ORIGINAL PAPER

DNA methylation-based classification of malformations of cortical development in the human brain

Samir Jabari[1] · Katja Kobow[1] · Tom Pieper[4] · Till Hartlieb[4,6] · Manfred Kudernatsch[5,6] · Tilman Polster[7] ·
Christian G. Bien[7] · Thilo Kalbhenn[8] · Matthias Simon[8] · Hajo Hamer[2] · Karl Rössler[3,9] · Martha Feucht[10] ·
Angelika Mühlebner[11,12] · Imad Najm[13,14] · José Eduardo Peixoto-Santos[15] · Antonio Gil-Nagel[16] ·
Rafael Toledano Delgado[16] · Angel Aledo-Serrano[16] · Yanghao Hou[17,18] · Roland Coras[1] · Andreas von Deimling[17] ·
Ingmar Blümcke[1]

Abstract
Malformations of cortical development (MCD) comprise a broad spectrum of structural brain lesions frequently associated with epilepsy. Disease definition and diagnosis remain challenging and are often prone to arbitrary judgment. Molecular classification of histopathological entities may help rationalize the diagnostic process. We present a retrospective, multi-center analysis of genome-wide DNA methylation from human brain specimens obtained from epilepsy surgery using EPIC 850 K BeadChip arrays. A total of 308 samples were included in the study. In the reference cohort, 239 formalin-fixed and paraffin-embedded (FFPE) tissue samples were histopathologically classified as MCD, including 12 major subtype pathologies. They were compared to 15 FFPE samples from surgical non-MCD cortices and 11 FFPE samples from post-mortem non-epilepsy controls. We applied three different statistical approaches to decipher the DNA methylation pattern of histopathological MCD entities, i.e., pairwise comparison, machine learning, and deep learning algorithms. Our deep learning model, which represented a shallow neuronal network, achieved the highest level of accuracy. A test cohort of 43 independent surgical samples from different epilepsy centers was used to test the precision of our DNA methylation-based MCD classifier. All samples from the test cohort were accurately assigned to their disease classes by the algorithm. These data demonstrate DNA methylation-based MCD classification suitability across major histopathological entities amenable to epilepsy surgery and age groups and will help establish an integrated diagnostic classification scheme for epilepsy-associated MCD.

Keywords Brain development · Cortical malformation · Epilepsy · Epigenetic · Deep learning

Introduction

Human brain malformations present with a broad spectrum of anatomo-pathological lesions [1], genetic alterations [26], and clinical phenotypes [53]. If the neocortical mantle is affected, a structural lesion is usually classified as malformation of cortical development (MCD), with focal epilepsy being a frequent clinical symptom [26]. Many patients with MCD and focal epilepsy do not respond to anti-seizure

medication. However, epilepsy surgery can be a curative treatment option [11, 42]. Recent studies in surgically resected human brain tissue demonstrated that MCD often result from prenatally acquired brain somatic mutations with or without additional germline mutation in developmental signaling pathways governing neuroepithelial proliferation, migration, and cell lineage differentiation [3, 43, 58]. The anatomo-pathological phenotype is likely dependent on the timing of the acquired brain somatic mutation, the targeted cell lineage, and the affected gene [22]. Disease definition and diagnosis of MCD remain challenging in everyday clinical practice. If surgical treatment is suggested, a definitive diagnosis of MCD should be established by histopathology review. Diagnostic terms for MCD categories and subtypes are often defined imprecisely, and histopathological criteria are prone to arbitrary judgment [10, 53]. Previous studies

---

Samir Jabari, Katja Kobow, Andreas von Deimling, and Ingmar Blümcke contributed equally.

✉ Katja Kobow
katja.kobow@uk-erlangen.de

Extended author information available on the last page of the article

have reported substantial inter- and intra-observer variability in the histopathological diagnosis of MCD, for example, in Focal Cortical Dysplasia (FCD) [10, 52]. The introduction of genetic biomarkers is a growing field but available only for a subgroup of MCD entities so far, i.e., FCD type 2 [2–4, 21, 22, 33, 50, 58] or mild malformations of cortical development with oligodendroglial hyperplasia (MOGHE) [14, 61]. Diagnostic discordance and uncertainty may confound the assignment of genetic variants to disease entities [14] and compromise decision-making in clinical practice as well as the interpretation and validity of clinical observations and trials.

Herein, we address DNA methylation as an objective molecular diagnostic biomarker that can be reliably detected and analyzed from archival human brain FFPE tissue [16, 60, 66]. The methylome in surgical brain tissue represents a combination of both somatically acquired DNA methylation changes, characteristics that reflect the cellular composition of the tissue as well as molecular memory marks in response to environmental or pathogenic cues, including seizures [23, 36–40]. DNA methylation profiling is highly robust and reproducible even from small samples and archival tissue, and such profiles have been widely used to classify CNS tumors successfully [16, 60]. Based on our previous work within single MCD entities [38, 40] (Holthausen et al. accepted in Epilepsia), we developed a comprehensive approach toward the DNA methylation-based classification of major MCD entities across all age groups.

Materials and methods

Reference cohort

We reviewed clinical, and MRI data of individuals who underwent surgery for the treatment of their focal pharmaco-resistant epilepsy and were diagnosed with FCD type 1A ($n = 12$), FCD type 2A ($n = 29$), FCD type 2B ($n = 29$), FCD type 3A ($n = 14$), FCD type 3B ($n = 15$, all with ganglioglioma), FCD type 3C ($n = 17$, six with Sturge–Weber Syndrome, four with arterio-venous malformations, and seven with cavernoma), FCD type 3D ($n = 15$, one with traumatic brain injury, five with Rasmussen encephalitis, four with perinatal stroke, and five not further specified), hemimegalencephaly (HME, $n = 6$), mild malformation of cortical development (mMCD, $n = 28$), mMCD with oligodendroglial hyperplasia in epilepsy (MOGHE, $n = 22$), polymicrogyria (PMG, $n = 33$), cortical tuber of tuberous sclerosis complex (TSC, $n = 19$), or temporal lobe epilepsy (TLE, $n = 15$). Based on MRI and histology, all 15 TLE patients were diagnosed with hippocampal sclerosis, but we used only histologically normal temporal neocortex. All cases included into this study have been extensively studied

at the microscopic level with Hematoyxlin–Eosin and Cresyl Violet – Luxol Fast Blue stainings available from all FFPE surgical tissue blocks. An immunohistochemistry panel of antibodies recommended for the neuropathology work-up of epilepsy surgery specimens [7, 9], including NeuN, MAP2, GFAP, Vimentin, neurofilament SMI32, Ki67, OLIG2, CD34, CD68 and CD45 epitopes were also made available for each case. Each diagnosis was finally agreed upon consensus by two of our coauthors (IB and RC) applying the International League Against Epilepsy (ILAE) classification system of 2011 [13] and 2013 [12]. An FFPE block containing a prototypic area of the lesion was selected for further processing. Four non-epilepsy autopsy control cases with no known neurological history were also included in the study. From some of these autopsy cases, temporal and frontal neocortex with micro-dissected gray and white matter were sampled and analyzed independently (CTRL, $n = 11$; Table 1, Supplement Table 1, online resource). We obtained written informed consent for molecular genetic investigations and publication of the results from all participating patients or their legal guardians. The Ethics Committee of the Medical Faculty of the Friedrich-Alexander-University (FAU) Erlangen-Nürnberg, Germany, approved this study within the framework of the EU project "DESIRE" (FP7, grant agreement #602,531; AZ 92_14B) and European Reference Network EpiCARE" (grant agreement #769,051; AZ 193_18B).

