nature cancer

Technical Report

https://doi.org/10.1038/s43018-026-01115-4

M-PACT leverages cell-free DNA methylomes to achieve robust classification of pediatric brain tumors

Received: 10 March 2025

Accepted: 9 January 2026

Published online: 17 February 2026

Check for updates

A list of authors and their affiliations appears at the end of the paper

Cerebrospinal fluid (CSF) liquid biopsies serve as a rich source of tumor-derived cell-free DNA (cfDNA) for evaluating persons with central nervous system (CNS) tumors. However, challenges stemming from trace cfDNA yields and low mutational burden have hindered sensitivity, whereas first-generation clinical assays have relied on genetic alterations as biomarkers. Leveraging the diagnostic utility of DNA methylation classification in CNS tumors, we developed M-PACT (methylation-based predictive algorithm for CNS tumors), a robust deep neural network that accurately classifies tumors from subnanogram-input cfDNA methylomes. Across embryonal CNS tumor benchmarking (n = 79) and validation (n = 58) cohorts, M-PACT achieved 92% and 88% accuracy, respectively. We further showcase M-PACT utility in nonembryonal CNS tumors, balanced tumor genomes and nonmalignant CSF. Beyond classification, this workflow enables methylation-based cellular deconvolution and sensitive copy-number variation detection. Altogether, we provide a blueprint for CNS tumor classification from low-input cfDNA methylomes, motivating prospective validation for future clinical implementation.

Pediatric brain tumors represent the leading cause of cancer-related death in children and encompass a heterogeneous group of low-grade and high-grade malignancies[1]. Over the past decade, advances in molecular profiling have provided foundational insights into disease biology and redefined tumor entities with novel biomarkers that now guide the World Health Organization diagnostic criteria for central nervous system (CNS) tumors[2,3]. Foremost, DNA methylation-based classification has become the cornerstone of pediatric brain tumor diagnostics, providing a molecular fingerprint that accurately discriminates between biologically distinct tumor subtypes associated with a wide range of clinical implications[4–6]. Current diagnostic pipelines, including histopathological evaluation and methylation classification, rely on the availability of tissue specimens. However, not all CNS tumors are amenable to neurosurgical intervention because of their high-risk locations; moreover, tumor biopsies may yield limited material and longitudinal tissue collection is not routinely conducted, especially in metastatic cases.

Liquid biopsies have the potential to overcome these challenges by providing a noninvasive source of tumor biomarkers (that is, circulating tumor cells and circulating tumor DNA (ctDNA)) in clinical oncology[7,8]. Compared to traditional tissue biopsies, liquid biopsies capture cells and/or genetic material released from anatomically distinct tumor subclones, thus offering more comprehensive insights into the cellular composition of malignancy[7,8]. While liquid biopsies from routine blood samples have revolutionized clinical management for common solid tumors and hematologic malignancies of adulthood[9–12], persons with CNS tumors encounter limitations with this approach, as the blood–brain barrier obstructs sensitive detection of tumor-derived material in plasma[13–17]. Cerebrospinal fluid (CSF) collected intraoperatively or, less invasively, through lumbar puncture (LP) or ventricular access has proven to be a rich source of disease analytes, namely ctDNA, in the adult and pediatric population with CNS tumors[14,16–18]. Prior studies have used divergent molecular assays, from targeted approaches (that is, droplet digital PCR and gene panels) to genome-wide sequencing (that is, low-coverage whole-genome sequencing (lcWGS) and Nanopore sequencing), demonstrating the feasibility of ctDNA profiling for risk stratification and disease monitoring[19–29]. Despite this promise, current

e-mail: paul.northcott@stjude.org

approaches lack coherence for ctDNA detection and are hampered by blind spots because of their reliance on high cell-free DNA (cfDNA) inputs, high-tumor-burden samples and genetic alterations such as copy-number variations (CNVs) or pathognomonic genetic mutations.

To address these challenges, while harnessing the diagnostic utility of DNA methylation, we developed M-PACT (methylation-based predictive algorithm for CNS tumors), a deep neural network (DNN) classifier that incorporates (1) CpG imputation; (2) a tumor enrichment algorithm; and (3) methylation-based deconvolution. Application of M-PACT to subnanogram-input cfDNA methylomes from the pediatric population with CNS tumors (n = 156) and nonmalignant donors (n = 58) demonstrated high accuracy and specificity in entity classification. Moreover, featured participant vignettes highlight the potential clinical utility of M-PACT in diagnosing CNS tumors from low-input samples, discriminating primary from secondary malignancies and monitoring disease status without prerequisite knowledge of underlying molecular landscapes. Collectively, we present a methylation classifier that overcomes current limitations in CSF liquid biopsies and holds potential as a valuable tool for cfDNA-based tumor classification.

Results

A methylation-based classifier for CNS tumor liquid biopsies

Prior efforts to implement DNA methylation-based tumor classification in liquid biopsies have been challenged by (1) low cfDNA inputs, which in turn yield shallow genome-wide sequencing coverage; (2) sparse CpG capture; and (3) low ctDNA fractions[23,30–32]. To address high CpG sparsity in cfDNA methylomes, we first created a network-based regression diffusion model to impute methylation values uncaptured by sequencing (Fig. 1a). Our imputation network, trained on a vast CNS tumor DNA methylation array reference cohort (n = 914), achieved high imputation accuracy even when provided values for only 100,000 CpGs (Extended Data Fig. 1a). We next introduced varying levels of tumor burden to simulate real-world samples for classifier training and validation. A comprehensive set of n = 3,195,000 in silico samples with diverse combinations of tumor fraction and CpG sparsity (ranging from 0.1 to 1.0) was generated from reference data representing 84 CNS tumor and 13 nontumor entities (Extended Data Fig. 1b). Three neural networks, which included a generic model and two models focused on low tumor fraction and sparse CpG recovery, attained a balanced accuracy of >0.9 after n = 100 epochs and were subsequently integrated into a three-layer ensemble model, yielding M-PACT (Fig. 1a and Extended Data Fig. 1c). Tumor signal and classification probabilities were enhanced by β regression of methylation signatures derived from nine nonmalignant cell classes (Methods) to subtract predicted normal background content from bulk methylation profiles. Following successful DNA methylation-based classification, tumor burden was estimated by deconvoluting methylation profiles into malignant and nonmalignant fractions for samples with tumor classification scores above 0.7 (Fig. 1a).

In the in silico test cohort (n = 1,152,000 simulated samples), M-PACT achieved average F_1 scores of 0.95 and 0.91 for tumor fractions above and below 0.5, respectively, and average F_1 scores of 0.96 and 0.86 for CpG sparsity below and above 0.5, respectively (Fig. 1b). In silico samples that were correctly classified by M-PACT showed a median tumor fraction of 0.32 and sparsity of 0.70 (that is, 30% of relevant CpGs captured), simulating cfDNA methylomes derived from real-world samples. In silico samples misclassified by M-PACT were driven by dual challenges of significantly lower tumor fractions (median misclassified: 0.15) and higher CpG sparsity (median misclassified: 0.81) (Fig. 1c,d).

Overall, M-PACT classification of in silico simulations was efficacious for samples with low tumor fraction and high CpG sparsity, motivating application to individual-derived cfDNA methylomes.

Enzymatic methylation sequencing (EM-seq) faithfully captures genome-wide CNVs from CSF liquid biopsies

To generate cfDNA methylomes from low-input, CSF-derived cfDNA specimens typically observed in pediatric populations with CNS tumors, we optimized the recently published EM-seq method[33] for cfDNA sequencing. Applying EM-seq to n = 268 cfDNA samples, minimal cfDNA inputs (median: 0.5 ng, range: 0.05–50 ng) yielded a median of 7 million CpG sites (range: <6,000 to >27,000,000) captured by ~4× genome-wide sequencing coverage (Extended Data Fig. 1d–k and Supplementary Table 1). We hypothesized that WGS of enzymatically converted cfDNA allows for both robust CNV detection equivalent to lcWGS[20,22,24,29] and accurate CNS tumor classification analogous to methods established for tumor tissue (Fig. 2a)[4].

To establish the feasibility of this approach in cfDNA-derived EM-seq data, we first characterized profiles using varying inputs (that is, 0.1–50 ng) of a high-yield cfDNA sample collected from an individual with choroid plexus papilloma. Tumor-associated CNVs detected by EM-seq were concordant with the tumor tissue-derived profile, as determined by DNA methylation array. M-PACT correctly classified cfDNA inputs as low as 0.1 ng despite substantial increases in data sparsity and duplication rate. Furthermore, DNA methylation patterns overlapping the transcriptionally active *OTX2* locus[34] were robustly recapitulated by EM-seq across the input range (Extended Data Fig. 2a–c). To mimic the diverse spectrum of ctDNA fractions expected in real-world samples, we subsequently mixed cfDNA from a high-tumor-fraction (94% ctDNA fraction based on CNVs) sonic hedgehog-activated medulloblastoma (MB, SHH) with a pool of nonmalignant cfDNA at decreasing tumor spike-in concentrations. Tumor-derived CNVs and CpG methylation patterns, exemplified at the *PTCH1* locus, were robustly detected and successfully classified by M-PACT for ctDNA fractions as low as 5% (Extended Data Fig. 2d,e). These results demonstrate the sensitivity of the EM-seq/M-PACT pipeline to accurately detect CNVs and achieve robust classification from sparse, low-tumor-fraction, low-input cfDNA profiles.

To systematically evaluate CNV detection from EM-seq and lcWGS data, we conducted an orthogonal comparison between the two platforms using samples where both sequencing assays were applied to the same input material. Using an optimized version of our previously published CNV caller (Methods)[20,35], we found high concordance between the assays, as exemplified by cases of MB, G3 with *MYCN* or *MYC* amplification, embryonal tumor with multilayered rosettes (ETMR) with C19MC alteration and atypical teratoid rhabdoid tumor (ATRT) with chr22q deletion (Fig. 2b). Across 110 samples from 72 participant CNS tumors (n = 105 CSF) and five nononcological controls (n = 5 CSF), genome-wide CNV profiles generated by the two cfDNA assays exhibited a median-scaled Manhattan similarity (Methods) of 0.94 (Fig. 2c). Estimation of ctDNA fraction based on CNVs demonstrated a significant correlation between EM-seq and lcWGS (r = 0.928) (Fig. 2d). Outliers in this comparison could largely be explained by differences in sequencing depth (Extended Data Fig. 3a–d).

These findings demonstrate that EM-seq enables reliable CNV-based ctDNA detection consistent with first-generation lcWGS assays[20,22,24,27,29].

M-PACT benchmarking for embryonal CNS tumor classification from cfDNA

To leverage the classification potential of EM-seq, we benchmarked M-PACT using ctDNA-positive CSF (n = 79 samples, n = 79 participants), as determined by the presence of CNVs. M-PACT demonstrated high classification accuracy, with 73/79 (92%) cfDNA samples matching the participant-derived tumor tissue methylation array classification (Fig. 3a,b). Three correctly classified MB, G3/4 samples exhibited subgroup switching along the known subgroup/subtype continuum[36–40]. Notably, most of the correctly classified samples had cfDNA inputs in the subnanogram range, highlighting the robustness of the EM-seq/M-PACT workflow. Correctly classified ctDNA samples clustered with their respective tumor methylation array references (Fig. 3c). Samples clustering with the nonmalignant background

https://doi.org/10.1038/s43018-026-01115-4

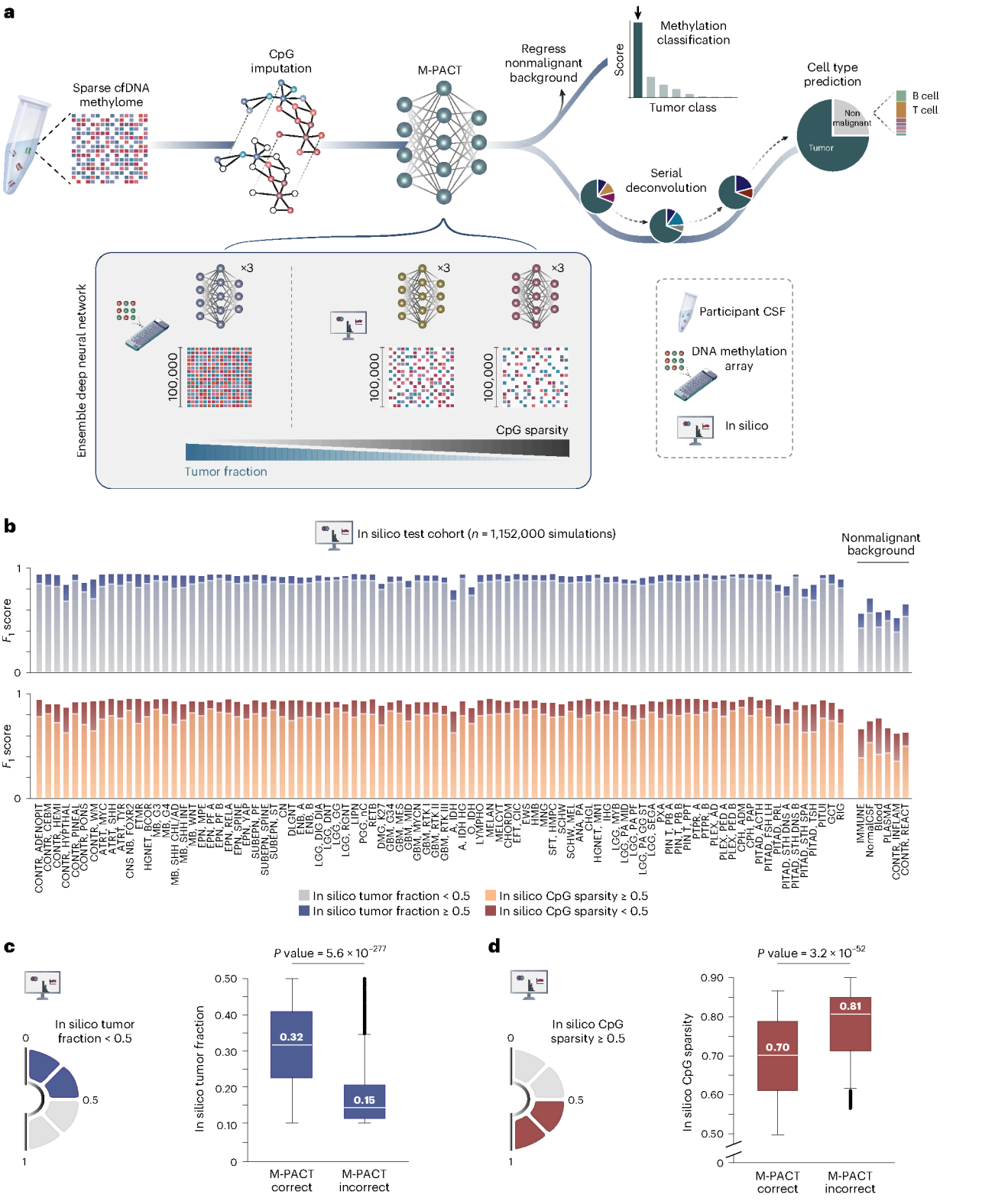


Fig. 1 | Development of a methylation-based classifier for CNS tumor liquid biopsies. a, Schematic workflow of the M-PACT computational pipeline. **b**, F_1 scores of M-PACT performance for the in silico test cohort ($n = 1{,}152{,}000$ samples) by simulated tumor fraction above or below 0.5 (top) and simulated CpG sparsity below or above 0.5 (bottom). **c**, Box plot comparing simulated tumor fraction of in silico samples correctly and incorrectly classified by M-PACT. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 5.6 \times 10^{-277}$ ($n = 1{,}152{,}000$ samples). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **d**, Box plot comparing simulated CpG sparsity of in silico samples correctly and incorrectly classified by M-PACT. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 3.2 \times 10^{-52}$ ($n = 1{,}152{,}000$ samples). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range.

