

The Effect of Unique Molecular Identifier Family Size Using Tumor-Informed Circulating Tumor DNA Analysis in Childhood Cancers

Downloaded from: https://research.chalmers.se, 2026-11-10 05:17 UTC

Citation for the original published paper (version of record):
Österlund, T., Andersson, D., Ibrahim, R. et al (2026). The Effect of Unique Molecular Identifier Family Size Using Tumor-Informed Circulating Tumor
DNA Analysis in Childhood Cancers. Journal of Molecular Diagnostics, In Press.
http://dx.doi.org/10.1016/j.jmoldx.2026.08.002

N.B. When citing this work, cite the original published paper.

research.chalmers.se offers the possibility of retrieving research publications produced at Chalmers University of Technology. It covers all kind of research output: articles, dissertations, conference papers, reports etc. since 2004. research.chalmers.se is administrated and maintained by Chalmers Library

(article starts on next page)

The Journal of Molecular Diagnostics, Vol. ■, No. ■, ■ 2026

ELSEVIER

the Journal of Molecular Diagnostics

jmdjournal.org

The Effect of Unique Molecular Identifier Family Size Using Tumor-Informed Circulating Tumor DNA Analysis in Childhood Cancers

Tobias Österlund,[*†‡] Daniel Andersson,[*] Raghda R. Ibrahim,[†§] Elisabeth Mellström,[†§¶] Jenny Lundmark,[‖] Junrui Li,[*] Stefan Filges,[*,**] Peter Micallef,[*‡] Manuel Luna Santamaría,[*†] Mandy Escobar,[*] Pia Mouhanna,[*††] Firaol T. Kebede,[*] Christoffer Vannas,[*‡‡§§] Ida Rahmqvist,[†§] Fani Pujol-Calderón,[†§] Agnes Dahlstrand Rudin,[†§] Wilma Franssila,[§] Åsa Torinsson Naluai,[¶¶‖‖***] Serik Sagitov,[†††] Joakim Håkansson,[‖¶¶] Henrik Fagman,[*‡‡‡] Martin Dalin,[†§¶] and Anders Ståhlberg[*†‡§§§]

From the Sahlgrenska Center for Cancer Research,[*] Department of Laboratory Medicine, Institute of Biomedicine, the Sahlgrenska Center for Cancer Research,[§] Department of Pediatrics, Institute of Clinical Sciences, the Department of Oncology,[§§] Institute of Clinical Sciences, the Department of Laboratory Medicine,[¶¶] Institute of Biomedicine, and the Core Facilities,[‖‖] Sahlgrenska Academy, the Wallenberg Centre for Molecular and Translational Medicine,[†] and the Science for Life Laboratory,[§§§] Institute of Biomedicine, University of Gothenburg, Gothenburg; the Departments of Clinical Genetics and Genomics,[‡] Oncology,[‡‡] and Clinical Pathology,[‡‡‡] the Childhood Cancer Centre,[¶] Queen Silvia Children's Hospital, and Biobank West,[***] Sahlgrenska University Hospital, Gothenburg; the RISE Research Institutes of Sweden,[‖] Division Material and Production, Medical Device and Diagnostics, Borås; Simsen Diagnostics AB,[**] Mölndal; the Department of Oncology,[††] Ryhov County Hospital, Jönköping; and Mathematical Sciences,[†††] Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden

Accepted for publication August 24, 2026.

Address correspondence to Anders Ståhlberg, Ph.D., or Martin Dalin, M.D., Ph.D., Sahlgrenska Center for Cancer Research, University of Gothenburg, Box 425, 405 30 Gothenburg, Sweden. E-mail: anders.stahlberg@gu.se or martin.dalin@gu.se.

Analysis of circulating tumor DNA (ctDNA) provides a molecular assessment that can complement routine imaging in childhood cancer management. Detailed monitoring of ctDNA levels may provide information on treatment efficacy and resistance and minimal residual disease and allows for early detection of relapse. Here, tumor-informed ctDNA analysis was applied to 90 blood plasma samples collected from eight children with malignant tumors. A total of 4 to 10 tumor-specific mutations per patient were assessed using SiMSen-Seq, a digital sequencing approach using unique molecular identifiers (UMIs). The effects of individual SiMSen-Seq assays and plasma samples were evaluated in relation to their impact on background error rate, number of detected target molecules, and mutant calling using different UMI family size cutoff settings. The use of at least two sequencing reads per UMI provided the best overall performance by generating the highest number of detected target molecules and hence the optimal chance to detect low-frequency mutations. Data were consistent between SiMSen-Seq assays and plasma samples, providing robust ctDNA profiling over time for all patients. In conclusion, the results show that optimal use of UMIs in tumor-informed ctDNA analysis enables sensitive molecular readout that can assist in management of childhood cancers. *(J Mol Diagn 2026, ■: 1—12; https://doi.org/10.1016/j.jmoldx.2026.08.002)*

Supported by the Adlerbertska Foundations (D.A.); Anna-Lisa och Bror Björnssons Stiftelse (D.A.); Assar Gabrielssons Foundation (T.Ö. and D.A.); Johan Jansson Foundation for Cancer Research (T.Ö. and D.A.); Magn. Bergvalls Stiftelse (D.A.); Region Västra Götaland (A.S.); Swedish Cancer Society 22-2080 (A.S.), 25-4425 (A.S.), and 25-4433 (M.D.); Swedish Childhood Cancer Foundation TJ2015-0058 (D.A.), PR2022-0030 (A.S.), PR2025-0064 (A.S.), KP2020-0002 (M.D.), KP2022-0007 (M.D.), ST-2019-0009 (M.D.), and KF2023-0003 (C.V.); Swedish Research Council 2021-01008 (A.S.); the Swedish Society for Medical Research (D.A.); the Swedish Society of Medicine (M.D.); Swedish governmental funding of clinical research (ALF) 1006009 (A.S.); Sweden's Innovation Agency 2020-04141 (J.H. and A.S.) and 2020-04733 (A.S.); the Swedish Agency for Economic and Regional Growth 20370391 (M.D. and A.S.); the Sjöberg Foundation 2025-1219 (A.S.); the Sarcoma patient union research fund, the memorial fund of Elvira Trané (C.V. and A.S.); Sahlgrenska University hospital funds (C.V.); the Gothenburg Medical Society GLS-1022335 (C.V.) and GLS-1000035 (C.V.); the Gothenburg Society of Medicine (C.V.); Wallenberg Centre for Molecular and Translational Medicine (M.D. and A.S.); and the Wilhelm and Martina Lundgren Foundation for Scientific Research (T.Ö. and D.A.).

T.Ö. and D.A. contributed equally to this work.

© 2026 The Author(s). Published by Elsevier Inc. on behalf of the Association for Molecular Pathology and American Society for Investigative Pathology.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0).
https://doi.org/10.1016/j.jmoldx.2026.08.002