Test cohort

The test cohort included 43 independent retrospective surgical samples, including a series of 18 patients provided through the epilepsy surgery program of the Cleveland Clinic, USA. These cases underwent independent iterative evaluation by 20 neuropathologists from 15 different countries and were previously published in the ILAE FCD agreement trial of histopathology and genetic testing [9]. Another 25 samples were provided through the European Epilepsy Brain Bank (EEBB). A clinical summary of all test samples is provided in Table 2, Supplement Table 1, online resource.

DNA extraction

A prototypic area within the center of the MCD lesion (neocortex) was identified on H&E slides as described above and macro-dissection performed by punch biopsy (pfm medical, Köln, Germany) or by hand (Fig. 1). DNA was extracted from formalin-fixed paraffin-embedded (FFPE) tissue using the Maxwell 16 FFPE Plus LEV DNA Kit (Promega, Madison, WI, USA), according to the manufacturer's instructions. DNA concentration was quantified using the Qubit dsDNA BR Assay kit (Invitrogen, Carlsbad, CA, USA).

Table 1 Clinical summary of the reference cohort

Diagnosis	n	∅ age at surgery	∅ age at onset	∅ duration	Sex (female/male)
FCD 1A	12	9.3 (±4.6)	2.2 (±3.1)	6.8 (±4.3)	7/5
FCD 2A	29	14.7 (±11.5)	3.5 (±4.1)	11.2 (±10.2)	13/16
FCD 2B	29	18.5 (±14.6)	3.6 (±4.0)	14.5 (±12.2)	13/16
FCD 3A	14	45.0 (±19.7)	9.4 (±11.2)	31.3 (±21.0)	9/5
FCD 3B	11(15)	34.5 (±14.7)	26.6 (±16.7)	7.91 (±8.8)	6/5
FCD 3C	17	23.5 (±18.9)	15.1 (±14.6)	7.6 (±8.9)	6/11
FCD 3D	15	15.4 (±11.0)	5.7 (±8.4)	6.8 (±5.4)	7/8
PMG	33	8.5 (±9.1)	1.6 (±2.7)	5.9 (±6.4)	10/23
HME	6	1.3 (±0.5)	0.1 (±0.2)	1.3 (±0.6)	2/4
TSC	19	5.5 (±6.9)	0 (±0)	5.5 (±6.9)	8/11
mMCD	28	24.6 (±17.6)	9.6 (±12.3)	11.6 (±13.3)	13/15
MOGHE	22	8.0 (±7.2)	8 (±7.3)	6.0 (±5.8)	8/12
TLE	15	37.0 (±15.3)	7.4 (±7.9)	29.6 (±16.6)	8/7
CTRL	4(11)	31.3 (±22.3)	n.a	n.a	3/1

CTRL control, *HME* hemimegalencephaly, *FCD* focal cortical dysplasia, *mMCD* mild malformation of cortical development, *MOGHE* mMCD with oligodendroglial hyperplasia in epilepsy, *PMG* polymicrogyria, *TLE* temporal lobe epilepsy, *TSC* tuberous sclerosis complex, *n.a.* not applicable

Table 2 Clinical summary of the test cohort

Diagnosis	N	∅ age at surgery	∅ age at onset	∅ duration	Sex (female/male)
FCD 1A	2	10.5 (±12.0)	4.5 (±6.4)	6.0 (±5.7)	1/1
FCD 2A	6	20.9 (±18.7)	4.4 (±3.4)	16.2 (±15.4)	1/5
FCD 2B	6	22.8 (±12.5)	6.6 (±6.1)	16.8 (±13.1)	0/6
FCD 3A	4	20.3 (±18.9)	3.8 (±4.3)	16.5 (±19.7)	2/2
FCD 3C	6	23.3 (±18.1)	14.2 (±12.5)	9.9 (±8.7)	3/3
FCD 3D	2	27.0 (±14.1)	2.5 (±3.5)	24.5 (±17.7)	0/2
mMCD	5	29.0 (±13.8)	17.0 (±8.9)	13.0 (±3.2)	2/3
MOGHE	5	9.7 (±10.4)	5.0 (±8.4)	4.7 (±2.8)	1/4
TLE	3	37.3 (±9.9)	17.0 (±11.4)	20.3 (±2.9)	0/3
TSC	4	3.6 (±1.9)	0.0 (±0.0)	3.6 (±1.9)	2/2

FCD focal cortical dysplasia, *mMCD* mild malformation of cortical development, *MOGHE* mMCD with oligodendroglial hyperplasia in epilepsy, *TLE* temporal lobe epilepsy, *TSC*– tuberous sclerosis complex