https://doi.org/10.1038/s43018-026-01115-4

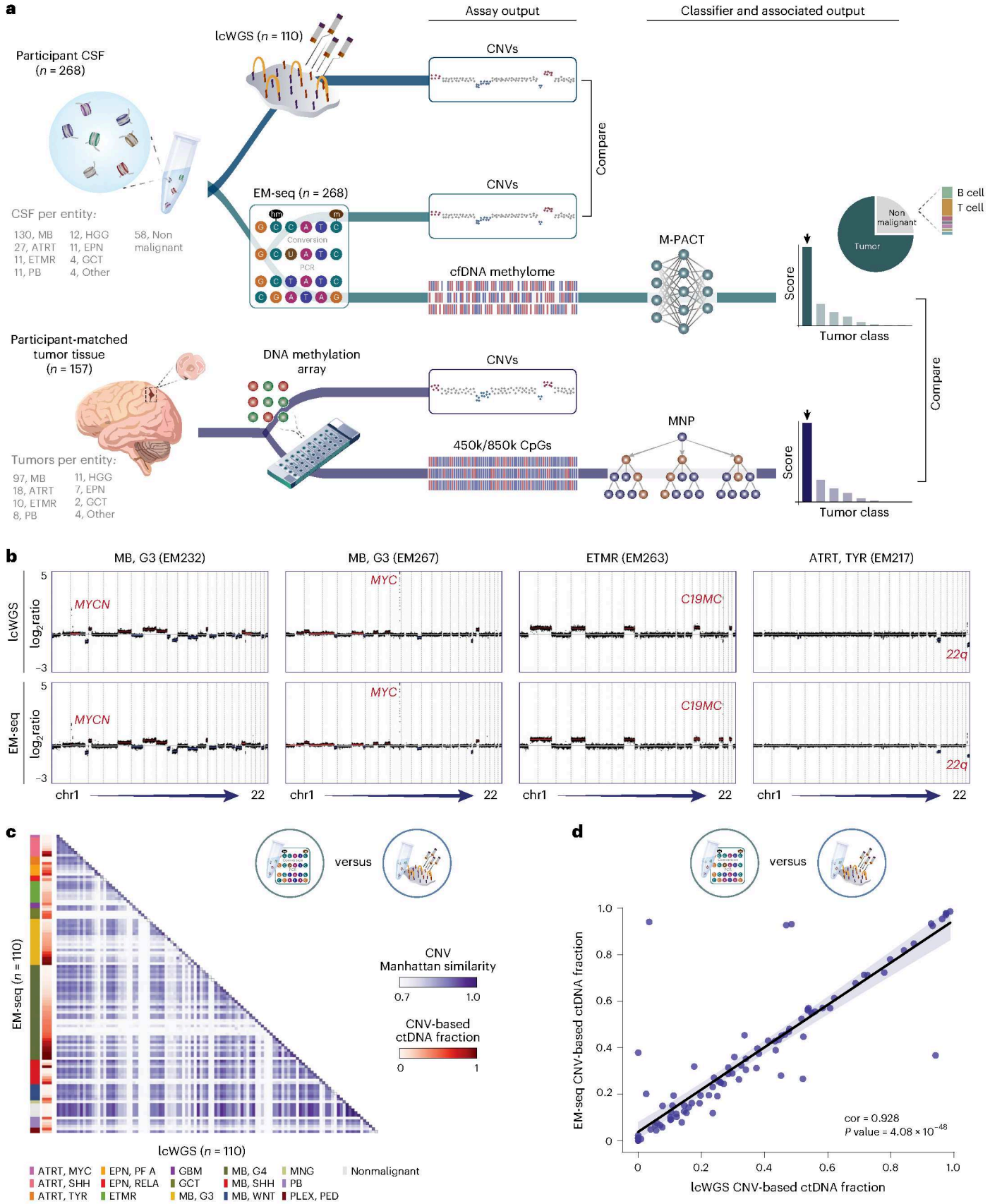


Fig. 2 | Accurate ctDNA detection using EM-seq. a, Schematic overview of CSF samples subjected to lcWGS and EM-seq for comparative CNV and DNA methylation analysis. **b**, Exemplary CNV plots for cfDNA samples sequenced by lcWGS (top) and EM-seq (bottom). **c**, Heat map of scaled Manhattan similarity for genome-wide CNVs in cfDNA samples analyzed by both EM-seq and lcWGS ($n = 110$). **d**, Scatter plot displaying CNV-based ctDNA fraction for EM-seq versus lcWGS. Pearson correlation = 0.928, $P = 4.08 \times 10^{-48}$; shading depicts the 95% confidence interval. Abbreviations: EPN, RELA/PF A, ependymoma, RELA-fusion positive/posterior fossa group A; GCT, germ cell tumor; HGG, high-grade glioma; MNG, meningioma; PB, pineoblastoma; PLEX, PED, choroid plexus tumor, pediatric.

reference still classified correctly ($n = 23$), emphasizing the sensitivity of M-PACT classification despite lower tumor fractions (median ctDNA fraction: 0.26). Across the benchmarking cohort, samples that were unclassifiable (that is, classification below threshold) or misclassified as 'nonmalignant background' showed a significantly lower ctDNA fraction (median incorrect: 0.13, correct: 0.32), while CpG sparsity remained comparable to that of correctly classified samples (median incorrect: 0.64, correct: 0.52) (Fig. 3d,e).

M-PACT validation across pediatric CNS tumor liquid biopsies

We subsequently applied the EM-seq/M-PACT workflow to a validation cohort of $n = 58$ CSF samples collected from $n = 48$ participants with embryonal CNS tumors. A correct M-PACT classification was achieved in 51/58 (88%) cfDNA samples (Fig. 3f). Consistent with benchmarking cohort results, samples incorrectly classified by M-PACT exhibited significantly lower ctDNA fraction (median incorrect: 0.1, correct: 0.39), with no significant difference in CpG sparsity (median incorrect: 0.13, correct: 0.34) (Fig. 3g,h). To expand our efforts beyond embryonal subtypes, we investigated M-PACT classification accuracy in an independent cohort ($n = 29$ CSF samples, $n = 23$ participants) encompassing high-grade glioma, ependymoma, germ cell tumor, choroid plexus tumor, meningioma and high-grade neuroepithelial tumor with *BCOR* alteration (Fig. 3i). M-PACT correctly classified 22/29 (76%) cfDNA methylomes with high probability (range: 0.71–0.99) (Fig. 3i). Samples incorrectly classified by M-PACT harbored significantly lower ctDNA fractions (median incorrect: 0.13, correct: 0.44) (Fig. 3j). This threshold is consistent with those observed in the benchmarking (Fig. 3d) and embryonal validation (Fig. 3g) cohorts. Additionally, incorrectly classified samples in the nonembryonal cohort exhibited significantly higher levels of CpG sparsity (median incorrect: 0.86, correct: 0.51) (Fig. 3k). Overall, application of M-PACT across the three EM-seq cohorts determined that a ctDNA fraction threshold of 0.15 yielded 95% classification accuracy (Extended Data Fig. 4a).

Taken together, successful results in benchmarking and validation cohorts establish the EM-seq/M-PACT pipeline as a reliable tool for cfDNA methylation-based classification across CNS tumor types.

Cross-platform application of M-PACT for cfDNA classification

Next, we sought to assess the compatibility of M-PACT across epigenetic profiling platforms, including Nanopore sequencing, whole-genome bisulfite sequencing (WGBS) and methylation array analysis. We generated sample-matched Nanopore sequencing data for $n = 21$ cfDNA samples (input range: 0.5–2.7 ng) correctly classified in the M-PACT benchmarking cohort (Fig. 3a–e). CNV profiles were highly concordant across EM-seq and Nanopore platforms (Fig. 4a). Comparisons between the data types revealed significant differences in CpG sparsity at 4× coverage (median EM-seq: 0.41, Nanopore: 0.99) (Fig. 4b). Using M-PACT, 18/21 (86%) Nanopore profiles were correctly classified (Fig. 4c), suggesting versatility of the M-PACT algorithm. These findings were further corroborated in published pediatric CNS tumor cohorts of (1) low-coverage Nanopore sequencing of CSF-derived cfDNA[23] (M-PACT: 32/51 (63%) correct versus NanoDx: 18/51 (35%) correct; Extended Data Fig. 4b,c); (2) Nanopore sequencing of tumor tissue-derived genomic DNA[41] (M-PACT: 142/184 (77%) correct versus published result: 148/184 (80%) correct; Extended Data Fig. 4d); and (3) WGBS of CSF-derived cfDNA[31] (M-PACT: 13/14 (93%) correct; Extended Data Fig. 4e). M-PACT classification accuracy was either consistent with or exceeded that reported in the original studies.

Moreover, we assessed applicability of M-PACT to widely available DNA methylation arrays, including profiles generated from tumor-derived genomic DNA and CSF-derived cfDNA. Comparison of tumor tissue classification results obtained with M-PACT versus the gold-standard Heidelberg brain tumor classifier (MNP)[4,6] ($n = 157$ tumor samples) confirmed the same tumor type in 95% of participant-matched cases (Extended Data Fig. 4f). In addition, using a published DNA methylation array dataset encompassing 574 independent CNS tumor profiles representing 37 entities, M-PACT achieved high concordance (that is, 96.8%) with a recently published neural network classifier (MLPnet)[42] (Extended Data Fig. 4g).

When orthogonally comparing the performance of M-PACT to MNP on cfDNA-based methylation array profiles (Fig. 4d), MNP correctly classified 15/24 (62.5%) CSF samples, whereas M-PACT successfully classified 20/24 (83.3%) (Fig. 4e). In addition, M-PACT assigned higher probability scores than MNP (MNP median: 0.68, M-PACT median: 0.98) (Fig. 4f).

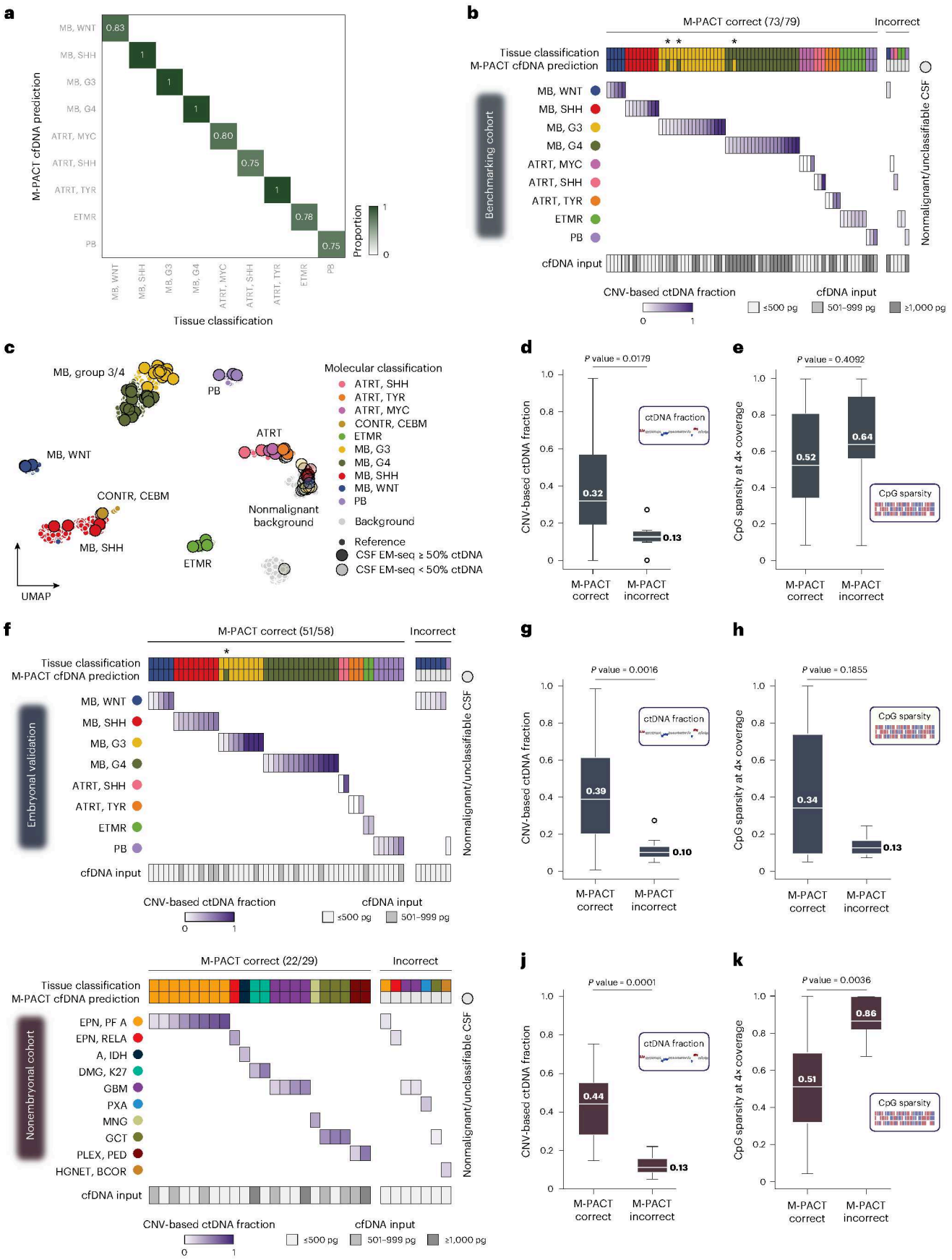
Across different state-of-the-art DNA methylation platforms, M-PACT demonstrated consistent utility for classifying diverse CNS tumor entities, while achieving accuracy comparable with or exceeding conventional classification algorithms.

Methylation-based binary classification enhances ctDNA sensitivity

While leveraging cfDNA methylomes allows for granular tumor classification by M-PACT, a binary readout (that is, tumor-positive versus tumor-negative) for ctDNA detection may increase sensitivity. To test

Fig. 3 | Benchmarking and validation of M-PACT for ctDNA classification. **a**, Confusion matrix summarizing M-PACT cfDNA predictions compared to tissue-based classifications. **b**, Heat map summarizing classification performance for M-PACT ($n = 73/79$ samples, 92% correct), annotated with EM-seq-derived CNV-based ctDNA fraction and cfDNA inputs. Asterisks indicate correctly classified samples with subgroup switching along the known MB, G3/4 continuum. **c**, Uniform manifold approximation and projection (UMAP) of cfDNA EM-seq profiles onto tumor and nonmalignant methylation array reference sets. **d**, Box plot comparing CNV-based ctDNA fraction for correctly and incorrectly classified CSF samples by M-PACT. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.0179$ ($n = 73$ and $n = 6$ samples, respectively). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **e**, Box plot comparing CpG sparsity for correctly and incorrectly classified CSF samples by M-PACT. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.4092$ ($n = 73$ and $n = 6$ samples, respectively). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **f**, Heat map summarizing classification performance for M-PACT ($n = 51/58$ samples, 88% correct) in the embryonal validation cohort, annotated with EM-seq-derived CNV-based ctDNA fraction and cfDNA inputs. Asterisk indicates a correctly classified sample with subgroup switching along the known MB, G3/4 continuum. **g**, Box plot comparing CNV-based ctDNA fraction for correctly and incorrectly classified CSF samples by M-PACT. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.0016$ ($n = 51$ and $n = 7$ samples, respectively). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **h**, Box plot comparing CpG sparsity for correctly and incorrectly classified CSF samples by M-PACT. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.1855$ ($n = 51$ and $n = 7$ samples, respectively). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **i**, Heat map summarizing classification performance for M-PACT ($n = 22/29$ samples, 76% correct) in the nonembryonal cohort, annotated with EM-seq-derived CNV-based ctDNA fraction and cfDNA inputs. **j**, Box plot comparing CNV-based ctDNA fraction for correctly and incorrectly classified CSF samples by M-PACT. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.0001$ ($n = 22$ and $n = 7$ samples, respectively). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **k**, Box plot comparing CpG sparsity for correctly and incorrectly classified CSF samples by M-PACT ($n = 22$ and $n = 7$ samples, respectively). Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.0036$. The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. Abbreviations: A, IDH, astrocytoma, isocitrate dehydrogenase mutant; DMG, K27, diffuse midline glioma, H3 K27 mutant; PXA, pleomorphic xanthoastrocytoma; HGNET, BCOR, high-grade neuroepithelial tumor, *BCOR* alteration; CONTR, CEBM, control tissue, cerebellar hemisphere.

