

# DNA methylation-based classification of central nervous system tumours

A list of authors and their affiliations appears in the online version of the paper.

Accurate pathological diagnosis is crucial for optimal management of patients with cancer. For the approximately 100 known tumour types of the central nervous system, standardization of the diagnostic process has been shown to be particularly challenging—with substantial inter-observer variability in the histopathological diagnosis of many tumour types. Here we present a comprehensive approach for the DNA methylation-based classification of central nervous system tumours across all entities and age groups, and demonstrate its application in a routine diagnostic setting. We show that the availability of this method may have a substantial impact on diagnostic precision compared to standard methods, resulting in a change of diagnosis in up to 12% of prospective cases. For broader accessibility, we have designed a free online classifier tool, the use of which does not require any additional onsite data processing. Our results provide a blueprint for the generation of machine-learning-based tumour classifiers across other cancer entities, with the potential to fundamentally transform tumour pathology.

The developmental complexity of the brain is reflected by the vast array of distinct brain tumour entities defined in the current WHO (World Health Organization) classification of central nervous system (CNS) tumours<sup>1</sup>. These tumours are clinically and biologically highly diverse, encompassing a wide spectrum from benign neoplasms, which can frequently be cured by surgery alone (for example, pilocytic astrocytoma), to highly malignant tumours that respond poorly to any therapy (for example, glioblastoma). Previous studies have reported substantial inter-observer variability in the histopathological diagnosis of many CNS tumours, for example, in diffuse gliomas<sup>2</sup>, ependymomas<sup>3</sup> and supratentorial primitive neuroectodermal tumours<sup>4</sup>. To address this, some molecular grouping has been introduced into the update of the WHO classification, but only for selected entities such as medulloblastoma. Furthermore, several single-gene tests based on DNA methylation analysis (for example, *MGMT* promoter methylation status), fluorescence *in situ* hybridization (for example, 1p/19q codeletion, *EGFR*, *MYC*, *MYCN*, *PDGFRA*, 19q13.42) or immunohistochemistry (for example, *CTNNB1* and *LIN28A*) that are required to cover the most important differential diagnoses have been shown to be difficult to standardize. Such diagnostic discordance and uncertainty may confound decision-making in clinical practice as well as the interpretation and validity of clinical trial results.

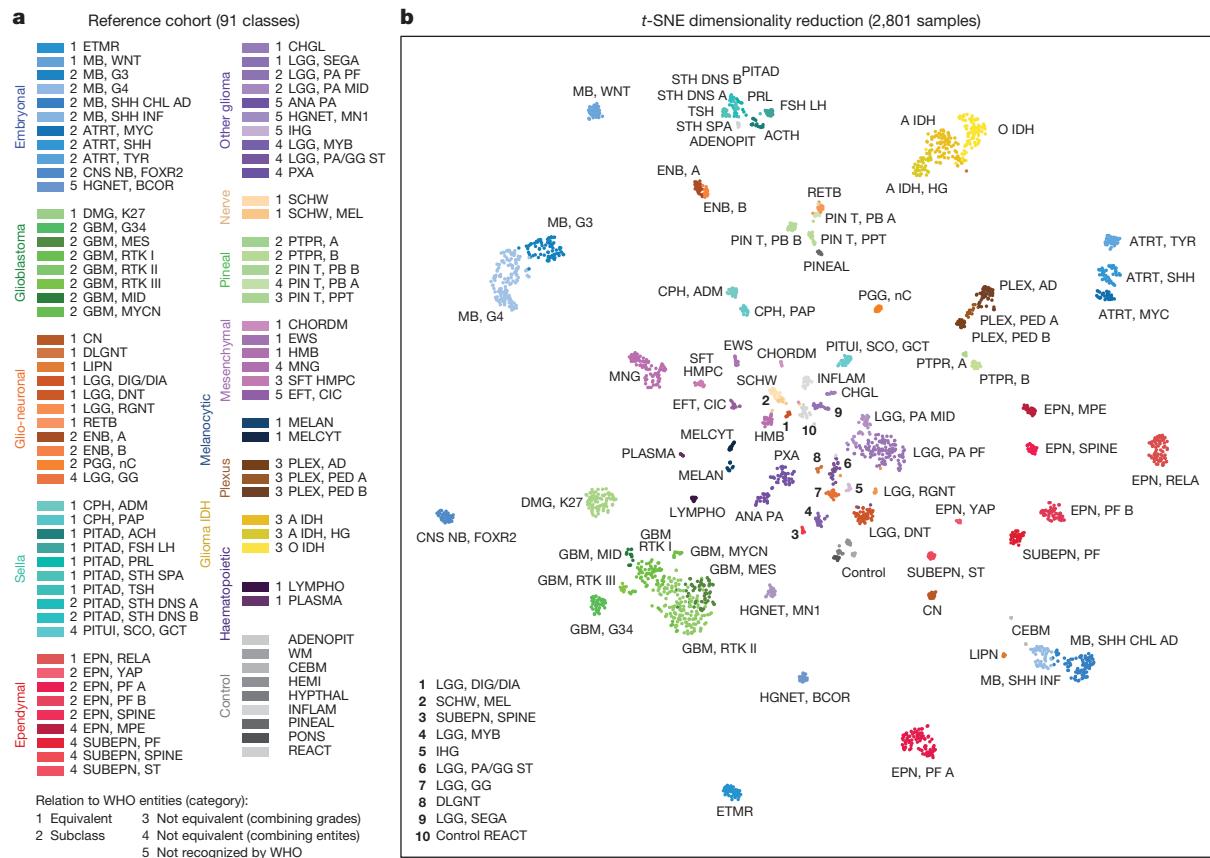
The cancer methylome is a combination of both somatically acquired DNA methylation changes and characteristics that reflect the cell of origin<sup>5,6</sup>. The latter property enables, for example, the tracing of the primary site of highly dedifferentiated metastases of cancers of unknown origin<sup>7</sup>. It has been convincingly shown that DNA methylation profiling is highly robust and reproducible even from small samples and poor quality material<sup>8</sup>, and such profiles have been widely used to sub-classify CNS tumours that were previously considered homogeneous diseases<sup>4,9–16</sup>. On the basis of this previous work within single entities, we present a comprehensive approach for the DNA methylation-based classification of all CNS tumour entities across age groups.

## CNS tumour reference cohort

To establish a comprehensive CNS tumour reference cohort, we generated genome-wide DNA methylation profiles using Infinium HumanMethylation450K BeadChip arrays (minimum of eight cases

per group) representing almost all WHO-defined neuroectodermal and sellar region tumours<sup>1</sup>. We further profiled mesenchymal tumours, melanoma, diffuse large B-cell lymphoma, plasmacytoma and six types of pituitary adenomas, in total comprising 76 histopathological entities and seven entity variants that occur in the CNS. All histopathological entities and variants were analysed by unsupervised clustering both within each entity and across histologically similar tumour entities, aiming to identify (i) distinct DNA methylation classes within one histopathological entity and (ii) DNA methylation classes comprising tumours displaying a varied histological phenotype. This iterative process led to the designation of 82 CNS tumour classes characterized by distinct DNA methylation profiles (Fig. 1a). Of these, 29 classes were equivalent to a single WHO entity (category 1); 29 classes represented subclasses within a WHO entity (category 2); in 8 classes, WHO grading was not fully recapitulated (category 3) and in 11 classes, the boundaries of methylation classes were not identical to the entity boundaries of WHO (category 4) (Fig. 1a). The remaining five represented DNA methylation classes that have not been defined by the WHO classification (category 5), three of which were recently described<sup>4</sup> as well as the not yet well-defined class of astrocytoma and one new subclass of infantile hemispheric glioma. There was evidence for several additional classes of rare tumours, with too few cases to be included at present. Taking the impact of the tumour microenvironment on the methylation profile into consideration, we included 47 tumour samples with a pronounced inflammatory or reactive tumour microenvironment, both of which have distinct methylation profiles. We additionally selected 72 samples that represent seven non-neoplastic CNS regions, resulting in a combined reference cohort of 2,801 samples from 91 classes (Fig. 1a) that was visualized using *t*-distributed stochastic neighbour embedding (*t*-SNE) dimensionality reduction<sup>17</sup> (Fig. 1b). This analysis further supported the separation of samples into the defined DNA methylation classes (see also Extended Data Fig. 1a, b; unprocessed IDAT files can be downloaded from the NCBI Gene Expression Omnibus (GEO), under accession number GSE109381). Supplementary Table 1 gives an overview of methylation class characteristics and Supplementary Table 2 shows case-by-case information of the reference samples.

A list of authors and their affiliations appears in the online version of the paper.



**Figure 1 | Establishing the DNA methylation-based CNS tumour reference cohort.** **a**, Overview of the 82 CNS tumour methylation classes and nine control tissue methylation classes of the reference cohort. The methylation classes are grouped by histology and colour-coded. Category 1 methylation classes are equivalent to a WHO entity, category 2 methylation classes are a subgroup of a WHO entity, category 3 methylation classes are not equivalent to a unique WHO entity with combining of WHO grades, category 4 methylation classes are not equivalent to a unique WHO entity with combining of WHO entities, and category 5 methylation classes are not recognized as a WHO entity. Full names and further details of the 91 classes are included in Supplementary Table 1. Embryonal tumours, shades of blue; glioblastomas, shades of green; other gliomas, shades of violet; ependymomas, shades of red; glio-neuronal tumours, shades of orange; IDH-mutated gliomas, shades of yellow; choroid plexus tumours, shades of brown; pineal region tumours, shades of mint green; melanocytic tumours, shades of dark blue; sellar region tumours, shades of cyan; mesenchymal tumours, shades of pink; nerve tumours, shades of beige; haematopoietic tumours, shades of dark purple; control tissues: shades of grey. **b**, Unsupervised clustering of reference cohort samples ( $n = 2,801$ ) using t-SNE dimensionality reduction. Individual samples are colour-coded in the respective class colour ( $n = 91$ ) and labelled with the class abbreviation. The colour code and abbreviations are identical to **a**.

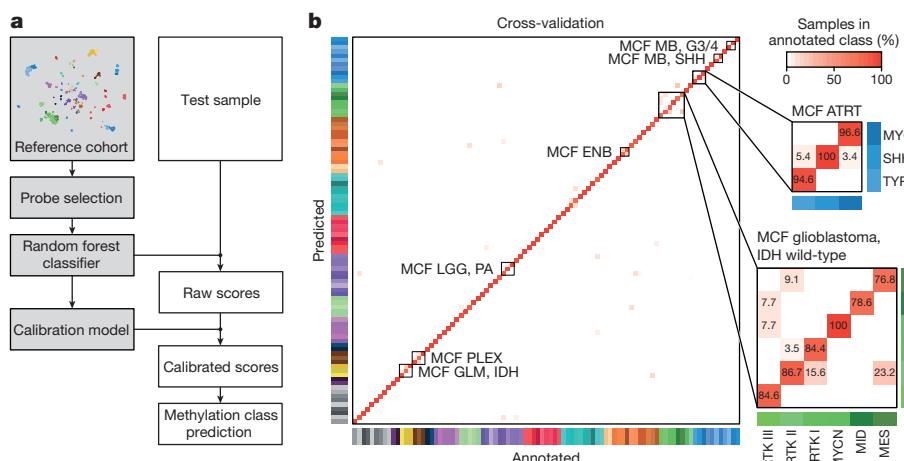
The stability of separation of methylation classes by t-SNE was analysed by iterative random down-sampling of the reference cohort and indicated a high stability of the groups (Extended Data Fig. 1c, d). Testing for confounding batch effects within our reference cohort did not reveal unexpected confounding factors (Extended Data Figs 2, 3a–c). For reference astrocytomas, oligodendroglomas and glioblastomas, we performed additional classification according to The Cancer Genome Atlas (TCGA) pan-glioma DNA methylation model<sup>18</sup>, which indicated a strong association between the TCGA classes LGM1–6 and specific classes that are defined in our reference cohort (Extended Data Fig. 3d and Supplementary Table 2).

## Classifier development

Application in routine diagnostics requires fast and reproducible classification of samples as well as a measure of confidence for the specific call. To this end, we used the random forest algorithm, which is a so-called ensemble method that combines the predictions of several ‘weak’ classifiers to achieve improved prediction accuracy<sup>19</sup>. Using this algorithm, we generated 10,000 binary decision trees, incorporating genome-wide information from all 2,801 reference samples and 91 methylation classes (Extended Data Fig. 4). Each of these trees assigns a given diagnostic sample to one of the 91 classes, resulting in an aggregate raw score (Fig. 2a). To obtain class probability estimates that can

be used to guide diagnostic decision-making, we fitted a multinomial logistic regression calibration model that transforms the raw score into a probability that measures the confidence in the class assignment (the calibrated score). The calibration enables the comparison of classifier results between classes despite a different raw score distribution (Extended Data Fig. 5a, b). Cross-validation of the random forest classifier resulted in an estimated error rate of 4.89% for raw scores and 4.28% for calibrated scores and an area under receiver operating characteristic curve of 0.99, indicating a high discriminating power (Fig. 2b and Extended Data Fig. 5c). The vast majority of cross-validation misclassifications occurred within eight groups of histologically and biologically closely related tumour classes, the distinction of which is currently without clinical impact (with the possible exception of choroid plexus tumours<sup>13</sup>; Fig. 2b). We therefore defined eight ‘methylation class families’, for which calibrated scores are summed up to a single score. This reduced the cross-validated error rate for the clinically relevant groupings to 1.14% (Fig. 2b and Extended Data Fig. 5c). Taking the maximum score for class assignment and using a multiclass approach<sup>20</sup>, overall sensitivity and specificity was 0.989 and 0.999, respectively (Extended Data Fig. 5c).

For application to diagnostic tumour samples, a threshold value for the prediction of a matching class is required. Using receiver operating characteristic curve analysis of the maximum calibrated scores we



**Figure 2 | Development and cross-validation of the DNA methylation-based CNS tumour classifier.** **a**, Schematic of principal classifier components (grey) and processing steps for individual test samples (white). The most informative probes are selected for training of the random forest classifier. The classifier produces raw scores that represent the number of decision trees that assign a test sample to a specific methylation class. To enable inter-class comparability, a calibration model is used, which transforms raw scores into calibrated scores. Calibrated

devised an optimal ‘common’ calibrated score threshold of  $\geq 0.9$  (Extended Data Fig. 5d, e). For subclasses within methylation class families, we defined a threshold value of  $\geq 0.5$  as sufficient for a valid prediction, as long as all family member scores add up to a total score of  $\geq 0.9$ . Single class specificity and sensitivity for the  $\geq 0.9$  threshold are provided in Supplementary Table 3.