Genome-wide DNA methylation profiling and data pre-procession

As described previously, samples were analyzed using Illumina Infinium MethylationEPIC 850 K BeadChip arrays [16, 38]. Briefly, DNA methylation data were generated at the Department of Neuropathology, Universitätsklinikum Heidelberg, Germany. We performed differential DNA methylation analysis using a self-customized Python wrapped cross R package pipeline described in [38] and publicly available at https://github.com/FAU-DLM/Methy lr2py. Additionally, we stratified quantile normalized data using the 'minfi' 'preprocessQuantile' function [25]. After that, probes targeting sex chromosomes, containing single-nucleotide polymorphisms, not uniquely matching, as well as known cross-reactive probes were removed [19]. Consequently, 433,213 CpG probes contained on the EPIC array were used for further analysis. Most significantly differentially methylated CpG between disease entities, as highlighted above, were identified by fitting a regression model with the disease as the target variable using the 'limma' R package [59]. All pairwise comparisons between disease groups were identified as contrasts and included in the analysis. Surrogate variable adjustments found one surrogate variable, which we corrected for ('sva' R package) [44]. We also corrected for the batch number, duration of epilepsy, white vs. gray matter, and the region the specimen originated from ('removeBatchEffect' from 'limma'). These factors were identified by calculating and plotting the Pearson's r coefficient as a correlation matrix for all variables of the first six principal components deriving from the corresponding M values (data not shown). Then unsupervised dimensionality reduction for cluster analysis on the data was performed. Uniform Manifold Approximation and Projection (UMAP) for general non-linear dimensionality reduction was used for visualization [49]. The following non-default parameters were used: init=random, min_dist=0.0, spread=3.0. To confirm the identified clusters from the previous step, we applied unsupervised learning using HDBSCAN as a clustering algorithm [48]. The following non-default parameters were used:
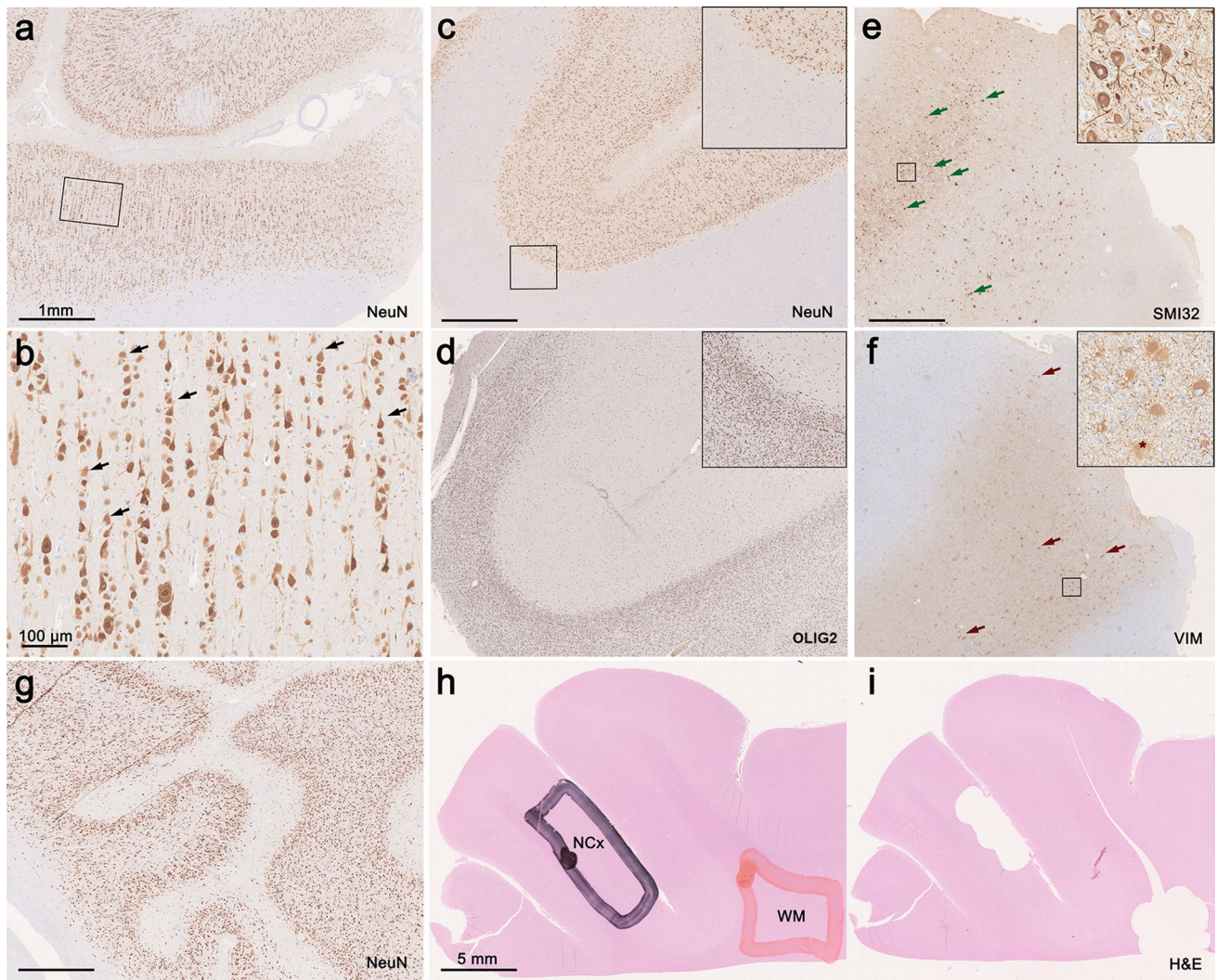
Fig. 1 Histopathological findings in representative MCD and control cases from the present cohort. **a–b** FCD1A with neocortex showing abundant radial organization of neurons (micro-columns, black arrows, NeuN immunohistochemistry). **c** In MOGHE the cortical ribbon shows no evidence for radial micro-columns or horizontal dyslamination. Instead, gray–white matter blurring with heterotopic neurons subjacent to white matter and (**d**) increase in OLIG2-immunoreactive oligodendroglial cells are detected. **e** In FCD2B, dysmorphic neurons accumulating non-phosphorylated neurofilament protein and lacking regular anatomic orientation (green arrows, SMI32 immunohistochemistry) as well as balloon cells characterized by large cell bodies occasionally presenting with multiple nuclei (asterisk) positively staining for Vimentin are present (VIM, dark red arrows). **g** In PMG, NeuN immunohistochemistry identifies abnormally folded sulci without pial opening. The cortical ribbon was thinned, mainly four-layered, and the gray to white matter boundaries were blurred with increased numbers of heterotopic neurons in the white matter. (**h–i**) Visualization of sampling method: Overview of H&E stained slides depicting selective sampling from regions of interest, e.g., neocortex or white matter, in a control sample. Scale bars are 1 mm if not shown otherwise

min_samples = 1, min_cluster_size = 4. We then re-ran the HDBSCAN algorithm within a loop while randomly down-sampling the data to five samples per disease group for a total of 100 times. We performed additional hierarchical cluster analysis of the corresponding normalized M values using python-based seaborns, clustermap plotting method [65]. The following non-default parameters were used: standard_scale = 1, method = 'ward'.

Machine and deep learning

'Scikit-learn', 'fastai', and 'Pytorch' were used as python-based packages to leverage machine and deep learning [30, 55, 56]. We split the processed methylation data from the steps above into an independent training, validation, and test set. Care was taken that disease classes were stratified across the sets evenly. The test set contained 43 fully

independent samples that were not used at any model training and validation stage.

Machine learning

Using 'Scikit-learn', various types of, classic machine learning algorithms were spot-checked on their performance on the dataset via stratified fivefold cross-validation. We separately trained an, extra trees classifier, a, nearest neighbor classifier, a, support vector classifier' and a, 'random forest classifier. We then modeled via stratified fivefold cross-validation an ensemble stacking model using, vecstack [32]. Stacked Generalization or "Stacking" is a two-step approach. The first step is to train base machine learning models on the dataset. Therefore, we used the same models as described above. The second step consists of training, a so-called meta-model on the predictions of the base models. This meta-model thereby tries to combine the base models predictions more robustly and accurately. We used the, XGBClassifier from, xgboost as our meta-model [18]. XGBoost stands for "Extreme Gradient Boosting" and is an implementation of gradient boosted decision trees. Boosting is an ensemble technique where new models are added to correct the errors made by existing models. Models are added sequentially until no further improvements can be made. Gradient boosting is an approach where new models are created that predict the residuals or errors of prior models and then added together to make the final prediction. It is called gradient boosting because it uses a gradient descent algorithm to minimize the loss when adding new models [47].

Deep learning

Utilizing, fastais tabular-learner, we modeled a deep linear neuronal network consisting of three subsequent layers. The first layer contained 500, the second layer 250, and the last layer only one neuron. In a stratified fivefold cross-validation manner, we then trained neuronal networks with a batch size of 32 by cycling the learning rate between 0.0001 and 0.08 for a total of four epochs which was identified by early stopping. Conventionally, the learning rate is decreased as the learning starts converging with time. It is helpful to oscillate the learning rate toward a higher learning rate as it may help get out of saddle points. This method was found to be most efficient in training neuronal networks [5, 63]. To be consistent with plots of the disease clusters, we transformed the predictions of the machine learning and neuronal network models into a two-dimensional space via UMAP dimensionality reduction.

Classifier performance measures

The performance of the resulting classifier predictions generated by the cross-validation for machine and deep learning models was evaluated by the balanced accuracy, precision, recall, F1 score, and the multiclass area under the receiver operating characteristic (ROC) curve (AUC). Results were plotted into a normalized confusion matrix. The balanced accuracy takes class imbalances into account. Precision and recall were chosen as additional metrics to measure how good samples were classified concerning the fraction of correctly and incorrectly classified samples. Precision is also known as the positive predictive value, and recall is also known as sensitivity. To easier assess these metrics during the training process, we additionally captured the harmonic mean between these scores: the F1 score.

