic profiling. *Proceedings of the National Academy of Sciences*, *119*(44), e2209852119. https://doi.org/10.1073/pnas.2209852119

Figure 1

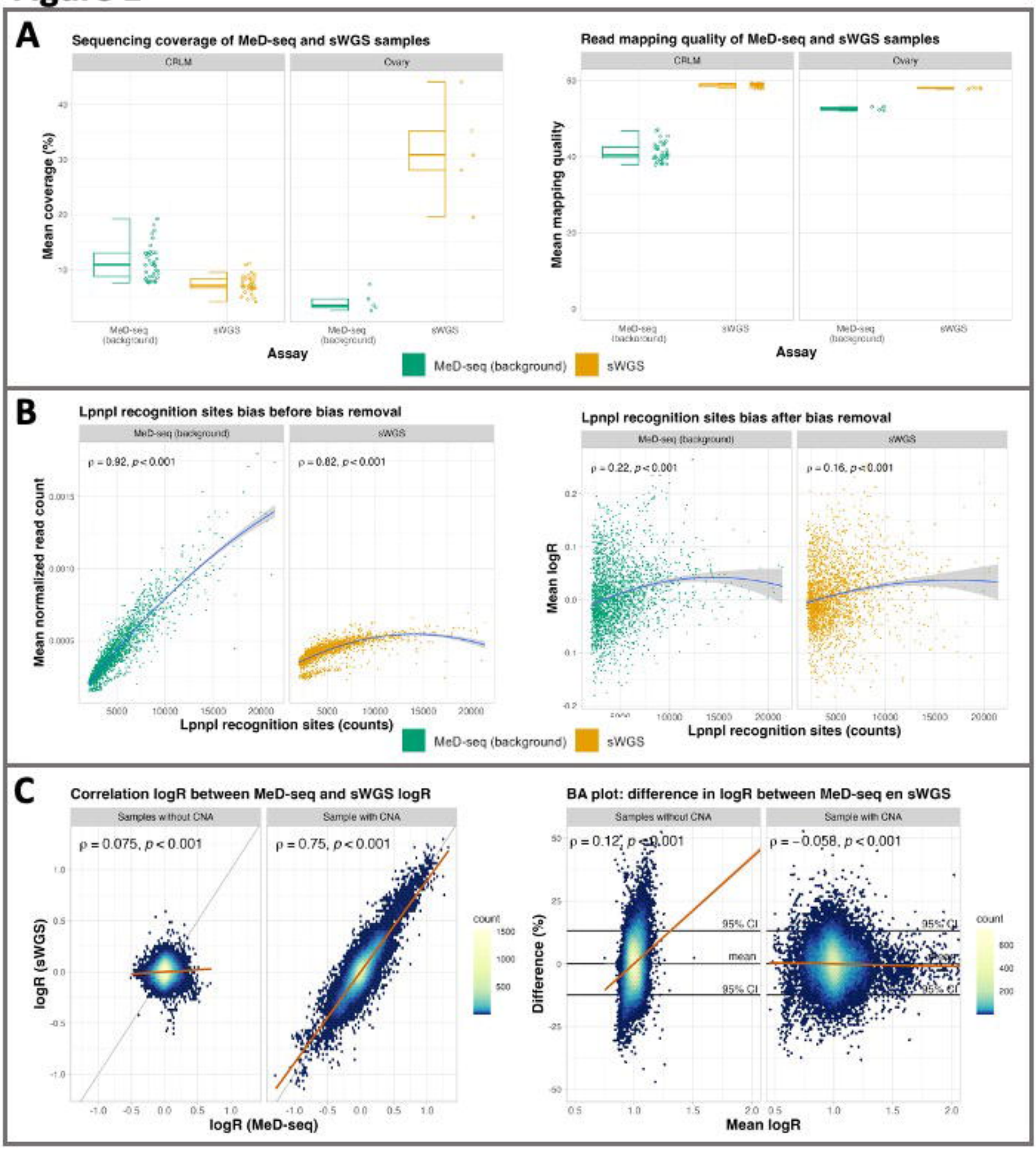
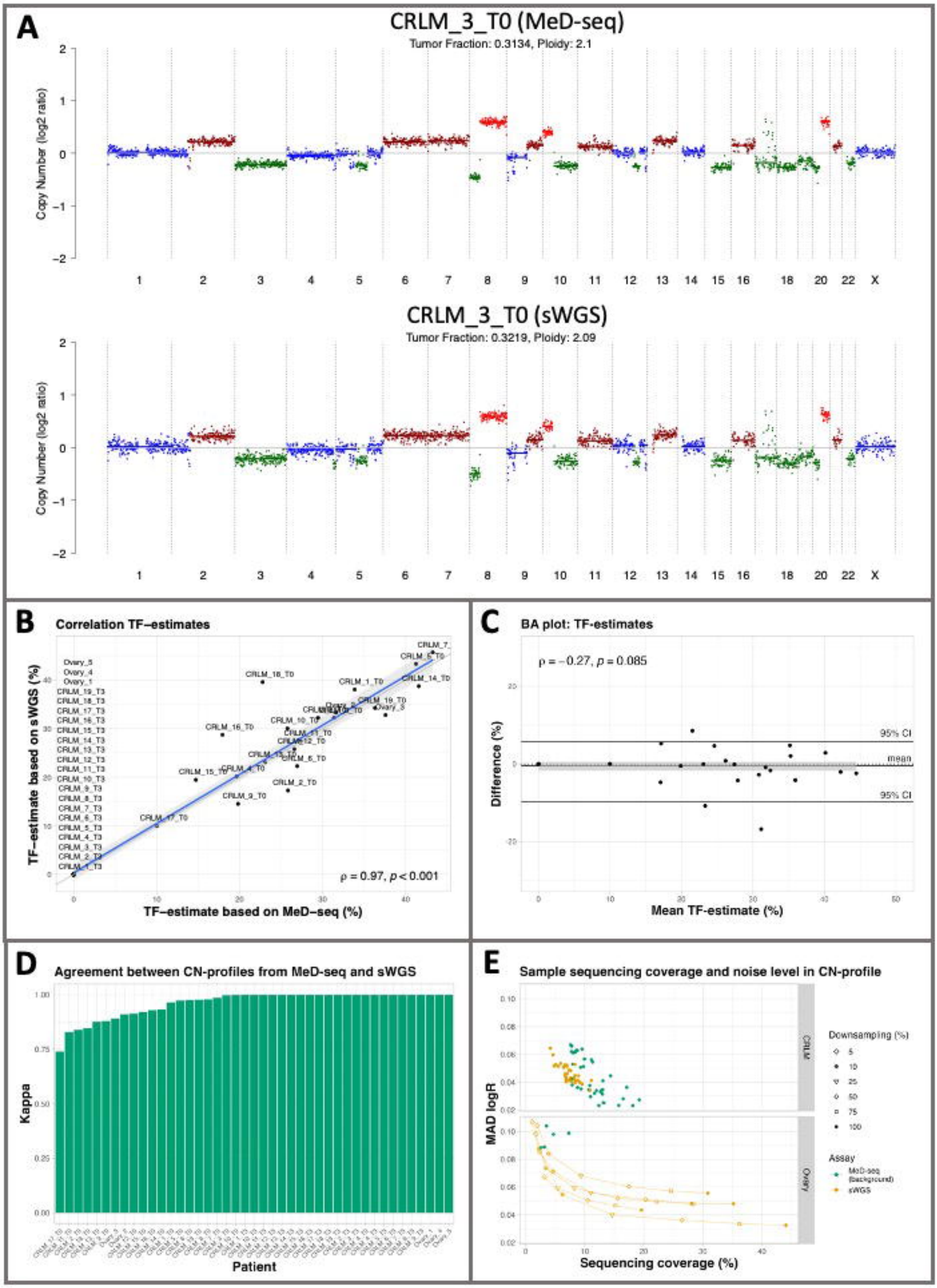