### Clinical implementation

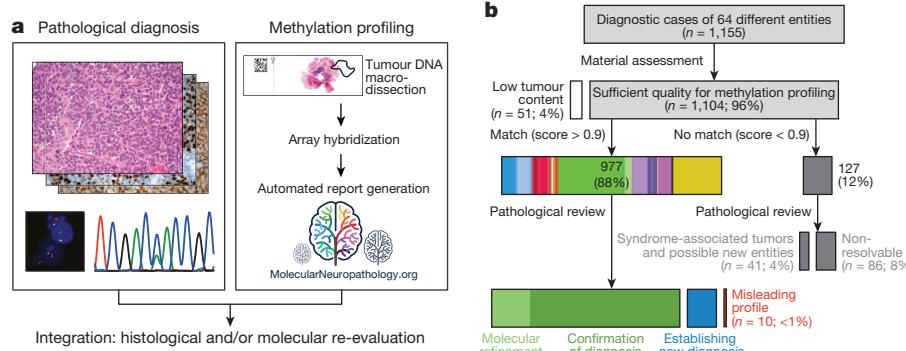
For evaluation of clinical utility, we prospectively analysed a series of 1,155 diagnostic CNS tumours in parallel with standard histopathological analyses (Fig. 3a, b). For 51 cases (4%) the material was not suitable for methylation profiling, mostly because the tumour cell content was too low or because of limited total material. Methylation profiling was performed for the remaining 1,104 samples and the cases were assigned as either ‘matching to a defined DNA methylation class’ (calibrated score  $\geq 0.9$ ) or as ‘no match’ cases (highest score  $<0.9$ ) (for a case-by-case list, see Supplementary Table 4). The investigated cases comprised 64 different histopathological entities from both adult (71%) and paediatric patients (29%). The spectrum of entities was enriched for rare and difficult to diagnose cases received for referral, and therefore did not exactly match the distribution seen in daily routine diagnostic practice. Histopathological evaluation was

scores represent an estimated probability measure of methylation class assignment. **b**, Heat map showing results of a threefold cross-validation of the random forest classifier incorporating information of  $n = 2,800$  biologically independent samples allocated to 91 methylation classes. Deviations from the bisecting line represent misclassification errors (using the maximum calibrated score for class prediction). Methylation class families (MCF) are indicated by black squares. The colour code and abbreviations are identical to Fig. 1a.

performed blinded to DNA methylation profiling results and included standard molecular testing.

In total, 88% of profiled samples ( $n = 977$  out of 1,104) matched to an established DNA methylation class with a calibrated classifier score  $\geq 0.9$  (Fig. 3b). For 838 of these (838 out of 1,104; 76%), results obtained by pathology and DNA methylation profiling were concordant. In 171 of the cases, an unambiguous molecular subgroup could be assigned, which would not have been available based on histopathology evaluation only (for example, molecular subgroups of medulloblastoma and ependymoma, many of which were included in the latest version of the WHO classification of CNS tumours<sup>1</sup>).

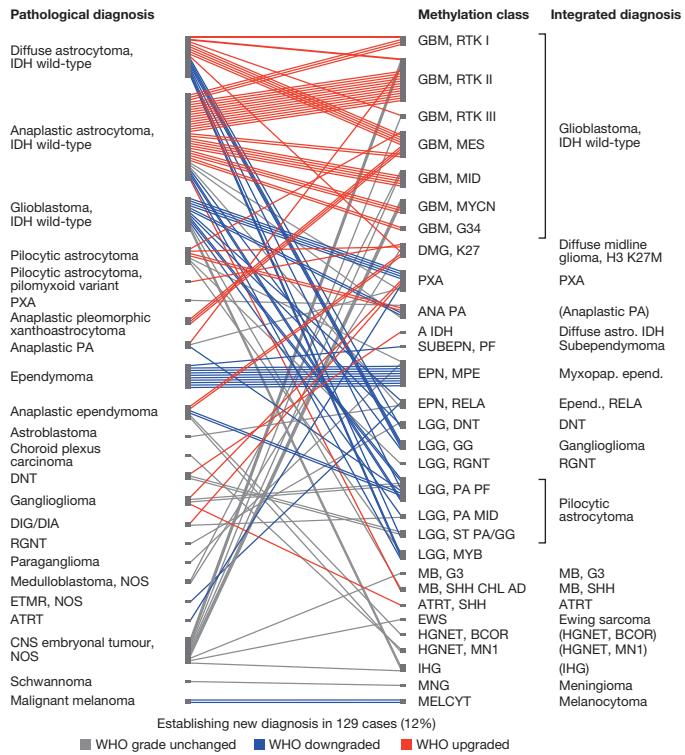
For the remaining 139 samples with a calibrated classifier score  $\geq 0.9$ , the DNA methylation class was not in accordance with the pathological diagnosis. These cases were histologically and molecularly re-evaluated, including additional molecular diagnostics (DNA copy-number profiling, targeted gene sequencing, gene panel sequencing<sup>21</sup> and gene-fusion analysis of a subset of cases, see Supplementary Table 5). This resulted in a revision of the initial histopathological diagnosis in 129 of the 139 cases (12% of all cases; Fig. 4) in favour of the predicted methylation class. In agreement with several recent reports<sup>16,22,23</sup>, several of these were IDH-wild-type astrocytomas and anaplastic astrocytomas reclassified



**Figure 3 | Implementation of the classifier in diagnostic practice.**

**a**, Classifier validation by an independent prospective cohort of diagnostic samples. Pathological diagnosis was established by current pathological standard according to the 2016 version of the WHO classification of CNS tumours and compared to classification by methylation profiling.

Cases were categorized as ‘confirmation of diagnosis’, ‘establishing new diagnosis’, ‘misleading profile’ or ‘no match to defined class’. **b**, Overview of methylation profiling result from 1,155 diagnostic samples and integration with pathological diagnosis.



**Figure 4 | Reassessment of discrepant cases and establishment of new diagnosis.** Discrepancy between pathological diagnosis (left) and methylation profiling (middle) was observed for 139 cases. For 129 cases, histological and molecular reassessment (Supplementary Table 5) resulted in a change in the initial diagnosis with formulation of a new integrated diagnosis (right). For 92 cases, this involved a change in the WHO grading, with both down- (blue) and upgrading (red). Integrated diagnoses in brackets are not recognized as a WHO entity. For methylation class abbreviations see Supplementary Table 1.

as IDH-wild-type glioblastomas. Establishing a new diagnosis had a profound clinical impact: a change in WHO grading was observed in 71% of these cases (92 out of 129), and both upgrading (41%, 53 out of 129) and downgrading (30%, 39 out of 129; Fig. 4). Discrepant results could not be resolved in only 10 cases (<1% of profiled cases), and the histopathological diagnosis was therefore retained.

To substantiate the impact in clinical practice, we contacted five external centres that have started to implement methylation profiling for diagnostic cases using our algorithm. In total, these centres analysed 401 diagnostic cases and in 50 cases (12%) a new diagnosis was established after methylation profiling, very closely recapitulating our rate of reclassification (Extended Data Fig. 6a and Supplementary Table 6). For individual centres, the rate of reclassification varied between 6% and 25%, which was most likely due to differences in the spectrum of investigated cases and more upfront molecular testing by some centres (Extended Data Fig. 6b and Supplementary Table 6).

Twelve per cent of tumours from the prospective cohort (127 out of 1,104) could not be assigned to a DNA methylation class using the rigid calibrated classifier score cutoff of  $\geq 0.9$  (Fig. 3b). To further clarify the role of these non-classifiable cases, we performed an unsupervised t-SNE analysis of the reference cohort together with the diagnostic cohort (Fig. 5a). Using this analysis, we found a high overlap of the classifiable cases with the reference cohort, whereas non-classifiable cases frequently fell in the periphery of the reference classes or were even completely separate from these and frequently grouped with other non-classifiable cases (Fig. 5a). This may indicate that such cases represent rare novel molecular entities that have not been recognized previously. An example for a likely novel CNS tumour entity is shown in Fig. 5b, c.

## Technical and inter-laboratory testing

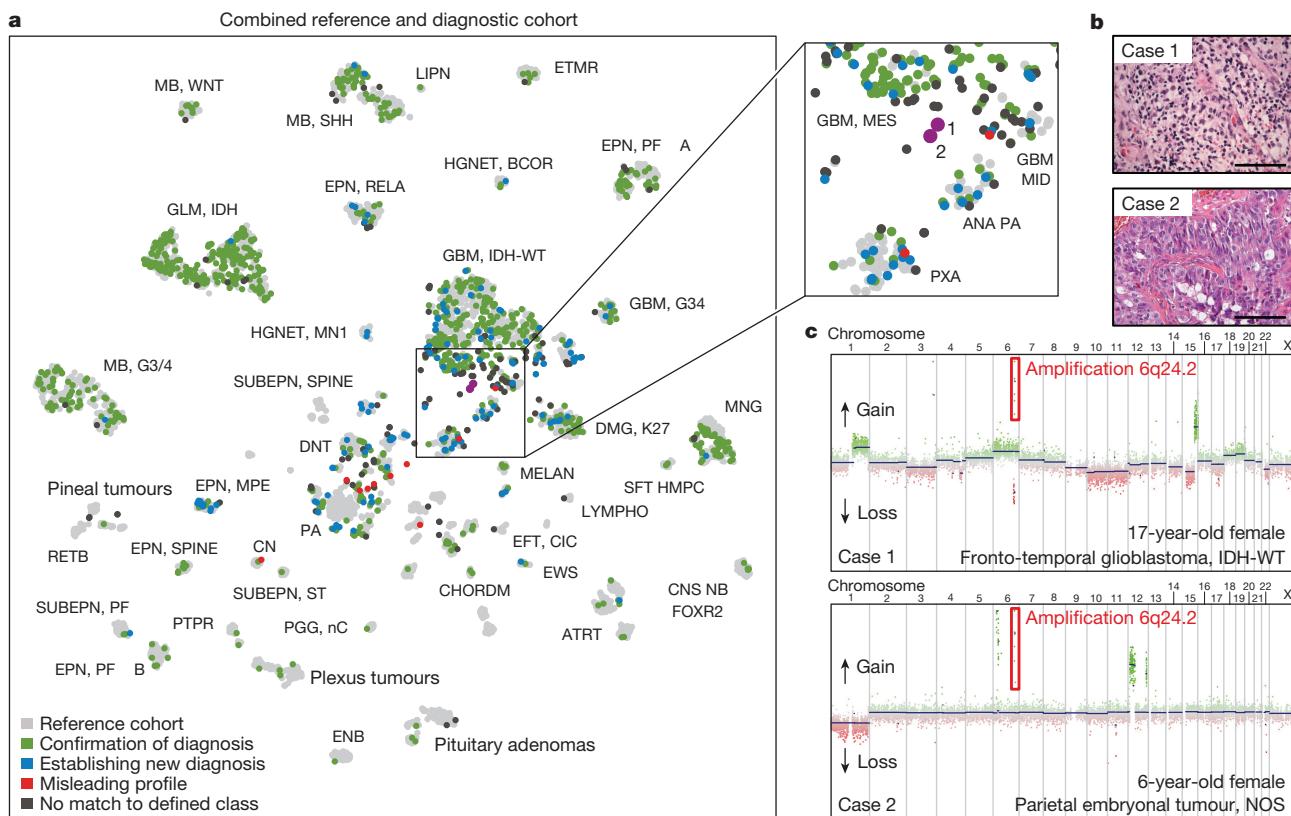
The technical robustness of the random forest classifier was investigated by inter-laboratory comparison. Results of two independent laboratories (starting from DNA extraction) were highly correlated, with only 2 out of 53 samples (4%) showing a classifier score slightly lower than 0.9 in one of the centres, whereas all other cases were classified identically (Extended Data Fig. 7a). Calculation of copy-number profiles was also stable across laboratories (Extended Data Fig. 7b). To ascertain forward compatibility with developing technologies, we further used the random forest classifier to investigate newer Infinium MethylationEPIC BeadChip DNA methylation arrays and high-coverage whole-genome bisulphite sequencing data. For all 16 samples from different CNS tumours that were profiled on both array platforms, raw scores (Extended Data Fig. 7c) and calibrated scores (data not shown) were highly correlated and running them through the classifying algorithm resulted in identical class-assignment for each case. Furthermore, for all 50 high-coverage whole-genome bisulphite sequencing samples (11 different CNS tumour entities), the highest prediction score was for the same class as with the Infinium HumanMethylation450K BeadChip array, suggesting that our approach is applicable to different DNA methylation profiling techniques with only slight adaptations (Extended Data Fig. 7d).

## Global dissemination of the platform

To ensure unrestricted community access to our classification system, we created a free web platform for data upload, automatic normalization, random forest classification and PDF report generation (<https://www.molecularneuropathology.org>). DNA copy-number profiles<sup>24</sup> and O<sup>6</sup>-methylguanine-DNA-methyltransferase (*MGMT*) promoter methylation status<sup>25</sup> are additionally provided, since they can be generated from the same data source—thus having the potential of replacing several time- and cost-intensive single-gene tests. A representative website report is shown in Extended Data Fig. 8. When data are uploaded, the data provider can choose to give consent that the data are allowed to be used for further classifier development. We expect that this web platform can therefore act as a hub for a worldwide cooperative network to continuously identify and track rare tumour classes so that they can eventually be added to the catalogue of known human cancers. In the first year after launch in December 2016, over 4,500 cases have been uploaded from over 15 participating centres. New biological insights are also likely to be gained based on the interrelationships of tumour classes, and by closer examination of how differential DNA methylation affects tumour biology.

## Discussion

We here demonstrate that DNA methylation-based CNS tumour classification using a comprehensive machine-learning approach is a valuable asset for clinical decision-making. In particular, the high level of standardization has great promise to reduce the substantial inter-observer variability observed in current CNS tumour diagnostics. Furthermore, in contrast to traditional pathology, where there is a pressure to assign all tumours to a described entity even for atypical or challenging cases, the objective measure that we provide here allows for ‘no match’ to a defined class. This information can also be of substantial value to highlight that a tumour is not a typical example of a given differential diagnosis, and may rather belong to a rarer, currently undefined class. We defined five categories of methylation classes that have different clinical implications. Category 1 can be directly translated to WHO entities. Category 2 represents subclasses of WHO entities. For all but ependymal tumours, subclassification currently has little clinical consequence and a translation back to the WHO class may be appropriate for clinical purposes. Category 3 reflects the fact that WHO grading cannot be fully recapitulated by methylation profiling for several classes. Further data are required to assess whether the methylation classes of this category may provide a more robust means of prognostication than histology alone, as has been demonstrated for



**Figure 5 | DNA methylation-based identification of potential new CNS tumour entities.** **a**, Unsupervised clustering of the combined reference ( $n = 2,801$ , grey) and diagnostic cohort ( $n = 1,104$ , coloured) using t-SNE dimensionality reduction. Abbreviated names indicate the reference cohort classes as in Fig. 1. The diagnostic samples are colour-coded as confirmation of diagnosis ( $n = 838$ , green), establishing new diagnosis ( $n = 129$ , blue), misleading profile ( $n = 10$ , red) and no match to defined class ( $n = 127$ , dark grey). The matching (green) and reclassified (blue) cases show high overlap with the reference cases. The non-classifiable (black) and the misleading (red) cases frequently fall in the periphery of the reference classes or are completely separate of these. The magnified area (right) highlights two non-classifiable cases (here in magenta for easier identification) that group together in the t-SNE representation.

several other classes<sup>4,9,11</sup>. In category 4, the WHO entity boundaries are not identical to the boundaries of the methylation classes. Until additional data on the exact boundaries become available, this category should be critically discussed in the clinical context and orthogonal testing should be undertaken whenever possible. Category 5 represents putative new entities that are currently not recognized by the WHO, and although limited data on these cases are currently available, the biological rationale for a novel class was considered to be strong.

A study in which reference pathology and molecular diagnostics, including DNA methylation profiling, are blinded for each other's results is currently ongoing for all childhood brain tumours diagnosed in Germany to objectivize the potential effect of re-classification on patient outcome (Molecular Neuropathology 2.0; <http://www.kitz-heidelberg.de/molecular-diagnostics>), with results due over the next few years.