Results

Methylation clusters define MCD and non-MCD

To establish a comprehensive MCD reference cohort, we generated genome-wide DNA methylation profiles using Infinium HumanMethylation850K BeadChip arrays (average group size 19; range 6–33 samples) from 239 surgical cases representing the majority of MCD disease entities (FCD 1A, 2A, 2B, 3A, 3B, 3C, 3D, PMG, HME, TSC, mMCD, MOGHE). We also included the intact temporal neocortex of 15 non-MCD TLE patients. Furthermore, we selected 11 samples from 4 autopsy cases representing non-neoplastic, non-MCD, non-epilepsy controls (CTRL; micro-dissected white matter and neocortex were studied individually, Fig. 1). Altogether, this resulted in a combined reference cohort of 265 samples (Table 1).

We performed unsupervised dimensionality reduction and hierarchical cluster analysis using 433,213 CpG probes. All disease groups in our analysis formed separate clusters in the UMAP dimensionality reduction characterized by distinct DNA methylation profiles (Fig. 2a). No confounding correlation with any other variable of our data was detected (e.g., sex, age at onset, age at surgery, lobe, neuronal proportion; Supplement Fig. 1, online resource). Unsupervised learning using HDBSCAN as a clustering algorithm (Fig. 2b) and hierarchical cluster analysis confirmed the separation of all samples at the disease level (Fig. 2c). To test the stability of identified clusters, we re-ran the HDBSCAN algorithm within a loop while randomly down-sampling the data to five samples per disease group for a total of 100 times. Thereby we demonstrated that the proximity of cases of the same class was preserved across iterations, indicative of high stability of methylation classes independent of the exact composition of the reference cohort (Fig. 2d).
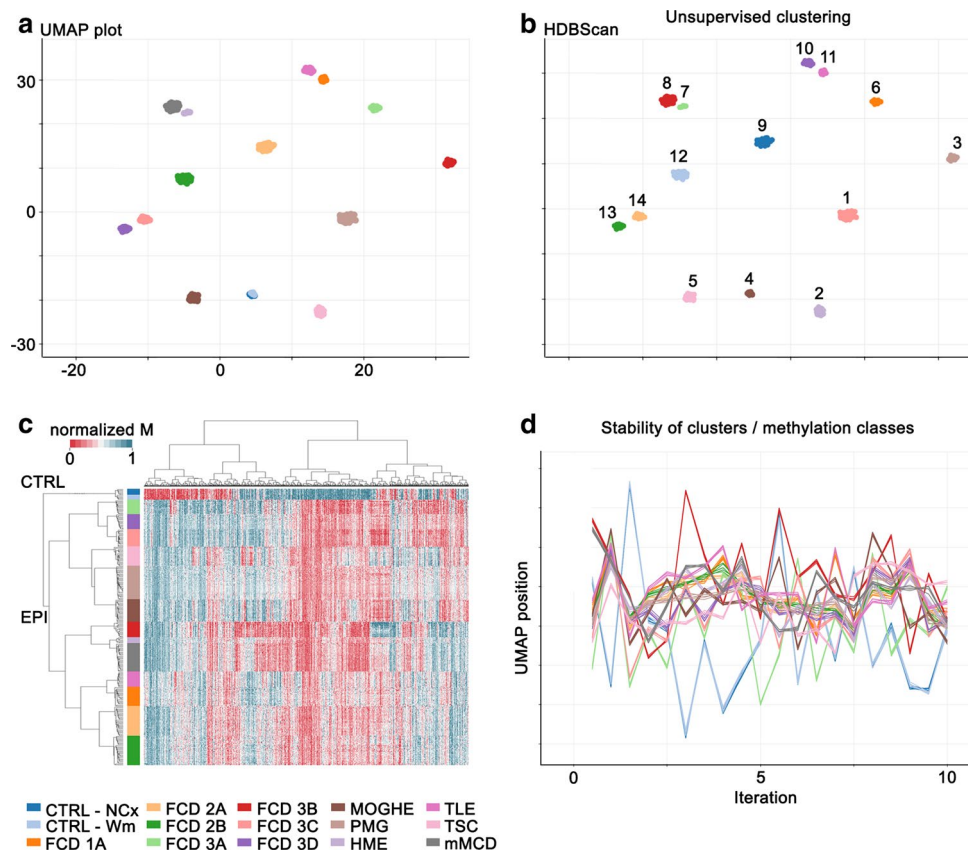
Fig. 2 Major MCD subtypes can be distinguished by their DNA methylation profiles. **a** UMAP plot for dimensionality reduction summarizing 12 MCD together with the TLE and control methylation classes of the reference cohort. Methylation classes reflect disease groups based on histology and are color-coded. **b** Confirmatory unsupervised identification of 14 clusters using HDBScan clustering algorithm (independent colors). This approach identified WM and NCx controls as a single uniform cluster. **c** Hierarchical cluster analysis summarizing DNA methylation profiles of 265 samples of the refer-

ence cohort. **d** X and Y coordinates of the first 10 of a total of 100 iterations of UMAP dimensionality reduction generated by random down-sampling to assess clustering stability. A line connects axis positions of individual cases. The depiction illustrates the proximity of cases of the same class across iterations, indicative of the high stability of methylation classes independent of the exact composition of the reference cohort. The color scheme for histopathological entities applies to **1–3**, except 1b)

Machine and deep learning can distinguish between histopathological entities

Future application in routine diagnostics requires fast, reproducible, and unbiased classification of samples. It also needs a measure of confidence for the specific call. We trained a, stacking machine learning algorithm (ML), a so-called ensemble method that combines the predictions of several 'weak' classifiers to improve prediction accuracy, and compared it to a three-layer shallow neuronal network (i.e., deep learning model, DL). Both ML and DL classifiers raw predictions were UMAP reduced to two dimensions, plotted, and showed excellent methylation class separation (Fig. 3a, b). When running fivefold cross-validation, the machine learning approach reached a balanced classification accuracy of 0.80, positive predictive value (i.e., precision) of 0.73, and sensitivity (i.e., recall) of 0.71, indicating already a good

discriminating power (Fig. 3c). However, the neuronal network approach outperformed the ML classifier on all metrics (balanced accuracy 0.94, positive predictive value of 0.98, and sensitivity of 0.98; Fig. 3d). Looking at the confidence of the classification decision for each sample, misclassified samples showed reduced confidence percentage scores for the machine learning and neuronal network approach (Fig. 3e, f), indicating that thresholding the classification confidence might be an appropriate method to minimize the method' error rate. Using Receiver Operating Characteristic (ROC) curve analysis, we devised an optimal threshold of ≥ 0.9 (Supplement Fig. 2, online resource).

Taken together, our findings provide evidence that methylation profiles are distinct for different epilepsy-associated disease entities and can be discriminated by machine and deep learning methods, which may help to rationalize disease classification and patient stratification.