https://doi.org/10.1038/s43018-026-01115-4



https://doi.org/10.1038/s43018-026-01115-4

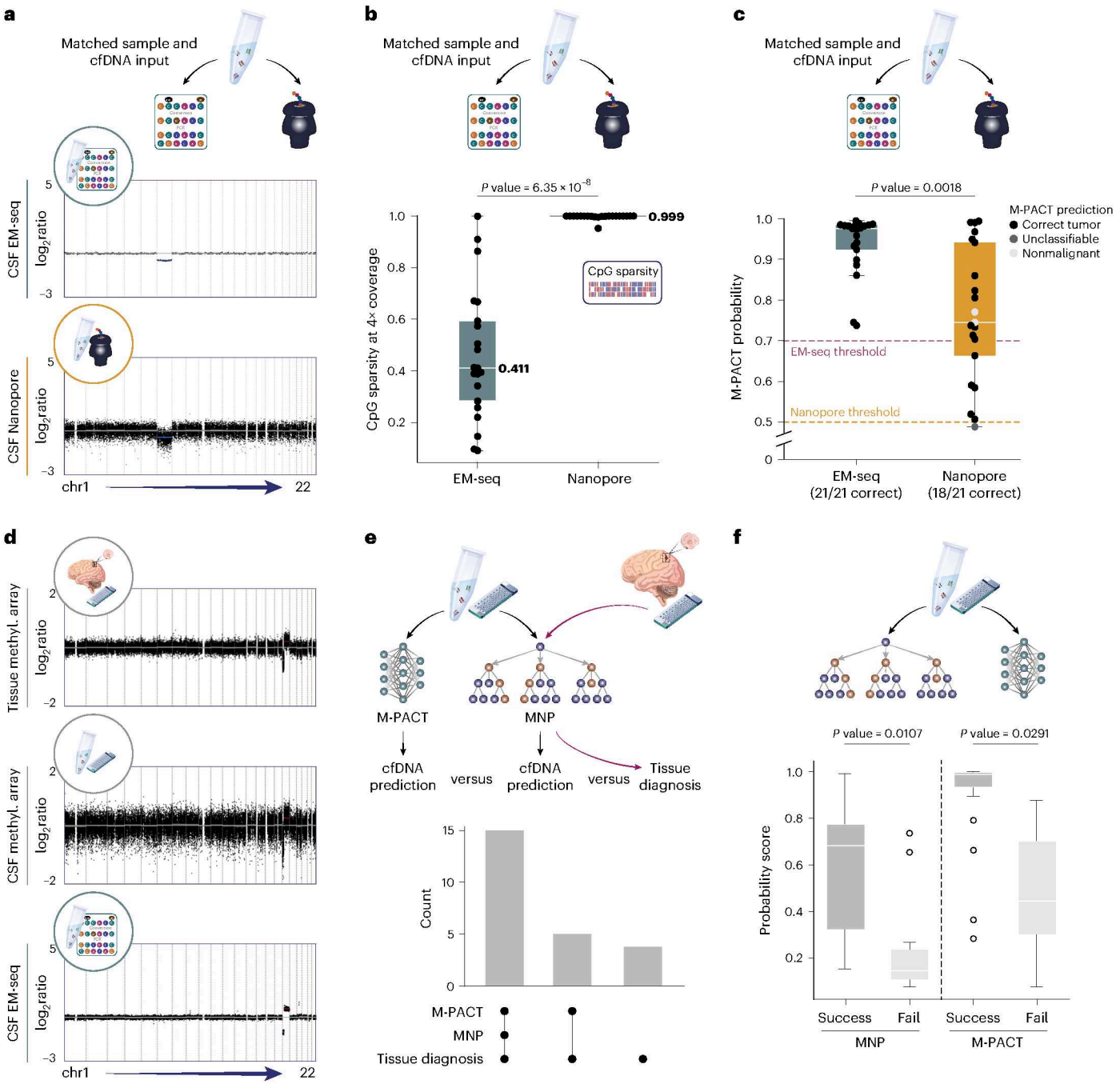


Fig. 4 | Cross-platform application of the M-PACT classifier. a, Exemplary CNV plots output from EM-seq and Nanopore sequencing of sample-matched cfDNA (MB, WNT (EM239)). **b**, Box plot of CpG sparsity for EM-seq and Nanopore cfDNA profiles. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 6.35 \times 10^{-8}$ ($n = 21$ samples). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **c**, Box plot of M-PACT probability for EM-seq and Nanopore cfDNA profiles, annotated with M-PACT predictions and platform-specific classification thresholds. Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.0018$ ($n = 21$ samples). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range. **d**, Exemplary CNV plots output from participant-matched tumor tissue DNA methylation array, cfDNA methylation array and cfDNA EM-seq (MB, G4 (EM087)). methyl., methylation. **e**, UpSet plot comparing cfDNA predictions by M-PACT versus MNP against tissue-based classification. **f**, Box plot comparing M-PACT versus MNP probability scores between correctly and incorrectly classified cfDNA samples ($n = 24$ samples). Statistical analysis was conducted using a two-sided Mann–Whitney *U*-test. $P = 0.0107$ (MNP), $P = 0.0291$ (M-PACT). The median is shown as a thick line, box limits are the 25th and 75th percentiles and whiskers denote 1.5× the interquartile range.

this, we developed an algorithm that compares the CSF-derived cfDNA methylome to participant-matched tumor tissue methylation array profile, nonmalignant control CSF and nonmalignant control brain tissue for a weighted correlation analysis, resulting in a combined *z* score (Extended Data Fig. 5a). Values above or below the combined *z*-score threshold of 1.2 determine the binary classification outcome (Methods). Across all CNV-positive cfDNA samples within the benchmarking, embryonal CNS tumor validation and nonembryonal CNS tumor cohorts, the binary classification algorithm detected ctDNA in 159/166 (96%) samples, compared to 146/166 (88%) by M-PACT. Samples rescued by the binary approach exhibited a median ctDNA fraction of 0.13 (Extended Data Fig. 5b–e). In summary, implementing a binary

https://doi.org/10.1038/s43018-026-01115-4

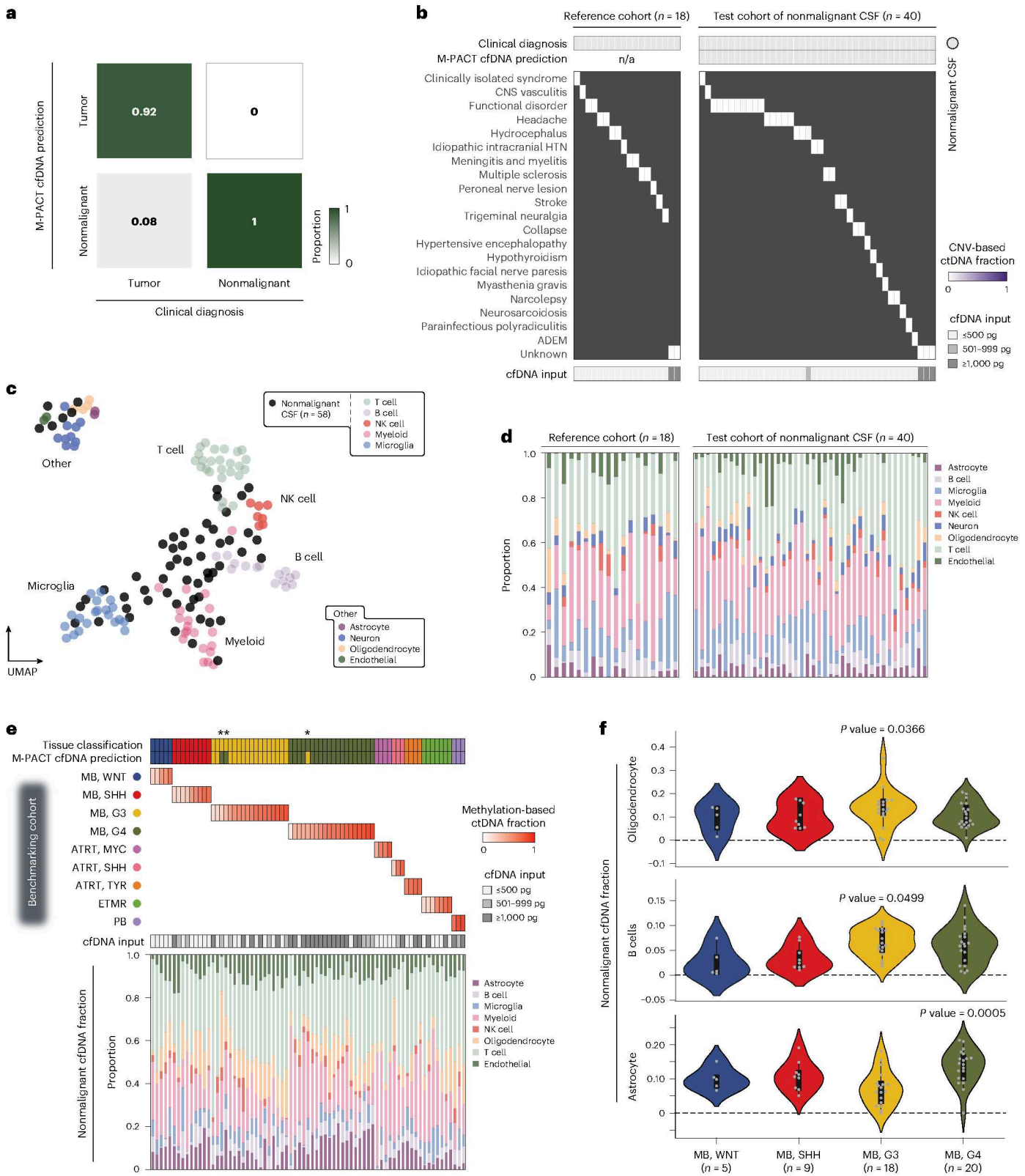


Fig. 5 | Deconvolution of CSF-derived cfDNA methylomes. a, Heat map summarizing M-PACT classification performance for the benchmarking cohort ($n = 79$ samples; Fig. 3) and nonmalignant control participants ($n = 40$ samples). **b**, Heat map summarizing M-PACT classification performance for nonmalignant CSF ($n = 40/40$ samples, 100% correct), annotated with CNV-based ctDNA fraction. HTN, hypertension; ADEM, acute disseminated encephalomyelitis. **c**, UMAP projection of nonmalignant cfDNA profiles onto DNA methylation-based reference set of nine nonmalignant cell classes. **d**, Stacked bar chart showing relative proportion of reference cell classes after deconvolution of nonmalignant CSF samples. **e**, Heat map depicting methylation-based ctDNA fraction derived from deconvolution of correctly classified CSF samples in the benchmarking cohort ($n = 73$ samples; Fig. 3). Asterisks indicate correctly classified samples with subgroup switching along the known MB, G3/4 continuum. Bottom, stacked bar plot indicating the proportion of nonmalignant reference cell classes after deconvolution. **f**, Violin plots showing the relative proportion of different nonmalignant cell classes across medulloblastoma subgroups (Kruskal–Wallis test). Violin plots extend to the minimum and maximum range of the data. Abbreviations: NK cell, natural killer cell.

classification strategy complements M-PACT for highly sensitive disease detection in samples with exceedingly low ctDNA fractions.

Deconvoluting tumor and nonmalignant cfDNA methylomes

Beyond tumor samples, we generated EM-seq profiles for a unique cohort of CSF samples collected from 58 donors without diagnosed malignancy (nonmalignant CSF; $n = 10$ pediatric, $n = 48$ adult). A subset of these samples was used as a reference for M-PACT development ($n = 18$) (Fig. 1a). M-PACT correctly classified all remaining 40 samples (40/40, 100%) as nonmalignant CSF (Fig. 5a,b). Nonmalignant cfDNA methylomes clustered with various reference cell classes, including lymphoid, myeloid and neural lineages (Fig. 5c). Implementation of a cellular deconvolution method using serial β regressions on nonmalignant profiles provided an estimate of the relative proportions of the nine nonmalignant reference cell classes (Fig. 5d). The same deconvolution approach was applied to correctly classified ctDNA-positive CSF samples from the benchmarking cohort (Fig. 3a–e), estimating methylation-based cellular fraction (Fig. 5e). Investigating the inferred prevalence and distribution of immune cell types between CSF compartments of consensus MB subgroups highlighted putative differences in proportions of oligodendrocytes, B cells and astrocytes (Fig. 5f).

In summary, the sensitive EM-seq/M-PACT pipeline allows for DNA methylation-based cell type deconvolution in low-input cfDNA samples from both nonmalignant controls and participants with CNS tumors.

Clinical utility of the M-PACT pipeline in real-world CSF liquid biopsies

To evaluate the translational potential of EM-seq/M-PACT, we integrated molecular cfDNA results with disease trajectories from the pediatric population with CNS tumors. First, we assessed a series of five CSF samples that were collected intraoperatively from independent donors. Genome-wide CNVs and M-PACT classifications from CSF were concordant with tumor tissue (Fig. 6a and Extended Data Fig. 6a,b). The calculated ctDNA fraction using M-PACT ranged from 57–88% ($n = 5$) in intraoperative liquid biopsies. These results indicate the utility of M-PACT at the diagnostic time point, which is particularly relevant for inoperable cases or cases yielding insufficient tumor tissue needed for molecular assessment.

In some pediatric CNS tumors, genomes lack discernable CNVs, rendering them incompatible with first-generation, CNV-dependent liquid biopsy assays. To evaluate the potential of epigenome-based ctDNA detection in such cases, we applied EM-seq/M-PACT to a series of CSF samples collected from participants with balanced tumor genomes ($n = 5$ donors). Remarkably, M-PACT accurately classified CNV-negative cfDNA profiles derived from participants diagnosed with MB, SHH ($n = 2$), MB, G4 ($n = 1$) and ATRT, SHH ($n = 2$), with high probability (range: 0.85–0.98). Methylation-based ctDNA fractions ranged from 3% to 68% (Fig. 6b and Extended Data Fig. 6c). In addition to the postoperative LP, longitudinal CSF samples, also collected through LP, were available for the MB, G4 participant (Fig. 6c). M-PACT classifications mirrored the participant's disease course, highlighting the utility of this method in the setting of balanced tumor genomes.

After treatment, discriminating between recurrence and secondary malignancies remains challenging on the basis of clinical and imaging features alone and surgical biopsy or resection is not routinely performed. As such, there is great demand for noninvasive, molecularly driven assays that can inform diagnosis and clinical management. To evaluate EM-seq/M-PACT in this context, we analyzed CSF collected from participants with uncertain recurrence versus secondary CNS tumor diagnoses. Participant EM268, diagnosed with a primary MB, WNT detectable in postoperative CSF collected through LP, presented with radiographic findings suggestive of disease recurrence in the fifth year of follow-up. The temporally matched EM-seq profile obtained through ventricular access demonstrated hallmark genetic alterations of a radiation-induced glioma, including *PDGFRA* amplification and focal *CDKN2A/B* deletion[43,44], and was classified as a glioblastoma (GBM) by M-PACT (Fig. 7a). Participant EM269, diagnosed with a balanced genome MB, SHH, presented with radiographic findings suggestive of either recurrence or secondary malignancy (Fig. 7b). EM-seq/M-PACT confirmed several high-level amplifications and focal deletions, including *CDKN2A/B*, as well as high-confidence classification as a GBM from intraoperative CSF.

Subgroup switching between primary and relapse MB has been reported in a rare subset of MB, G3/4 (ref. 36). Among our series of participants with MB with longitudinal CSF, participant EM229 exhibited subgroup switching along the MB, G3/4 biological continuum[37–40] in the off-therapy LP (Fig. 7c,d).

Given the variability of CSF sampling sites across clinical contexts, we further evaluated whether M-PACT performance and ctDNA fraction differ between cranial and lumbar collections. Highly comparable M-PACT probability scores (median cranial: 0.96; lumbar: 0.94) were observed across the two anatomical locations, where cranial CSF ($n = 44$) exhibited a median ctDNA fraction of 0.42 compared to 0.28 in lumbar CSF ($n = 114$; two-sided Mann–Whitney U-test; $P = 0.212$) (Extended Data Fig. 6d,e).

Collectively, we showcase relevant clinical applications of M-PACT for diagnosis, disease monitoring and longitudinal tumor characterization.

Discussion

Although prior studies have demonstrated the feasibility of ctDNA detection from CSF using heterogeneous molecular assays, current limitations in the field arise from dependence on genetic alterations, as well as high-tumor-burden and high-cfDNA inputs, faltering in the setting of low-sample inputs often seen in pediatric neurooncology[19–29]. To address these limitations, we developed M-PACT, a tumor-agnostic DNN that accurately classifies low-input, low-tumor-burden cfDNA methylomes generated using EM-seq. Application of M-PACT to liquid biopsies collected from a broad range of pediatric CNS tumor entities resulted in concordant classifications with gold-standard tissue diagnostics. These technical advances represent an essential step toward routine adoption of liquid biopsies into clinical practice.