A uniform implementation of the classification algorithm holds great promise for standardization of tumour diagnostics across centres and across clinical trials. Furthermore, the digital nature of methylation data facilitates easy exchange and will allow aggregation of extensive tumour libraries. This will probably result in the detection of exceptionally rare tumour classes and a continued refinement of classifiers. Inclusion of new classes will allow a prompt translation into diagnostic practice, almost certainly resulting in a more dynamic tumour classification. In our experience, adaptation of this technique in diagnostic laboratories

**b**, Both highlighted non-classifiable cases occurred in female children, and had primitive neuroectodermal histology (glioblastoma- or embryonal tumour-like characteristics). Histology was assessed by three independent pathologists with similar results. Scale bars, 200  $\mu\text{m}$ . **c**, Both cases shared a high-level amplification of chromosome 6q24.2 (common amplified region chr6:144,293–144,649,987). The common region includes only 5 protein-coding genes: *LTV1* (*LTV1* ribosome biogenesis factor), *ZC2HC1B* (zinc finger C2HC-type containing 1B), *PLAGL1* (*PLAG1* like zinc finger 1), *SF3B5* (splicing factor 3b subunit 5) and *STX11* (syntaxin 11). This amplification was not observed in any of the other tumours from the reference or diagnostic cohort. Copy-number analysis was performed once using copy-number information derived from the methylation array data.

is relatively straightforward. Extended Data Fig. 9 summarizes a sample workflow for diagnostic implementation. We expect that the principle of using DNA methylation signatures as part of a combined histology and molecular tumour classification will improve diagnostic accuracy not only in neuropathology, but will also serve as a blueprint in other fields of tumour pathology.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 May 2017; accepted 13 February 2018.

Published online 14 March 2018.

1. Louis, D. N., Ohgaki, H., Wiestler, O. D. & Cavenee, W. K. *WHO Classification of Tumours of the Central Nervous System* revised 4th edn (IARC, 2016).
2. van den Bent, M. J. Interobserver variation of the histopathological diagnosis in clinical trials on glioma: a clinician's perspective. *Acta Neuropathol.* **120**, 297–304 (2010).
3. Ellison, D. W. *et al.* Histopathological grading of pediatric ependymoma: reproducibility and clinical relevance in European trial cohorts. *J. Negat. Results Biomed.* **10**, 7 (2011).
4. Sturm, D. *et al.* New brain tumor entities emerge from molecular classification of CNS-PNETs. *Cel.* **164**, 1060–1072 (2016).
5. Fernandez, A. F. *et al.* A DNA methylation fingerprint of 1628 human samples. *Genome Res.* **22**, 407–419 (2012).
6. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510**, 537–541 (2014).

7. Moran, S. *et al.* Epigenetic profiling to classify cancer of unknown primary: a multicentre, retrospective analysis. *Lancet Oncol.* **17**, 1386–1395 (2016).
8. Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125**, 913–916 (2013).
9. Sturm, D. *et al.* Hotspot mutations in *H3F3A* and *IDH1* define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* **22**, 425–437 (2012).
10. Reuss, D. E. *et al.* Adult IDH wild type astrocytomas biologically and clinically resolve into other tumor entities. *Acta Neuropathol.* **130**, 407–417 (2015).
11. Pajtler, K. W. *et al.* Molecular classification of ependymal tumors across all CNS compartments, histopathological grades, and age groups. *Cancer Cell* **27**, 728–743 (2015).
12. Lambert, S. R. *et al.* Differential expression and methylation of brain developmental genes define location-specific subsets of pilocytic astrocytoma. *Acta Neuropathol.* **126**, 291–301 (2013).
13. Thomas, C. *et al.* Methylation profiling of choroid plexus tumors reveals 3 clinically distinct subgroups. *Neuro-oncol.* **18**, 790–796 (2016).
14. Mack, S. C. *et al.* Epigenomic alterations define lethal CIMP-positive ependymomas of infancy. *Nature* **506**, 445–450 (2014).
15. Johann, P. D. *et al.* Atypical teratoid/rhabdoid tumors are comprised of three epigenetic subgroups with distinct enhancer landscapes. *Cancer Cell* **29**, 379–393 (2016).
16. Wiestler, B. *et al.* Integrated DNA methylation and copy-number profiling identify three clinically and biologically relevant groups of anaplastic glioma. *Acta Neuropathol.* **128**, 561–571 (2014).
17. Van Der Maaten, L. & Hinton, G. H. Visualizing data using t-SNE. *J. Mach. Learn. Res.* **9**, 2579–2605 (2008).
18. Ceccarelli, M. *et al.* Molecular profiling reveals biologically discrete subsets and pathways of progression in diffuse glioma. *Cell* **164**, 550–563 (2016).
19. Breiman, L. Random forests. *Mach. Learn.* **45**, 5–32 (2001).
20. Sokolova, M. & Lapalme, G. A systematic analysis of performance measures for classification tasks. *Inf. Process. Manage.* **45**, 427–437 (2009).
21. Sahm, F. *et al.* Next-generation sequencing in routine brain tumor diagnostics enables an integrated diagnosis and identifies actionable targets. *Acta Neuropathol.* **131**, 903–910 (2016).
22. Weller, M. *et al.* Molecular classification of diffuse cerebral WHO grade II/III gliomas using genome- and transcriptome-wide profiling improves stratification of prognostically distinct patient groups. *Acta Neuropathol.* **129**, 679–693 (2015).
23. The Cancer Genome Atlas Research Network. Comprehensive, integrative genomic analysis of diffuse lower-grade gliomas. *N. Engl. J. Med.* **372**, 2481–2498 (2015).
24. Hovestadt, V. & Zapatka, M. conumee: enhanced copy-number variation analysis using Illumina methylation arrays. v.1.4.2 R package v.0.99.4 <http://www.bioconductor.org/packages/release/bioc/html/conumee.html> (2015).
25. Bady, P., Delorenzi, M. & Hegi, M. E. Sensitivity analysis of the MGMT-TP27 model and impact of genetic and epigenetic context to predict the MGMT methylation status in gliomas and other tumors. *J. Mol. Diagn.* **18**, 350–361 (2016).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We thank U. Lass, A. Habel, I. Oezen for technical and administrative support, the Microarray unit of the Genomics and Proteomics Core Facility (DKFZ) for methylation services, the German Glioma Network and the Neuroonkologische Arbeitsgemeinschaft for sharing their data. This research was supported by the DKFZ-Heidelberg Center for Personalized Oncology (DKFZ-HIPO\_036), the German Childhood Cancer Foundation (DKS 2015.01), an Illumina Medical Research Grant, the DKTK joint funding project ‘Next Generation Molecular Diagnostics of Malignant Gliomas’, the A Kids’ Brain Tumour Cure (PLGA) Foundation, the Brain Tumour Charity (UK) for the Everest Centre for Paediatric Low-Grade Brain Tumour Research, the Friedberg Charitable Foundation and the Sohn Conference Foundation (to M. Snuderl and M. Karajannis), the RKA-Förderpool (Project 37) and Stichting Kinderen Kankervrij and Stichting AMC Foundation (to E. Aronica), NIH/NCI 5T32CA163185 (to A.O.), NIH/NCI Cancer Center Support Grant P30 CA008748 to MSKCC, the Luxembourg National Research Fund (FNR PEARL P16/BM/11192868 to M.M.) and the National Institute of Health Research (NIHR) UCLH/UCL Biomedical Research Centre (S.Bra.).

**Author Contributions** D.C. and D.T.W.J. composed the reference cohort and defined methylation classes; M.Si. and V.Ho. developed and technically validated the classification algorithm; D.Sc. developed the classification website; D.C., D.T.W.J., M.Si., A.Ben., V.Ho., D.Sc., D.Sti., M.Z., A.v.D. and S.M.P. developed additional methodology and software; D.C., D.T.W.J., M.Si., D.Stu., C.Ko., F.Sa., L.C., D.E.R., A.Kr., A.K.W., K.H., L.S., P.N.H., K.H.P., J.Schi., G.R., M.Pri., W.B., F.Se., H.W., T.M., O.W., S.Bre., M.S.-R., D.H., A.Ku., C.M.K., H.L.M., S.Ru., K.V.H., M.C.F., A.Gn., G.F., S.T., G.C., C.-M.M., M.G., T.P., M.B., J.D., M.PI., A.U., W.W., M.M., C.Har., C.H.-M., M.H., A.Kor., A.v.D. and S.M.P. performed the prospective cohort analysis; P.N.H., K.H.P., H.D., B.K.G., J.H., S.F., P.W., Z.J., T.A., S.Bra. generated and collected the external centre data; K.W.P., A.O., N.W.E., A.K.B., R.C., A.Hö., E.H., R.Be., J.Schi., O.S., K.W., P.V., M.Pa., P.T., D.L., E.A., F.G., E.R., W.S., C.G., F.J.R., A.Bec., M.Pre., C.Hab., R.Bj., J.C., M.F., M.D., S.H., V.Ha., S.Ro., J.R.H., P.K., B.W.K., M.L., B.L., C.M., R.K., Z.K., F.H., A.Koc., A.J., C.Ke., H.M., W.M., U.P., M.Pri., N.G.G., P.H.D., A.P., C.J., T.S.J., B.R., T.P., J.Schr., G.S., M.Wes., G.R., P.W., M.Wel., V.P.C., I.B., A.Hu., N.J., P.A.N., W.P., A.Ga., G.W.R., M.D.T., M.R., M.A.K., M.M., C.Har., K.A., U.S., R.Bu., P.L., M.K., C.H.-M., D.W.E., M.H., S.Bra., A.Kor., A.v.D. and S.M.P. provided reference cohort material and data; K.L., M.B.-H., M.Sc. and R.F. performed methylation profiling; J.Se., K.K., A.T., M.K. and M.Sn. performed technical validation experiments; A.v.D. and S.M.P. supervised the project. The manuscript underwent an internal collaboration-wide review process. All authors approved the final version of the manuscript.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to S.M.P. ([s.pfister@dkfz.de](mailto:s.pfister@dkfz.de)) or A.v.D. ([Andreas.vonDeimling@med.uni-heidelberg.de](mailto:Andreas.vonDeimling@med.uni-heidelberg.de)).

**Reviewer Information** *Nature* thanks S. Pomeroy, M. L. Suva, R. Verhaak and S. Yip for their contribution to the peer review of this work.

David Capper<sup>1,2,3,4\*</sup>, David T. W. Jones<sup>5,6,\*</sup>, Martin Sill<sup>5,6,7\*</sup>, Volker Hovestadt<sup>8,9,\*</sup>, Daniel Schrimpf<sup>1,2</sup>, Dominik Sturm<sup>5,6,9</sup>, Christian Koelsche<sup>1,2</sup>, Felix Sahm<sup>1,2</sup>, Lukas Chavez<sup>5,6</sup>, David E. Reuss<sup>1,2</sup>, Annekathrin Kratz<sup>1,2</sup>, Anniika K. Wefers<sup>1,2</sup>, Kristin Huang<sup>1,2</sup>, Kristian W. Pajtler<sup>5,6,9</sup>, Leonille Schweizer<sup>1,3</sup>, Damian Stichel<sup>1,2</sup>, Adriana Olar<sup>10,11,12</sup>, Nils W. Engel<sup>13,14</sup>, Kerstin Lindenbergs<sup>2</sup>, Patrick N. Harter<sup>15,16</sup>, Anne K. Braczynski<sup>15,16</sup>, Karl H. Plate<sup>15,16</sup>, Hildegard Dohmen<sup>17</sup>, Boyan K. Garvalov<sup>17</sup>, Roland Coras<sup>18</sup>, Annett Hölsken<sup>18</sup>, Ekkehard Hewer<sup>19</sup>, Melanie Bewerunge-Hudler<sup>20</sup>, Matthias Schick<sup>20</sup>, Roger Fischer<sup>20</sup>, Rudi Beschorner<sup>21</sup>, Jens Schittenhelm<sup>21</sup>, Ori Staszewski<sup>22</sup>, Khalida Wan<sup>23</sup>, Pascale Varlet<sup>24</sup>, Melanie Pages<sup>24</sup>, Petra Temming<sup>25</sup>, Dietmar Lohmann<sup>26</sup>, Florian Seif<sup>5,9,27</sup>, Hendrik Witt<sup>5,6,9</sup>, Till Milde<sup>5,9,27</sup>, Olaf Witt<sup>5,9,27</sup>, Eleonora Aronica<sup>28,29,30</sup>, Felice Giangaspero<sup>31,32</sup>, Elisabeth Rushing<sup>33</sup>, Wolfram Scheurlen<sup>34</sup>, Christoph Geisenberger<sup>35,36</sup>, Fausto J. Rodriguez<sup>37</sup>, Albert Becker<sup>38</sup>, Matthias Preusser<sup>39</sup>, Christine Haberle<sup>40</sup>, Rolf Bjerkvig<sup>41,42</sup>, Jane Cryan<sup>43</sup>, Michael Farrell<sup>43</sup>, Martina Deckert<sup>44</sup>, Jürgen Hench<sup>45</sup>, Stephan Frank<sup>45</sup>, Jonathan Serrano<sup>46</sup>, Kasthuri Kannan<sup>46</sup>, Aristotelis Tsirigos<sup>46</sup>, Wolfgang Brück<sup>47</sup>, Silvia Hofer<sup>48</sup>, Stefanie Brehmer<sup>49</sup>, Marcel Seiz-Rosenthalen<sup>49</sup>, Daniel Hägggi<sup>49</sup>, Volkmar Hans<sup>50,51</sup>, Stephanie Rozsnoki<sup>52</sup>, Jordan R. Hansford<sup>53,54,55</sup>, Patricia Kohlhof<sup>56</sup>, Bjarne W. Kristensen<sup>57</sup>, Matt Lechner<sup>58</sup>, Beatriz Lopez<sup>59</sup>, Christian Mawrin<sup>60</sup>, Ralf Ketter<sup>61</sup>, Andreas Kulozik<sup>5,9</sup>, Ziad Khatib<sup>62</sup>, Frank Heppner<sup>3,63,64</sup>, Arend Koch<sup>3</sup>, Anne Jouvet<sup>65</sup>, Catherine Keohane<sup>66</sup>, Helmut Mühlleisen<sup>67</sup>, Wolf Mueller<sup>68</sup>, Ute Pohl<sup>69</sup>, Marco Prinz<sup>22,70</sup>, Axel Benner<sup>7</sup>, Marc Zapata<sup>8</sup>, Nicholas G. Gottardo<sup>71,72,73</sup>, Pablo Hernández-Díaz<sup>74</sup>, Christof M. Kramm<sup>75</sup>, Hermann L. Müller<sup>76</sup>, Stefan Rutkowski<sup>77</sup>, Katja von Hoff<sup>74,77</sup>, Michael C. Fröhwald<sup>78</sup>, Astrid Gnekow<sup>78</sup>, Gudrun Fleischhack<sup>25</sup>, Stephan Tippelt<sup>25</sup>, Gabriele Calaminus<sup>79</sup>, Camelia-Maria Monoranu<sup>80</sup>, Arie Perry<sup>81</sup>, Chris Jones<sup>82</sup>, Thomas S. Jacques<sup>83</sup>, Bernhard Radlwimmer<sup>8</sup>, Marco Gessi<sup>38</sup>, Torsten Pietsch<sup>38</sup>, Johannes Schramm<sup>84</sup>, Gabriele Schackert<sup>85</sup>, Manfred Westphal<sup>86</sup>, Guido Reifenberger<sup>87,88</sup>, Pieter Wesseling<sup>89,90</sup>, Michael Weller<sup>91</sup>, Vincent Peter Collins<sup>92</sup>, Ingmar Blümcke<sup>18</sup>, Martin Bendszus<sup>93</sup>, Jürgen Debus<sup>94</sup>, Annie Huang<sup>95</sup>, Nada Jabado<sup>96</sup>, Paul A. Northcott<sup>97</sup>, Werner Paulus<sup>52</sup>, Amar Gajjar<sup>98</sup>, Giles W. Robinson<sup>98</sup>, Michael D. Taylor<sup>99</sup>, Zane Jaunmuktane<sup>100,101,102</sup>, Marina Ryzhova<sup>103</sup>, Michael Platten<sup>104</sup>, Andreas Unterberg<sup>35</sup>, Wolfgang Wick<sup>105</sup>, Matthias A. Karajannis<sup>106</sup>, Michel Mittelbronn<sup>15,16,107,108,109,110</sup>, Till Acker<sup>17</sup>, Christian Hartmann<sup>111</sup>, Kenneth Aldape<sup>112</sup>, Ulrich Schüller<sup>14,113,114,115</sup>, Rolf Buslei<sup>18,116</sup>, Peter Lichter<sup>8</sup>, Marcel Kooi<sup>5,6</sup>, Christel Herold-Mende<sup>35</sup>, David W. Ellison<sup>117</sup>, Martin Hasselblatt<sup>52</sup>, Matija Snuderl<sup>118</sup>, Sebastian Brandner<sup>100,102</sup>, Andrey Korshunov<sup>1,2</sup>, Andreas von Deimling<sup>1,2,§</sup> & Stefan M. Pfister<sup>5,6,9</sup>