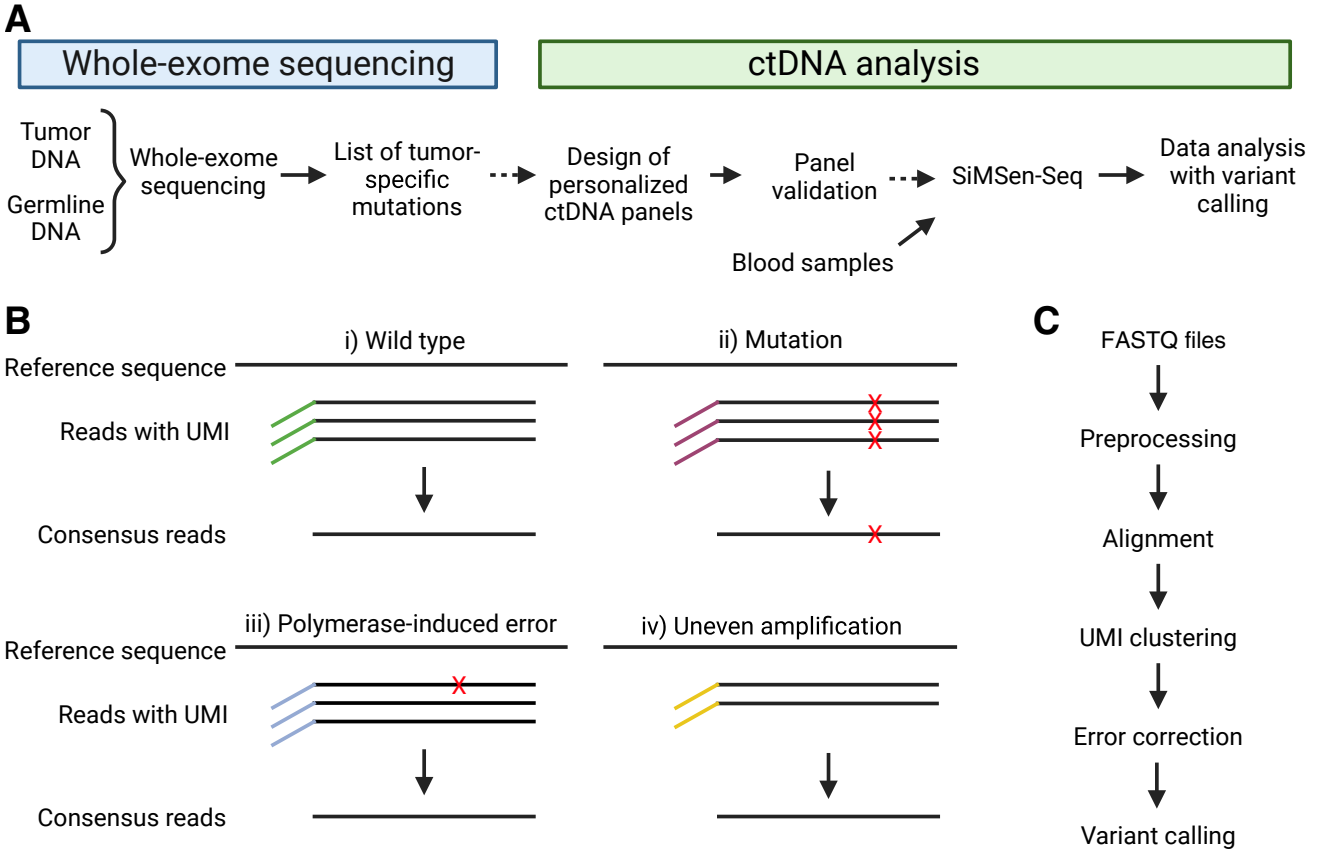


Figure 1 Workflow of tumor-informed ctDNA analysis and use of UMIs. **A:** To identify somatic variants, tumor tissue and germline DNA are analyzed with whole-exome sequencing. A list of patient-specific tumor variants serves as the basis for SiMSen-Seq panel design, later used to detect ctDNA in patient blood plasma samples. **B:** The concept of UMIs. Four different scenarios are shown: i) wild type, ii) true mutation, iii) polymerase-induced error, and iv) uneven amplification. Reads are aligned to the reference genome and grouped into UMI families. All reads in each UMI family are collapsed to form a consensus read by selecting the most common nucleotide at each aligned position, where red X is a true or false nonreference nucleotide variant. The error-corrected consensus reads are used for variant calling and downstream analysis. The UMI family size is the minimal number of raw reads that are needed to form a consensus read for each UMI family. **C:** Schematic overview of the bioinformatic pipeline UMIErrorCorrect using sequencing data with UMIs.

Despite significant achievements in the management of childhood cancers during the past decades, a clinical need remains for improved analytical methods to evaluate treatment efficacy and detect relapse at an early stage. Imaging techniques are the gold standard for monitoring solid cancers but have limitations in sensitivity and specificity. Complementary liquid biopsy analyses can lead to increased detection rates and provide molecular information to help guide treatment decisions and improve the timing of radiological examinations. In particular, circulating tumor DNA (ctDNA) has shown promise as a sensitive biomarker in children with malignant tumors and can be detected even before neoplastic cells produce metastatic lesions.[1–4] The level of ctDNA correlates to tumor burden,[5–7] and specific mutations in resistance genes may provide information about the choice of treatment.[8–10]

In tumor-informed ctDNA analysis, the risk of false-positive results is low compared with agnostic approaches because only pre-identified tumor-specific mutations are assessed. To identify tumor-specific mutations, the primary tumor and matched germline DNA are typically analyzed by whole-genome or whole-exome sequencing. These mutations are used to design the personalized ctDNA panels used to analyze plasma samples collected at diagnosis and subsequent time points (Figure 1A). The development of tumor-informed ctDNA analysis is particularly relevant in childhood cancers because they are characterized by few recurrent mutations.[11,12] Standard sequencing approaches typically have a clinical sensitivity to detect variants with allele frequencies of approximately 1% to 5%.[13,14] However, many applications based on ctDNA analysis require higher sensitivity.[15,16] Digital sequencing approaches allow detection of variant alleles with frequencies of <0.1% and can be divided into two major groups: PCR-based approaches, such as SiMSen-Seq,[17] Safe-SeqS,[18] and BiSeqS[19]; and ligation-based strategies, such as DupSeq,[20] TARDIS,[21] CODEC,[22] and GeneBits.[23] These methods use unique molecular identifiers (UMIs) that are introduced during library construction.[24] The UMI is typically a 6 to 12 nucleotides long randomized DNA sequence that uniquely labels each target DNA molecule. After sequencing, reads with identical UMIs are collapsed into one consensus read

ARTICLE IN PRESS

because they all originate from the same DNA molecule. By assessing each original target molecule multiple times, it is possible to correct for both polymerase-induced errors and quantification biases (Figure 1, B and C).[25,26] However, the effects of applying different cutoffs for UMI family size (ie, number of reads with identical UMI) remain poorly studied in tumor-informed ctDNA analysis.

Here, tumor-informed ctDNA analysis was applied in eight children with malignant tumors using SiMSen-Seq to evaluate the impact of different UMI family size cutoffs. Personalized SiMSen-Seq panels were designed on the basis of mutations identified by whole-exome sequencing of the primary tumor and were used to assess ctDNA levels in longitudinally collected blood plasma samples. The effects of individual SiMSen-Seq assays and samples were evaluated in relation to their impact on background error rate, number of detected target molecules, and mutant calling. The data provide bioinformatics insights for the use of UMIs in tumor-informed ctDNA analysis using SiMSen-Seq in childhood tumor types.

Materials and Methods

Patient Cohort

Written informed consent was obtained from all participants, patients, and/or legal guardians. The study was approved by the regional ethical review board in Gothenburg (Diary number 054-15, 485-16, and 655-17), with amendments approved by the regional ethical review board (Diary number T795-16 and T525-18) and the Swedish Ethical Review Authority (Diary number 2019-06285, 2021-04895, and 2023-01300-02). The study was conducted in compliance with the Declaration of Helsinki. For Patients 1 and 2, ctDNA results have been published with focus on clinical outcome.[27]

Whole-Exome Sequencing

Tumor DNA was extracted using GeneRead FFPE DNA kit, and germline DNA from the cell fraction of a matched blood sample was extracted using DNeasy Blood and Tissue kit [catalog number (cat. no.) 180134 and 69504, respectively; both from Qiagen, Hilden, Germany]. Library preparation was performed using TruSeq DNA PCR-Free library prep kit (Illumina, San Diego, CA) and target enrichment using SureSelect clinical research exome version 2 (Agilent Technologies, Santa Clara, CA). Whole-exome sequencing was performed at Clinical Genomics Gothenburg, Science for Life Laboratory, University of Gothenburg (Gothenburg, Sweden), using NextSeq 500 (Illumina) with a coverage of 90× and 30× for tumor and germline DNA, respectively.

Somatic variants were identified by comparing tumor DNA with germline DNA for each individual. Raw fastq-files were quality checked using FastQC version 0.12.1, and read-through sequencing adapters were trimmed off with cutadapt version 1.9[28] using a Phred score cutoff of 20. These reads were used as input to the Sarek pipeline version 2.0.0 for somatic variant calling.[29] The Sarek pipeline aligned reads to the hg19 reference genome for identification of single-nucleotide variants, insertions, and deletions using the combined results from three different somatic variant callers: Mutect2 version 3.8-0,[30] Strelka2 version 2.8.2,[31] and Freebayes version 1.1.0.[32] Exonic variants were ranked on the basis of the variant allele frequency, also including mutations reported in the Catalogue of Somatic Mutations in Cancer database version 85.[33] All selected candidate mutations for ctDNA analysis were manually inspected in Integrative Genomics Viewer version 2.16.2[34] using the binary alignment map files and passed validation. A summary of all called variants is shown in Supplemental Table S1. Mutation selection was primarily based on variant allele frequency, while also considering known cancer-associated variants when available. The genomic coordinates of the mutations were converted to the hg38 reference genome using hgLiftOver (*https://genome.ucsc.edu/cgi-bin/hgLiftOver*, last accessed May 7, 2026) for further analysis.