Current DNA methylation-based CNS tumor classifiers (for example, MNP and MLPnet) have been predominantly evaluated on tumor tissue, where DNA input and tumor fraction are rarely limiting[4,5,45]. Head-to-head comparisons of M-PACT versus MNP and MLPnet on tissue-derived DNA methylation array profiles achieved high classification concordance. Furthermore, application of M-PACT to Nanopore sequencing, WGBS and DNA methylation array datasets generated from CSF-derived cfDNA yielded high classification accuracy across platforms, underscoring the versatility of M-PACT.

Studies aiming to classify CNS tumors from cfDNA methylomes have shown suboptimal classification accuracy and required nanogram-level cfDNA inputs[23,30–32]. In contrast, our workflow achieves highly accurate CNS tumor classification from subnanogram cfDNA inputs, an order-of-magnitude lower than recent studies[23,46], typical of what is encountered in the routine clinical setting[14,16,29]. Additionally, our workflow consistently detects CNVs with high concordance to lcWGS[20,22,24,27,29,35,47]. In cases of balanced tumor genomes, M-PACT reliably classified the corresponding cfDNA methylomes, overcoming a major limitation of CNV-based approaches. This advancement broadens the scope of tumors evaluable by liquid biopsy, independent of prerequisite genetic alterations. One notable limitation of our EM-seq/M-PACT workflow concerns the inability to detect gene-level mutations. Future assay and computational advances may overcome this incompatibility that is currently achieved by targeted sequencing approaches.

In commonly diagnosed adult solid tumors, sequencing-based assays applied to plasma liquid biopsies have transformed standard of care, detecting ctDNA fractions as low as 0.001–1% (refs. 48,49). However, in CNS tumors, ctDNA detection from plasma is notoriously inferior to CSF; thus, CSF has emerged as the most promising compartment for disease monitoring[22,29,50,51]. Unlike plasma, total cfDNA yields in CSF are

https://doi.org/10.1038/s43018-026-01115-4

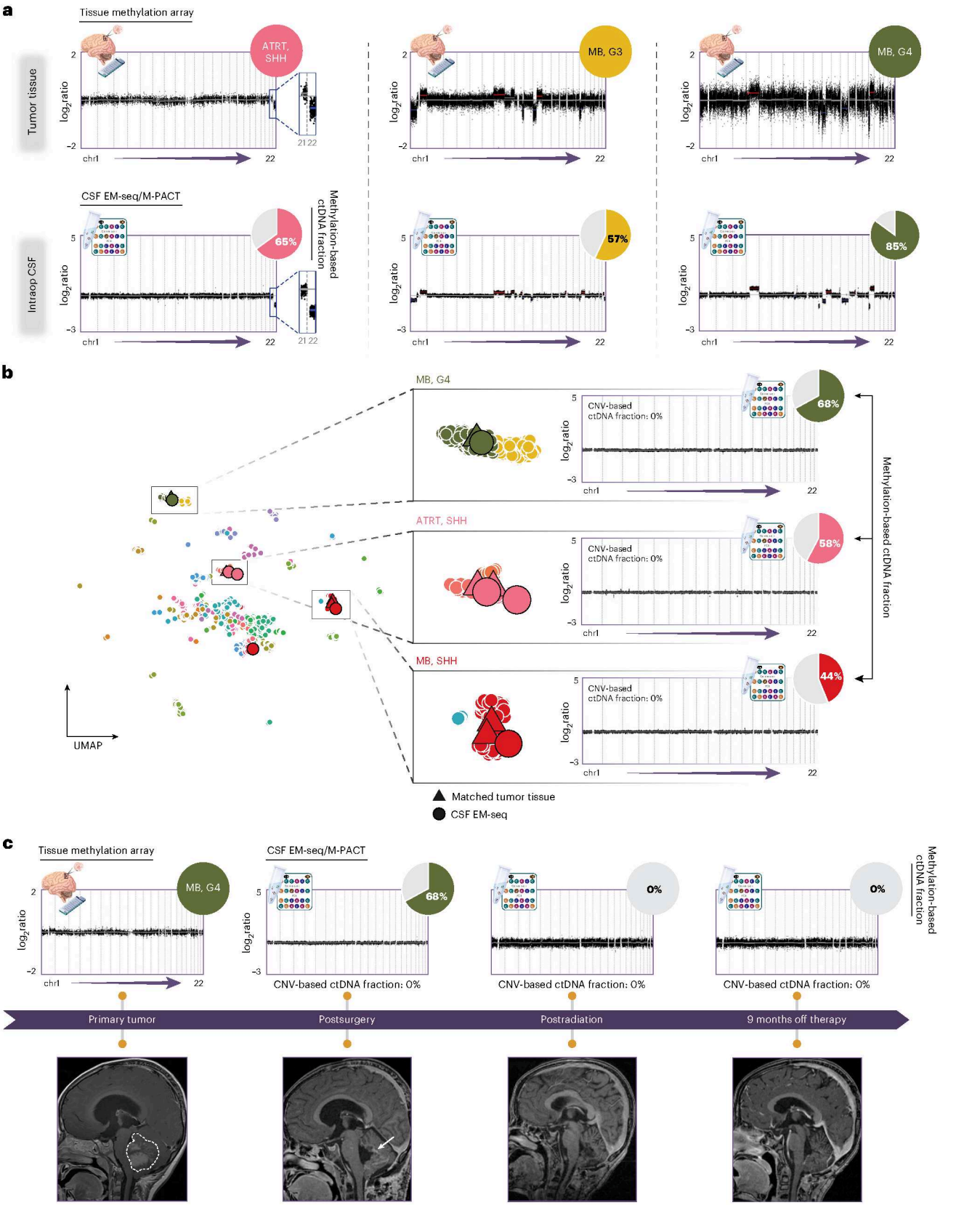


Fig. 6 | Diagnosis and monitoring of pediatric CNS tumors using EM-seq/M-PACT. a, Case vignettes showing accurate diagnosis of embryonal tumors from intraoperative CSF (ATRT, SHH (EM105), MB, G3 (EM076) and MB, G4 (EM095)). Pie charts show cfDNA deconvolution of methylation-based ctDNA fraction (colored segment) and nonmalignant fraction (gray segment). **b**, UMAP projection of cfDNA profiles collected in cases with balanced tumor genomes onto tumor methylation array reference sets. EM-seq-derived CNV profiles and pie charts indicating methylation-based ctDNA fraction are displayed (MB, G4 (EM243), ATRT, SHH (EM214) and MB, SHH (EM275)). **c**, Longitudinal disease monitoring of participant EM243 diagnosed with a copy-number-neutral MB, G4 using EM-seq/M-PACT.

https://doi.org/10.1038/s43018-026-01115-4

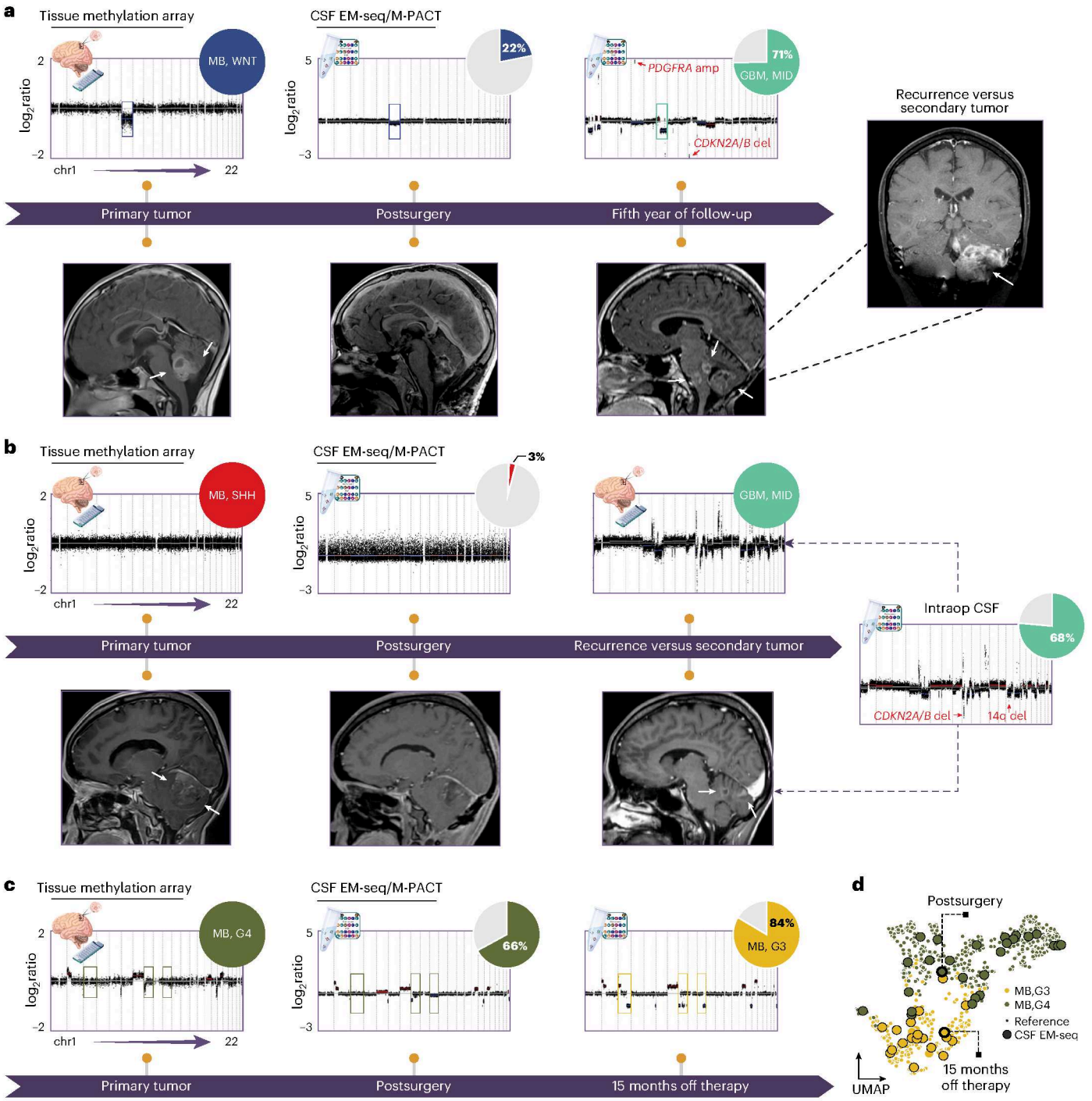


Fig. 7 | Detection of secondary malignancies and tumor evolution in serial CSF. **a**, Detection of a secondary glioma from CSF collected in the fifth year of follow-up after multimodal therapy for a primary MB, WNT (EM268). **b**, Detection of a secondary glioma from CSF collected in the second year of follow-up after multimodal therapy for a primary MB, SHH (EM269). **c**, Tumor evolution of MB, G4 (primary tumor methylation array and postsurgery CSF) into MB, G3 (15 months off therapy) (EM229). **d**, UMAP highlighting MB subgroup switching across postsurgery CSF and 15 months off therapy (EM229). Abbreviations: amp, amplification; del, deletion.

typically orders-of-magnitude lower, greatly reducing the sensitivity of ctDNA detection. Most sequencing-based CSF liquid biopsy studies reported to date have set conservative thresholds of ~5–10% tumor burden[20,24,52]. In the current study, similar thresholds were achieved using either CNV-based or methylation-based methods. Implementation of a binary classification algorithm that makes use of participant-matched tumor methylation array profiles rescued CSF samples below the M-PACT detection limit. Future advancements in sample preparation, sequencing chemistry and bioinformatics will be required to enhance sensitivity of ctDNA detection in neurooncology. Despite these current limitations, CSF liquid biopsies have begun to demonstrate clinical utility for disease surveillance in the pediatric population with brain tumors[20,21,29]. With the optimization of technical and analytical advances such as EM-seq/M-PACT, this trajectory will undoubtedly continue.

In the diagnostic setting, M-PACT accurately classified intraoperative CSF-derived cfDNA that was concordant with reference

tissue-based DNA methylation array classification. Although the sample numbers studied herein are relatively modest, our findings support the use of this method in the preoperative setting, where CSF-based diagnosis may enable presurgical medical intervention, identification of actionable targets and more informed surgical approaches. For participants where surgical resection is not feasible, CSF liquid biopsies offer a promising alternative for accurate diagnosis.

In the monitoring setting, EM-seq/M-PACT reliably detected CNVs and shifts in epigenetic landscapes from a single, low-input cfDNA specimen. Harnessing these strengths, this approach holds clinical potential to characterize evolving, spatially distinct clones and surveil for disease recurrence. In addition, there is an ongoing debate on the optimal sampling location for CSF liquid biopsies. Collection sites for CSF sampling vary by tumor subtype and reflect established divergences in clinical management (for example, preoperative ventricular sampling for pineoblastoma versus LP for staging medulloblastoma). The application of EM-seq/M-PACT in controlled studies with matched cranial versus distant sampling could shed light on the most suitable anatomical location for CSF collection[53–55].

Current imaging technologies often lack the sensitivity and granularity required to distinguish tumor recurrence from secondary malignancy, which may present similarly in the radiographic setting. While outcomes for secondary gliomas remain poor[43,44,56,57], molecular diagnosis from CSF liquid biopsies could provide a critical window for timely adjustment of therapeutic strategies and trial enrollment, with the potential to improve outcomes.

Deconvolution approaches have characterized the cellular origins of malignant and nonmalignant cfDNA in plasma liquid biopsies, including the composition of immune cell landscapes[58–60]. Similar approaches have yet to be conducted in the CSF liquid biopsy setting. In the current study, M-PACT estimated the relative proportions of cfDNA derived from nonmalignant cell types. Future applications investigating ctDNA dynamics and the immunological response to both standard-of-care and experimental therapies, including cellular therapies, represent an untapped opportunity to decipher the longitudinal dynamics of treatment response and resistance mechanisms in neurooncology. With ongoing investigations into how treatment can influence the immune response in the CNS compartment, cellular deconvolution by M-PACT holds promise in deciphering immunosuppressive versus antitumor microenvironment states. This offers a valuable tool for informing immunotherapy strategies in the clinical setting.

In summary, we established a highly sensitive workflow that overcomes key challenges in pediatric neurooncology CSF liquid biopsies. Establishing a CNS tumor classifier optimized for low-input, low-tumor-burden methylomes represents a foundational step toward translating DNA methylation-based tumor classification from tissues to liquid biopsies. The methodology yields accurate classification across diverse CNS tumor entities, with the added value of distinguishing primary from secondary malignancies. To assess the full breadth of clinical utility for this CSF liquid biopsy workflow, validation in prospective clinical trials is warranted.

Methods

Participant samples and data

Participant material and clinical data including matched MRI results of the liquid biopsy cohort (n = 268 CSF samples, n = 214 donors) were obtained from multiple centers under protocols approved by their respective institutional review boards (IRBs). These centers include St. Jude Children's Hospital (IRB: Pro00008912, Pro00004251), Hopp Children's Cancer Center (IRB: S-531/2020, S-502/2013), Medical University of Vienna (IRB: 1244/2016), Tampere University Hospital (IRB: R13050), Erasmus Medical Center (IRB: OC2004_011) and University Hospital Kiel (IRB: D569-23). Written consent was obtained from each participant or their legal representative. CSF samples were collected during routine clinical procedures at diagnosis, during therapy and/or at follow-up. Specifically, CSF was collected in sterile tubes (1) preoperatively through ventricular access (reservoir, external ventricular drain or endoscopic third ventriculostomy); (2) intraoperatively at varying locations depending on surgical approaches; (3) postoperatively during staging through LP or through ventricular access; or (4) at follow-up through LP or ventricular access.

Study design

The main study cohort included pediatric participants diagnosed with a primary CNS tumor, each having at least one banked CSF sample available for molecular analysis (n = 210 CSF samples, n = 156 participants). Matched tumor tissues for these participants analyzed on Illumina methylation array served as the reference standard for validating and interpreting CSF results. CSF samples from individuals without malignancies (n = 58) were included as a nononcological control panel. Clinical data were extracted independently of liquid biopsy results to ensure unbiased analyses.