<sup>1</sup>Department of Neuropathology, University Hospital Heidelberg, Heidelberg, Germany. <sup>2</sup>Clinical Cooperation Unit Neuropathology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>3</sup>Charité — Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Department of Neuropathology, Berlin, Germany. <sup>4</sup>German Cancer Consortium (DKTK), Partner Site Berlin, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>5</sup>Hopp Children's Cancer Center at the NCT Heidelberg (KITE), Heidelberg, Germany. <sup>6</sup>Division of Pediatric Neurooncology, German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>7</sup>Division of Biostatistics, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>8</sup>Division of Molecular Genetics, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>9</sup>Department of Pediatric Oncology, Hematology and Immunology, University Hospital Heidelberg, Heidelberg, Germany. <sup>10</sup>Department of Pathology and Laboratory Medicine, Medical University of South Carolina, Charleston, South Carolina 29425, USA. <sup>11</sup>Department of Neurosurgery, Medical University of South Carolina, Charleston, South Carolina 29425, USA. <sup>12</sup>Hollings Cancer Center, Charleston, South Carolina 29425, USA. <sup>13</sup>Department of Oncology and Hematology with Sections Bone Marrow Transplant and Pneumology, Hubertus Wald Tumorzentrum/University Cancer Center Hamburg, University Medical Center Hamburg, Hamburg, Germany. <sup>14</sup>Center for Neuropathology and Prion Research, Ludwig-Maximilians-University, Munich, Germany. <sup>15</sup>Institute of Neurology (Edinger Institute), Goethe-University Frankfurt am Main, Frankfurt am Main, Germany. <sup>16</sup>German Cancer Consortium (DKTK), Partner Site Frankfurt/Mainz, Frankfurt am Main, German Cancer Research Center (DKFZ) Heidelberg, Germany. <sup>17</sup>Institute of Neuropathology, University of Giessen, Giessen, Germany. <sup>18</sup>Neuropathological Institute, University Hospital Erlangen, Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany. <sup>19</sup>Institute of Pathology, University of Bern, Bern, Switzerland. <sup>20</sup>Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>21</sup>Institute of Pathology and Neuropathology, Department of Neuropathology, University Hospital Tübingen, Tübingen, Germany. <sup>22</sup>Institute of Neuropathology, Medical Center-University of Freiburg, Faculty of Medicine, University of Freiburg, Germany. <sup>23</sup>Department of Translational Molecular Pathology, University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. <sup>24</sup>Department of Neuropathology, Centre Hospitalier Sainte Anne, Paris, France. <sup>25</sup>Pediatrics III, Pediatric Oncology and Hematology, University Hospital Essen, Essen, Germany. <sup>26</sup>Eye Cancer Research Group, Faculty of Medicine, University of Duisburg-Essen, Essen, Germany. <sup>27</sup>CCU Pediatric Oncology (G340), German Cancer Research Center (DKFZ) and German Cancer Consortium (DKTK), Heidelberg, Germany. <sup>28</sup>Department of (Neuro) Pathology, Academic Medisch Centrum (AMC), University of Amsterdam, Amsterdam, The Netherlands. <sup>29</sup>Swammerdam Institute for Life Sciences, Center for Neuroscience, University of Amsterdam, Amsterdam, The Netherlands. <sup>30</sup>Stichting Epilepsie Instellingen Nederland (SEIN), Amsterdam, The Netherlands. <sup>31</sup>Department of Radiological, Oncological and Anatomopathological Sciences, Sapienza University, Rome, Italy. <sup>32</sup>IRCCS Neuromed, Pozzilli, Italy. <sup>33</sup>Department of Neuropathology, University Hospital Zurich, Zurich, Switzerland. <sup>34</sup>Cnopsf'sche

Kinderklinik Nuremberg, Nuremberg, Germany. <sup>35</sup>Department of Neurosurgery, Heidelberg University Hospital, Heidelberg, Germany. <sup>36</sup>Hubrecht Institute-KNAW (Royal Netherlands Academy of Arts and Sciences), Uppsalaalaan 8, 3584 CT Utrecht, The Netherlands. <sup>37</sup>Division of Neuropathology of the Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. <sup>38</sup>Department of Neuropathology, University of Bonn, Bonn, Germany. <sup>39</sup>Department of Medicine I, Comprehensive Cancer Center Vienna, CNS Unit (CCC-CNS), Medical University of Vienna, Vienna, Austria. <sup>40</sup>Institute of Neurology, Medical University of Vienna, Vienna, Austria. <sup>41</sup>Department of Biomedicine, University of Bergen, Bergen, Norway. <sup>42</sup>NORLUX Neuro-Oncology Laboratory, Department of Oncology, Luxembourg Institute of Health, Luxembourg, Luxembourg. <sup>43</sup>Department of Neuropathology, Beaumont Hospital, Dublin, Ireland. <sup>44</sup>Department of Neuropathology, University Hospital of Cologne, Cologne, Germany. <sup>45</sup>Department of Neuropathology, Institute of Pathology, Basel University Hospital, Basel, Switzerland. <sup>46</sup>NYU Langone Medical Center, New York, New York, USA. <sup>47</sup>Institute of Neuropathology, University Medical Center Göttingen, Göttingen, Germany. <sup>48</sup>Division of Oncology, Luzerner Kantonsspital, Luzern, Switzerland. <sup>49</sup>Department of Neurosurgery, University Medical Center Mannheim, University of Heidelberg, Mannheim, Germany. <sup>50</sup>Institut für Neuropathologie, Evangelisches Krankenhaus Bielefeld gGmbH, Bielefeld, Germany. <sup>51</sup>Institut für Neuropathologie, Universitätsklinikum Essen, Essen, Germany. <sup>52</sup>Institute of Neuropathology, University Hospital Münster, Münster, Germany. <sup>53</sup>Children's Cancer Centre, Royal Children's Hospital, University of Melbourne, Melbourne, Victoria, Australia. <sup>54</sup>Murdoch Children's Research Institute, University of Melbourne, Melbourne, Victoria, Australia. <sup>55</sup>Department of Pediatrics, University of Melbourne, Melbourne, Victoria, Australia. <sup>56</sup>Institute for Pathology, Katharinenhospital Stuttgart, Stuttgart, Germany. <sup>57</sup>Department of Pathology, Odense University Hospital, Department of Clinical Research, University of Southern Denmark, Odense, Denmark. <sup>58</sup>University College London Cancer Institute and University College London Hospitals, London, UK. <sup>59</sup>Department of Pathology, University of Virginia, Charlottesville, Virginia, USA. <sup>60</sup>Institute of Neuropathology, Otto-von-Guericke-University, Magdeburg, Germany. <sup>61</sup>Department of Neurosurgery, University Hospital Saarland, Homburg, Saar, Germany. <sup>62</sup>Nicklaus Children's Hospital Brain Institute, Miami, Florida 33155, USA. <sup>63</sup>Cluster of Excellence, NeuroCure, Berlin, Germany. <sup>64</sup>Berlin Institute of Health (BIH), Berlin, Germany. <sup>65</sup>Département de Pathologie et Neuropathologie, Hôpital Neurologique, Hôpices Civils de Lyon, Lyon, France. <sup>66</sup>Department of Neuropathology, Cork University Hospital, Cork, Ireland. <sup>67</sup>Department of Pathology, Ludwigsburg Hospital, Ludwigsburg, Germany. <sup>68</sup>Department of Neuropathology, Leipzig University, Leipzig, Germany. <sup>69</sup>Department of Cellular Pathology, Queen's Hospital, Romford, UK. <sup>70</sup>BIOS Centre for Biological Signalling Studies, University of Freiburg, Freiburg, Germany. <sup>71</sup>Department of Pediatric Oncology and Haematology, Princess Margaret Hospital for Children, GPO Box D184, Perth, Western Australia 6840, Australia. <sup>72</sup>Telethon Kids Institute, University of Western Australia, PO Box 855, Perth, Western Australia 6872, Australia. <sup>73</sup>School of Paediatrics and Child Health, University of Western Australia, GPO Box D184, Perth, Western Australia 6840, Australia. <sup>74</sup>Department of Pediatric Oncology/Hematology, Charité-Universitätsmedizin Berlin, Berlin, Germany. <sup>75</sup>Division of Pediatric Hematology and Oncology, University Medical Center Göttingen, Göttingen, Germany. <sup>76</sup>Department of Pediatrics and Pediatric Hematology/Oncology, Klinikum Oldenburg AöR, Medical Campus University Oldenburg, 26133 Oldenburg, Germany. <sup>77</sup>Department for Pediatric Hematology and Oncology, University Hospital Hamburg-Eppendorf, Hamburg, Germany. <sup>78</sup>Children's Hospital Augsburg, Swabian Children's Cancer Centre, Augsburg, Germany. <sup>79</sup>Department of Pediatric Hematology/Oncology, University of Bonn Medical Center, Bonn, Germany. <sup>80</sup>Department of Neuropathology, Institute of Pathology, Comprehensive Cancer Center (CCC) Mainfranken, University of Würzburg, Würzburg, Germany. <sup>81</sup>Department of Pathology, University of California San Francisco, San Francisco, California, USA. <sup>82</sup>Division of Molecular Pathology, Institute of Cancer Research, London, UK. <sup>83</sup>Developmental Biology and Cancer Programme, UCL Great Ormond Street Institute of Child Health and Histopathology Department, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK. <sup>84</sup>Medical Faculty, University of Bonn Medical School, Bonn, Germany. <sup>85</sup>Department of Neurosurgery, University Hospital Carl Gustav Carus, Technische Universität Dresden, Dresden, Germany. <sup>86</sup>Department of Neurosurgery, University Hospital Hamburg-Eppendorf, Hamburg, Germany. <sup>87</sup>Department of Neuropathology, Heinrich Heine University Düsseldorf, Düsseldorf, Germany. <sup>88</sup>German Cancer Consortium (DKTK), partner site Essen/Düsseldorf, German Cancer Research Center (DKFZ), Heidelberg, Germany. <sup>89</sup>Department of Pathology, Princess Máxima Center for Pediatric Oncology and University Medical Center Utrecht, Utrecht, The Netherlands. <sup>90</sup>Department of Pathology, VU University Medical Center, Amsterdam, The Netherlands. <sup>91</sup>Department of Neurology, University Hospital and University of Zurich, Zurich, Switzerland. <sup>92</sup>Department of Pathology, Division of Molecular Histopathology, University of Cambridge, Cambridge, UK. <sup>93</sup>Department of Neuroradiology, Heidelberg University Hospital, Heidelberg, Germany. <sup>94</sup>Department of Radiation Oncology, Heidelberg University Hospital, Heidelberg, Germany. <sup>95</sup>Department of Pediatrics, Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario, Canada. <sup>96</sup>Division of Hematology/Oncology, McGill University, Montreal, Quebec, Canada. <sup>97</sup>Department of Developmental Neurobiology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>98</sup>Department of Oncology, St Jude Children's Research Hospital, Memphis, Tennessee, USA. <sup>99</sup>Division of Neurosurgery, Arthur and Sonia Labatt Brain Tumor Research Centre, Hospital for Sick Children, University of Toronto, Toronto, Ontario, Canada. <sup>100</sup>Division of Neuropathology, UCL Hospitals, Institute of Neurology, University College London, Queen Square, WC1N 3BG London, UK. <sup>101</sup>Department of Molecular Neuroscience, Institute of Neurology, University College London, Queen Square, WC1N 3BG London, UK. <sup>102</sup>Department of Neurodegeneration, Institute of Neurology, University College London, Queen Square, WC1N 3BG London, UK. <sup>103</sup>NN Burdenko Neurosurgical Institute, Moscow, Russia. <sup>104</sup>Department of Neurology, Universitätsmedizin Mannheim, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany. <sup>105</sup>Department of Neurology, Heidelberg University Hospital, Heidelberg, Germany. <sup>106</sup>Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. <sup>107</sup>NORLUX Neuro-Oncology Laboratory, Luxembourg Institute of Health (LIH), Luxembourg, Luxembourg. <sup>108</sup>Luxembourg Centre for Systems Biomedicine (LCSB), University

of Luxembourg, Luxembourg, Luxembourg.<sup>109</sup> Laboratoire national de santé (LNS), Dudelange, Luxembourg.<sup>110</sup> Luxembourg Centre of Neuropathology (LCNP), Luxembourg, Luxembourg.<sup>111</sup> Department of Neuropathology, Hannover Medical School (MHH), Hannover, Germany.<sup>112</sup> Department of Pathology, University of Toronto, Toronto, Ontario, Canada.<sup>113</sup> Institute of Neuropathology, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany.<sup>114</sup> Research Institute Children's Cancer Center, Hamburg, Germany.<sup>115</sup> Department of Pediatric Hematology and Oncology, University Medical Center, Hamburg-Eppendorf, Hamburg, Germany.<sup>116</sup> Section Neuropathology, Institute of Pathology, Sozialstiftung Bamberg, Klinikum

am Bruderwald, Bamberg, Germany.<sup>117</sup> Department of Pathology, St Jude Children's Research Hospital, Memphis, Tennessee, USA.<sup>118</sup> Division of Neuropathology, Department of Pathology, NYU Langone Medical Center, New York, New York, USA.

†Present addresses: Department of Pathology and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts, USA (V.Ho.); Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA (V.Ho.).

\*These authors contributed equally to this work.

§These authors jointly supervised this work.

## METHODS

**Data reporting.** No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

**Patient material.** Patient material and clinical data of the retrospective reference cohort (total,  $n = 2,801$ ) were obtained from the National Center for Tumour Diseases (NCT) in Heidelberg and supplemented with samples from additional centres (Supplementary Table 2) according to protocols approved by the Institutional Review Boards of the appropriate centres with written consent obtained from each patient. Tumours were histopathologically re-assessed according to the current WHO classification<sup>1</sup>. Areas with highest tumour cell content ( $\geq 70\%$ ) were selected for DNA extraction. Subsets of the reference cohort have been previously published<sup>4,6,9–16,26–33</sup>. Additional patient characteristics can be found in Supplementary Table 2. The prospectively assessed clinical cohort was analysed as part of the National Center for Tumour Diseases Precision Oncology Program according to procedures approved by the Institutional Review Board at the Medical Faculty Heidelberg. All patients gave written consent for diagnostic procedures, comprising molecular testing, including methylation profiling. For all the above human research participants, all relevant ethical regulations were followed. Additional patient characteristics can be found in Supplementary Table 4. Details of the online-analysed cohort of the five additional centres can be found in Supplementary Table 6. Usage of the data was according to protocols approved by the Institutional Review Boards of the University of Basel, Frankfurt am Main University Hospital, University Medical Center Utrecht and Princess Máxima Center for Paediatric Oncology Utrecht, Giessen University Hospital and University College London Hospitals.