Blood Sampling and Cell-Free DNA Extraction

Patient blood (approximately 8.5 mL) was collected in cf-DNA/cf-RNA Preservative Tubes (number 63960; Norgen Biotek, Thorold, ON, Canada). Plasma was isolated according to manufacturer's recommendations by centrifugation at 430 × *g* at room temperature for 20 minutes and subsequently stored at −80°C. Plasma samples from Patients 1 to 6 were centrifuged a second time at 4500 × *g* at 4°C. Cell-free DNA extraction was performed using QIAamp Circulating Nucleic Acid Kit (cat. no. 51104; Qiagen), according to the manufacturer's protocol. DNA concentration was quantified using Qubit 3.0 Fluorometer (Thermo Fisher Scientific, Waltham, MA) with Qubit 1× dsDNA High Sensitivity Assay Kit (cat. no. Q33231; Thermo Fisher Scientific). Cell-free DNA was concentrated using Vivacon 500, 30,000 molecular weight cutoff centrifugal units (cat. no. VN01H22; Sartorius, Göttingen, Germany) to 6 to 15 μL. Samples were stored at −80°C.

SiMSen-Seq and ctDNA Analysis

Digital sequencing was performed using SiMSen-Seq that consists of two rounds of PCR, where target DNA is UMI labeled in the first PCR and then amplified with sequencing adapters in the second PCR (Supplemental Figure S1).[17] After PCRs, libraries are purified and sequenced. Here, target primers were designed with annealing temperature of 60°C and amplicon size ranging from 59 to 105 bp, using either Primer-BLAST[35] or PanelPlex (DNA Software, Plymouth, MI). The personalized panels were validated using human genomic DNA (cat. no. 11691112001; Roche Diagnostics, Mannheim, Germany) as described

previously.[36] In total, 73 assays were designed, of which 7 were excluded and 17 were redesigned, generating 66 assays dived into 8 panels. The UMI was 12 nucleotides long, minimizing the risk for UMI collisions between detected molecules.[24] Detailed information about the selected mutations in the SiMSen-Seq panels is shown in Supplemental Table S1.

Briefly, cfDNA was barcoded in two to three parallel 15-μL reactions, containing 4 μL concentrated cfDNA, 1× Platinum SuperFi Buffer, 0.15 U Platinum SuperFi DNA polymerase (both cat. no. 12351050; Thermo Fisher Scientific), 0.2 mmol/L dNTPs (cat. no D7295; Sigma-Aldrich, St. Louis, MO), 40 nmol/L of each primer (Ultramers; Integrated DNA Technologies, Coralville, IA), and 0.5 mol/L L-carnitine inner salt (cat. no. C0158; Sigma-Aldrich). Extracted cfDNA from healthy individuals from a preanalytical study[37] and human genomic DNA extracted from blood cells (cat. no. 11691112001; Roche Diagnostics) were used as control DNA. The thermal program was performed with a T100 Thermal cycler (Bio-Rad, Hercules, CA): 3 minutes at 98°C, followed by three cycles of amplification (98°C for 10 seconds, 62°C for 6 minutes, and 72°C for 30 seconds), 15 minutes at 65°C, and 15 minutes at 95°C. Thirty microliters Tris-EDTA buffer (pH 8.0; cat. no. AM9858; Invitrogen, Carlsbad, CA) containing 45 ng/μL *Streptomyces griseus* protease (cat. no. P5147; Sigma-Aldrich) was added to each reaction at the start of the 15-minute incubation step at 65°C. Fifteen microliters of the barcoded product was subsequently amplified in a 60-μL adapter reaction, containing 1× Q5 High-Fidelity Master Mix (cat. no. M0494L; New England Biolabs, Ipswich, MA) and 400 nmol/L of indexed Illumina adapter primers (desalted; Sigma-Aldrich) using the following thermal program on a T100 Thermal cycler: 3 minutes at 98°C, followed by 27 cycles of amplification (98°C for 10 seconds, 80°C for 1 second, 72°C for 30 seconds, and 76°C for 30 seconds, all with ramping at 0.2°C/second). Libraries were purified using Agencourt AMPure XP system (cat. no. A63881; Beckman Coulter, Brea, CA) at a beads/sample ratio of 1:1 or by using Pippin Prep (Sage Science, Beverly, MA) with a 2% agarose ethidium-free gel cassette (cat. no. CEF2010; Sage Science) with the target range of 210 to 300 bp, according to the manufacturers' instructions. Library quality and quantity were assessed with a Fragment Analyzer (Agilent Technologies) using the HS NGS Fragment kit (cat. no. DNF-474; Agilent Technologies). Purified libraries were pooled to enable similar sequencing depth across all samples. Pooled libraries were quantified by quantitative real-time PCR on a CFX384 Real-Time System (Bio-Rad) using a modified version of the NEBNext Library Quant Kit for Illumina (cat. no. E7630; New England Biolabs). Briefly, 2 μL of the purified and diluted libraries was quantified in 10-μL reactions, containing 1× TATAA SYBR GrandMaster Mix (cat. no. TA01-3750; TATAA Biocenter, Gothenburg, Sweden) and 400 nmol/L of each Illumina adapter primer (forward primer: 5′-AATGATACGGCGACCACCGA-3′ and reverse primer: 5′-CAAGCAGAAGACGGCATACGA-3′). The concentration was estimated with adjustment to the average library size. Sequencing was performed with a MiniSeq (Illumina) using a High Output Reagent Kit (150 cycles; cat. no. FC-420-1002; Illumina) or with a NextSeq (Illumina) using a NextSeq 500/550 Mid Output Reagent Kit version 2.5 (150 cycles; cat. no. 20024904; Illumina) containing 10% to 20% PhiX Control version 3 (cat. no. FC-110-3001; Illumina) with a final library concentration of 1.5 to 1.8 pmol/L. The effect of sequencer on error rates was neglectable for UMI-corrected data (Supplemental Figure S2). SiMSen-Seq data were analyzed using the UmiErrorCorrect pipeline version 0.29[38] with default parameters: UMI length, 12; spacer length, 16; and Q-value, ≥10. The reads were aligned to the hg38 reference genome. Browser extensible data files with genomic coordinates of the patient-specific assays were used to constrain variant calling to specified genomic regions and for annotation. For the *OSBPL7* assay (Patient 7), two mutations were assessed together.

Theoretical Number of UMIs per DNA Molecule in SiMSen-Seq

In SiMSen-Seq, UMIs are incorporated during the barcoding PCR as part of the forward primer (Supplemental Figure S1A). Only complete molecules with adapter sequences at both ends will be amplified in the second adapter PCR (Supplemental Figure S1B). The three rounds of the barcoding PCR produce eight possible molecule outcomes with six different UMIs (UMI 1 to 6): i) None, ii) Only 4, iii) Only 2, iv) Only 5, v) Only 1, vi) 1 and 6, vii) Only 3, viii) None.

The 100% PCR efficiency and dilution proportion p, for example, $p = 1/3$, is assumed. The number X of specific UMIs per molecule is a random variable with possible values $x = 0, 1, 2, 3, 4, 5, 6$. Here, six probabilities $P(X = x)$ and the expected value

$$\mu = E(X) \tag{1}$$

can be computed for an arbitrary p.

Observe that

$$X = Y + Z \tag{2}$$

where Y and Z are two independent random variables defined as

- Y is the number of unique UMIs involving 2, 3, 4, 5, and
- Z is the number of unique UMIs involving 1 and 6.

The random variable Y takes values $y = 0, 1, 2, 3, 4$ according to the binomial distribution Bin(4,p):

$$P(Y=0) = q^4$$

$$P(Y=1) = 4pq^3$$

$$P(Y=2) = 6p^2q^2$$

$$P(Y=3)=4p^3q$$

$$P(Y=4)=p^4 \quad (3)$$

where $q = 1 - p$ is the probability of not being sampled.

The random variable Z takes values $z = 0, 1, 2$, with probabilities:

$$P(Z=0)=q^2$$

$$P(Z=1)=pq$$

$$P(Z=2)=p \quad (4)$$

Since

$$E(Y)=4p \text{ and } E(Z)=2p+pq=p(3-p) \quad (5)$$

then

$$\mu=4p+p(3-p)=p(7-p) \quad (6)$$

In particular,

$$\text{if } p=1/3, \mu=20/9\approx 2.222 \text{ is obtained.} \quad (7)$$

Thus, for $p = 0$ and $p = 1$, $\mu = 0$ and $\mu = 6$, respectively, are obtained.