Methylation array analysis

Genomic DNA extracted from frozen or formalin-fixed paraffin-embedded tumor tissues underwent bisulfite conversion using the EZ DNA methylation kit (Zymo). Samples were analyzed on Illumina Infinium 450k and 850k methylation arrays (Illumina) according to the manufacturer's instructions. Gold-standard CNS tumor classifications were obtained using a DNA methylation-based web platform (https://app.epignostix.com/; versions 11b4, 11b6 and 12b4)[4]. CNVs and methylation β values were calculated using the MethylVerse Python package (version 1.0) for version hg38 of the human genome.

cfDNA isolation from CSF

Participant CSF samples stored at −80 °C were thawed on ice before cfDNA extraction using NucleoSnap cfDNA (Macherey-Nagel), Quick-cfDNA/RNA serum and plasma (Zymo) or Maxwell RSC ccfDNA extraction (Promega) kits according to the manufacturer's instructions. Briefly, CSF was centrifuged to pellet and remove any debris, with subsequent protein digestion and lysis of CSF supernatant fractions. cfDNA was captured on silica membranes, washed, dried and eluted in elution buffer or nuclease-free water. Quality control was performed using one of the following pipelines: cfDNA quantification by qPCR using established primers for *ALU* sequences (*ALU*115-F′: CCTGAGGTCAGGAGTTCGAG, *ALU*115-R′: CCCGAGTAGCTGGGATTACA) followed by fragment length assessment on the high-sensitivity D1000 TapeStation platform (Agilent Technologies)[20,35] or total DNA quantification on the Qubit 3.0 fluorometer using the double-stranded DNA high-sensitivity assay kit (Thermo Fisher) followed by fragment length assessment on the high-sensitivity DNA Bioanalyzer platform (Agilent Technologies)[47].

EM-seq

NEBNext EM-seq technology (New England Biolabs)[33] was used for enzymatic conversion of cfDNA CpG sites and library preparation for methylation sequencing. CpG-methylated pUC19 (0.1 ng μl^{-1}) and unmethylated lambda DNA (2 ng μl^{-1}) controls were sheared to 150–200 bp using the Covaris S2 instrument (Covaris) and spiked into libraries as control for conversion performance. Participant-derived cfDNA was not subjected to manual fragmentation. The EM-seq protocol was optimized for cfDNA inputs ranging from 0.005 to 50 ng. After optimization, libraries were prepared with a median of 0.5 ng of cfDNA, if available, or up to 44 µl of cfDNA if the concentration was below the limit of detection. There were no technical dropouts. In brief, end repair, dA tailing and adaptor ligation were performed using NEBNext DNA Ultra II reagents. Ligated samples were next subjected to TET2-catalyzed oxidation to preserve 5-methylcytosine and 5-hydroxymethylcytosine modifications. Upon DNA denaturation with

formamide, nonmethylated cytosines were deaminated to uracil by the APOBEC3A enzyme. Libraries were subsequently amplified with unique NEBNext dual index primers for 12–18 PCR cycles depending on cfDNA input and eluted in 20 µl of elution buffer. Fragment distributions and library concentrations were quantified using the high-sensitivity D5000 kit on TapeStation (Agilent Technologies) or high-sensitivity DNA kit on Bioanalyzer (Agilent Technologies). Libraries pooled in equimolar amounts at 10 nM with a targeted insert size of 50–400 bp were subjected to 150-bp paired-end sequencing on the NovaSeq 6000 (Illumina) at ~4× coverage. For a subset of samples, EM-seq libraries were analyzed on the Illumina methylation array, as described above, by directly inputting them at the DNA restoration step, followed by classification using MNP (version 12.5) or M-PACT.

Serial titration of ctDNA

To evaluate the sensitivity of M-PACT ctDNA classification, an MB, SHH CSF sample with 94% CNV-based tumor fraction was mixed with a pool of nonmalignant, human-derived cfDNA at varying ctDNA concentrations: 94%, 50%, 25%, 10%, 5%, 1% and 0.03%. Subsequent EM-seq library preparation and Illumina sequencing were performed as described above.

Nanopore sequencing

cfDNA samples (input range: 0.5–2.7 ng) were prepared using SQK-NBD114.24 (Oxford Nanopore Technologies) and subsequently sequenced on a PromethION 2 Solo, according to the manufacturer's instructions. Read length distributions and read counts per barcode were captured in real time with MiniKNOW. Samples were demultiplexed using the highly accurate model (hac) as implemented in Dorado (version 0.9.1).

Analysis of EM-seq genome-wide cfDNA methylome data

Genome-wide EM-seq methylomes were mapped to the human reference genome (hg38) with the addition of methylated pUC19 and unmethylated lambda DNA sequences using Biscuit (version 1.4.0)[61]. The methylation status of CpGs was determined by methyldackel (version 0.5.1). Resulting β values at CpGs located on autosomes with 1× coverage were used for all downstream analyses conducted with the methyl (version 1.0) Python package. CNVs and CNV-based tumor fraction estimates were determined using the cfdna (version 2.0.0) Python package as previously published[20,35].

Analysis of Nanopore cfDNA methylome data

Nanopore sequencing reads were mapped using the high-accuracy model of Dorado version 1 and CpG methylation determined by modkit version 1. Hydroxymethylated CpG sites were excluded from analysis; the ratio of methylated reads and the total coverage (that is, β values) were used for downstream analysis. CNV analyses were conducted using the cfdna CNV caller version 2.4.5 with the Nanopore parameter set to true.

Methylation imputation

An EPIC methylation array reference cohort of 914 samples was constructed from methylation profiles associated with CNS tumors (GSE215240) and immune cells (GSE167998)[5,62]. A CpG correlation network based on the Euclidean distance between CpG β values across all samples was developed using the scikit-network (version 0.33.1) Python package. The resulting network served as the foundation for a network diffusion regression model to predict β values of missing CpGs from sparse methylomes. Imputation was validated on a cohort of 950 EPIC arrays (GSE276299)[42].

Neural network-based ctDNA classification

A reference methylation cohort of 5,014 methylation profiles encompassing 4,996 methylation arrays from CNS tumors (GSE90496, GSE109379, GSE215240, GSE183798 and GSE175543), control blood (GSE89278) and immune cells (GSE167998), in addition to 18 nonmalignant CSF (current study), was assembled[4,5,62–65]. Methylation profiles were divided into 97 categories for classification. A series of multilayer perceptron (MLP) DNNs were trained and integrated into an ensemble classifier to predict these categories from sparse, low-tumor-fraction methylomes. HM450 methylation array probes were ordered by importance scores through implementation of a random forest classifier. These scores were used to partition the 450,000 CpGs into their most informative features, resulting in three sets of 100,000 CpGs with predictive value and removal of nonpredictive CpGs.

The first MLP classifier in the ensemble model was trained directly on the dense methylation cohort using scikit-learn (version 1.5.0). Specifically, three neural networks were trained as an ensemble model on separate folds of CpGs, each using 100,000 CpGs with the last layer of each model concatenated yielding a unified prediction.

For the second and third MLP classifiers in the ensemble model, the reference methylation cohort was used to create a simulated cohort for training. Each simulation was created by first generating a methylation profile from a randomly weighted average of all samples belonging to the same class (for example, ETMR). A nonmalignant methylation profile was generated using a randomly weighted average of all samples categorized as blood, control CSF, immune cells, plasma, control reactive tissue and control inflammatory tissue. Averaging nonmalignant and tumor profiles at different proportions of weighted averages, we simulated samples harboring a range of tumor purities. Lastly, CpGs were randomly removed from simulated samples to mimic sparse methylomes. The second and third classifiers were each trained on >3 million total simulations comprised of 36,000 per class. Similar to the first classifier, these two classifiers comprised three individual neural networks trained as an ensemble model on distinct folds of 100,000 CpGs, with the last layer of each model concatenated yielding a unified prediction. This approach reduces the likelihood of confidently called false positives because of the concordance required between models.

Layering the three models together, an ensemble classifier was created, with class prediction reported by the model with the highest prediction probability score.

Methylation deconvolution and regression

A reference WGBS cohort of nonmalignant background signatures spanning 21 different cell types merged into nine cell classes (GSE186458) and nonmalignant CSF (this study) were curated for methylation-based deconvolution[66]. A signature matrix was created by calculating the average methylation profile for each nonmalignant class. Additionally, a tumor methylation profile was added to the signature matrix by averaging reference methylation array profiles. The observed methylation (β_{obs}) at each CpG site is expressed as a weighted (ω) sum of the reference (β_{ref}) and tumor components (β_{tumor}).

$$\beta_{obs} = \omega \times \beta_{ref} + (1 - \omega) \times \beta_{tumor}$$

A Huber regression was implemented on the 4,000 most variable CpGs after logit transformation of the signature matrix and those captured in the sample. Regression coefficients were scaled to sum to 1 and used as the predicted cell type proportion. The regression was conducted serially by removing all signature cell types with a predicted proportion of 0 and rerun iteratively until all cell types contained nonzero proportion predictions. The same signature matrix was used to regress out nonmalignant background. A simple linear regression was implemented, with residuals subtracted from the sample methylation profile. All β values below 0 were set to 0 and those above 1 were set to 1.

CNV comparisons

lcWGS data were generated as previously described[20,29,35,67] and analyzed using the cfdna (version 2.0.0) Python package[20,35]. Comparisons between sample-matched CNV profiles for lcWGS and EM-seq samples

were calculated using a scaled Manhattan similarity on the basis of the total number of common 100-kb bins between samples. The scaled Manhattan distances were determined by dividing the observed Manhattan distance by the total number of features (for example, common 100-kb bins) and then converted into a similarity metric as 1 − scaled distance.

Binary classification algorithm

Creation of a methylation-based binary classification algorithm (for example, tumor versus not tumor) was implemented as a tumor-informed approach to determine whether an observed CSF-derived methylome is more similar to the primary tumor or non-malignant control populations (for example, control CSF and control brain tissue). First, a Pearson correlation was calculated using the 300,000 CpGs and weighted by importance scores derived from the M-PACT classifier, yielding an observed similarity between tumor and CSF-derived methylome. Next, the same correlation metric was calculated between the CSF-derived methylome and all control CSF samples in the cohort ($n = 58$), resulting in a background distribution. Comparison of the observed similarity between the CSF-derived methylome and the primary tumor with the background distribution resulted in a z score. The same comparison was conducted with control brain methylomes ($n = 22$) and resulting z scores were combined using Stouffer's method. Lastly, combined z scores greater than or equal to 1.2 (80th percentile) were classified as tumor, while those under 1.2 were classified as not tumor.

External validation datasets

Raw fast5 files were obtained from a previously published[23] CSF-derived low-coverage Nanopore sequencing cohort for validation of the M-PACT classifier. The cohort was reprocessed using Dorado (version 0.9.1) and modkit (version 1.0.0) for determination of CpG methylation status. CNVs were called using the cfdna CNV caller (version 2.4.5) to determine ctDNA positivity and compared to previously published NanoDx classifications. Additionally, a CSF-derived WGBS cohort[31] was compiled (GSE178169 and GSE178666) from raw sequencing files and processed identical to our EM-seq cohort.

Statistics and reproducibility

Two-sided Fisher's exact tests and Mann–Whitney U-tests were performed in an unpaired manner using the scipy package in Python for comparison of categorical and continuous variables, respectively. When appropriate, P values were corrected for multiple comparisons using the Benjamini–Hochberg correction for false discovery rate as implemented in the statsmodels Python package. Data distributions were assumed to be normal but were not formally tested. For all analyses, P values below 0.05 were considered statistically significant. No statistical methods were conducted before sample collection. Cohort selection was determined by sample availability and not randomized. No data were excluded from analyses. Study investigators were blinded to clinical annotations.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

Microarray and sequencing data that support the findings of this study were deposited to the Gene Expression Omnibus (GSE292312) and European Genome–Phenome Archive (EGAS50000001365), respectively. Sequencing data (EGAS50000001365) are under controlled access because of participant privacy and governed by the EGAC00001002368 Data Access Committee with the corresponding data usage agreement (https://ega-archive.org/datasets/EGAD00001008277). Data access agreements will be reviewed and responded to within 30 days of application. Previously published lcWGS data that were reanalyzed here are available under accession number EGAS00001005592. Source data are provided with this paper.

Code availability

All algorithms are available from GitHub (https://github.com/kyles-smith) and have been packaged into a Python module distributed through PyPI. Additionally, the classifier has been made accessible through an online portal (https://mpact.stjude.org) and is available for download (https://doi.org/10.5281/zenodo.15091136)[68].

References

1. Siegel, R. L., Miller, K. D., Fuchs, H. E. & Jemal, A. Cancer statistics, 2022. *CA Cancer J. Clin.* **72**, 7–33 (2022).
2. Louis, D. N. et al. The 2021 WHO classification of tumors of the central nervous system: a summary. *Neuro Oncol.* **23**, 1231–1251 (2021).
3. Pfister, S. M. et al. A summary of the inaugural WHO classification of pediatric tumors: transitioning from the optical into the molecular era. *Cancer Discov.* **12**, 331–355 (2022).
4. Capper, D. et al. DNA methylation-based classification of central nervous system tumours. *Nature* **555**, 469–474 (2018).
5. Sturm, D. et al. Multiomic neuropathology improves diagnostic accuracy in pediatric neuro-oncology. *Nat. Med.* **29**, 917–926 (2023).
6. Capper, D. et al. Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. *Acta Neuropathol.* **136**, 181–210 (2018).
7. Ignatiadis, M., Sledge, G. W. & Jeffrey, S. S. Liquid biopsy enters the clinic—implementation issues and future challenges. *Nat. Rev. Clin. Oncol.* **18**, 297–312 (2021).
8. Alix-Panabieres, C. & Pantel, K. Liquid biopsy: from discovery to clinical application. *Cancer Discov.* **11**, 858–873 (2021).
9. Jongen-Lavrencic, M. et al. Molecular minimal residual disease in acute myeloid leukemia. *N. Engl. J. Med.* **378**, 1189–1199 (2018).
10. Anagnostou, V. et al. ctDNA response after pembrolizumab in non-small cell lung cancer: phase 2 adaptive trial results. *Nat. Med.* **29**, 2559–2569 (2023).
11. Tie, J. et al. Circulating tumor DNA analysis guiding adjuvant therapy in stage II colon cancer. *N. Engl. J. Med.* **386**, 2261–2272 (2022).
12. Bruggemann, M. et al. Standardized next-generation sequencing of immunoglobulin and T-cell receptor gene recombinations for MRD marker identification in acute lymphoblastic leukaemia; a EuroClonality-NGS validation study. *Leukemia* **33**, 2241–2253 (2019).
13. Bettegowda, C. et al. Detection of circulating tumor DNA in early- and late-stage human malignancies. *Sci. Transl. Med.* **6**, 224ra24 (2014).
14. Liu, A. P., Northcott, P. A., Robinson, G. W. & Gajjar, A. Circulating tumor DNA profiling for childhood brain tumors: technical challenges and evidence for utility. *Lab. Invest.* **102**, 134–142 (2022).
15. Meng, Y. et al. MR-guided focused ultrasound liquid biopsy enriches circulating biomarkers in patients with brain tumors. *Neuro Oncol.* **23**, 1789–1797 (2021).
16. Miller, A. M. & Karajannis, M. A. Current role and future potential of CSF ctDNA for the diagnosis and clinical management of pediatric central nervous system tumors. *J. Natl Compr. Canc. Netw.* **20**, 1363–1369 (2022).
17. Soffietti, R. et al. Liquid biopsy in gliomas: a RANO review and proposals for clinical applications. *Neuro Oncol.* **24**, 855–871 (2022).
18. Berzero, G., Pieri, V., Mortini, P., Filippi, M. & Finocchiaro, G. The coming of age of liquid biopsy in neuro-oncology. *Brain* **146**, 4015–4024 (2023).