Fig. 3 Machine and deep learning models can be trained to distinguish disease entities based on DNA methylation. **a**, **b** UMAP plots showing methylation classes based on ML and DL models. **c**, **d** Precision-recall curve to quantify the efficiency of our multiclass prediction task by ML and DL models. **e**, **f** Performance of the ML and DL models to discriminate 14 classes from the validation and test datasets presented as normalized confusion matrices. The vertical axis indicates the true (annotated) disease class of a sample, and the horizontal axis represents the predicted class. Precision of the DL model is almost 100% in all classes except controls, where, after correction for neuronal proportion, *CTRL* NCx (dark blue) and CTRL—Wm (light blue) form a single methylation class. Color scheme as in Fig. 1. *DL* deep learning; *ML* – machine learning; *UMAP* uniform manifold approximation and projection
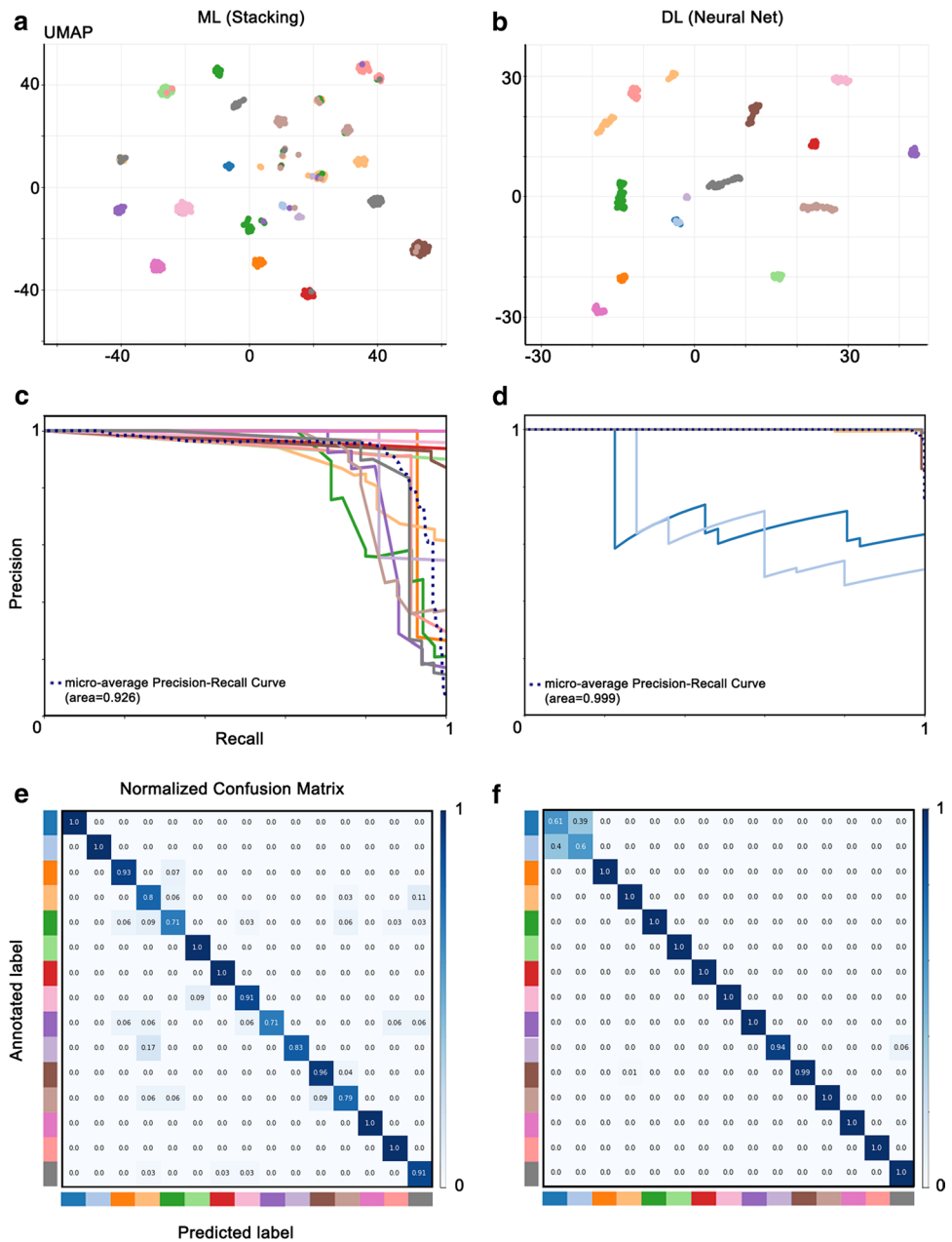
Disease classification in an independent test cohort

Next, we tested both models against an independent test cohort (*n*=43), including 18 samples obtained from the most recent ILAE FCD agreement trial [9]. These difficult-to-classify surgical brain samples obtained from pediatric and adult focal epilepsy patients had undergone multiple rounds of histopathological evaluations by 13 international expert neuropathologists to achieve an agreement on the diagnosis and were now analyzed for DNA methylation. Clinical data for the entire reference cohort are summarized in Table 2 and Supplement table 1, online resource. Methylation profiling and data analysis were performed as for the reference cohort, and test cohort

cases were assigned as either 'matching to a defined DNA methylation class' (score ≥ 0.9) or as 'no match' cases (highest score < 0.9). All profiled samples of the test cohort matched to an established DNA methylation class in both ML (Fig. 4a) and DL models with a classifier score ≥ 0.9 (Fig. 4b). However, only in the DL model were the results obtained by pathology and DNA methylation profiling concordant.

Fig. 4 Model testing with independent samples. Mapping of independent test cohort (black circle) to methylation classes identified by our (**a**) ML and (**b**) DL models. Only in the DL model were fully concordant results by pathology and DNA methylation profiling obtained
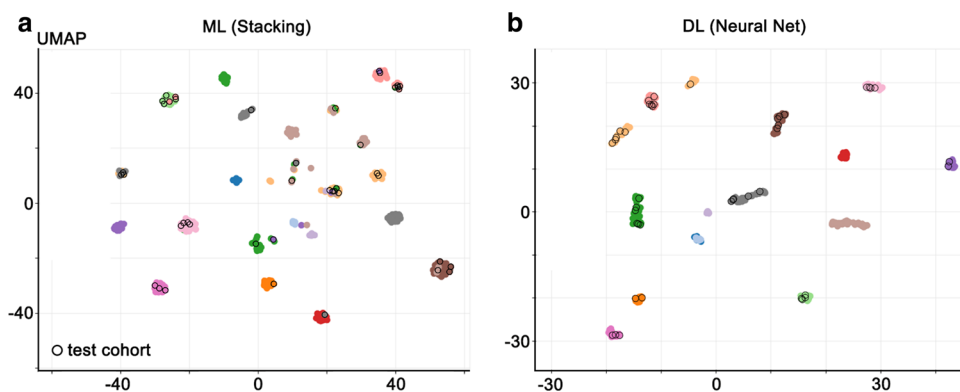
Discussion

Array-based DNA methylation profiling of formalin-fixed and paraffin-embedded human tissue samples has become a valuable tool to inform histopathology diagnosis in brain tumors [15, 16, 35, 60, 67]. Our data now suggest a practical application also in the diagnostic arena of epilepsy surgery and difficult-to-diagnose brain malformations. We studied a series of 308 cases with pharmaco-resistant epilepsy that underwent surgical treatment and were diagnosed with histopathologically confirmed MCD. This cohort covered the 12 most common MCD subtypes [11] and also our control categories of non-MCD epilepsy and non-epilepsy postmortems. We demonstrated that DNA methylation profiling distinguished epilepsy tissues from controls and specifically separated all 12 MCD subtypes. These pathology-associated methylation classes could be further discriminated by machine and deep learning algorithms.