**Data generation, processing and random forest classifier generation.** Samples were analysed using Illumina Infinium HumanMethylation450 BeadChip (450k) arrays according to the manufacturer's instructions. To investigate stability across platforms a selection of samples were additionally assessed using the successor Methylation BeadChip (EPIC) array or whole-genome bisulphite sequencing (WGBS, generated and analysed as described<sup>6</sup>). Array data analysis was performed using R v.3.2.0<sup>34</sup>, using a number of packages from Bioconductor<sup>35</sup> and other repositories. A random forest<sup>19</sup> classifier compatible with both 450k and EPIC platforms was trained, and a calibration model that calculates class probabilities from random forest scores was devised.

**Methylation array processing.** The 450k array was used to obtain genome-wide DNA methylation profiles for tumour samples and normal control tissues, according to the manufacturer's instructions (Illumina). DNA methylation data were generated at the Genomics and Proteomics Core Facility of the DKFZ (Heidelberg, Germany) and the NYU Langone Medical Center (New York, USA). Data were generated from both freshly frozen and formalin-fixed paraffin-embedded (FFPE) tissue samples. For most freshly frozen samples,  $>500$  ng of DNA was used as input material, whereas 250 ng of DNA was used for most FFPE tissues. On-chip quality metrics of all samples were carefully controlled. Copy-number variation analysis from 450k methylation array data was performed using the conumee<sup>24</sup> Bioconductor package v.1.3.0. Two sets of 50 control samples displaying a balanced copy-number profile from both male and female donors were used for normalization.

Raw signal intensities were obtained from IDAT files using the minfi Bioconductor package v.1.14.0<sup>36</sup>. Each sample was individually normalized by performing a background correction (shifting of the 5% percentile of negative control probe intensities to 0) and a dye-bias correction (scaling of the mean of normalization control probe intensities to 10,000) for both colour channels. Subsequently, a correction for the type of material tissue (FFPE or frozen) was performed by fitting univariate, linear models to the log<sub>2</sub>-transformed intensity values (removeBatchEffect function, limma package v.3.24.15). The methylated and unmethylated signals were corrected individually. Estimated batch effects were also used to adjust diagnostic samples or test samples within the cross-validation. Beta values were calculated from the retransformed intensities using an offset of 100 (as recommended by Illumina). To test for possible confounding batch effects within our pre-processed reference cohort dataset (after adjusting for FFPE versus frozen material) we applied the sva algorithm<sup>37,38</sup>. We found no significant surrogate variable (data not shown).

The following filtering criteria were applied: removal of probes targeting the X and Y chromosomes ( $n = 11,551$ ), removal of probes containing a single-nucleotide polymorphism (dbSNP132 Common) within five base pairs of and including the targeted CpG site ( $n = 7,998$ ), probes not mapping uniquely to the human reference genome (hg19) allowing for one mismatch ( $n = 3,965$ ), and probes not included on the Illumina EPIC array ( $n = 32,260$ ). In total, 428,799 probes targeting CpG sites were kept for further analysis.

**Unsupervised analysis.** Pairwise Pearson correlation was calculated for all 2,801 reference samples by selecting the 32,000 most variably methylated probes

(s.d.  $> 0.228$ ; Extended Data Fig. 1a). The same probes were used for principal component analysis (PCA). For PCA, pairwise probe covariances of centred beta values were calculated. Eigenvalue decomposition was performed using the eigs function of the RSpectra package version 0.12. The number of non-trivial components was determined by comparing eigenvalues to the maximum eigenvalue of a PCA using randomized beta values (shuffling of sample labels per probe) (Extended Data Fig. 1b). Principal component scores for all non-trivial components ( $n = 94$ ) were used for t-SNE analysis<sup>17</sup> (Rtsne package v.0.11, Fig. 1b). The following non-default parameters were used: theta = 0, pca = F, max\_iter = 2500. A similar approach was used for the combined analysis of reference and diagnostic cases (Fig. 5a).

**The random forest algorithm.** The random forest<sup>19</sup> algorithm is a so-called ensemble method that combines the predictions of several weak classifiers to achieve improved prediction accuracy. The random forest algorithm uses binary decision trees (classification and regression trees, CART<sup>39</sup>) as weak classifiers (Extended Data Fig. 4). Each of these trees is a sequence of binary splitting rules that are learned by recursive binary splitting. The CART algorithm starts with all samples assigned to a 'root' node and tries to find the variable, for example, a measured CpG probe, and a corresponding cutoff that results in the purest split into the different classes. To measure this gain in class 'purity', the Gini index is used. To fit a tree, the CART algorithm iteratively repeats these steps until no further improvements can be made. To predict the class of a new diagnostic case, the binary splitting rules are compared with the new data starting in the root node down to one of the leaf nodes. The tree then predicts or votes for the class of that leaf node. Decision trees have the advantage that they are non-parametric and do not rely on any distributional assumptions. The main disadvantages of decision trees are that they often tend to overfit the data and that they have a weak prediction performance. To improve the prediction accuracy, the random forest algorithm combines thousands of trees by bootstrap aggregation (bagging). In brief, each tree is fitted using training datasets that are generated by drawing bootstrap samples. In addition, at each node only a random subset of the available variables is used to find an optimal splitting rule. This additional source of randomization allows selecting variables with lower predictive value. This feature guarantees that the resulting trees are decorrelated, that is, they use different variables to find an optimal prediction rule. Taking the majority vote over thousands of bootstrap-aggregated and decorrelated trees greatly improves the prediction accuracy of the random forest.

**Classifier development.** To train the random forest classifier, the randomForest R package<sup>40</sup> was used. First, the most important features (probes) were selected by applying the random forest algorithm to the beta values of all filtered 428,799 probes. For efficient computation, the probes were split into 43 sets of approximately 10,000 probes. For each set, 100 trees were fitted using 654 randomly sampled candidate features at each split (mtry parameter, square root of 428,799, as would be used by default when not splitting into sets). To take the imbalanced methylation class sizes into account, a down-sampling strategy was followed that ensures an identical number of samples per class (parameter sampsize = rep(8, 91)), eight reflecting the minimum number of cases in the 91 classes<sup>41</sup>. For all other parameters the default settings were used. This procedure was repeated 100 times, essentially fitting 10,000 trees per probe. Finally, features are selected by the permutation-based variable importance measure as implemented in the randomForest R package<sup>40</sup>. The importance measure is the class-specific mean decrease in classification accuracy when the feature is permuted. We selected features by ranking them using the minimal rank of the variable importance measures across all classes.

The final random forest classifier was trained by fitting 10,000 trees with the parameter mtry = 100 using beta values of the 10,000 probes selected during feature selection. Imbalanced class sizes were accounted for by down-sampling (as described above), and for all other parameters the default settings were used. An overview of the processes is given in Extended Data Fig. 4.

**Classifier cross-validation.** Overfitting of the training data is a typical problem expected when training classifiers on high-dimensional data. As it often cannot be avoided, the typical strategy to deal with this problem is to evaluate the model accuracy on an independent test dataset or apply cross-validation methods<sup>42</sup>. Because some of the newly defined methylation groups presented in this work cannot be diagnosed by classical histopathological methods or other established molecular assays, an independent test set to assess model accuracy is not available. Therefore, the accuracy of the presented random forest model with the accompanying calibration model was evaluated by a threefold, nested cross-validation. For this, the reference dataset was split into three equally sized parts. In each cross-validation iteration, two-thirds of the data were used to train a random forest classifier in the same way as the random forest classifier for the complete dataset was trained. Then, the remaining one-third of the data were used for predictions using this random forest classifier. After the third iteration of the cross-validation is completed, each

of the 2,801 reference samples has been predicted by an independent random forest classifier, that is, where the sample was not used for estimating batch effects, performing variable selection, or training of the classifier.

**Classifier score calibration.** The classification scores generated by our multiclass random forest model (that is, the proportion of trees voting for a class) performed well when they were used to assign the correct class labels, but they do not reflect well calibrated class probabilities<sup>43</sup>. Furthermore, the distribution of the random forest scores varied between classes, which made inter-class comparisons difficult. Moreover, to evaluate a diagnostic classification, the uncertainties associated with an individual prediction in terms of confidence scores or estimated class probabilities are needed.

To obtain scores that are comparable between classes and that are improved estimates of the certainty of individual predictions, we performed a classification score recalibration by mapping the original scores to more accurate class probabilities<sup>44,45</sup>. To find such a mapping, an L2-penalized, multinomial, logistic regression model was fitted, which takes the methylation class as response variable and the random forest scores as explanatory variables. The R package *glmnet*<sup>46</sup> was used to fit this model. In addition, the model was fitted by incorporating a small ridge-penalty (L2) on the likelihood to prevent overfitting, as well as to stabilize estimation in situations in which classes are perfectly separable. The amount of this regularization, that is, the penalization parameter, is determined by running a tenfold cross-validation and choosing the largest value that lies within one standard error of the minimum cross-validation error. Independent random forest scores are needed to fit this model, that is, the scores need to be generated by a random forest classifier that was not trained using the same samples, otherwise the random forest scores will be systematically biased and not comparable to scores of unseen cases. As such, random forest scores generated by the threefold cross-validation are used.

To validate the class predictions generated by using the recalibrated scores of the calibration model, a nested threefold cross-validation loop is incorporated into the main threefold cross-validation that validated the random forest classifier (Extended Data Fig. 4). Within each cross-validation run this nested threefold cross-validation is applied to generate independent random forest scores, which are then used to train a calibration model. The predicted random forest scores resulting from predicting the one-third test data of the outer cross-validation loop are then recalibrated by applying the calibration model that was fitted on the random forest scores generated during the nested cross-validation. A similar cross-validation scheme was used previously<sup>47</sup> to validate estimated classification probabilities.

**Classifier performance measures.** Performances of the resulting classifier predictions and scores generated by the cross-validation were assessed by the misclassification error, multiclass area under receiver operating characteristic (ROC) curve (AUC) and the multiclass Brier score. The misclassification error measures the frequency of falsely assigned class labels when using the maximum of the random forest scores or re-calibrated scores as a cutoff to determine the predicted class, that is, the majority vote. To measure the AUC for our multiclass random forest the generalization of the AUC for multiclass classification problems<sup>48</sup> was used. To measure how well the resulting random forest scores and recalibrated scores perform when used as class probabilities, the multiclass Brier score<sup>42,43,49</sup> was used. The Brier score is the mean-squared difference between the actual and the predicted class probability and thus measures the same characteristic as the mean squared error measures for a continuous forecast.

**Methylation class families.** We observed that the majority of misclassification errors occurred within eight groups of histologically and biologically closely related tumour classes. We therefore defined eight methylation class families (MCF). Since calibrated scores represent class probabilities, it is possible to apply the addition rule of well calibrated class probability estimates<sup>43</sup> within one MCF to get a class probability for the MCF.

**Threshold analysis.** Finding an optimal cutoff for diagnostic tests usually involves finding an optimal trade-off between sensitivity and specificity. If there are no preferences regarding specificity or sensitivity, the optimal cutoff is chosen by the upper left corner of the ROC curve or by maximizing the Youden index (specificity + sensitivity – 1). In an application like the one described here, where the cost of false negative is that a tumour cannot be classified and the cost of a false positive is a falsely predicted methylation class, a threshold with high specificity is preferred. ROC analysis is typically defined for binary classification problems. Finding a threshold for multiclass classifiers either involves performing a ROC analysis for each class resulting in class-wise individual thresholds or finding some common threshold for all classes.

The calibrated MC/MCF scores (here referring to MCF and methylation classes (MC) that are not assigned to a MCF) are already validated probability estimates for the methylation class with a direct interpretation, that is, we expect among all samples with scores of approximately 0.9 that 10% are falsely predicted. Applying an additional threshold is not required from a statistical point of view, but is desired in clinical practice. In addition, owing to calibration, scores are comparable across

classes and it is thus reasonable to define a common threshold for all classes instead of finding the optimal cutoff for each individual methylation class.

To determine a common threshold for the calibrated MC/MCF scores, we performed a ROC analysis of the maximum calibrated MC/MCF scores calculated via cross-validation. For this ROC analysis, we defined a new binary class, that is, samples correctly classified during the cross-validation using the maximum calibrated MC/MCF score for classification were considered as ‘classifiable’ and samples falsely classified using this score were considered ‘non-classifiable’.

Following this ROC analysis approach, we determined a cutoff of 0.836 that maximizes the Youden index with a specificity of 93.8% and sensitivity of 93.4% (Extended Data Fig. 5d, e). A maximum specificity of 100% with a sensitivity of 82.7% can be achieved with a threshold of 0.958. Bootstrapped 95% confidence intervals (grey area in Extended Data Fig. 5d) demonstrate the uncertainty of sensitivity and specificity estimates, especially in the upper left corner of the ROC figure, where the considered thresholds are located.

Both thresholds have been determined by cross-validation on our training data of high quality, but real life diagnostic samples were found to achieve slightly lower scores, due to a number of factors that we cannot control, such as lower overall sample quality and lower tumour purity compared to the samples in our reference cohort. Therefore, we decided to lower the maximum specificity threshold to allow a wider spectrum of samples to become a match. For this, we chose a threshold of  $\geq 0.9$ , which lies in the middle between the Youden index and the threshold for maximum specificity.

**Comparison to TCGA pan-glioma methylation classes.** To compare our methylation-based classification of CNS tumours with described methylation classes of brain tumours by The Cancer Genome Atlas (TCGA) project, we downloaded the pre-processed methylation dataset that has previously been published<sup>18</sup>, including methylation data of 418 low-grade glioma and 377 glioblastoma samples, which had been analysed using the Illumina 450k array or 27k array platforms. To classify our samples according to the TCGA pan-glioma DNA methylation classification, we trained a random forest classifier on this dataset using the 1,300 CpG probe signature provided by the authors and using the default settings of the random forest algorithms implemented in the R package *randomForest*. The results of this classification for astrocytomas, oligodendrogiomas and glioblastomas are shown in Extended Data Fig. 3d and are provided on a case-by-case basis in Supplementary Tables 2, 4.