19. Cantor, E. et al. Serial H3K27M cell-free tumor DNA (cf-tDNA) tracking predicts ONC201 treatment response and progression in diffuse midline glioma. *Neuro Oncol.* **24**, 1366–1374 (2022).
20. Liu, A. P. Y. et al. Serial assessment of measurable residual disease in medulloblastoma liquid biopsies. *Cancer Cell* **39**, 1519–1530.e4 (2021).
21. Miller, A. M. et al. Next-generation sequencing of cerebrospinal fluid for clinical molecular diagnostics in pediatric, adolescent and young adult brain tumor patients. *Neuro Oncol.* **24**, 1763–1772 (2022).
22. Pages, M. et al. Liquid biopsy detection of genomic alterations in pediatric brain tumors from cell-free DNA in peripheral blood, CSF, and urine. *Neuro Oncol.* **24**, 1352–1363 (2022).
23. Afflerbach, A. K. et al. Classification of brain tumors by Nanopore sequencing of cell-free DNA from cerebrospinal fluid. *Clin. Chem.* **70**, 250–260 (2024).
24. O'Halloran, K. et al. Low-pass whole-genome and targeted sequencing of cell-free DNA from cerebrospinal fluid in pediatric patients with central nervous system tumors. *Neurooncol. Adv.* **5**, vdad077 (2023).
25. Madlener, S. et al. Clinical applicability of miR517a detection in liquid biopsies of ETMR patients. *Acta Neuropathol.* **145**, 843–846 (2023).
26. Stepien, N. et al. Proof-of-concept for liquid biopsy disease monitoring of *MYC*-amplified group 3 medulloblastoma by droplet digital PCR. *Cancers* **15**, 2525 (2023).
27. O'Halloran, K. et al. Low-pass whole genome sequencing of cell-free DNA from cerebrospinal fluid: a focus on pediatric central nervous system tumors. *Clin. Chem.* **71**, 87–96 (2025).
28. O'Halloran, K. et al. Targeted detection of sequence variants in cell-free DNA from cerebrospinal fluid in pediatric central nervous system tumors. *Front. Oncol.* **14**, 1513073 (2024).
29. Fischer, T. T. et al. Ultra-low-input cell-free DNA sequencing for tumor detection and characterization in a real-world pediatric brain tumor cohort. *Acta Neuropathol. Commun.* **13**, 134 (2025).
30. Cornelli, L. et al. Diagnosis of pediatric central nervous system tumors using methylation profiling of cfDNA from cerebrospinal fluid. *Clin. Epigenetics* **16**, 87 (2024).
31. Li, J. et al. Reliable tumor detection by whole-genome methylation sequencing of cell-free DNA in cerebrospinal fluid of pediatric medulloblastoma. *Sci. Adv.* **6**, eabb5427 (2020).
32. Zuccato, J. A. et al. Cerebrospinal fluid methylome-based liquid biopsies for accurate malignant brain neoplasm classification. *Neuro Oncol.* **25**, 1452–1460 (2023).
33. Vaisvila, R. et al. Enzymatic methyl sequencing detects DNA methylation at single-base resolution from picograms of DNA. *Genome Res.* **31**, 1280–1289 (2021).
34. Hill, A. D. et al. Single-nucleus RNA-seq dissection of choroid plexus tumor cell heterogeneity. *EMBO J.* **43**, 6766–6791 (2024).
35. Liu, A. P. Y., Smith, K. S., Kumar, R., Robinson, G. W. & Northcott, P. A. Low-coverage whole-genome sequencing of cerebrospinal-fluid-derived cell-free DNA in brain tumor patients. *STAR Protoc.* **3**, 101292 (2022).
36. Kumar, R. et al. Clinical outcomes and patient-matched molecular composition of relapsed medulloblastoma. *J. Clin. Oncol.* **39**, 807–821 (2021).
37. Smith, K. S. et al. Unified rhombic lip origins of group 3 and group 4 medulloblastoma. *Nature* **609**, 1012–1020 (2022).
38. Williamson, D. et al. Medulloblastoma group 3 and 4 tumors comprise a clinically and biologically significant expression continuum reflecting human cerebellar development. *Cell Rep.* **40**, 111162 (2022).
39. Hendrikse, L. D. et al. Failure of human rhombic lip differentiation underlies medulloblastoma formation. *Nature* **609**, 1021–1028 (2022).
40. Tao, R. et al. Arrested development: the dysfunctional life history of medulloblastoma. *Genes Dev.* **39**, 4–17 (2025).
41. Kuschel, L. P. et al. Robust methylation-based classification of brain tumours using Nanopore sequencing. *Neuropathol. Appl. Neurobiol.* **49**, e12856 (2023).
42. Tran, Q. T. et al. Comparison of DNA methylation based classification models for precision diagnostics of central nervous system tumors. *NPJ Precis. Oncol.* **8**, 218 (2024).
43. Deng, M. Y. et al. Radiation-induced gliomas represent H3-/IDH-wild type pediatric gliomas with recurrent *PDGFRA* amplification and loss of *CDKN2A/B*. *Nat. Commun.* **12**, 5530 (2021).
44. Trkova, K. et al. Clinical and molecular study of radiation-induced gliomas. *Sci. Rep.* **14**, 3118 (2024).
45. Vermeulen, C. et al. Ultra-fast deep-learned CNS tumour classification during surgery. *Nature* **622**, 842–849 (2023).
46. Brandl, B. et al. Rapid brain tumor classification from sparse epigenomic data. *Nat. Med.* **31**, 840–848 (2025).
47. Maass, K. K. et al. From sampling to sequencing: a liquid biopsy pre-analytic workflow to maximize multi-layer genomic information from a single tube. *Cancers* **13**, 3002 (2021).
48. Widman, A. J. et al. Ultrasensitive plasma-based monitoring of tumor burden using machine-learning-guided signal enrichment. *Nat. Med.* **30**, 1655–1666 (2024).
49. Dao, J. et al. Using cfDNA and ctDNA as oncologic markers: a path to clinical validation. *Int. J. Mol. Sci.* **24**, 13219 (2023).
50. Escudero, L. et al. Circulating tumour DNA from the cerebrospinal fluid allows the characterisation and monitoring of medulloblastoma. *Nat. Commun.* **11**, 5376 (2020).
51. De Mattos-Arruda, L. et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat. Commun.* **6**, 8839 (2015).
52. Adalsteinsson, V. A. et al. Scalable whole-exome sequencing of cell-free DNA reveals high concordance with metastatic tumors. *Nat. Commun.* **8**, 1324 (2017).
53. Madlener, S. et al. Detection of H3F3A K27M or BRAF V600E in liquid biopsies of brain tumor patients as diagnostic and monitoring biomarker: impact of tumor localization and sampling method. *Acta Neuropathol.* **149**, 5 (2025).
54. Orzan, F. et al. Liquid biopsy of cerebrospinal fluid enables selective profiling of glioma molecular subtypes at first clinical presentation. *Clin. Cancer Res.* **29**, 1252–1266 (2023).
55. Gajjar, A. et al. Comparison of lumbar and shunt cerebrospinal fluid specimens for cytologic detection of leptomeningeal disease in pediatric patients with brain tumors. *J. Clin. Oncol.* **17**, 1825–1828 (1999).
56. van Tilburg, C. M. et al. The pediatric precision oncology INFORM registry: clinical outcome and benefit for patients with very high-evidence targets. *Cancer Discov.* **11**, 2764–2779 (2021).
57. Berlanga, P. et al. The European MAPPYACTS trial: precision medicine program in pediatric and adolescent patients with recurrent malignancies. *Cancer Discov.* **12**, 1266–1281 (2022).
58. Conway, A. M. et al. A cfDNA methylation-based tissue-of-origin classifier for cancers of unknown primary. *Nat. Commun.* **15**, 3292 (2024).
59. Sun, T. et al. Systematic evaluation of methylation-based cell type deconvolution methods for plasma cell-free DNA. *Genome Biol.* **25**, 318 (2024).
60. Moss, J. et al. Comprehensive human cell-type methylation atlas reveals origins of circulating cell-free DNA in health and disease. *Nat. Commun.* **9**, 5068 (2018).
61. Zhou, W. et al. BISCUIT: an efficient, standards-compliant tool suite for simultaneous genetic and epigenetic inference in bulk and single-cell studies. *Nucleic Acids Res.* **52**, e32 (2024).

62. Salas, L. A. et al. Enhanced cell deconvolution of peripheral blood using DNA methylation for high-resolution immune profiling. *Nat. Commun.* **13**, 761 (2022).
63. Sievers, P. et al. Pediatric-type high-grade neuroepithelial tumors with *CIC* gene fusion share a common DNA methylation signature. *NPJ Precis. Oncol.* **7**, 30 (2023).
64. DeSisto, J. et al. Comprehensive molecular characterization of pediatric radiation-induced high-grade glioma. *Nat. Commun.* **12**, 5531 (2021).
65. van Dijk, S. J. et al. Effect of prenatal DHA supplementation on the infant epigenome: results from a randomized controlled trial. *Clin. Epigenetics* **8**, 114 (2016).
66. Loyfer, N. et al. A DNA methylation atlas of normal human cell types. *Nature* **613**, 355–364 (2023).
67. Fischer, T. T. et al. Cerebrospinal fluid liquid biopsy guides differential diagnosis of relapsed medulloblastoma versus secondary glioma: a case report of a pediatric patient enrolled on a PDGFRA inhibitor trial. *Neurooncol. Adv.* **7**, vdaf110 (2025).
68. Smith, K. MethylationAnnotations_v1. *Zenodo* https://doi.org/10.5281/zenodo.15091136 (2025).

Acknowledgements

This study was principally supported by the Verein unser_kind (J.G.), the Forschungsgesellschaft für Cerebrale Tumore (J.G.), Ein Kiwi gegen Krebs (K.W.P. and K.K.M.) and the American Lebanese Syrian Associated Charities (P.A.N.). P.A.N. is the recipient of the Emerging Leader Award (Mark Foundation for Cancer Research) and the Robert J. Arceci Innovation Award (St. Baldrick's Foundation), and acknowledges additional funding from the Brain Tumor Funders' Collaborative and the National Cancer Institute (P01CA096832, R01CA270785 and R01CA259372). Additional funding was provided by the FWF der Wissenschaftsfonds (D.S., J4353-B28), the OeNB Jubiläumsfonds (D.S., 18535), the Physician Researcher Pathway Scholarships of the Medical University of Vienna (N.S.), the City of Vienna Fund for Innovative Interdisciplinary Cancer Research (N.S., 21080), the CCP Starter Grant 2020 (N.S.), the Oncomine Clinical Research Grant 2020 (S.M.), Emil Aaltonen Foundation (J.H.), Competitive State Research Financing of the Expert Responsibility area of Tampere University Hospital (J.H.), the Finnish Ministry of Social Affairs and Health (J.H.), Tampere University Foundation Trust (J.H.), the Väre Foundation for Pediatric Cancer Research (K.N. and J.H.) and The Foundation for Pediatric Research (K.N. and J.H.). K.K.M. is a recipient of the Robert Connor Dawes Scientific Fellowship of the National Brain Tumor Charity. We express our sincerest gratitude to all participants, families, clinical care teams, operations staff and research personnel from participating institutions for their contributions. We thank the St. Jude Biorepository for storing and managing participant specimens, the St. Jude Hartwell Center for Biotechnology and the High-Throughput Sequencing unit of the Genomics and Proteomics Core Facility at the German Cancer Research Center (DKFZ) for providing next-generation sequencing services, the Microarray Core Facility (DKFZ) for providing the Illumina methylation array and related services and the Omics IT and Data Management Core Facility (DKFZ) for data management infrastructure. We also thank research nurses L. Jalonen and M. Nieminen at Tampere University Hospital for their critical contributions to sample collection and A. Patel (St. Jude), J. Freese (University Medical Center Hamburg-Eppendorf) and U. Schüller (University Medical Center Hamburg-Eppendorf) for their assistance with Nanopore studies.

Author contributions

Conceptualization, K.S.S., T.T.F., K.H., K.W.P., K.K.M. and P.A.N. Data curation, K.S.S., T.T.F., K.H., A.K., H.L., D.S., N. Stepien, S.V., S.M., C.H., S.K.D., P.R.B., M.T.S., B.A.O., J.H. and K.N. Methodology, K.S.S., T.T.F., K.W.P., K.K.M. and P.A.N. Investigation, K.S.S., T.T.F., K.H., A.K., H.L., D.S., T.S., N. Stepien, N. Schwarz, T.W. and S.M. Formal analysis, K.S.S., T.T.F., K.H. and A.K. Visualization, K.S.S., T.T.F., K.H. and A.K. Software, K.S.S. Resources, S.M., C.H., S.A.U., J.B., H.H., J.D., F.L., S.M.P., E.H., A.G., G.W.R., J.H., K.N., J.G., K.W.P., K.K.M. and P.A.N. Supervision, J.G., K.W.P., K.K.M. and P.A.N. Writing—original draft, K.S.S., T.T.F., K.H., A.K. and P.A.N. Writing—review and editing, all authors. Funding acquisition, J.G., K.W.P., K.K.M. and P.A.N.

Competing interests

The authors declare no competing interests.

Additional information

Extended data is available for this paper at https://doi.org/10.1038/s43018-026-01115-4.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s43018-026-01115-4.

Correspondence and requests for materials should be addressed to Paul A. Northcott.

Peer review information *Nature Cancer* thanks Krishna Bhat, Alexandra Miller and Benjamin Purow for their contribution to the peer review of this work.

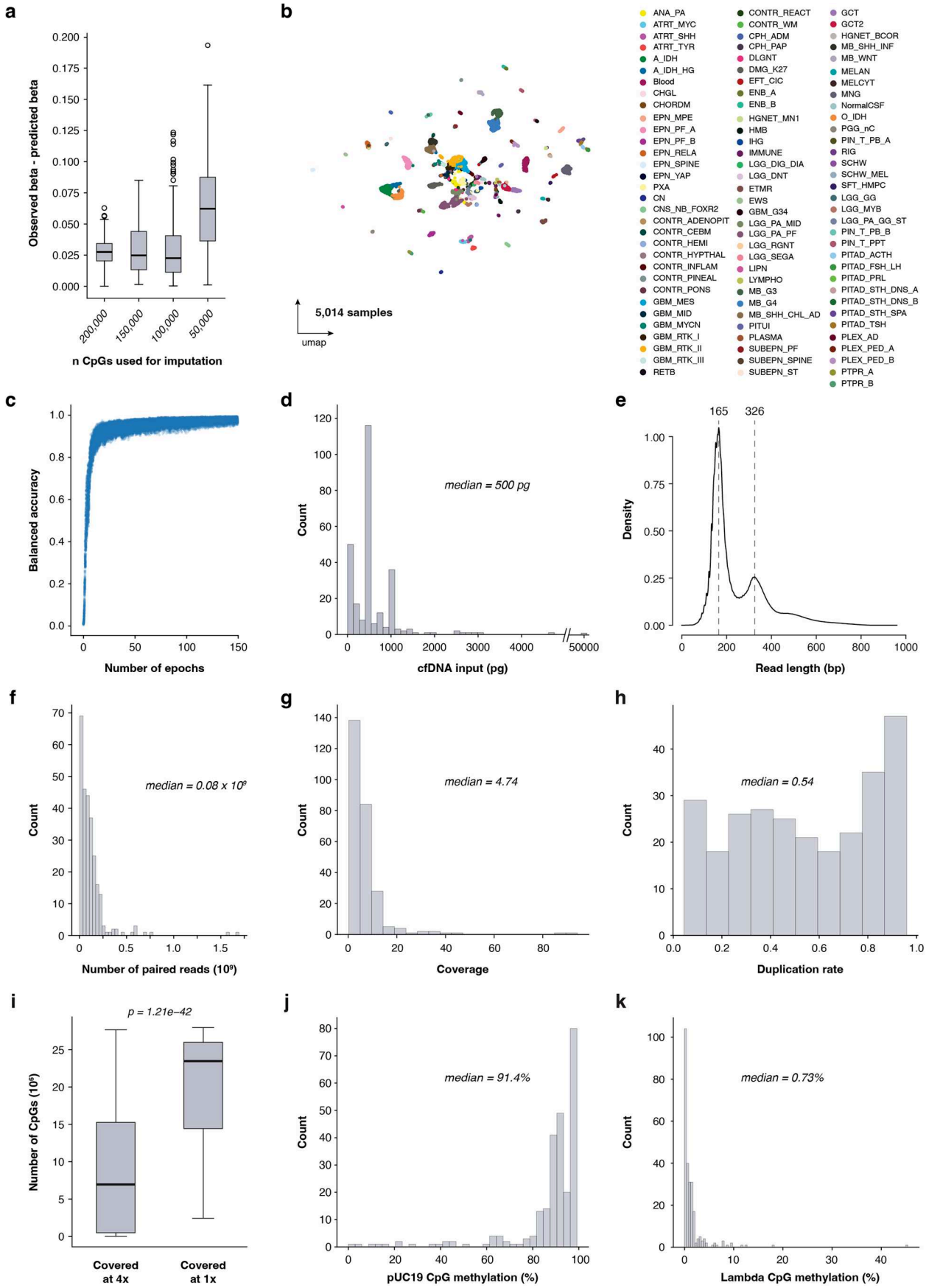
Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Kyle S. Smith[1,2,21], Tom T. Fischer[3,4,5,6,7,21], Katie Han[1,2,21], Anna Kostecka[1,2], Hong Lin[1,2], Daniel Senfter[8], Taha Soliman[1,2], Natalia Stepien[8], Stefanie Volz[3,4,5,6], Nathalie Schwarz[3,4,5,6], Tatjana Wedig[3,4,5,6], Sibylle Madlener[8], Christine Haberler[9], Sandeep K. Dhanda[10], Santhosh A. Upadhyaya[10], Patrick R. Blackburn[11], Maria T. Schmook[12], Judith de Bont[13,14], Hannu Haapasalo[15], Justina Dargvainiene[16], Frank Leypoldt[16,17], Stefan M. Pfister[3,4,5,6], Esther Hulleman[13,14], Brent A. Orr[11], Amar Gajjar[10], Giles W. Robinson[10], Joonas Haapasalo[15,18], Kristiina Nordfors[19,20], Johannes Gojo[8,22], Kristian W. Pajtler[3,4,5,6,22], Kendra K. Maass[3,4,5,6,22] & Paul A. Northcott[1,2,22] ✉