A
Sequencing coverage of MeD-seq and sWGS samples
CRLM
Ovary
Mean coverage (%)
MeD-seq (background)
sWGS
Assay
Read mapping quality of MeD-seq and sWGS samples
Mean mapping quality
MeD-seq (background)
sWGS
B
Lpnpl recognition sites bias before bias removal
MeD-seq (background)
sWGS
ρ = 0.92, p < 0.001
ρ = 0.82, p < 0.001
Mean normalized read count
Lpnpl recognition sites (counts)
Lpnpl recognition sites bias after bias removal
ρ = 0.22, p < 0.001
ρ = 0.16, p < 0.001
Mean logR
C
Correlation logR between MeD-seq and sWGS logR
Samples without CNA
Sample with CNA
ρ = 0.075, p < 0.001
ρ = 0.75, p < 0.001
logR (sWGS)
logR (MeD-seq)
count
BA plot: difference in logR between MeD-seq en sWGS
ρ = 0.12, p < 0.001
ρ = −0.058, p < 0.001
95% CI
mean
Difference (%)
Mean logR

Figure 2


A
CRLM_3_T0 (MeD-seq)
Tumor Fraction: 0.3134, Ploidy: 2.1
Copy Number (log2 ratio)
CRLM_3_T0 (sWGS)
Tumor Fraction: 0.3219, Ploidy: 2.09
Copy Number (log2 ratio)
B
Correlation TF–estimates
TF–estimate based on sWGS (%)
TF–estimate based on MeD–seq (%)
ρ = 0.97, p < 0.001
C
BA plot: TF-estimates
ρ = −0.27, p = 0.085
Difference (%)
Mean TF-estimate (%)
95% CI
mean
95% CI
D
Agreement between CN-profiles from MeD-seq and sWGS
Kappa
Patient
E
Sample sequencing coverage and noise level in CN-profile
MAD logR
Sequencing coverage (%)
CRLM
Ovary
Downsampling (%)
Assay
MeD-seq (background)
sWGS