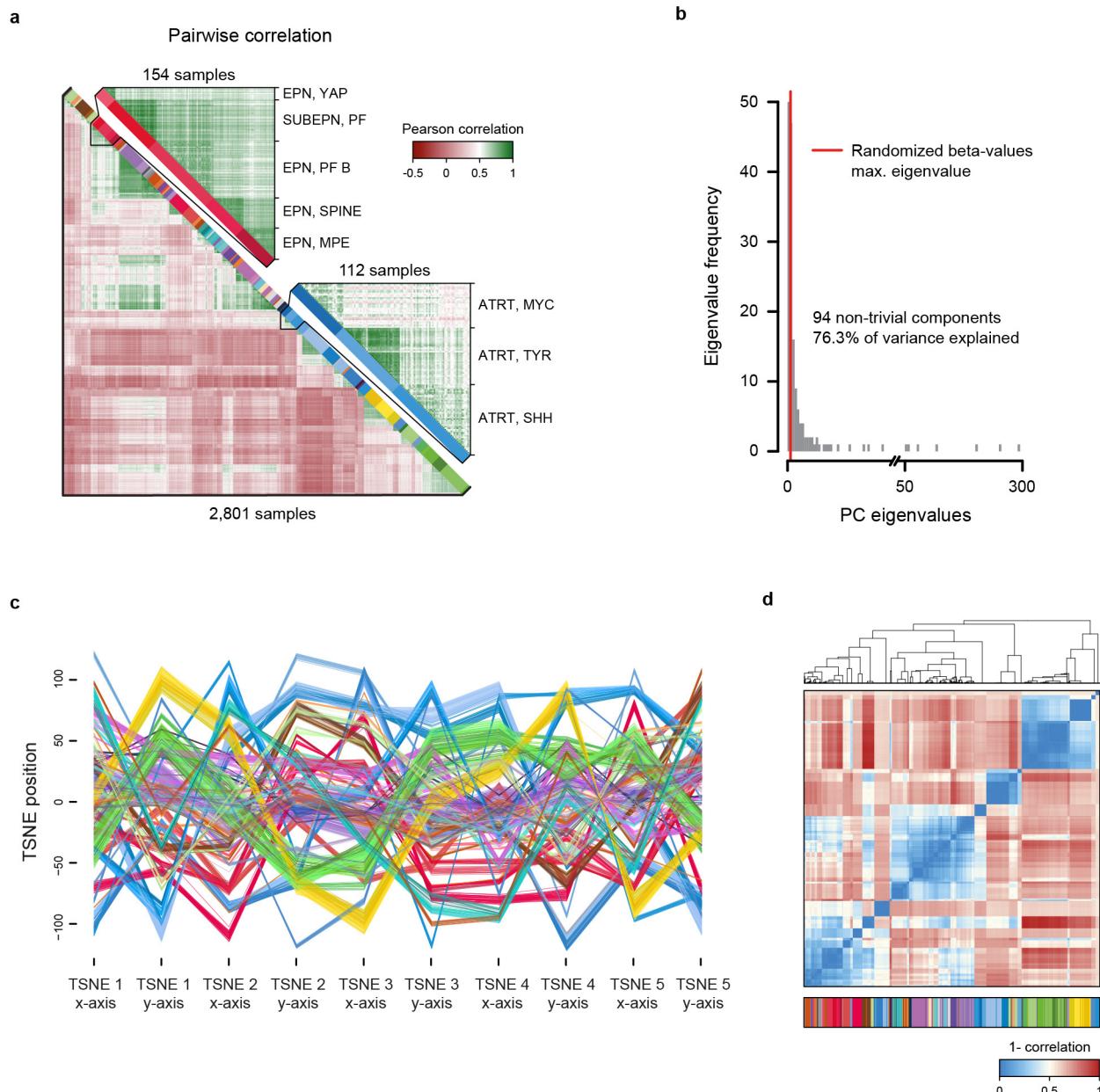
**Estimating tumour purity from DNA methylation data.** Because of the subjective nature of histological assessment of tumour purity, we additionally used the previously published dataset<sup>18</sup> to train a random forest regression (continuous response variable) model to predict tumour purity<sup>51</sup>. This random forest was trained on the 1,000 most important CpG probes for purity estimation selected also by a random forest (similar to the variable selection described for the random forest classifier). The out-of-bag (that is, random forest trees in which the respective sample, for which purity is predicted, was not used for training) mean squared error of the final model was 0.015, indicating that this model was able to yield reasonable predictions of tumour purity from methylation data (Extended Data Fig. 3a–c). The estimated tumour purities for individual cases are given in Supplementary Tables 2, 4.

**Code availability.** The generated code is available from the corresponding authors upon reasonable request for non-commercial use.

**Data availability.** The complete methylation values required for the construction of the classifier (reference set) as well as the prospective cohort (validation set) have been deposited in GEO (GSE109381). Supplementary Tables 2 (reference cohort) and 4 (prospective validation cohort) include the IDAT-file names for assignment to patient characteristics. Source Data for Figs 1b, 2b, 3b, 4, 5a, c and Extended Data Figs 1c, 2a–f, 3a–d, 5a, b, d, e, 6 and 7a, c, d are provided with the online version of the paper.

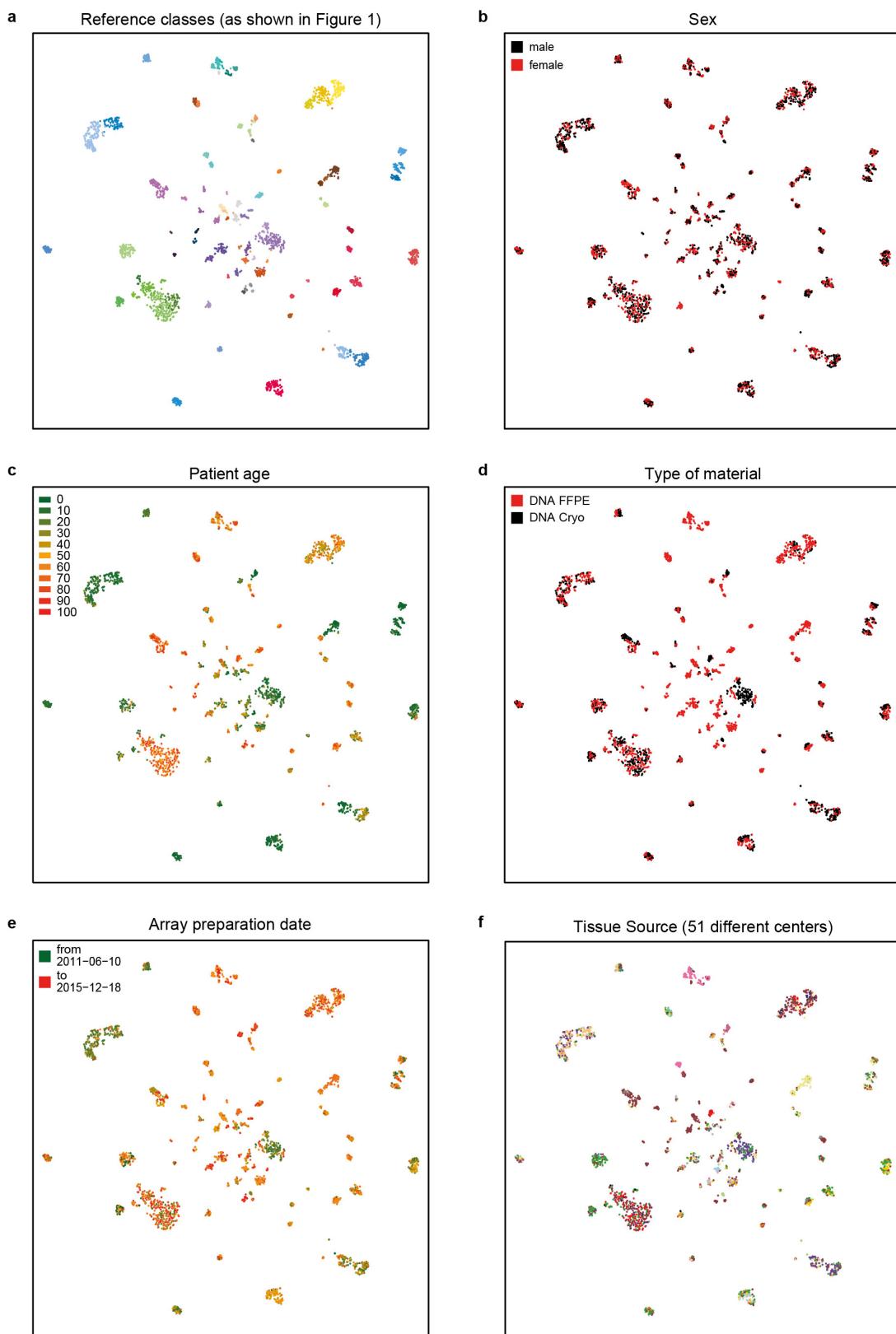
26. Korshunov, A. *et al.* Histologically distinct neuroepithelial tumors with histone H3 G34 mutation are molecularly similar and comprise a single nosologic entity. *Acta Neuropathol.* **131**, 137–146 (2016).
27. Korshunov, A. *et al.* Embryonal tumor with abundant neuropil and true rosettes (ETANTR), ependymoblastoma, and medulloepithelioma share molecular similarity and comprise a single clinicopathological entity. *Acta Neuropathol.* **128**, 279–289 (2014).
28. Hölsken, A. *et al.* Adamantinomatous and papillary craniopharyngiomas are characterized by distinct epigenomic as well as mutational and transcriptomic profiles. *Acta Neuropathol. Commun.* **4**, 20 (2016).
29. Heim, S. *et al.* Papillary Tumor of the pineal region: a distinct molecular entity. *Brain Pathol.* **26**, 199–205 (2016).
30. Koelsche, C. *et al.* Melanotic tumors of the nervous system are characterized by distinct mutational, chromosomal and epigenomic profiles. *Brain Pathol.* **25**, 202–208 (2015).
31. Jones, D. T. *et al.* Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat. Genet.* **45**, 927–932 (2013).
32. Jones, D. T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488**, 100–105 (2012).

33. Pietsch, T. et al. Prognostic significance of clinical, histopathological, and molecular characteristics of medulloblastomas in the prospective HIT2000 multicenter clinical trial cohort. *Acta Neuropathol.* **128**, 137–149 (2014).
34. R Core Team. R: A Language and Environment for Statistical Computing. <http://www.R-project.org/> (R Foundation for Statistical Computing, Vienna, Austria, 2016).
35. Huber, W. et al. Orchestrating high-throughput genomic analysis with Bioconductor. *Nat. Methods* **12**, 115–121 (2015).
36. Aryee, M. J. et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369 (2014).
37. Leek, J. T. & Storey, J. D. Capturing heterogeneity in gene expression studies by surrogate variable analysis. *PLoS Genet.* **3**, e161 (2007).
38. Leek, J. T. & Storey, J. D. A general framework for multiple testing dependence. *Proc. Natl Acad. Sci. USA* **105**, 18718–18723 (2008).
39. Breiman, L. *Classification and Regression Trees* (Chapman & Hall/CRC, 1984).
40. Liaw, A. & Wiener, M. Classification and Regression by randomForest. *R News* **2/3**, 18–22 (2002).
41. Chen, C., Liaw, A. & Breiman, L. *Using Random Forest to Learn Imbalanced Data*. Report 666 (Univ. California, Berkeley, 2004).
42. Kim, K. I. & Simon, R. Overfitting, generalization, and MSE in class probability estimation with high-dimensional data. *Biom. J.* **56**, 256–269 (2014).
43. Simon, R. Class probability estimation for medical studies. *Biom. J.* **56**, 597–600 (2014).
44. Bühlmann, P. & Yu, B. Calibrating Random Forests. In *Proc. 7th International Conference on Machine Learning and Applications* 121–126 (ICMLA, 2008).
45. Smola, A. J. *Advances in Large Margin Classifiers* (MIT press, 2000).
46. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
47. Appel, I. J., Gronwald, W. & Spang, R. Estimating classification probabilities in high-dimensional diagnostic studies. *Bioinformatics* **27**, 2563–2570 (2011).
48. Hand, D. J. & Till, R. J. A simple generalisation of the area under the ROC curve for multiple class classification problems. *Mach. Learn.* **45**, 171–186 (2001).
49. Brier, G. W. Verification of forecasts expressed in terms of probability. *Mon. Weath. Rev.* **78**, 1–3 (1950).
50. Carter, S. L. et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* **30**, 413–421 (2012).



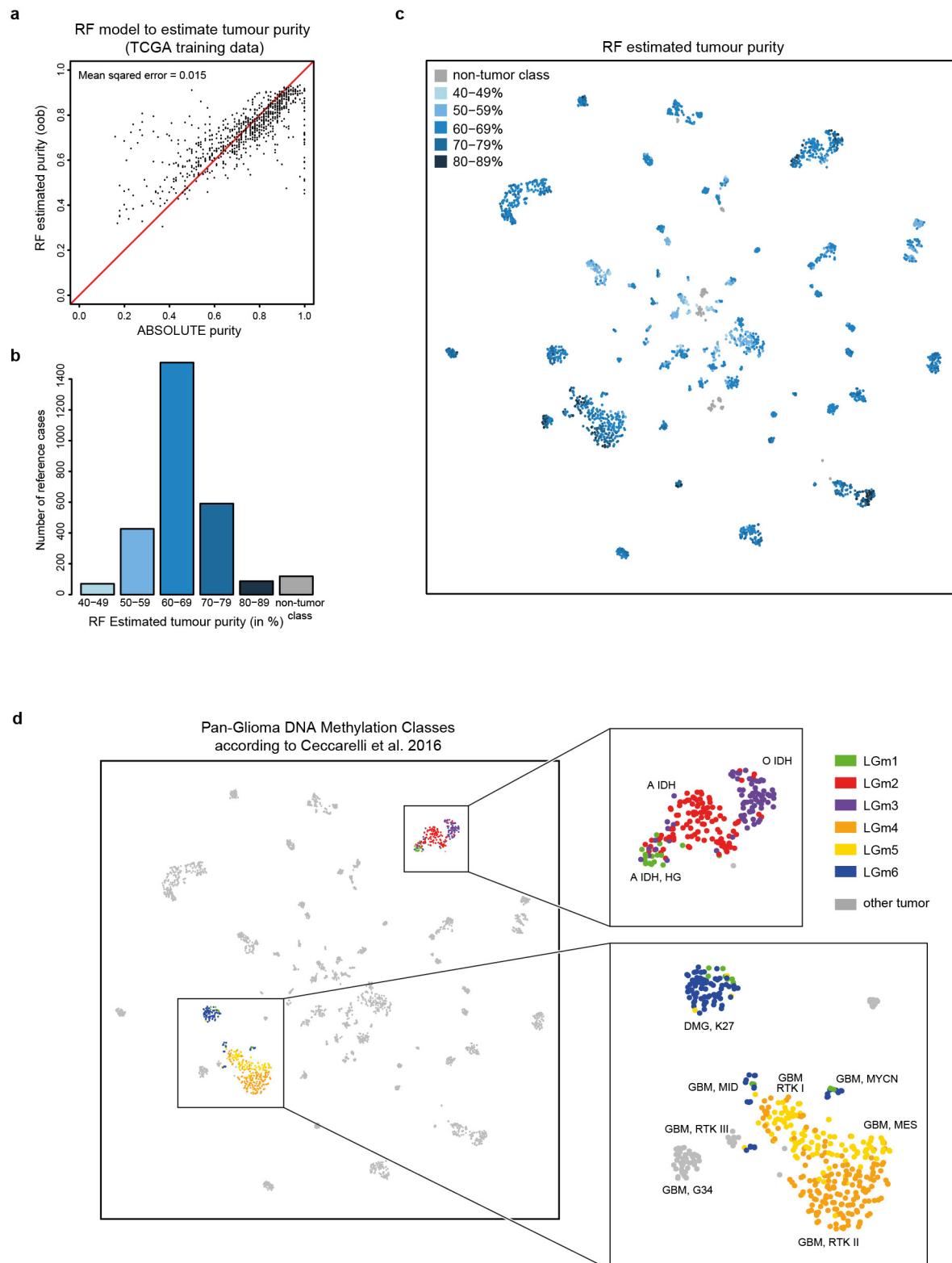
**Extended Data Figure 1 | Unsupervised clustering of the DNA methylation-based reference cohort.** **a**, Heat map showing the pairwise Pearson correlation (bottom left) of the 32,000 most variably methylated CpG probes of all 2,801 biologically independent samples of the reference cohort. A detailed view of closely related ependymal classes (top right) and the three subclasses identified in atypical teratoid rhabdoid tumours (ATRTs) (bottom right) indicates higher correlation within classes. The colour code and abbreviations are identical to Fig. 1a. **b**, Eigenvalue frequencies of a PCA using the 32,000 most variably methylated CpG probes of all 2,801 biologically independent samples as in a. The number of non-trivial components was determined by comparing eigenvalues to the maximum eigenvalue of a PCA using randomized beta values

(shuffling of sample labels per probe). **c**, x and y coordinates of the first five of a total of 500 iterations of t-SNE dimensionality reduction generated by random down-sampling to 90% of the 2,801 biologically independent samples to assess clustering stability. Axis positions of individual cases are connected by a line coloured according to the colour code of Fig. 1a. The depiction illustrates the close proximity of cases of the same class across iterations, indicative of a high stability independent of the exact composition of the reference cohort. **d**, Pairwise correlation of x and y coordinates between 2,801 biologically independent samples over all iterations of the down-sampling analysis demonstrates a very high correlation within classes (average correlation 0.982), indicating a high stability of the t-SNE analysis.