[1]Center of Excellence in Neuro-Oncology Sciences (CENOS), St. Jude Children's Research Hospital, Memphis, TN, USA. [2]Department of Developmental Neurobiology, St. Jude Children's Research Hospital, Memphis, TN, USA. [3]Hopp Children's Cancer Center (KiTZ), Heidelberg, Germany. [4]Department of Pediatric Oncology, Hematology and Immunology, Heidelberg Faculty of Medicine, Heidelberg University, Heidelberg University Hospital, Heidelberg, Germany. [5]Division of Pediatric Neurooncology, German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Heidelberg, Germany. [6]National Center for Tumor Diseases (NCT), NCT Heidelberg, a partnership between DKFZ and Heidelberg University Hospital, Heidelberg, Germany. [7]Heidelberg Biosciences International Graduate School (HBIGS), Faculty of Biosciences, Heidelberg University, Heidelberg, Germany. [8]Department of Pediatrics and Adolescent Medicine, Comprehensive Center for Pediatrics and Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria. [9]Department of Neurology, Division of Neuropathology and Neurochemistry, Medical University of Vienna, Vienna, Austria. [10]Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA. [11]Department of Pathology, St. Jude Children's Research Hospital, Memphis, TN, USA. [12]Division of Neuroradiology and Musculoskeletal Radiology, Department of Biomedical Imaging and Image-Guided Therapy, Medical University of Vienna, Vienna, Austria. [13]Departments of Pediatric Oncology/Hematology, Cancer Center Amsterdam, Amsterdam University Medical Centers, Amsterdam, The Netherlands. [14]Princess Máxima Center for Pediatric Oncology, Utrecht, The Netherlands. [15]Department of Pathology, Fimlab Laboratories Ltd, Tampere, Finland. [16]Institute of Clinical Chemistry, University Hospital Schleswig-Holstein, Kiel/Lübeck, Kiel, Germany. [17]Department of Neurology, University Hospital Schleswig-Holstein, Kiel, Kiel, Germany. [18]Department of Neurosurgery, Tampere University Hospital and Tampere University, Tampere, Finland. [19]Tampere Center for Child, Adolescent, and Maternal Health Research, Faculty of Medicine and Health Technology, Tampere University, Tampere, Finland. [20]Department of Pediatric Hematology and Oncology and Tays Cancer Center, Tampere University Hospital, Tampere, Finland. [21]These authors contributed equally: Kyle S. Smith, Tom T. Fischer, Katie Han. [22]These authors jointly supervised this work: Johannes Gojo, Kristian W. Pajtler, Kendra K. Maass, Paul A. Northcott. ✉e-mail: paul.northcott@stjude.org

https://doi.org/10.1038/s43018-026-01115-4

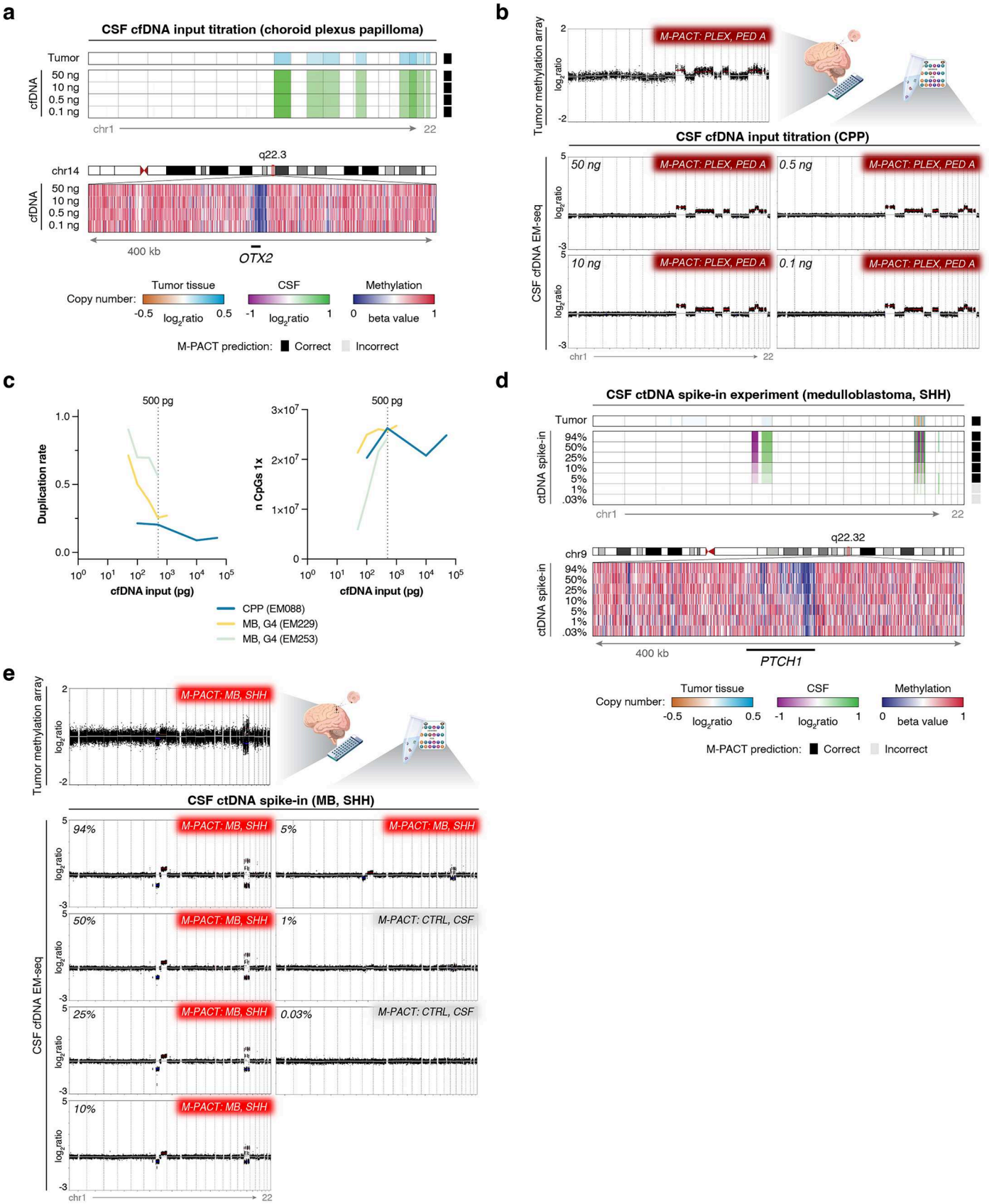


Extended Data Fig. 1 | See next page for caption.

https://doi.org/10.1038/s43018-026-01115-4

Extended Data Fig. 1 | Methods development for a methylation-based classifier for CNS tumor liquid biopsies and pre-analytical and quality control metrics for the cfDNA EM-seq cohort. a, Boxplot summarizing the difference between observed and predicted β-values across a range of 50,000 to 200,000 CpGs used to impute the 850,000 CpGs included in the Illumina EPICv1 (n = 914 samples). The median is shown as a thick line; box limits are 25th and 75th percentiles; whiskers denote 1.5 times the interquartile range. **b**, UMAP of the reference dataset of n = 5,014 methylation arrays encompassing 84 CNS tumor and 13 nontumor entities. **c**, Balanced accuracy of classification performance against the number of training epochs for one of the three neural networks subsequently integrated into the three-layer ensemble model. **d**, Histogram showing the distribution of quantifiable cfDNA inputs (pg). **e**, Density plot showing fragment size distribution of cfDNA sequencing reads. **f**, Histogram showing the distribution of number of paired reads. **g**, Histogram showing the distribution of sequencing coverage after de-duplication. **h**, Histogram showing the distribution of duplication rates. **i**, Box plot showing the number of CpGs with 4x versus 1x coverage based on a median sequencing depth of 4.74x (n = 268 samples); two-sided Mann-Whitney U-test; $P = 1.21 \times 10^{-42}$. The median is shown as a thick line; box limits are 25th and 75th percentiles; whiskers denote 1.5 times the interquartile range. **j**, Histogram showing the distribution of percent CpG methylation for available pUC19 control DNA spike-ins at 4x coverage. **k**, Histogram showing the distribution of percent CpG methylation for available lambda control DNA spike-ins at 4x coverage.

https://doi.org/10.1038/s43018-026-01115-4

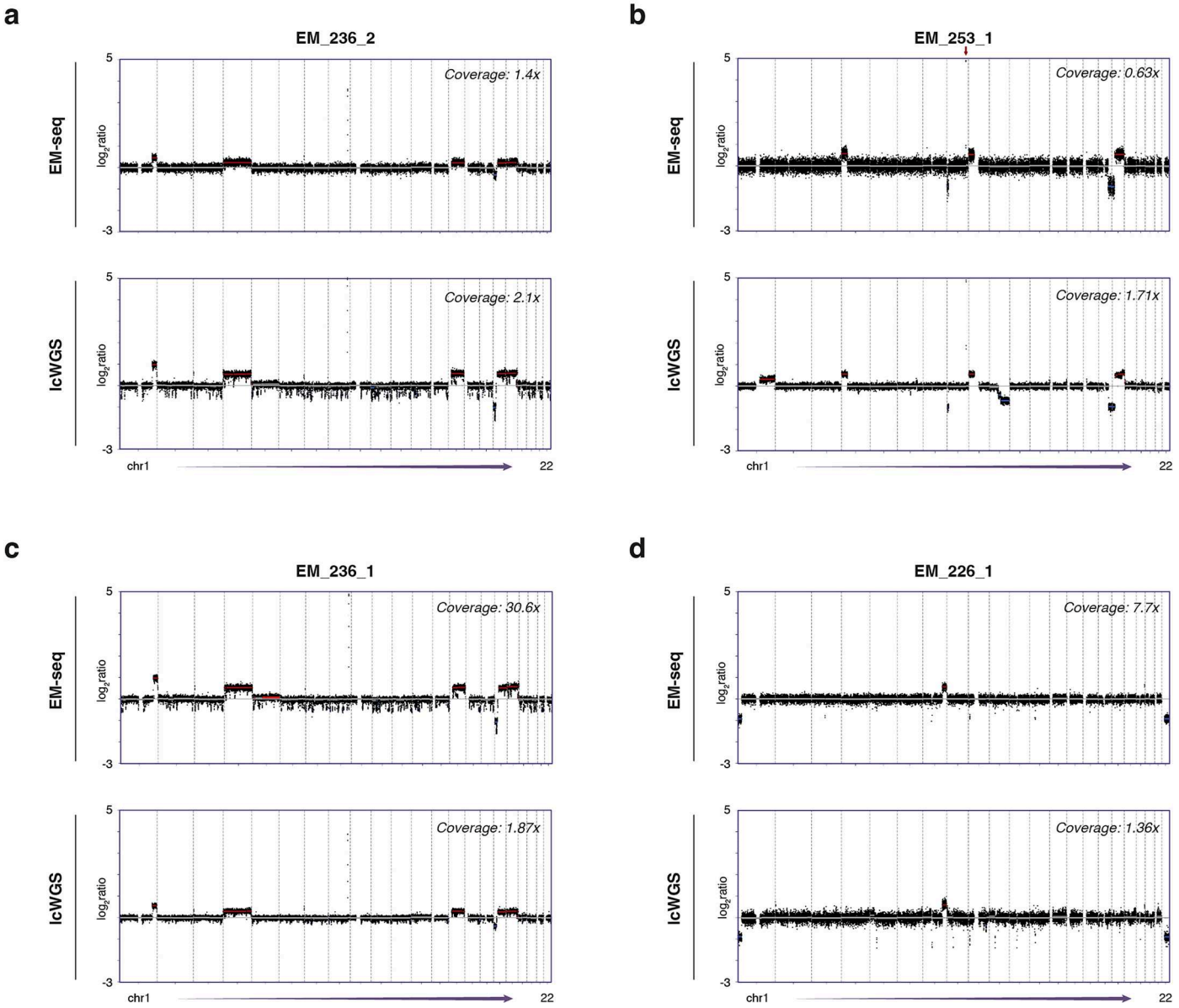


Extended Data Fig. 2 | See next page for caption.

https://doi.org/10.1038/s43018-026-01115-4

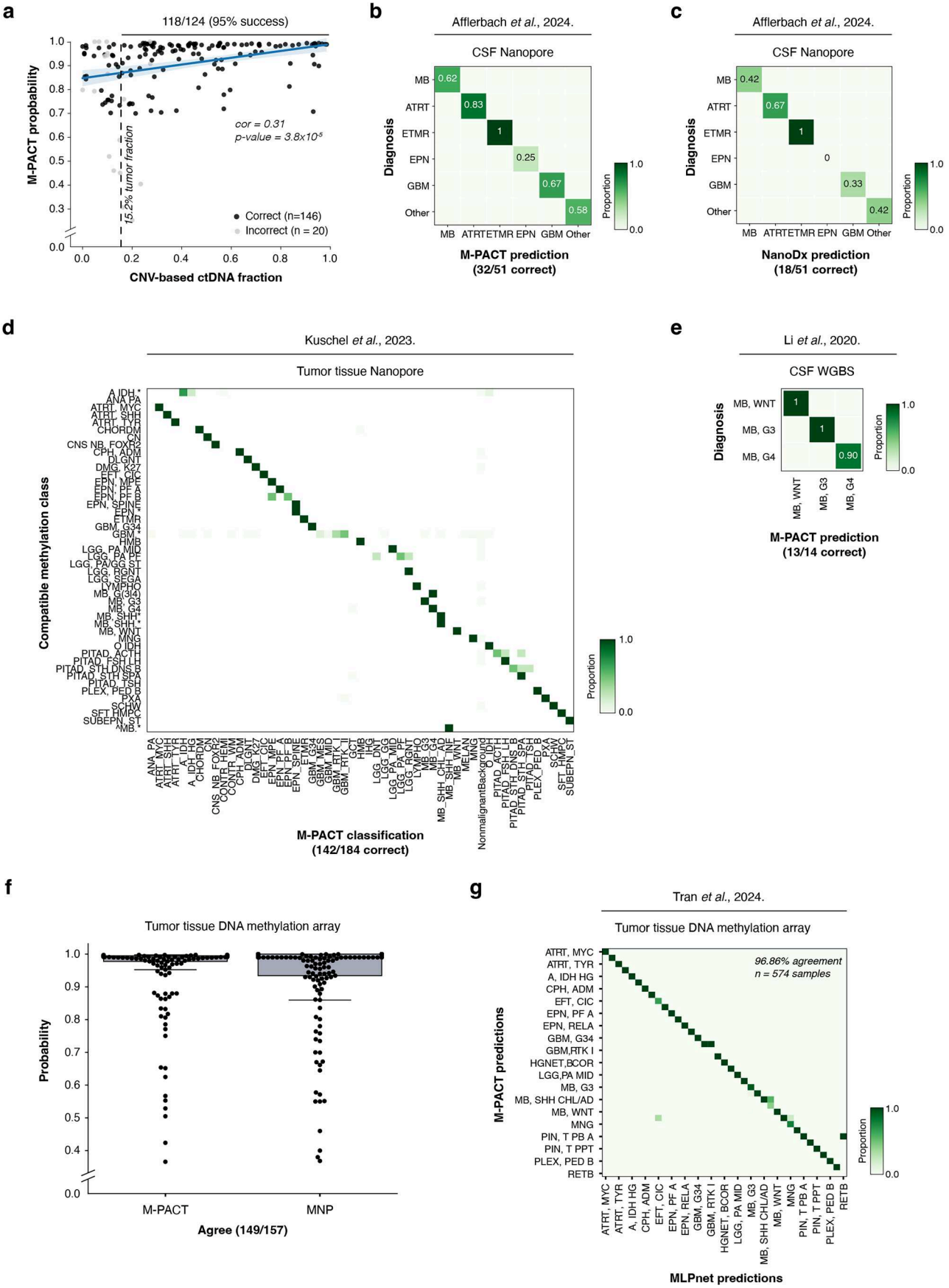
Extended Data Fig. 2 | Establishing EM-seq for low-input cfDNA samples. **a**, Genome-wide copy number variation (CNV) profiles annotated with M-PACT predictions (top) and methylation profiles (exemplified at the *OTX2* locus, bottom) at decreasing inputs of CSF-derived cfDNA from a choroid plexus papilloma patient (EM088). **b**, CNV plots from patient-matched tumor tissue methylation array and cfDNA EM-seq with associated M-PACT classifications for the cfDNA input titration experiment. **c**, Line plot showing duplication rate (left) and number of CpGs recovered at 1x coverage (right) against a range of cfDNA inputs for three patient CSF samples (EM088, EM229, EM253). **d**, Genome-wide CNV profiles annotated with M-PACT predictions (top) and methylation profiles (exemplified at the *PTCH1* locus, bottom) from the ctDNA spike-in experiment at decreasing fractions of a MB, SHH (EM270) sample. **e**, CNV plots from patient-matched tumor tissue methylation array and cfDNA EM-seq with associated M-PACT classifications for the ctDNA spike-in experiment.