Shown below is the general probability distribution of X followed by the special case $p = 1/3$, $q = 2/3$:

$$P(X=0)=q^4q^2=q^6\approx 0.0878$$

$$P(X=1)=q^4pq+4pq^3q^2=5pq^5\approx 0.2195$$

$$P(X=2)=q^4p+4pq^3pq+6p^2q^2q^2=pq^4+10p^2q^4\approx 0.2853$$

$$P(X=3)=4pq^3p+6p^2q^2pq+4p^3qq^2=4p^2q^3+10p^3q^3\approx 0.2414$$

$$P(X=4)=6p^2q^2p+4p^3qpq+p^4q^2=6p^3q^2+5p^4q^2\approx 0.1262$$

$$P(X=5)=4p^3qp+p^4pq=4p^4q+p^5q\approx 0.0357$$

$$P(X=6)=p^4p=p^5\approx 0.0041 \quad (8)$$

Computing the SD σ of X by taking the square root of the variance

$$\sigma^2=\text{Var}(X)=\text{Var}(Y)+\text{Var}(Z) \quad (9)$$

then

$$\text{Var}(Y)=4pq \quad (10)$$

and

$$\text{Var}(Z)=E(Z^2)-(E(Z))^2=pq+4p-(2p+pq)^2=p(5-p)-p^2(3-p)^2 \quad (11)$$

Yielding

$$\sigma^2=4p(1-p)+p(5-p)-p^2(3-p)^2=p(9-5p)-p^2(3-p)^2 \quad (12)$$

For the special case $p = 1/3$ and $q = 2/3$, $\sigma^2 = 1.654321$ is obtained, giving

$$\sigma\approx 1.286. \quad (13)$$

Results

Patients and Tumor-Informed ctDNA Analysis

Eight children diagnosed with rhabdomyosarcoma, osteosarcoma, Wilms tumor, synovial sarcoma, or germ cell tumor were longitudinally monitored with tumor-informed ctDNA analysis using SiMSen-Seq (Table 1 and Figure 1A). To identify each patient's somatic mutations, primary tumor DNA and germline DNA derived from leukocytes were analyzed with whole-exome sequencing. Up to 10 single-nucleotide variants were selected per patient to design personalized SiMSen-Seq panels. The panels were validated using human reference DNA, and poorly performing assays were redesigned or excluded. The final personalized ctDNA panels targeted between 4 and 10 tumor-specific mutations (Table 1). Detailed information about called mutations and SiMSen-Seq panels is shown in Supplemental Table S1.

Sequencing Error Correction Is Independent of UMI Family Size Cutoff

The distribution of PCR- and sequencing-induced errors in tumor-informed ctDNA analysis of 90 plasma samples collected from the eight patients was calculated. The background error rate decreased on average 55 times using UMIs compared with no UMIs (0.007% versus 0.366%), using a UMI family size of $\geq$2 (ie, a minimum of two raw reads with the same UMI was required to form a consensus read). Increasing the UMI family size cutoff showed no significant change in average background error rate, which ranged between 0.004% and 0.006% for UMI family size cutoffs between 2 and 30 (Figure 2A). Most of all raw sequencing errors were transversions (ie, a change from a purine to a pyrimidine or vice versa; 92.2%), whereas transitions (ie, a change from a purine to another purine or a change from a pyrimidine to another pyrimidine; 7.8%), deletions (0.5%), and insertions (none) were less common. After UMI correction, most substitution errors were corrected, leading to a smaller difference in frequency between transversions and transitions (58.6% versus 41.4%) (Figure 2B).

The average background error rate among all 66 assays was 0.007% using a UMI family size of $\geq$2. Only five of these assays showed average error rates >0.01%, of which two assays (*SPN* and *ZNF827*) had an average error rate >0.02% (Figure 2C). The error-corrected sequencing data for most of the assays displayed a pattern where most nucleotide positions in the amplicon were free from background errors. Figure 2D shows the average background error rate across all patient-specific assays for each individual sample. Most samples displayed a low average background error rate (<0.02%), but four samples (Patient 7 at time points 0, 35, 63, and 137 days) displayed higher background error rates (0.03% to 0.08%). The increased

Table 1 Patient, Sample, and Assay Characteristics

Patient identifier	Tumor type	Analyzed mutations, *n*	Blood samples, *n*
Patient 1	Rhabdomyosarcoma	10	11
Patient 2	Rhabdomyosarcoma	10	10
Patient 3	Osteosarcoma	4	15
Patient 4	Wilms tumor	10	7
Patient 5	Synovial sarcoma	10	15
Patient 6	Wilms tumor	7	9
Patient 7	Osteosarcoma	10	14
Patient 8	Germ cell tumor	5	9
Total		66	90

error rate can be explained by the *SPN* assay that performed poorly in these four samples, providing few detected molecules. This in combination with few individual molecules with random background errors that typically occur in some assays generated higher error rates. When comparing UMI-corrected background error rates for UMI family size cutoffs between 2 and 30 using a three-way analysis of variance, assay and sample showed significant differences ($P < 10^{-20}$) (Supplemental Table S2), whereas the UMI family size cutoff did not (Supplemental Tables S2 and S3). For comparison, the error rates using cfDNA generated from healthy individuals and genomic DNA from blood cells were assessed (Supplemental Figure S3). The level of errors was similar for the cfDNA derived from healthy individuals, but significantly lower for genomic DNA that represents intact molecules that have not been exposed to the same biological degradation processes as cfDNA.

The effect of UMI error correction is exemplified in a mutation-positive sample (*PCDHGA4*, at day 0 in Patient 1) and a mutation-negative sample (*CHD5*, during follow-up after successful treatment in Patient 6) in Figure 2E. For the *PCDHGA4* assay, the average background error rate was reduced 37 times from 0.185% to 0.005% with UMI error correction. Here, 1226 of 409,877 total raw reads (0.299%) showed the expected mutation, which is in the same range as the average background error level without UMI error correction. Using UMIs, 26 of 8671 consensus reads (0.300%) displayed the expected mutation, which was 60 times above the error-corrected background noise. Thus, the use of UMIs reduced experimentally induced errors without affecting the mutant allele frequency, demonstrating the advantage of digital sequencing. The average background error rate in the *CHD5* assay was 0.474% without and 0.004% with UMI error correction (ie, corresponding to a 119 times reduction in error rate). The *CHD5* assay displayed 4 of 339,374 total raw reads with the expected mutation without using UMIs, but no mutated reads were detected using UMIs. This indicated no ctDNA molecules for *CHD5* in this sample and that the mutated reads without using UMIs were false positives. In line with this, the patient remained free from disease with a follow-up time of 5 years.

Number of Detected Molecules Is Dependent on UMI Family Size Cutoff

In some digital sequencing approaches, the number of original target molecules may be directly estimated from the number of consensus reads. In SiMSen-Seq, each double-stranded DNA molecule generates on average 2.22 UMIs, assuming 100% target amplification efficiency (see *Materials and Methods* for details). As expected, the number of detected DNA molecules decreased with increasing UMI family size cutoff (Figure 3A). In total, 19.1% of the DNA molecules, assessed by fluorometric analysis after cell-free DNA extraction, were detected using UMI family size of $\geq$2, whereas only 9.1% of the molecules were detected using a UMI family size of $\geq$30. The number of molecules detected by SiMSen-Seq correlated linearly with the amount of cell-free DNA used as input in the library construction, showing that DNA input is a major factor for intersample variations in number of detected molecules (Figure 3B). The number of detected molecules was also assay dependent. The differences between the best and worst performing assay within patient-specific SiMSen-Seq panels ranged from 1.4 times for Patient 3 to 19 times for Patient 7 (Figure 3C). The average number of detected molecules per sample and time point ranged from 22.5 to 8070 (Figure 3D). It was concluded that a UMI family size of $\geq$2 was optimal to minimize error rates, while simultaneously maximizing number of detected molecules.

Clinical Assessment of ctDNA during Treatment and Follow-Up

To monitor treatment efficacy and detect minimal residual disease as well as relapse, ctDNA was assessed over time for each patient (Figure 4). For each sample, the total number of ctDNA molecules per mL plasma was considered for all assays together using a UMI family size of $\geq$2. A sample was defined as ctDNA positive if at least one mutation was called (ie, with a Q-score of $\geq$10 at the position of interest).