**Extended Data Figure 2 | Unsupervised clustering is not biased by a range of possible confounding factors.** **a**, *t*-SNE representations of the 2,801 biologically independent samples constituting the reference cohort as shown in Fig. 1b overlaid with potentially confounding factors (**b–f**). **b**, Distribution of patient sex among the classes illustrates equal or near equal distribution of many classes, but also an expected enrichment for one sex in some classes (for example, female in meningioma or CNS

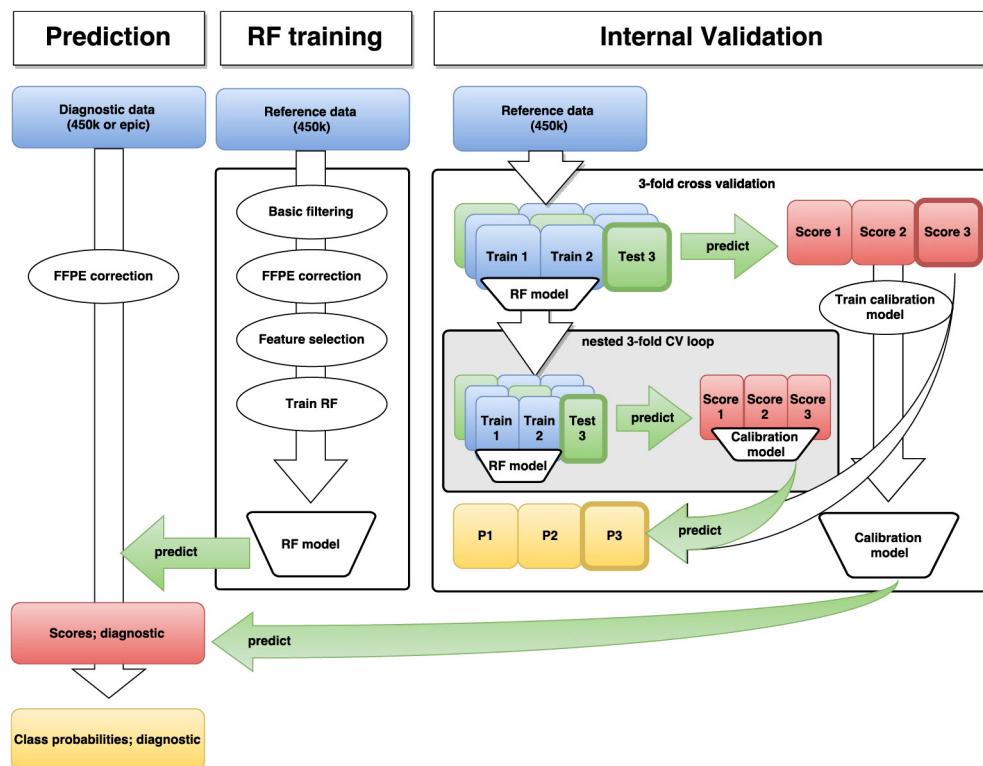
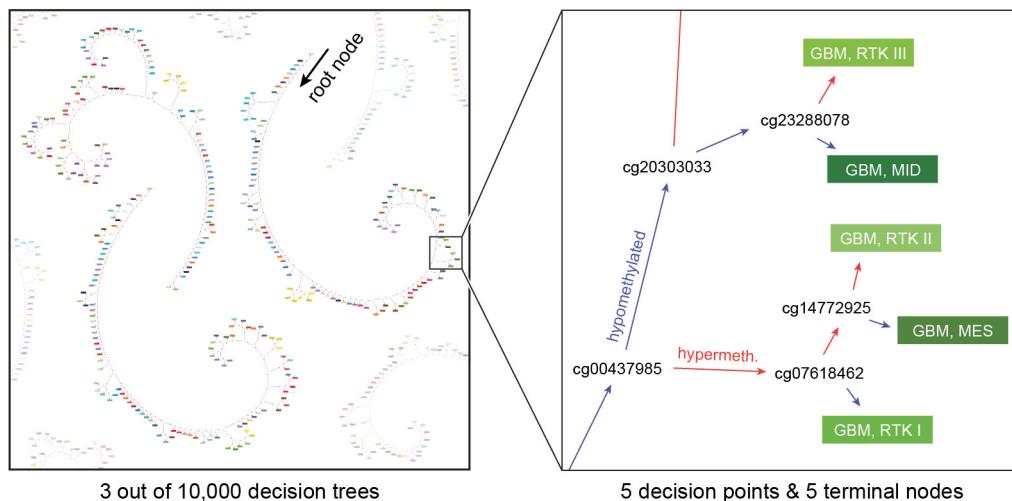
high-grade neuroepithelial tumours with MN1 alteration). **c**, Patient age illustrates the expected age distribution of many tumour classes. **d–f**, The slightly uneven distribution of type of material (for example, pilocytic astrocytoma or meningioma) (**d**), array preparation date (**e**) and tissue source (**f**) are related to the specifics of assembling the reference cohort and do not indicate an apparent confounding effect on the unsupervised clustering.



Extended Data Figure 3 | See next page for caption.

**Extended Data Figure 3 | Estimation of tumour purity and relation to TCGA pan-glioma methylation classes.** **a**, A random forest model was trained to predict absolute tumour purity estimates<sup>50</sup> using the TCGA pan-glioma dataset (795 biologically independent samples)<sup>18</sup>. The plot shows absolute purity estimates and out-of-bag random forest tumour purity predictions (that is, using only random forest trees for which the respective sample was not involved in the training). The estimated mean squared error is 0.015, indicating that this model is able to yield reasonable predictions of tumour purity from methylation data. **b**, The distribution of random forest predicted purity in the reference dataset (2,801 biologically independent samples). Purity estimates have been transformed into five categories indicated by different shades of blue. The exact case-by-case values are provided in Supplementary Table 2. The median estimated purity in the reference cohort is 66% (range 42% to 87%) and 78% of samples have an estimated purity of at least 60%. **c**, *t*-SNE representation of the reference cohort (2,801 biologically independent samples) overlaid with random forest predicted purity categories. Methylation classes are generally composed of mixed tumour purity categories. Tumour purity shows some association with the WHO grade (WHO I median tumour purity 60%, range 39–77%; WHO II median 66%, range 43–80%; WHO

III median 68%, range 54–84%; WHO IV median 69%, range 49–87%). A further association of tumour purity with the composition of classes in the unsupervised *t*-SNE analysis was not evident. **d**, *t*-SNE representation of the reference cohort (2,801 biologically independent samples) overlaid with predicted TCGA pan-glioma DNA methylation classes according to the previously published dataset<sup>18</sup>. Pan-glioma methylation classes were predicted by training a random forest on the previously published dataset<sup>18</sup>, which included methylation data of 418 low-grade glioma and 377 glioblastoma samples that were acquired using the Illumina 450k and 27k platforms. The random forest algorithm was trained using the 1,300 CpG signature as described in ref. 18 and using the default settings of the random forest algorithm implemented in the R package randomForest. Pan-glioma class prediction was only performed for subsets of mostly adult astrocytomas, oligodendroglomas and glioblastomas (magnified areas) included in the previously published dataset<sup>18</sup>. LGm1, LGm2 and LGm3 show a high overlap with the methylation classes A IDH HG, A IDH and O IDH, respectively. LGm4 shows the highest overlap with methylation class GBM RTK II. LGm5 shows the highest overlap with methylation classes GBM MES and GBM RTK I. LGm6 show the highest overlap with DMG K27, GBM MID and GBM MYCN.

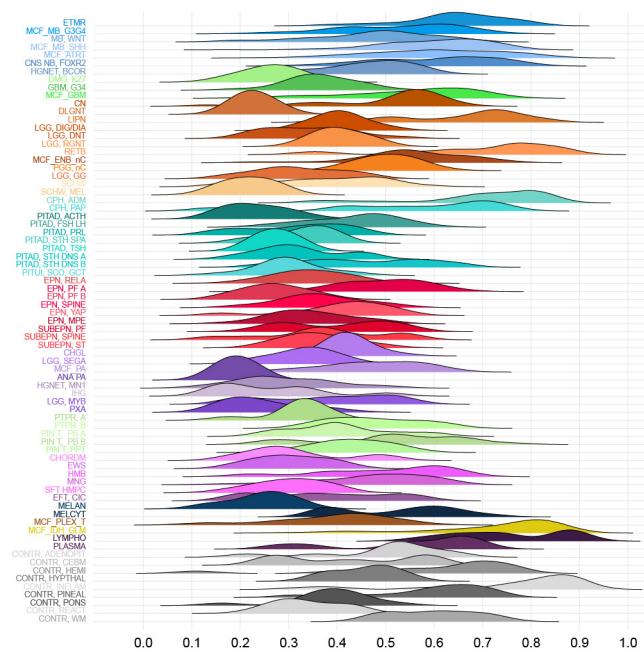
**a****b**

**Extended Data Figure 4 | Development of the random forest classifier.**  
**a**, The random forest training consists of four steps. First, basic filtering of probes that were not included on the EPIC array, probes located on the X and Y chromosomes, probes affected by single nucleotide polymorphisms, and probes not mapping uniquely to the genome was performed. In the second step, the probe-wise batch effects between samples from FFPE and frozen material were estimated and adjusted by a linear model approach. In the third step, feature selection was performed by training a random forest algorithm using all probes and selecting the 10,000 probes with highest variable importance measure. In the last step, the final random forest is trained using only the 10,000 selected probes. The validation of the random forest classifier involves a threefold nested cross-validation. In the outer loop of the cross-validation, the complete random forest training procedure consisting of four steps as described above are applied to the training data and the resulting random forest is used to predict the test data to generate random forest scores. In the inner loop of the cross-validation a threefold cross-validation is applied to

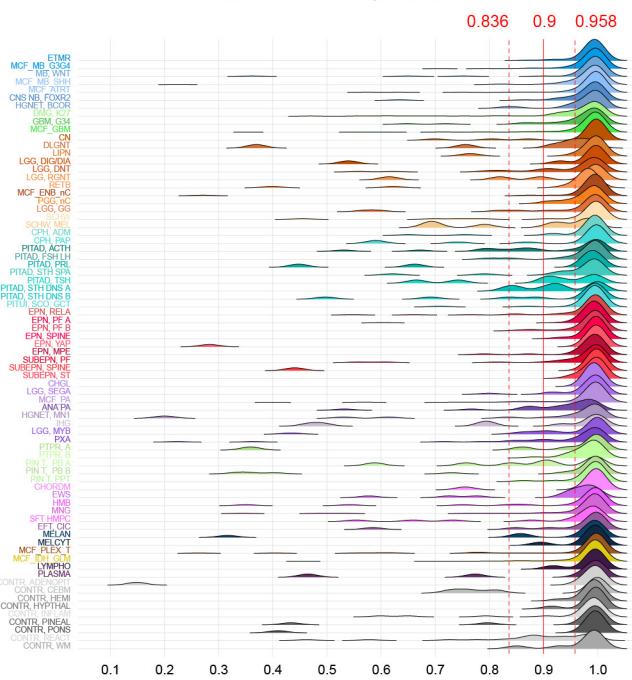
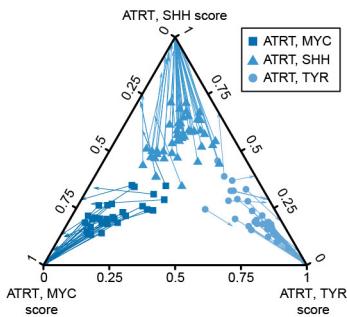
training data of the outer loop in order to generate random forest scores independent of the test data in the outer loop. These scores are then used to fit a calibration model, that is, a L2-penalized, multinomial, logistic regression that takes the random forest scores of the test data in the outer cross-validation loop to estimate tumour class probabilities (P1, P2, P3). To fit a calibration model to estimate class probabilities of diagnostic samples using all data in the reference set, the random forest scores generated in the outer cross-validation loop were used.  
**b**, Schematic depiction of three example binary decision trees of the random forest classifier (left), and magnification on five example decisions nodes relevant for glioblastoma classification (right). For prediction, a diagnostic sample enters the root node of each of the 10,000 trees. At every decision node, the decision path is determined on the methylation level of a single CpG, until it reaches a terminal node that provides the class prediction. The joint class prediction of all trees represents the raw prediction score. The colour code and abbreviations are identical to Fig. 1a.

**a**

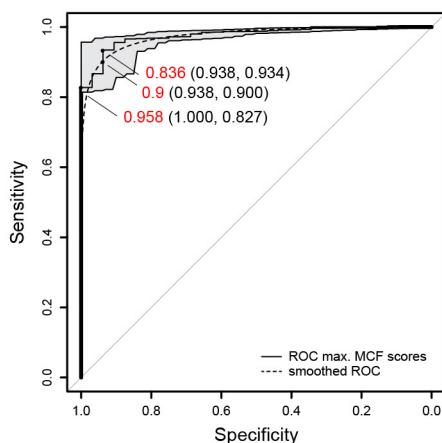
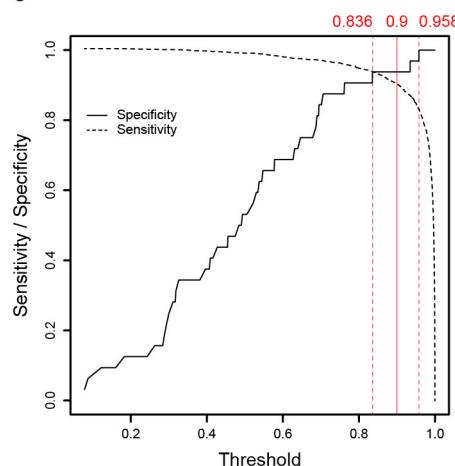
Raw scores for correctly classified cases



Calibrated scores for correctly classified cases

**b****c**

	Error rate	AUC	Brier score	Sensitivity	Specificity
Raw score	4.89%	0.9996	0.444	0.951	0.999
Calibrated score	4.28%	0.9998	0.070	0.957	0.999
Calibrated MCF score	1.14%	0.9999	0.024	0.989	0.999

**d****e**

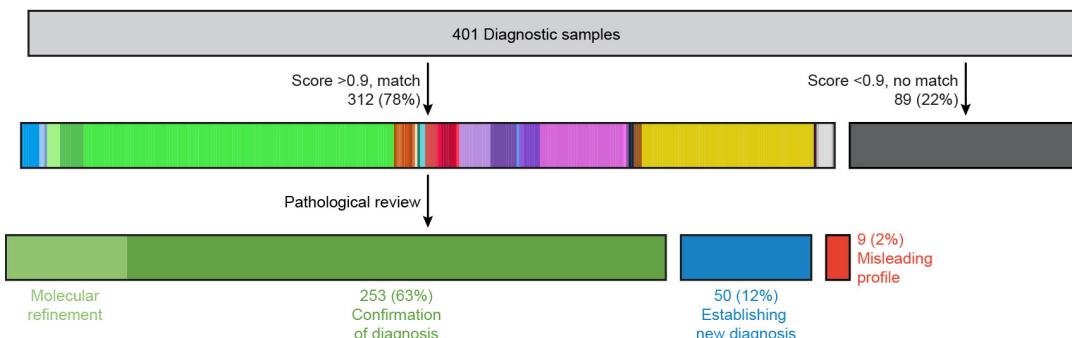
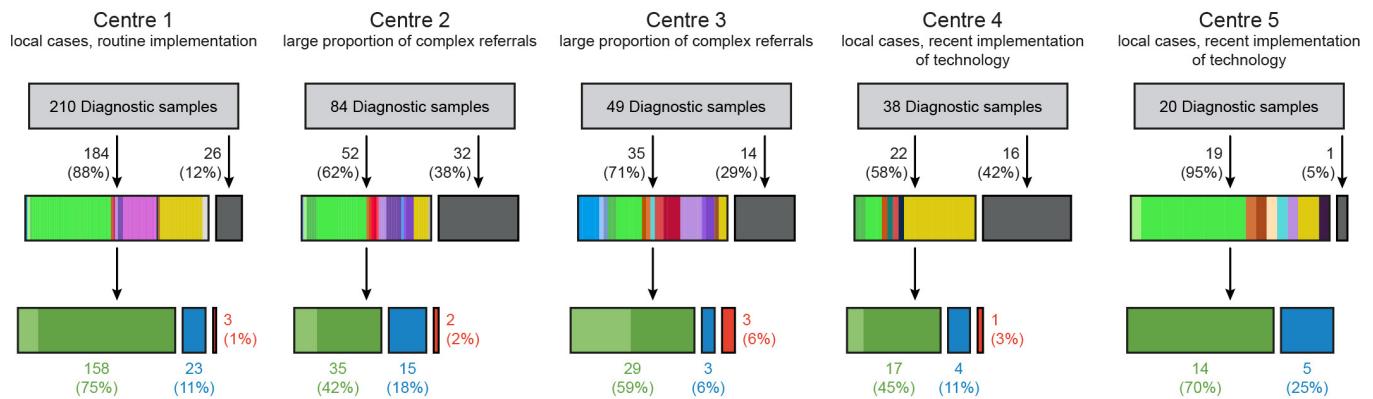
Extended Data Figure 5 | See next page for caption.