https://doi.org/10.1038/s43018-026-01115-4



Extended Data Fig. 3 | Evaluation of discordant CNV profiles between EM-seq and lcWGS. a-d, CNV profiles derived from EM-seq (top) and lcWGS (bottom) with divergent estimations in CNV-based ctDNA fraction.

https://doi.org/10.1038/s43018-026-01115-4

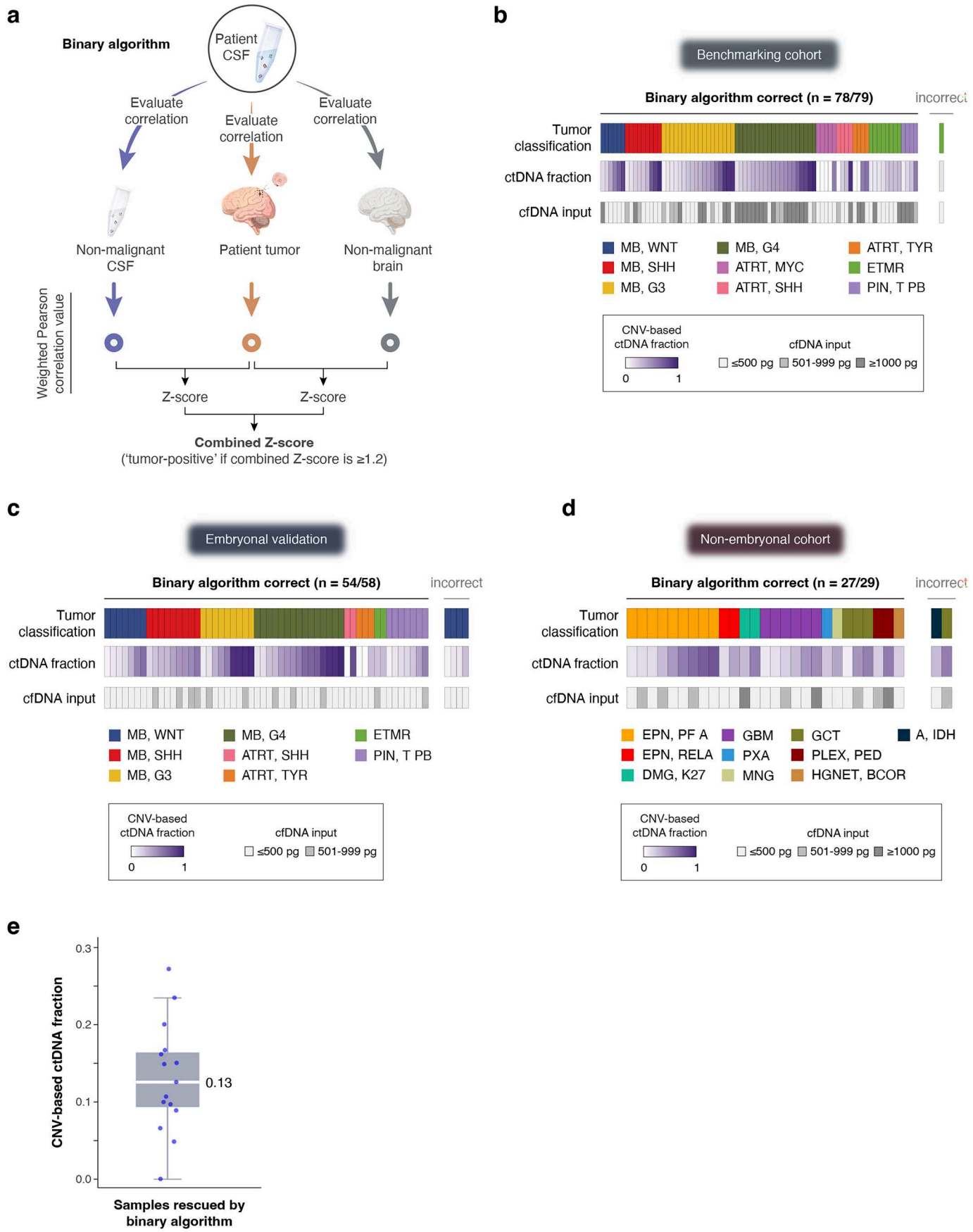


Extended Data Fig. 4 | See next page for caption.

https://doi.org/10.1038/s43018-026-01115-4

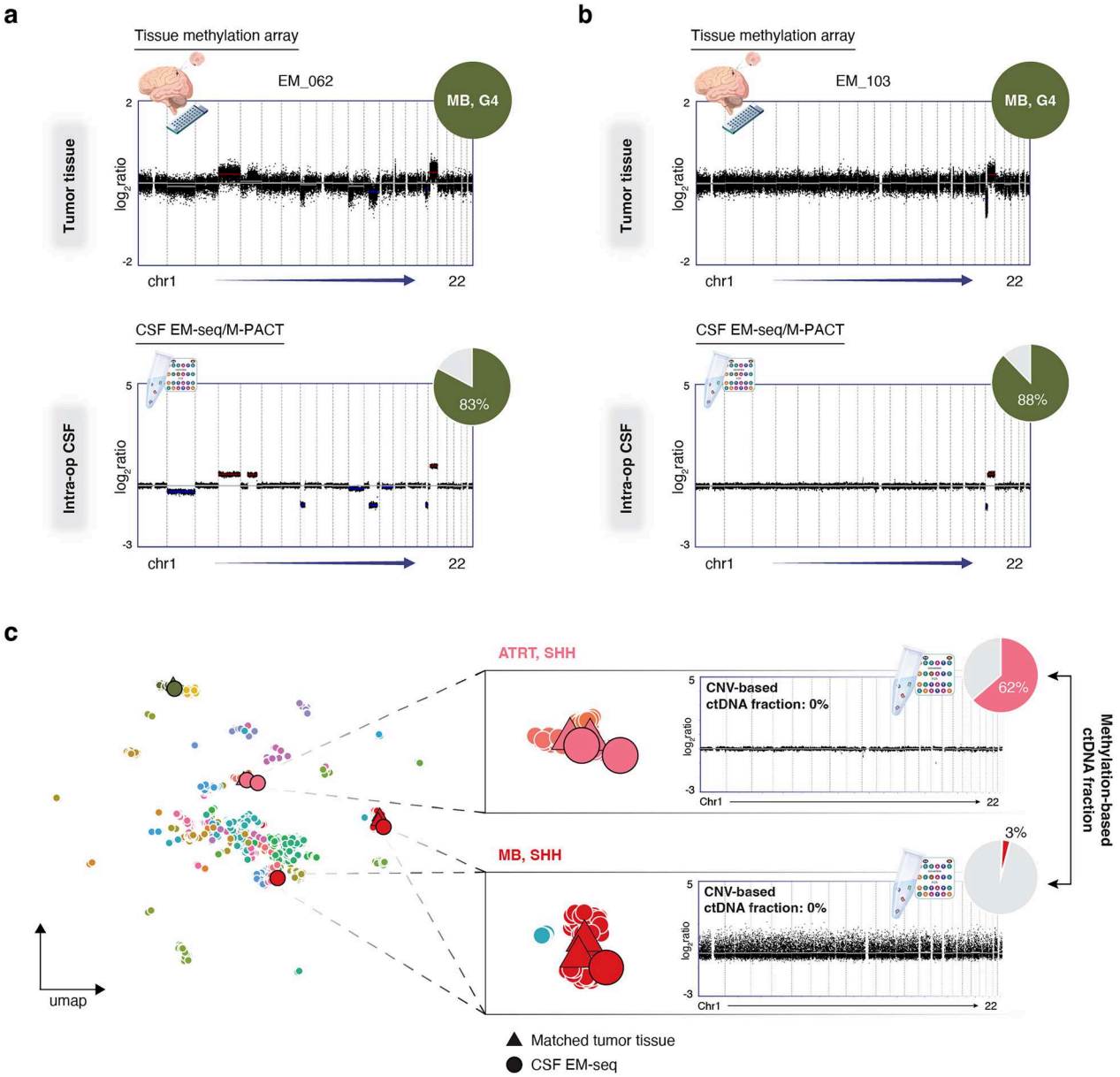
Extended Data Fig. 4 | ctDNA fraction threshold for accurate M-PACT classification and cross-platform validation of M-PACT across DNA methylation datasets. a, Scatter plot showing M-PACT probability against CNV-based ctDNA fraction, colored by correct versus incorrect M-PACT prediction (Pearson correlation). ctDNA fraction threshold of 0.15 yielded 95% classification accuracy across benchmarking, embryonal validation, and non-embryonal CNS tumor cohorts. Shading depicts the 95% confidence interval. **b**, Confusion matrix summarizing M-PACT performance in a cohort of cfDNA samples sequenced with Nanopore[23]. **c**, Confusion matrix summarizing NanoDx performance in the same cfDNA Nanopore dataset. **d**, Confusion matrix summarizing M-PACT performance in a tumor tissue-derived Nanopore dataset against predictions provided in the original study[41]. **e**, Confusion matrix summarizing M-PACT performance in a CSF whole-genome bisulfite sequencing dataset[31]. **f**, Box plot comparing classification probability scores for M-PACT versus the Heidelberg Brain Tumor classifier MNP for $n = 157$ tumor tissues processed on DNA methylation array. The median is shown as a thick line; box limits are 25th and 75th percentiles; whiskers denote 1.5 times the interquartile range. **g**, Confusion matrix summarizing M-PACT performance compared to the MLPnet classifier in a cohort of tumor tissue DNA methylation arrays[42]. For brevity and readability, only a subset of entities is labeled.

https://doi.org/10.1038/s43018-026-01115-4



Extended Data Fig. 5 | Development of a methylation-based binary classification method for CSF-derived cfDNA. a, Schematic of the binary classification strategy. **b-d**, Heatmaps summarizing the performance of the binary algorithm in the benchmarking cohort (**b**), embryonal validation cohort (**c**), and non-embryonal cohort (**d**). **e**, Box plot of CNV-based ctDNA fractions for CSF samples rescued by the binary algorithm ($n = 15$ samples). The median is shown as a thick line; box limits are 25th and 75th percentiles; whiskers denote 1.5 times the interquartile range.

https://doi.org/10.1038/s43018-026-01115-4



Extended Data Fig. 6 | See next page for caption.

https://doi.org/10.1038/s43018-026-01115-4

Extended Data Fig. 6 | Additional examples of classifying intra-operative CSF and balanced tumor genomes and evaluation of M-PACT performance and ctDNA fraction across sampling sites. a,**b**, Case vignettes showing accurate diagnosis of embryonal tumors from intra-operative CSF. Pie charts show methylation-based ctDNA fraction (colored segment) and nonmalignant cfDNA fraction (gray segment). **c**, UMAP projection of CSF samples collected in cases with balanced tumor genomes onto tumor methylation array reference profiles. EM-seq derived CNV profiles and pie charts illustrating methylation-based ctDNA fraction are displayed (ATRT, SHH [EM219], MB, SHH [EM269]). **d**, Box plot summarizing M-PACT probability for cfDNA derived from lumbar (n = 114 samples) versus cranial (n = 44 samples) CSF; two-sided Mann-Whitney U-test; P = 0.615. **e**, Box plot summarizing CNV-based ctDNA fraction for cfDNA derived from lumbar (n = 114 samples) versus cranial (n = 44 samples) CSF; two-sided Mann-Whitney U-test; P = 0.212.

nature portfolio

Corresponding author(s): Paul A. Northcott

Last updated by author(s): Dec 7, 2025

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	DNA methylation data was generated using Illumina Infinium MethylationEPIC arrays according to the manufacturer's instructions. EM-seq converted reads were generated according to the manufacturer's instructions and sequenced on a Nova-seq X. Nanopore sequencing was prepared using SQK-NBD114.24 (Oxford Nanopore Technologies) and subsequently sequenced on a PromethION 2 Solo, according to manufacturer's instructions.
Data analysis	Data processing code has been made available at https://github.com/kylessmith/MethylVerse and https://github.com/kylessmith/cfdna. Both software packages are publicly available for installation via the Python Package Index (pypi).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Microarray and sequencing data that support the findings of this study have been deposited in GEO (GSE292312) and EGA (EGAS50000001365), respectively. Data under the EGAS50000001365 is under controlled access due to patient privacy and governed by the EGAC00001002368 Data Access Committee with the corresponding data usage agreement (https://ega-archive.org/datasets/EGAD00001008277). Data access agreements will be reviewed and responded to within 30 days of application. Previously published lcWGS data that were re-analyzed here are available under accession code EGAS00001005592. Source data are provided with this paper.

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	No sex or gender analysis was conducted, as these annotations were not available.
Reporting on race, ethnicity, or other socially relevant groupings	Annotations on patient race or ethnicity were not available.
Population characteristics	We review genomics data from a series of 214 patients, 156 with CNS tumors and 58 with non-malignant diagnoses.
Recruitment	Participants were recruited based on their respective institutional guidelines and study protocols.
Ethics oversight	The protocols were approved by each institution's Institutional Review Board, and written informed consent was provided by all patients and/or their legal guardians.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Sample size was defined by sample availability. There was no sample-size calculation performed.
Data exclusions	No data were excluded from the analyses.
Replication	Each CSF sample was considered unique as they were either collected from different individuals or different clinical timepoints.
Randomization	Samples were included in each cohort (i.e., benchmarking cohort, embryonal validation cohort, non-embryonal cohort) based on sample availability and broad tumor type categorization (i.e., embryonal vs. non-embryonal CNS tumor).
Blinding	All sample processing, data collection, and analyses were performed without knowledge of clinical annotations.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☐	☒ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT00085202 (SJMB03); NCT02114229 (SJATRT); NCT03434262 (SJDAWN)
Study protocol	Treatment of Patients With Newly Diagnosed Medulloblastoma, Supratentorial Primitive Neuroectodermal Tumor, or Atypical Teratoid Rhabdoid Tumor (NCT00085202; NCT02114229; NCT03434262)
Data collection	Data collection was conducted in accordance with availability.
Outcomes	NA

Plants

Seed stocks	NA
Novel plant genotypes	NA
Authentication	NA

Magnetic resonance imaging

Experimental design

Design type	Standard of care clinical imaging.
Design specifications	Routine, off-the shelf, FDA approved imaging sequences.
Behavioral performance measures	NA

Acquisition

Imaging type(s)	Anatomical and diffusion-weighted imaging.
Field strength	Typically 3T, occasionally 1.5T.
Sequence & imaging parameters	Precontrast 2D sagittal T1-weighted gradient echo, axial T2-weighted fast spin-echo, axial T1-weighted gradient echo, axial echo planar diffusion-weighted and postcontrast 2D axial T1-weighted, axial T2-weighted FLAIR and 3D T1-weighted (MP-RAGE) sequences.
Area of acquisition	Whole brain imaging.

Diffusion MRI ☒ Used ☐ Not used

Parameters: echo planar DWI (b: 0, 500, 1000)

Preprocessing

Preprocessing software	No preprocessing for the purposes of this study.
Normalization	NA
Normalization template	NA
Noise and artifact removal	NA
Volume censoring	NA

Statistical modeling & inference

Model type and settings	NA
Effect(s) tested	NA

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016)	NA
Correction	NA

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☒	☐ Multivariate modeling or predictive analysis