All five patients with an available baseline sample (day 0) were ctDNA positive before the start of treatment, ranging from 1.05 to 3855 ctDNA molecules per mL plasma. For Patient 1, the first three samples were collected at days 0, 1, and 2 of the first cycle of chemotherapy. All these samples were ctDNA positive, for 6, 10, and 10 assays each, and the total ctDNA level across all 10 assays peaked at 55.5 mutated molecules per mL plasma. After initiation of treatment, all samples became ctDNA negative, correlating with a good radiological response and complete regression at the follow-up after cessation of treatment.

Patient 2 was ctDNA positive in baseline sample, with a ctDNA level of 1.05 mutated molecules per mL plasma. After treatment initiation, the ctDNA levels were continuously below the detection limit until a sample collected 1 day before the planned tumor surgery indicated residual

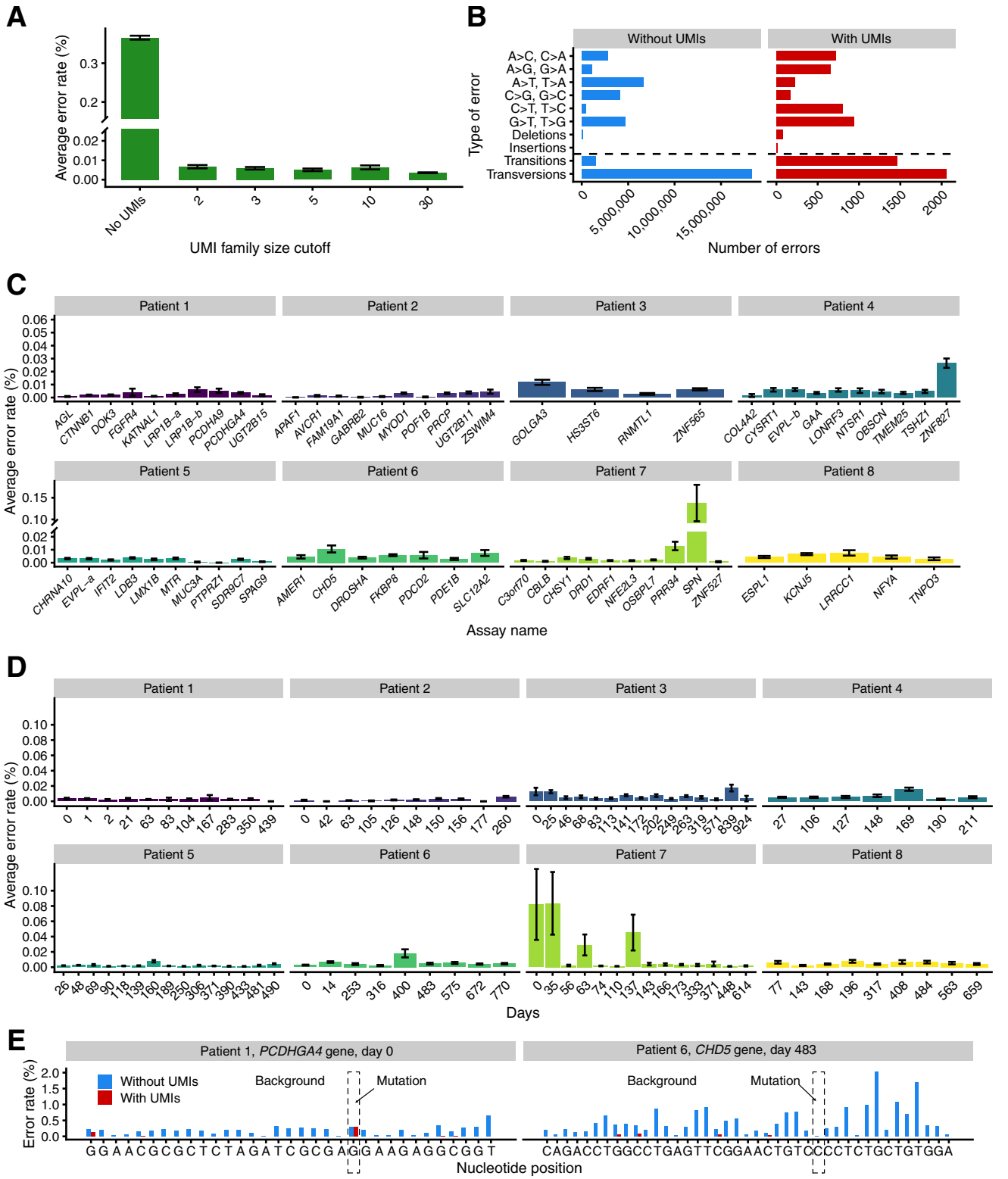


Figure 2 Error rates for tumor-informed ctDNA analysis using SiMSen-Seq. **A:** The average background error rate per position for different UMI family sizes. One-way analysis of variance showed no significant difference in average background error rates for UMI family size cutoffs 2, 3, 5, 10, and 30 ($P = 0.503$). Pairwise *t*-tests showed no significant differences between any of the UMI family size cutoffs (adjusted $P > 0.34$). **B:** Type of sequencing errors with and without UMI correction. The total number of background errors is shown across all patients, samples, and time points. The **dashed line** separates total number of transition and transversion errors from the individual error types. **C:** The average background error rate per assay when averaging over all samples. **D:** The average background error rate per sample when averaging over all assays. **E:** Example of the sequencing error rates per nucleotide position in target assays for two different samples and assays. The error rates are calculated as the sum of all nonreference variants divided by the total number of detected molecules at each nucleotide position. The target mutation positions are marked with **dashed boxed areas**. The left sample was ctDNA positive based on the *PCDHGA4* assay, whereas no ctDNA was detected in the right sample based on the *CHD5* assay. **B**—**E:** UMI family size of ≥ 2 was used as cutoff. Data are shown as means ± SEM (**A**, **C**, and **D**). $n = 572$ to 778 (**A**); $n = 7$ to 15 (**C**); $n = 4$ to 10 (**D**).