Histopathology diagnosis of epilepsy surgery brain tissue poses a particular challenge to everyday clinical practice, especially in FCD [13]. This has been demonstrated several times in international agreement trials with kappa values of 0.4968 for mMCD and FCD 1 in one study [17] and 0.7824 in another study testing the ILAE classification scheme of 2011 [20]. The inter-observer agreement may vary from poor ($k = 0.16$) to good ($k = 0.68$) depending on the additional amount of information being available for the neuropathologist, i.e., immunohistochemistry or gene panel analysis [9]. In fact, the FCD classification scheme has been continuously modified and adapted to address this issue [13, 54]. Difficult-to-anticipate anatomical landmarks in not well preserved or presented surgical specimens and loosely described histopathology features remain the major obstacles to date [52]. While immuno-histochemical markers were introduced and recommended in 2016 by an ad hoc Task Force of the ILAE on diagnostic methods [7], it was not yet included in the FCD consensus classification scheme. Moreover, the small number of epilepsy surgery cases in an individual center requires continuous training of the neuropathologist, but only a few opportunities exist to

attend specialized training programs [57]. Hence, developing an easy-to-use and FFPE-compatible diagnostic tool is of great importance to enhance the diagnostic yield in MCD, overcome inter-observer variability and standardize MCD diagnostics across centers and clinical trials [9].

DNA methylation analysis already fosters detection and molecular characterization of more specific and new disease entities in the broad group of brain tumors, particularly those characterized by specific pathogenic variants or treatable by targeted therapies [28, 34, 41, 51, 64, 68]. We assume that the disease classification of MCD will also show such a dynamic adaptation, with more molecular genetic data becoming available over time. An integrated phenotype–genotype classification scheme has already been proposed for FCD, mainly Type 2, where brain somatic mutations in MTOR and GATOR signaling have been repeatedly identified [3, 8, 22]. Another practical example is MOGHE, which was first described histopathologically in 2017, specified further by a characteristic MRI signature, and finally revealed brain somatic mutations in the UDP-galactose transporter gene *SLC35A2* [8, 14, 27, 61]. In the present study, MOGHE cases showed a specific DNA methylation cluster, distinct from the clinically most challenging differential diagnosis of FCD 1A (Holthausen et al. accepted in Epilepsia) or other mMCD and non-lesional focal epilepsy [14].

Further, we recently identified polymicrogyria (PMG) with mosaic trisomy of the long arm of chromosome 1 as a molecularly defined MCD subgroup [38]. Its specific DNA methylation signature and copy number profile clinically associated with a unilateral frontal or hemispheric PMG without hemimegalencephaly, a severe form of intractable epilepsy with seizure onset in the first months of life, and severe developmental delay. Thus, it was to represent a distinct subtype within the spectrum of PMG disorders.

Yet another ongoing interest and research area has been low-grade developmental brain tumors associated with early-onset epilepsy, with many new categories implemented in each novel WHO classification scheme [6]. DNA methylation revealed distinct molecular signatures

for many of these new brain tumor entities [24, 31, 62], including papillary glioneuronal tumors [29] and, more recently also isomorphic diffuse glioma [66]. However, epilepsy surgery tissue is distinct from CNS tumor samples and imposes specific challenges to be addressed (see Supplement Fig. 3, online resource, for details). First, cortical malformations obtained from epilepsy surgery usually contain low-level mosaicism of affected cells mixed with normally developed neurons and glial cells. MCD also result from pathogenic variants at variable sites of the affected genes compared to more frequent hot-spot mutations in brain tumors. Many MCD pathologies completely lack any known driver mutation. While tumors are considered to develop from single cells of origin by clonal evolution so that all cells within the tumor harbor the same mutation, epileptic tissue fails to show that pattern. Even neighboring neurons in the normal brain may carry a genetic profile much different from each other [45, 46].

In contrast, the genomic DNA methylation in bulk epileptic brain tissue has been highly specific to the seizure phenotype across species and model systems irrespective of cellular composition and appeared further specific for etiology and histopathology [23, 38–40]. While previous studies analyzed only small sample cohorts focusing on specific pathologies, e.g., FCD or PMG, the present study is the first comprehensive description of diagnostically valuable DNA methylation signatures across the broad spectrum of MCD and all age groups. Continuing efforts for molecular characterization of epilepsy surgery tissue may in future enhance our understanding of, e.g., hemimegalencephaly, which remains a solely macroscopic diagnosis based on MRI so far, or other heterogeneous and not yet well-defined diagnostic entities, e.g., FCD Type 1, and non-lesional tissue. The inclusion of new diagnostic MCD entities based on such an advanced molecular diagnostic workup will, however, require a careful review to advance clinical patient management and precision medicine in the arena of epileptology.

Data availability The complete methylation data required for constructing the classifier, i.e., the reference and the independent test cohorts, have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo) under accession numbers GSE185090 and GSE156374. Supplement Table 1 provides IDAT-file names and assignments to patient characteristics (excluding previously published samples [38]).

Declarations

References

1. Adle-Biassette H, Golden JA, Harding B (2018) Chapter 6 - developmental and perinatal brain diseases. In: Kovacs GG, Alafuzoff I (eds) Handb clin neurol. Elsevier, pp 51–78

2. Alcantara D, Timms AE, Gripp K, Baker L, Park K, Collins S et al (2017) Mutations of AKT3 are associated with a wide spectrum of developmental disorders including extreme megalencephaly. Brain 140:2610–2622

3. Baldassari S, Ribierre T, Marsan E, Adle-Biassette H, Ferrand-Sorbets S, Bulteau C et al (2019) Dissecting the genetic basis of focal cortical dysplasia: a large cohort study. Acta Neuropathol 138:885–900

4. Baulac S, Ishida S, Marsan E, Miquel C, Biraben A, Nguyen DK et al (2015) Familial focal epilepsy with focal cortical dysplasia due to DEPDC5 mutations. Ann Neurol 77:675–683

5. Bengio Y (2012) Practical recommendations for gradient-based training of deep architectures neural networks: tricks of the trade. Springer

6. Blumcke I, Aronica E, Becker A, Capper D, Coras R, Honavar M et al (2016) Low-grade epilepsy-associated neuroepithelial tumours - the 2016 WHO classification. Nat Rev Neurol 12:732–740

7. Blumcke I, Aronica E, Miyata H, Sarnat HB, Thom M, Roessler K et al (2016) International recommendation for a comprehensive neuropathologic workup of epilepsy surgery brain tissue: a consensus task force report from the ILAE commission on diagnostic methods. Epilepsia 57:348–358

8. Blumcke I, Cendes F, Miyata H, Thom M, Aronica E, Najm I (2021) Toward a refined genotype-phenotype classification scheme for the international consensus classification of Focal Cortical Dysplasia. Brain Pathol 31:e12956

9. Blümcke I, Coras R, Busch RM, Morita-Sherman M, Lal D, Prayson R et al (2021) Toward a better definition of focal cortical dysplasia: an iterative histopathological and genetic agreement trial. Epilepsia 62:1416–1428

10. Blumcke I, Spreafico R (2011) An international consensus classification for focal cortical dysplasias. Lancet Neurol 10:26–27

11. Blumcke I, Spreafico R, Haaker G, Coras R, Kobow K, Bien CG et al (2017) Histopathological findings in brain tissue obtained during epilepsy surgery. N Engl J Med 377:1648–1656

12. Blumcke I, Thom M, Aronica E, Armstrong DD, Bartolomei F, Bernasconi A et al (2013) International consensus classification of hippocampal sclerosis in temporal lobe epilepsy: a task force report from the ILAE commission on diagnostic methods. Epilepsia 54(7):1315–1329