**Extended Data Figure 5 | Comparison of raw and calibrated classifier scores and threshold definition.** **a**, Density plots illustrating the distribution of raw and calibrated classifier scores for samples correctly classified during cross-validation ( $n = 2,701$  independent biological samples for raw and  $n = 2,769$  independent biological samples for calibrated), depicted for each methylation class or methylation class family. Score calibration results in a harmonization of score distribution and allows the establishment of a shared classification threshold. Three thresholds for maximizing specificity (0.958), maximizing the Youden index (0.836), and the cutoff used in this study (0.9) are indicated by red lines (see also **d** and **e**). **b**, Multivariate score calibration illustrated as a ternary plot showing scores of the three ATRT subclasses (MYC, SHH and TYR; together  $n = 112$  independent biological samples). Arrows indicate transformation of the scores for individual samples by the calibration model, which increases the discrimination between the three subclasses. **c**, The accuracy of prediction of the random forest classifier constructed of  $n = 2,801$  biologically independent samples (measured by misclassification error, AUC, Brier score, multiclass sensitivity and specificity) is improved by score calibration and by combining classes into MCF). **d**, To determine

a common threshold for the calibrated MCF scores, we performed a ROC analysis of the maximum calibrated MCF scores of all  $n = 2,801$  biologically independent samples calculated via cross-validation. For this ROC analysis, we defined a new binary class, that is, samples correctly classified during the cross-validation using the maximum calibrated MCF score for classification were considered as ‘classifiable’ ( $n = 2,769$ ) and samples that got falsely classified using this score were considered ‘non-classifiable’ ( $n = 32$ ). Three thresholds for different sensitivity and specificity are highlighted in the ROC curve: a threshold of 0.958 achieving a maximum specificity of 1 with a sensitivity of 0.827, a threshold of 0.836 obtaining a maximum Youden index with specificity 0.938 and sensitivity 0.934, and our recommended threshold of 0.9 that results in a specificity of 0.938 and a sensitivity of 0.9. Bootstrapped 95% confidence intervals for estimated sensitivity and specificity are indicated in grey. **e**, Sensitivity and specificity for all possible thresholds applied to cross-validated maximum MCF classifier scores of all  $n = 2,801$  biologically independent samples. Three thresholds for maximizing specificity (0.958), maximizing the Youden index (0.836) and 0.9 are highlighted by red lines.

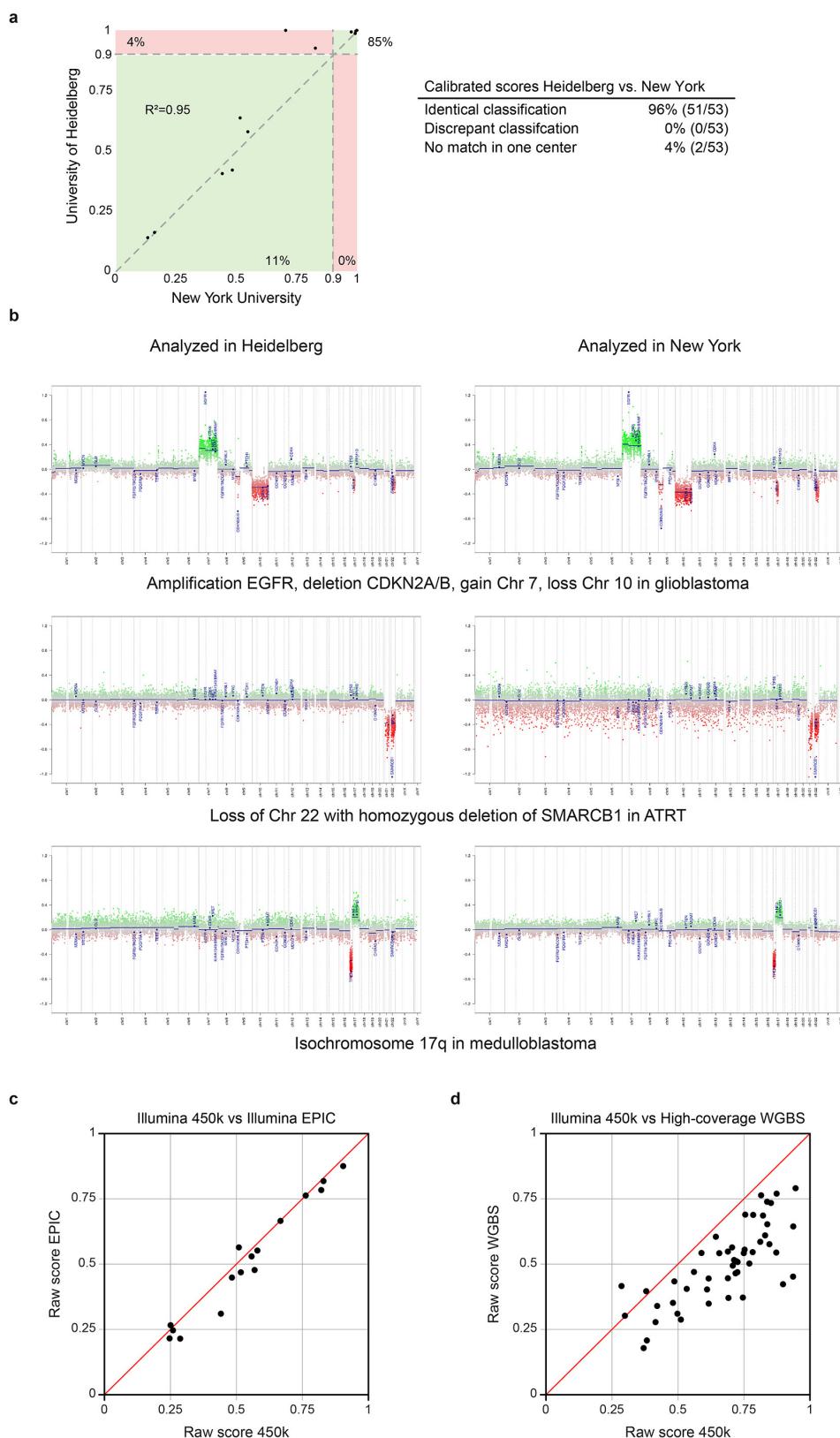
**a**

## All 5 centres combined

**b**

**Extended Data Figure 6 | Diagnostic utility of the DNA methylation-based classifier, assessed at different centres.** **a**, Implementation of the DNA methylation classifier by five external centres. In total, 401 independent biological samples were analysed. 78% matched to an established class with a cut-off score of  $\geq 0.9$  (class colours as in Fig. 1a).

A new diagnosis was established in 12% of cases. **b**, Depiction of individual centre results, illustrating the different composition of samples included in the analysis, variation in the rate of non-matching cases, and of cases for which a new diagnosis was established. Case-by-case details are provided in Supplementary Table 6.



Extended Data Figure 7 | See next page for caption.

**Extended Data Figure 7 | Inter-centre and inter-platform reproducibility of the DNA methylation-based classification.**

**a**, Calibrated scores of 53 independent biological samples representing diagnostic CNS tumour cases analysed at the University of Heidelberg and at the New York University pathology department. Both laboratories performed independent DNA extraction, array hybridization and data analysis. Cases falling into green areas were classified identically in both centres (96%); cases in the red area were non-classifiable in one centre (4%). None of the 53 samples was assigned to a different methylation class by the two centres. **b**, Copy-number profiles calculated from the array data generated at both centres were highly comparable and allowed identification of chromosomal gains, losses, amplifications and deletions.

Calculations and interpretation were performed once at each centre.

**c**, Plot of maximum raw classification scores of 16 different tumour samples generated using both 450k and EPIC arrays. All cases fall close to the bisecting line (red) indicating a high concordance of the scores. Further, the methylation class prediction was identical for all samples.

**d**, The CNS tumour classifier also performs well with data generated by WGBS. The plot shows classifier scores calculated from WGBS and 450k arrays of 50 cases comprising 11 different brain tumour entities (bisecting line in red). Methylation beta values were calculated from high-coverage WGBS data (>10 fold average coverage) and run through the CNS tumour classifier and plotted against the same case analysed using 450k arrays. The highest class prediction score was identical in all cases.



Extended Data Figure 8 | Example of the PDF report of an IDH wild-type glioblastoma sample.

Timeline working days (total days) day of week	Process	Comments
Day 1 (1) Tuesday	Material assessment	Methylation profiling requires high tumour content (>70% tumor cells of total cells) Hematoxylin and eosin staining for identification of optimal area. Scraping of tumour cell rich areas from unstained slides (10 slides of 10 µm each) or punch extraction from paraffin block (either 1.5 mm or 3 mm diameter). Cases with low tumor content are avoided when possible.
Day 2 (2) Wednesday	DNA extraction (1/2)	Day one of automated DNA extraction (Maxwell 16 FFPE plus LEV DNA purification kit, AS1135, Promega).
Day 3 (3) Thursday	DNA extraction (2/2) DNA quality control Bisulfite conversion (1/2)	Day two of DNA extraction. Concentration measurement using Qubit (Qubit dsDNA BR assay kit, Q32853, Invitrogen). Illumina FFPE QC kit only used optionally (e.g. for highly necrotic cases). Day one of bisulfite conversion of 250ng (in exceptional cases as little as 50-100ng) (Zymo EZ DNA methylation kit; D5002, Zymo).
Day 4 (4) Friday	Bisulfite conversion (2/2) DNA Restoration	Day two of bisulfite conversion. FFPE DNA Restore (Illumina). DNA cleanup. Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual. Freezing over weekend.
Day 5 (7) Monday	Whole genome amplification	Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual.
Day 6 (8) Tuesday	Fragmentation, precipitation Resuspension Array hybridization (1/2)	Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual.
Day 7 (9) Wednesday	Array hybridization (2/2) Washing and staining Scanning of array Technical quality control	Part of methylation array kit provided by Illumina. Performed strictly according to instruction manual. Technical quality control using assay internal system controls following the instruction manual. Eliminate bad samples.
Day 8 (10) Thursday	Data upload to Webportal Integration with pathological findings Writing of reports	Upload to <a href="http://www.molecularneuropathology.org">www.molecularneuropathology.org</a> Gender check. Checking for genotype matches if established. Exclusion of case swapping. Integration of pathological findings with methylation profiling. Possibly reconsideration of histology and possibly additional molecular investigations required.

Extended Data Figure 9 | Example work flow and timeline of diagnostic methylation profiling.

## Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see [Reporting Life Sciences Research](#). For further information on Nature Research policies, including our [data availability policy](#), see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

### ► Experimental design

#### 1. Sample size

Describe how sample size was determined.

The present work focuses on predictive modelling and no inferential statistic was performed. The minimal sample size to include a tumor class into the model was determined empirically by training and testing models with different sample sizes. The minimal class size of 8 allowed us to include rare tumor classes without losing prediction performance.

#### 2. Data exclusions

Describe any data exclusions.

Tumor content was required to be above 70% (as described in the Methods), otherwise data was not generated. This criteria was pre-established.

#### 3. Replication

Describe whether the experimental findings were reliably reproduced.

The separation of samples into the defined DNA methylation classes was reliably reproduced by iterative random downsampling of the reference cohort. The rate of establishment of a new diagnosis by methylation profiling was confirmed by the data of the external centres. The interlaboratory comparison demonstrated a reliable reproduction of the results of the original laboratory.

#### 4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

The construction of the methylation classifier reference cohort was done in a supervised fashion to recapitulate the entities established in the WHO classification of tumours of the central nervous system, no randomization was performed. For the clinical implementation in the prospective samples also no randomization was performed as all cases with sufficient material were subjected to the analysis. For the technical validation samples were also not randomized, instead 51 samples of a wide selection of histological classes were chosen to increase the validity for a broader range of tumours.

#### 5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

The initial pathological diagnosis of the prospective series was done fully blinded as the methylation data was not generated before the finalisation of pathological diagnosis of a given case (usually two weeks after the pathological diagnosis).

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

## 6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a Confirmed

- The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
- A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- A statement indicating how many times each experiment was replicated
- The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
- A description of any assumptions or corrections, such as an adjustment for multiple comparisons
- The test results (e.g. *P* values) given as exact values whenever possible and with confidence intervals noted
- A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
- Clearly defined error bars

See the web collection on [statistics for biologists](#) for further resources and guidance.

## ► Software

Policy information about [availability of computer code](#)

### 7. Software

Describe the software used to analyze the data in this study.

R: A language and environment for statistical computing.

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). [Nature Methods guidance for providing algorithms and software for publication](#) provides further information on this topic.

## ► Materials and reagents

Policy information about [availability of materials](#)

### 8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

Distribution of material of human research participants is restricted

### 9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

No antibodies were used.

### 10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

No eukaryotic cell lines were used.

b. Describe the method of cell line authentication used.

does not apply

c. Report whether the cell lines were tested for mycoplasma contamination.

does not apply

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by [ICLAC](#), provide a scientific rationale for their use.

does not apply

## ► Animals and human research participants

Policy information about [studies involving animals](#); when reporting animal research, follow the [ARRIVE guidelines](#)

### 11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

No animals were used in the study.

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

Reference cohort: tumor samples of 2801 individual research participants: 1278 female, 1466 male, 57 sex not available; age range 0-93 years, median 24 years;  
Prospective cohort: tumor samples of 1104 individual research participants: 481 female, 591 male, 32 sex not available; age range 0-85 years, median 38 years.  
External centre cohort: tumor samples of 401 individual research participants: 202 female, 199 male; Age range 0-86 years, median 53 years.