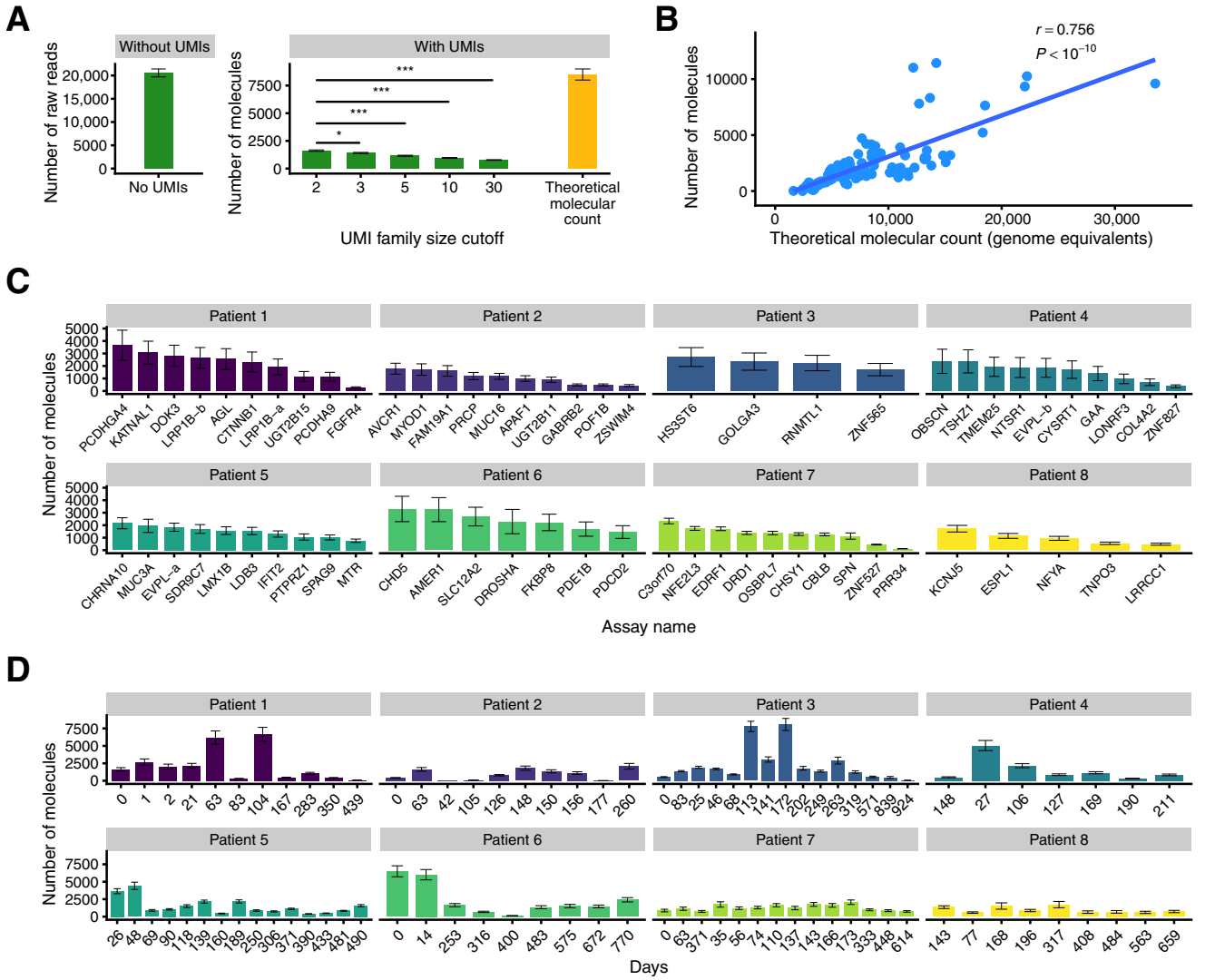


Figure 3 Number of detected molecules. **A:** Number of molecules in relation to UMI family size cutoff. For the data without UMI correction, the average number of raw reads per assay and sample is shown. For UMI-corrected data, the estimated number of molecules per target and sample is shown assuming that each input DNA molecule generates on average 2.22 UMI families (for details, see *Materials and Methods*). Pairwise *t*-test was performed. *False discovery rate <0.05; ***false discovery rate <0.001. The yellow bar shows estimated number of DNA molecules based on fluorometric analysis of input cell-free DNA in SiMSen-Seq, where 1 ng human DNA corresponds to 310 haploid genome equivalents. **B:** Correlation between detected number of molecules using SiMSen-Seq with UMI family size of ≥2 and the amount of input cell-free DNA per sample assessed using fluorometric analysis. Each dot refers to the maximum number of detected molecules in a panel for any individual assay. The linear fit is to guide the eye. The Pearson correlation coefficient (*r*) with *P* value is shown. **C:** The average number of detected molecules per assay, averaging all samples for each assay using UMI family size of ≥2. **D:** The average number of detected molecules per sample, averaging across all assays for each sample using UMI family size of ≥2. Data are shown as means ± SEM (**A**, **C**, and **D**). $n = 572$ to 778 (**A**); $n = 90$ (**B**); $n = 7$ to 15 (**C**); $n = 4$ to 10 (**D**).

disease with 11.0 mutated molecules per mL plasma. Three samples collected after the surgery were ctDNA negative. In Patients 3 to 7, ctDNA was positive in the first samples collected before or at the beginning of the treatment and negative in the following samples collected during and after treatment. In Patient 8, tumor surgery was performed before enrollment in the study, and all samples collected during and after adjuvant chemotherapy treatment were ctDNA negative.

Patients 2 and 5 experienced disease relapse 6 to 7 months after collection of the last blood samples, whereas the other six patients remained free from disease at their latest follow-up. Mutation calling using higher UMI family size cutoffs was also performed (Supplemental Figure S4). Overall, the number of detected mutations remained the same or decreased with the higher UMI family size cutoffs.

Discussion

In this study, the effects of UMI family size in tumor-informed ctDNA analysis with SiMSen-Seq in childhood

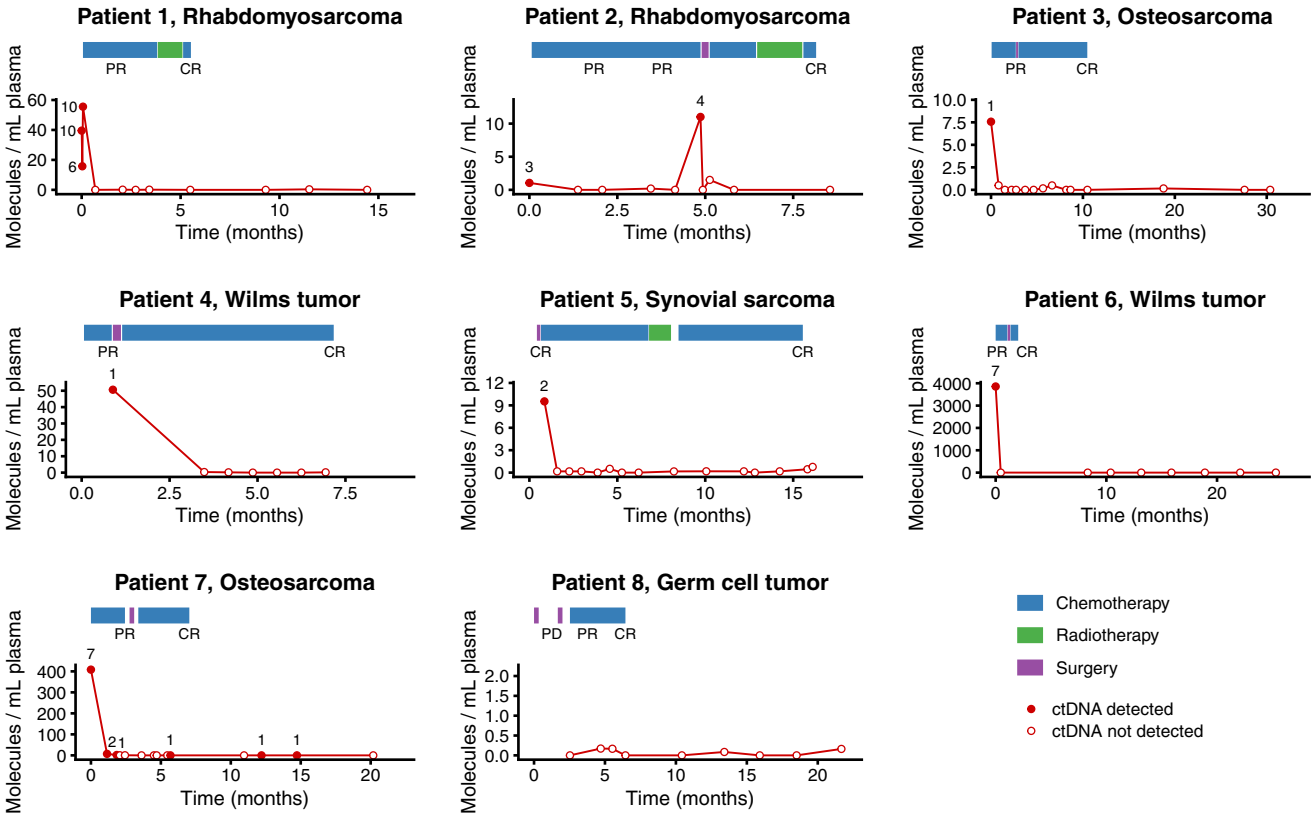


Figure 4 ctDNA levels in patients with childhood cancer. The total number of detected ctDNA molecules per mL plasma over time across all patients. Time in relation to start of treatment (baseline) is shown on the *x* axis. The *y* axis shows the total number of ctDNA molecules detected across all assays divided by the plasma volume used for input DNA in SiMSen-Seq analysis. **Filled circle** represents ≥1 assay that was statistically positive using a Q-score ≥10. The number of ctDNA-positive assays is shown above the circles. **Open circles** indicate no ctDNA detected. **Colored bars** above the graphs show the applied treatment and results of radiological assessments. UMI family size of ≥2 was used. CR, complete response; PD, progressive disease; PR, partial response.

cancers were determined. Childhood malignancies are characterized by few recurrent mutations in comparison to most adult cancers, making agnostic panels that depend on shared hotspot variants insufficient. Several ctDNA studies in pediatric cancers have focused on detecting structural variants, such as copy number variations and translocations.[6,39,40] Many approaches focus on individual patient-specific mutations using digital PCR, with only a few adopting tumor-informed approaches using sequencing panels.[4,27,41,42] The detection of ctDNA in blood plasma requires high sensitivity because the frequency of mutations can be <0.1% in cancer management applications, including treatment monitoring, prognostication, and detection of resistance mutations and minimal residual disease and early relapse detection. To ensure reliable and reproducible ctDNA analysis, the experimental procedure and assays should be carefully validated and standardized.[43,44] In this study, it is shown that UMIs efficiently correct for experimentally induced sequencing errors, enabling the analysis of individual cell-free DNA molecules and thus facilitating the detection of ctDNA at low allele frequencies. The sensitivity to detect ctDNA also improves with increasing number of assessed mutations.[45] SiMSen-Seq panel design is flexible regarding number of monitored mutations, single-nucleotide variants, and short insertions/deletions, and allows for addition of generic assays (eg, assays related to treatment resistance, such as *ALK* sequences in neuroblastoma undergoing targeted therapies).[46] Hence, digital sequencing with SiMSen-Seq can be applied to several applications within childhood cancer management.