13. Blumcke I, Thom M, Aronica E, Armstrong DD, Vinters HV, Palmini A et al (2011) The clinicopathologic spectrum of focal cortical dysplasias: a consensus classification proposed by an ad hoc Task Force of the ILAE Diagnostic Methods Commission. Epilepsia 52:158–174

14. Bonduelle T, Hartlieb T, Baldassari S, Sim NS, Kim SH, Kang HC et al (2021) Frequent SLC35A2 brain mosaicism in mild malformation of cortical development with oligodendroglial hyperplasia in epilepsy (MOGHE). Acta Neuropathol Commun 9:3

15. Capper D, Engel NW, Stichel D, Lechner M, Gloss S, Schmid S et al (2018) DNA methylation-based reclassification of olfactory neuroblastoma. Acta Neuropathol 136:255–271

16. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D et al (2018) DNA methylation-based classification of central nervous system tumours. Nature 555:469–474

17. Chamberlain WA, Cohen ML, Gyure KA, Kleinschmidt-DeMasters BK, Perry A, Powell SZ et al (2009) Interobserver and intraobserver reproducibility in focal cortical dysplasia (malformations of cortical development). Epilepsia 50:2593–2598

18. Chen T, Guestrin C (2016) XGBoost: a scalable tree boosting system. Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining

19. Chen YA, Lemire M, Choufani S, Butcher DT, Grafodatskaya D, Zanke BW et al (2013) Discovery of cross-reactive probes and polymorphic CpGs in the Illumina Infinium HumanMethylation450 microarray. Epigenetics 8:203–209

20. Coras R, de Boer OJ, Armstrong D, Becker A, Jacques TS, Miyata H et al (2012) Good interobserver and intraobserver agreement in the evaluation of the new ILAE classification of focal cortical dysplasias. Epilepsia 53:1341–1348

21. D'Gama AM, Geng Y, Couto JA, Martin B, Boyle EA, LaCoursiere CM et al (2015) Mammalian target of rapamycin pathway mutations cause hemimegalencephaly and focal cortical dysplasia. Ann Neurol 77:720–725

22. D'Gama AM, Woodworth MB, Hossain AA, Bizzotto S, Hatem NE, LaCoursiere CM et al (2017) Somatic mutations activating the mTOR pathway in dorsal telencephalic progenitors cause a continuum of cortical dysplasias. Cell Rep 21:3754–3766

23. Debski KJ, Pitkanen A, Puhakka N, Bot AM, Khurana I, Harikrishnan KN et al (2016) Etiology matters - genomic DNA methylation patterns in three rat models of acquired epilepsy. Sci Rep 6:25668

24. Deng MY, Sill M, Sturm D, Stichel D, Witt H, Ecker J et al (2020) Diffuse glioneuronal tumour with oligodendroglioma-like features and nuclear clusters (DGONC) - a molecularly defined glioneuronal CNS tumour class displaying recurrent monosomy 14. Neuropathol Appl Neurobiol 46:422–430

25. Fortin JP, Triche TJ Jr, Hansen KD (2017) Preprocessing, normalization and integration of the Illumina HumanMethylationEPIC array with minfi. Bioinformatics 33:558–560

26. Guerrini R, Dobyns WB (2014) Malformations of cortical development: clinical features and genetic causes. Lancet Neurol 13:710–726

27. Hartlieb T, Winkler P, Coras R, Pieper T, Holthausen H, Blumcke I et al (2019) Age-related MR characteristics in mild malformation of cortical development with oligodendroglial hyperplasia and epilepsy (MOGHE). Epilepsy Behav 91:68–74

28. Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M et al (2005) MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med 352:997–1003

29. Hou Y, Pinheiro J, Sahm F, Reuss DE, Schrimpf D, Stichel D et al (2019) Papillary glioneuronal tumor (PGNT) exhibits a characteristic methylation profile and fusions involving PRKCA. Acta Neuropathol 137:837–846

30. Howard J, Gugger S (2020) Fastai: a layered API for deep learning. Information 11:108

31. Huse JT, Snuderl M, Jones DT, Brathwaite CD, Altman N, Lavi E et al (2017) Polymorphous low-grade neuroepithelial tumor of the young (PLNTY): an epileptogenic neoplasm with oligodendroglioma-like components, aberrant CD34 expression, and genetic alterations involving the MAP kinase pathway. Acta Neuropathol 133:417–429

32. Ivanov I (2016) Vecstack. https://github.com/vecxoz/vecstack

33. Jansen LA, Mirzaa GM, Ishak GE, O'Roak BJ, Hiatt JB, Roden WH et al (2015) PI3K/AKT pathway mutations cause a spectrum of brain malformations from megalencephaly to focal cortical dysplasia. Brain 138:1613–1628

34. Jaunmuktane Z, Capper D, Jones DTW, Schrimpf D, Sill M, Dutt M et al (2019) Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single centre. Acta Neuropathol Commun 7:24

35. Jurmeister P, Bockmayr M, Seegerer P, Bockmayr T, Treue D, Montavon G, Vollbrecht C, Arnold A, Teichmann D, Bressem K et al (2019) Machine learning analysis of DNA methylation profiles distinguishes primary lung squamous cell carcinomas from head and neck metastases. Sci Transl Med 11

36. Kiese K, Jablonski J, Hackenbracht J, Wrosch JK, Groemer TW, Kornhuber J et al (2017) Epigenetic control of epilepsy target genes contributes to a cellular memory of epileptogenesis in cultured rat hippocampal neurons. Acta Neuropathol Commun 5:79

37. Kobow K, Blumcke I (2012) The emerging role of DNA methylation in epileptogenesis. Epilepsia 53(Suppl 9):11–20

38. Kobow K, Jabari S, Pieper T, Kudernatsch M, Polster T, Woermann FG et al (2020) Mosaic trisomy of chromosome 1q in human brain tissue associates with unilateral polymicrogyria, very early-onset focal epilepsy, and severe developmental delay. Acta Neuropathol 140:881–891

39. Kobow K, Kaspi A, Harikrishnan KN, Kiese K, Ziemann M, Khurana I et al (2013) Deep sequencing reveals increased DNA methylation in chronic rat epilepsy. Acta Neuropathol 126:741–756

40. Kobow K, Ziemann M, Kaipananickal H, Khurana I, Muhlebner A, Feucht M et al (2019) Genomic DNA methylation distinguishes subtypes of human focal cortical dysplasia. Epilepsia 60:1091–1103

41. Kowalewski A, Durślewicz J, Zdrenka M, Grzanka D, Szylberg Ł (2020) Clinical relevance of BRAF V600E mutation status in brain tumors with a focus on a novel management algorithm. Target Oncol 15:531–540

42. Lamberink HJ, Otte WM, Blumcke I, Braun KPJ, Writi EEBB, Grp S et al (2020) Seizure outcome and use of antiepileptic drugs after epilepsy surgery according to histopathological diagnosis: a retrospective multicentre cohort study. Lancet Neurol 19:748–757

43. Lee JH, Huynh M, Silhavy JL, Kim S, Dixon-Salazar T, Heiberg A et al (2012) De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly. Nat Genet 44:941–945

44. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD (2012) The sva package for removing batch effects and other unwanted

variation in high-throughput experiments. Bioinformatics (Oxford, England) 28:882–883