There are multiple bioinformatic strategies for handling UMIs, and digital sequencing approaches often use their own tailored solutions. The most common approach, also applied in SiMSen-Seq, is to sequence each molecule multiple times and generate consensus reads. In DupSeq, complementary UMIs are bioinformatically related to each other to model each double-stranded DNA molecule and to correct for errors using reads of both strands.[20] There are also other methods that correct for sequencing errors without using UMIs, such as CAPP-seq,[47] ERASE-seq,[48] and Signera,[49] where sequencing errors are corrected by generating a background error model of healthy donor plasma. With SiMSen-Seq, the absolute number of ctDNA molecules in relation to blood plasma volume can be assessed using UMIs. After sequencing, approximately

19% of all cell-free DNA molecules estimated by Qubit were detected. The number of detected molecules is within the expected range for all analyzed samples. This number might be affected by overestimation of the DNA concentration by fluorometric analysis, here Qubit, which has been reported when carrier RNA is used during extraction of cell-free DNA.[50] The main losses depend on cell-free DNA fragmentation, where many target molecules cannot be amplified by standard sequencing primer pairs.[45] In addition, some DNA is lost during the concentration step preceding SiMSen-Seq, the amplification efficiencies are most likely not 100% because of variations in the specific assays, and presence of inhibitors carried over from the plasma will all affect the number of detected molecules.

Here, it is demonstrated that a UMI family size of ≥ 2 is optimal to minimize error rates while maximizing the number of detected molecules. A low cutoff is also important from a cost perspective because of the deep sequencing required by digital sequencing. For example, a SiMSen-Seq panel consisting of 10 individual assays analyzing 10 ng DNA will correspond to 3100 DNA molecules per sequence.[51] If all DNA molecules are detected, they will generate on average 6882 UMIs per assay (3100×2.22), and each UMI needs to be sequenced at least twice. Consequently, at least 137,640 reads per sample are required. Increasing the UMI family size cutoff will linearly increase the number of sequencing reads needed.

The use of UMIs efficiently corrects most technical errors. However, as polymerase-induced errors may occur during the first barcoding PCR step in SiMSen-Seq,[17] it cannot be determined if the small remaining noise is of experimental or biological origin. Sequencing errors in the UMI itself are not corrected in the same way as in the target sequence, as the UMI information is used for the error correction of the target sequence. Instead, the UMI sequences for each target are extracted in the first step of the bioinformatics pipeline and clustered on the basis of sequence similarity to ensure that highly similar UMIs are grouped together.[52] In the applied UMIErrorCorrect pipeline, the default setting for the edit distance threshold is set to 1 (ie, one mismatch between UMIs is allowed). The UMI length in relation to the amount of DNA input is an important factor for successful error correction, as UMI collisions may occur (ie, two different DNA molecules are labeled with identical UMIs) because of lack of diversity in the UMI pool. Longer UMI sequences help solve this issue as they allow for an increased number of possible combinations.[24] The use of UMIs can at both experimental and bioinformatic levels impact ctDNA detection as well as the estimated number of cell-free DNA molecules. Variant calling is based on a statistical model in which the background error is described using a β-binomial distribution.[38] The model has a high sensitivity for detection of mutations with low variant allele frequencies, with the total number of molecules and the number of mutated molecules at each position serving as key variables. Consequently, detection of low variant allele frequency variants is facilitated by a larger number of observed molecules. At low variant allele frequencies, no mutated ctDNA molecules may be present in the sample because of sampling ambiguity.[45] To address this issue, several mutations for each patient can be targeted, increasing the probability to detect ctDNA molecules. The limitations of this study are that data cover only one method and that the effects of UMI parameters beyond family size are not assessed, such as edit distance, UMI length, or structure. The number of patients with specific cancers is also low. Hence, the findings have technical rather than clinical value.

In conclusion, this study demonstrates the feasibility of using tumor-informed ctDNA analysis in childhood cancers. There are several experimental and bioinformatic factors that can influence the sensitivity of ctDNA assessments. Here, the effects of assay, sample, and UMI family size cutoff on background error rate and the number of detected molecules are shown. The generated results show that a UMI family size of ≥ 2 should be applied for best performance in tumor-informed ctDNA analysis using SiMSen-Seq.

Acknowledgments

We thank Mimmi Sörensen Larsson and the other research coordinators at the Pediatric Cancer Research Center, Sahlgrenska University Hospital (SU), for planning and coordinating patient enrollment and blood sampling, as well as clinical staff at the Childhood Cancer Center, SU, for patient enrollment and collection of blood samples; and acknowledge Clinical Genomics Gothenburg, Science for Life Laboratory, Sahlgrenska Academy, University of Gothenburg, and Center for Medical Genomics, Department of Clinical Genetic and Genomics, SU, for providing sequencing assistance. This study was conducted using professional biobank services from Biobank West and Biobank Sweden.

Author Contributions

T.Ö., D.A., M.D., and A.S. conceptualized the study; T.Ö., D.A., R.R.I., E.M., P.Mi., M.L.S., M.E., P.Mo., F.T.K., C.V., I.R., F.P.C., A.D.R., W.F., J.H., H.F., M.D., and A.S. curated data; T.Ö., D.A., M.D., and A.S. performed formal analysis; T.Ö., D.A., J.H., C.V., M.D., and A.S. acquired funding; T.Ö., D.A., R.R.I., E.M., J.Lu., J.Li, S.F., P.Mi., M.L.S., M.E., P.Mo., F.T.K., C.V., I.R., F.P.C., A.D.R., W.F., Å.T.N., S.S., J.H., H.F., M.D., and A.S. performed investigations; D.A., R.R.I., J.Lu., J.Li, S.F., P.Mi., M.L.S., M.E., P.Mo., F.T.K., I.R., F.P.C., A.D.R., and W.F. developed methods; J.H., M.D., and A.S. obtained resources; T.Ö. obtained software; M.D. and A.S. supervised the study; T.Ö. and D.A. performed visualization; T.Ö., D.A., and A.S. wrote the original draft; and T.Ö., D.A., R.R.I., E.M., J.Lu.,

ARTICLE IN PRESS

J.Li, S.F., P.Mi., M.L.S., M.E., P.Mo., F.T.K., C.V., I.R., F.P.C., A.D.R., W.F., Å.T.N., S.S., J.H., H.F., M.D., and A.S. reviewed and edited the manuscript.

Disclosure Statement

A.S. invented the SiMSen-Seq technology, which is protected by a patent (US serial number 15/552,618). S.F. is employed by and is a shareholder of Simsen Diagnostics AB. A.S. declares stock ownership in Tulebovaasta, Iscaff Pharma, and Simsen Diagnostics AB and is a board member in Tulebovaasta.

Supplemental Data

Supplemental material for this article can be found at *https://doi.org/10.1016/j.jmoldx.2026.08.002*.