45. Linnarsson S (2015) A tree of the human brain. Science 350:37–37

46. Lodato Michael A, Woodworth Mollie B, Lee S, Evrony Gilad D, Mehta Bhaven K, Karger A et al (2015) Somatic mutation in single human neurons tracks developmental and transcriptional history. Science 350:94–98

47. Mason L, Baxter J, Bartlett P, Frean M (1999) Boosting algorithms as gradient descent. NIPS

48. McInnes L, Healy J, Astels S (2017) hdbscan: Hierarchical density based clustering. J Open Source Softw 2:205

49. McInnes L, Healy J, Melville J (2018) Umap: uniform manifold approximation and projection for dimension reduction. arXiv preprint

50. Mirzaa GM, Campbell CD, Solovieff N, Goold C, Jansen LA, Menon S et al (2016) Association of MTOR mutations with developmental brain disorders, including megalencephaly, focal cortical dysplasia, and pigmentary mosaicism. JAMA Neurol 73:836–845

51. Mueller T, Stucklin ASG, Postlmayr A, Metzger S, Gerber N, Kline C et al (2020) Advances in targeted therapies for pediatric brain tumors. Curr Treat Options Neurol 22:43

52. Najm IM, Sarnat HB, Blumcke I (2018) Review: the international consensus classification of focal cortical dysplasia - a critical update 2018. Neuropathol Appl Neurobiol 44:18–31

53. Oegema R, Barakat TS, Wilke M, Stouffs K, Amrom D, Aronica E et al (2020) International consensus recommendations on the diagnostic work-up for malformations of cortical development. Nat Rev Neurol 16:618–635

54. Palmini A, Najm I, Avanzini G, Babb T, Guerrini R, Foldvary-Schaefer N et al (2004) Terminology and classification of the cortical dysplasias. Neurology 62:S2–S8

55. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G et al (2019) PyTorch: an imperative style high-performance deep learning library. NeurIPS

56. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O et al (2011) Scikit-learn: machine learning in Python. J Mach Learn Res 12:2825–2830

57. Peixoto-Santos JE, Blumcke I (2021) Neuropathology of the 21th century for the latin American epilepsy community. Seizure 90:51–59

58. Ribierre T, Deleuze C, Bacq A, Baldassari S, Marsan E, Chipaux M et al (2018) Second-hit mosaic mutation in mTORC1 repressor DEPDC5 causes focal cortical dysplasia-associated epilepsy. J Clin Invest 128:2452–2458

59. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W et al (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47

60. Sahm F, Schrimpf D, Stichel D, Jones DTW, Hielscher T, Schefzyk S et al (2017) DNA methylation-based classification and grading system for meningioma: a multicentre, retrospective analysis. Lancet Oncol 18:682–694

61. Schurr J, Coras R, Rossler K, Pieper T, Kudernatsch M, Holthausen H et al (2017) Mild malformation of cortical development with oligodendroglial hyperplasia in frontal lobe epilepsy: a new clinico-pathological entity. Brain Pathol 27:26–35

62. Sievers P, Appay R, Schrimpf D, Stichel D, Reuss DE, Wefers AK et al (2019) Rosette-forming glioneuronal tumors share a distinct DNA methylation profile and mutations in FGFR1, with recurrent co-mutation of PIK3CA and NF1. Acta Neuropathol 138:497–504

63. Smith LN (2017) Cyclical learning rates for training neural networks. 2017 IEEE Winter conference on applications of computer vision (WACV). 464–472

64. Suwala AK, Stichel D, Schrimpf D, Kloor M, Wefers AK, Reinhardt A et al (2021) Primary mismatch repair deficient IDH-mutant astrocytoma (PMMRDIA) is a distinct type with a poor prognosis. Acta Neuropathol 141:85–100

65. Waskom ML (2021) Seaborn: statistical data visualization. J Open Source Softw 6:3021

66. Wefers AK, Stichel D, Schrimpf D, Coras R, Pages M, Tauziede-Espariat A et al (2020) Isomorphic diffuse glioma is a morphologically and molecularly distinct tumour entity with recurrent gene fusions of MYBL1 or MYB and a benign disease course. Acta Neuropathol 139:193–209

67. Weidema ME, van de Geer E, Koelsche C, Desar IME, Kemmeren P, Hillebrandt-Roeffen MHS et al (2020) DNA methylation profiling identifies distinct clusters in angiosarcomas. Clin Cancer Res 26:93–100

68. Yan H, Parsons DW, Jin G, McLendon R, Rasheed BA, Yuan W et al (2009) IDH1 and IDH2 mutations in gliomas. N Engl J Med 360:765–773

Authors and Affiliations

Samir Jabari[1] · Katja Kobow[1] · **Tom Pieper[4] · Till Hartlieb[4,6] · Manfred Kudernatsch[5,6] · Tilman Polster[7] · Christian G. Bien[7] · Thilo Kalbhenn[8] · Matthias Simon[8] · Hajo Hamer[2] · Karl Rössler[3,9] · Martha Feucht[10] · Angelika Mühlebner[11,12] · Imad Najm[13,14] · José Eduardo Peixoto-Santos[15] · Antonio Gil-Nagel[16] · Rafael Toledano Delgado[16] · Angel Aledo-Serrano[16] · Yanghao Hou[17,18] · Roland Coras[1] · Andreas von Deimling[17] · Ingmar Blümcke[1]**

1. Department of Neuropathology, Affiliated Partner of the ERN EpiCARE, Universitätsklinikum Erlangen, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany

2. Department of Neurology, Epilepsy Center, Universitätsklinikum Erlangen, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany

3. Department of Neurosurgery, Universitätsklinikum Erlangen, Friedrich-Alexander University Erlangen-Nürnberg (FAU), Erlangen, Germany

4. Center for Pediatric Neurology, Neurorehabilitation and Epileptology, Vogtareuth, Germany

5. Center for Neurosurgery and Epilepsy Surgery, Schön Klinik Vogtareuth, Vogtareuth, Germany

6. Research Institute, Rehabilitation, Transition, Palliation", PMU Salzburg, Salzburg, Austria

7   Department of Epileptology (Krankenhaus Mara), Medical School, Bielefeld University, Bielefeld, Germany

8   Department of Neurosurgery - Epilepsy Surgery, Evangelisches Klinikum Bethel, Universitätsklinikum OWL, Bielefeld University, Bielefeld, Germany

9   Department of Neurosurgery, Medical University Vienna, Vienna, Austria

10 Department of Pediatrics and Adolescent Medicine, Affiliated Partner of the ERN EpiCARE, Medical University Vienna, Vienna, Austria

11 Department of Pathology, University Medical Center Utrecht, Utrecht, The Netherlands

12 Department of (Neuro) Pathology, Amsterdam UMC, Location AMC, Amsterdam, The Netherlands

13 Charles Shor Epilepsy Center, Cleveland Clinic, Cleveland, OH, USA

14 Department of Neurology, Cleveland Clinic, Cleveland, OH, USA

15 Department of Neurology and Neurosurgery, Paulista Medical School, UNIFESP, Sao Paulo, Brazil

16 Epilepsy Program, Hospital Ruber Internacional, Madrid, Spain

17 Department of Neuropathology, German Cancer Research Center (DKFZ), Universitätsklinikum Heidelberg, and CCU Neuropathology, Heidelberg, Germany

18 Institute of Pathology and Southwest Cancer Center, Southwest Hospital, Third Military Medical University (Army Medical University), Chongqing, China