References

1. Eguchi-Ishimae M, Tezuka M, Kokeguchi T, et al. Early detection of the PAX3-FOXO1 fusion gene in circulating tumor-derived DNA in a case of alveolar rhabdomyosarcoma. *Genes Chromosomes Cancer*. 2019;58:521−529.
2. Lyskjær I, Kara N, De Noon S, et al. Osteosarcoma: novel prognostic biomarkers using circulating and cell-free tumour DNA. *Eur J Cancer*. 2022;168:1−11.
3. Christodoulou E, Yellapantula V, O'Halloran K, et al. Combined low-pass whole genome and targeted sequencing in liquid biopsies for pediatric solid tumors. *NPJ Precis Oncol*. 2023;7:21.
4. Rahmqvist I, Engström E, Mellström E, et al. Personalized circulating tumor DNA analysis for sensitive disease monitoring and detection of relapse in neuroblastoma. *Biomark Res*. 2024;12:148.
5. Bettegowda C, Sausen M, Leary RJ, et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci Transl Med*. 2014;6:224ra224.
6. Klega K, Imamovic-Tuco A, Ha G, et al. Detection of somatic structural variants enables quantification and characterization of circulating tumor DNA in children with solid tumors. *JCO Precis Oncol*. 2018;2018:PO.17.00285.
7. Doculara L, Trahair TN, Bayat N, Lock RB. Circulating tumor DNA in pediatric cancer. *Front Mol Biosci*. 2022;9:885597.
8. Zill OA, Banks KC, Fairclough SR, et al. The landscape of actionable genomic alterations in cell-free circulating tumor DNA from 21,807 advanced cancer patients. *Clin Cancer Res*. 2018;24:3528−3538.
9. Ramkissoon LA, Pegram W, Haberberger J, et al. Genomic profiling of circulating tumor DNA from cerebrospinal fluid to guide clinical decision making for patients with primary and metastatic brain tumors. *Front Neurol*. 2020;11:544680.
10. Berko ER, Witek GM, Matkar S, et al. Circulating tumor DNA reveals mechanisms of lorlatinib resistance in patients with relapsed/refractory ALK-driven neuroblastoma. *Nat Commun*. 2023;14:2601.
11. Sweet-Cordero EA, Biegel JA. The genomic landscape of pediatric cancers: implications for diagnosis and treatment. *Science*. 2019;363:1170−1175.
12. Gröbner SN, Worst BC, Weischenfeldt J, et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555:321−327.
13. Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. Accuracy of next generation sequencing platforms. *Next Gener Seq Appl*. 2014;1:1000106.
14. Xu H, DiCarlo J, Satya RV, Peng Q, Wang Y. Comparison of somatic mutation calling methods in amplicon and whole exome sequence data. *BMC Genomics*. 2014;15:244.
15. Dang DK, Park BH. Circulating tumor DNA: current challenges for clinical utility. *J Clin Invest*. 2022;132:12.
16. Wan JC, Massie C, Garcia-Corbacho J, et al. Liquid biopsies come of age: towards implementation of circulating tumour DNA. *Nat Rev Cancer*. 2017;17:223−238.
17. Ståhlberg A, Krzyzanowski PM, Jackson JB, Egyud M, Stein L, Godfrey TE. Simple, multiplexed, PCR-based barcoding of DNA enables sensitive mutation detection in liquid biopsies using sequencing. *Nucleic Acids Res*. 2016;44:e105.
18. Kinde I, Wu J, Papadopoulos N, Kinzler KW, Vogelstein B. Detection and quantification of rare mutations with massively parallel sequencing. *Proc Natl Acad Sci U S A*. 2011;108:9530−9535.
19. Mattox AK, Wang Y, Springer S, et al. Bisulfite-converted duplexes for the strand-specific detection and quantification of rare mutations. *Proc Natl Acad Sci U S A*. 2017;114:4733−4738.
20. Kennedy SR, Schmitt MW, Fox EJ, et al. Detecting ultralow-frequency mutations by duplex sequencing. *Nat Protoc*. 2014;9:2586−2606.
21. McDonald BR, Contente-Cuomo T, Sammut SJ, et al. Personalized circulating tumor DNA analysis to detect residual disease after neoadjuvant therapy in breast cancer. *Sci Transl Med*. 2019;11:eaax7392.
22. Bae JH, Liu R, Roberts E, et al. Single duplex DNA sequencing with CODEC detects mutations with high sensitivity. *Nat Genet*. 2023;55:871−879.
23. Broche J, Kelemen O, Sekar A, et al. GeneBits: ultra-sensitive tumour-informed ctDNA monitoring of treatment response and relapse in cancer patients. *J Transl Med*. 2025;23:964.
24. Andersson D, Kebede FT, Escobar M, Österlund T, Ståhlberg A. Principles of digital sequencing using unique molecular identifiers. *Mol Aspects Med*. 2024;96:101253.
25. Kou R, Lam H, Duan H, et al. Benefits and challenges with applying unique molecular identifiers in next generation sequencing to detect low frequency mutations. *PLoS One*. 2016;11:e0146638.
26. Kivioja T, Vähärautio A, Karlsson K, et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat Methods*. 2012;9:72−74.
27. Rahmqvist I, Dahlstrand Rudin A, Mellström E, et al. Patient-specific sequencing panels enable sensitive circulating tumor DNA analysis in rhabdomyosarcoma independent of genetic profile. *NPJ Precis Oncol*. 2025;9:327.
28. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 2011;17:10−12.
29. Garcia M, Juhos S, Larsson M, et al. Sarek: a portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Res*. 2020;9:63.
30. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31:213−219.
31. Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15:591−594.
32. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing: Preprint. Posted online July 20, 2012. arXiv 1207.3907. https://doi.org/10.48550/arXiv.1207.3907.
33. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue of Somatic Mutations in Cancer. *Nucleic Acids Res*. 2019;47:D941−D947.
34. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform*. 2013;14:178−192.
35. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinform*. 2012;13:1−11.
36. Ståhlberg A, Krzyzanowski PM, Egyud M, Filges S, Stein L, Godfrey TE. Simple multiplexed PCR-based barcoding of DNA for ultrasensitive mutation detection by next-generation sequencing. *Nat Protoc*. 2017;12:664.

37. Andersson D, Kristiansson H, Luna Santamaría M, et al. Evaluation of automatic cell free DNA extraction metrics using different blood collection tubes. *Sci Rep*. 2025;15:19364.
38. Österlund T, Filges S, Johansson G, Ståhlberg A. UMIErrorCorrect and UMIAnalyzer: software for consensus read generation, error correction, and visualization using unique molecular identifiers. *Clin Chem*. 2022;68:1425−1435.
39. Ruhen O, Lak NSM, Stutterheim J, et al. Molecular characterization of circulating tumor DNA in pediatric rhabdomyosarcoma: a feasibility study. *JCO Precis Oncol*. 2022;6:e2100534.
40. Shah AT, Azad TD, Breese MR, et al. A comprehensive circulating tumor DNA assay for detection of translocation and copy-number changes in pediatric sarcomas. *Mol Cancer Ther*. 2021;20:2016−2025.
41. Fu Y, Xu Y, Liu W, et al. Tumor-informed deep sequencing of ctDNA detects minimal residual disease and predicts relapse in osteosarcoma. *EClinicalMedicine*. 2024;73:102697.
42. Miller AR, Shah T, Strawser CN, et al. MRD4U: a path to development for personalized liquid biopsy for children with central nervous system tumors. *BMC Cancer*. 2025;25:1365.
43. Lockwood CM, Borsu L, Cankovic M, et al. Recommendations for cell-free DNA assay validations: a joint consensus recommendation of the Association for Molecular Pathology and College of American Pathologists. *J Mol Diagn*. 2023;25:876−897.
44. Pascual J, Attard G, Bidard FC, et al. ESMO recommendations on the use of circulating tumour DNA assays for patients with cancer: a report from the ESMO Precision Medicine Working Group. *Ann Oncol*. 2022;33:750−768.
45. Johansson G, Andersson D, Filges S, et al. Considerations and quality controls when analyzing cell-free tumor DNA. *Biomol Detect Quantif*. 2019;17:100078.
46. Ek T, Ibrahim RR, Vogt H, et al. Long-lasting response to lorlatinib in patients with ALK-driven relapsed or refractory neuroblastoma monitored with circulating tumor DNA analysis. *Cancer Res Commun*. 2024;4:2553−2564.
47. Chaudhuri AA, Chabon JJ, Lovejoy AF, et al. Early detection of molecular residual disease in localized lung cancer by circulating tumor DNA profiling. *Cancer Discov*. 2017;7:1394−1403.
48. Kamps-Hughes N, McUsic A, Kurihara L, et al. ERASE-Seq: leveraging replicate measurements to enhance ultralow frequency variant detection in NGS data. *PLoS One*. 2018;13:e0195272.
49. Reinert T, Henriksen TV, Christensen E, et al. Analysis of plasma cell-free DNA by ultradeep sequencing in patients with stages I to III colorectal cancer. *JAMA Oncol*. 2019;5:1124−1131.
50. van Dessel LF, Vitale SR, Helmijr JC, et al. High-throughput isolation of circulating tumor DNA: a comparison of automated platforms. *Mol Oncol*. 2019;13:392−402.
51. Piovesan A, Pelleri MC, Antonaros F, Strippoli P, Caracausi M, Vitale L. On the length, weight and GC content of the human genome. *BMC Res Notes*. 2019;12:1−7.
52. Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res*. 2017;27:491−499.