Methylation-based algorithms for diagnosis: experience from neuro-oncology

Jessica C Pickles, Thomas J Stone, Thomas S Jacques

Developmental Biology and Cancer Research & Teaching Department, UCL Great Ormond Street Institute of Child Health, London, UK and Department of Histopathology, Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK

Corresponding author: Thomas S Jacques

Address: UCL GOS Institute for Child Health, 30 Guilford Street, London WC1N 1EH, UK

Tel: +44 (0)20 7829 8663

Email: t.jacques@ucl.ac.uk

Word Count: 3905

Running title: Methylation based tumour diagnosis

The authors declare no conflict of interest relevant to the published work.

Key words: CNS tumours, classification, pathology, methylation, diagnosis, algorithm

Abstract

Brain tumours are the commonest tumour-related cause of death in young people. Survivors are at risk of significant disability, at least in part, related to the effects of treatment. Therefore, there is a need for precise diagnosis that stratifies patients for the most suitable treatment, matched to the underlying biology of their tumour. While traditional histopathology has been accurate in predicting treatment responses in many cases, molecular profiling has revealed a remarkable, previously unappreciated, level of biological complexity in the classification of these tumours. Amongst different molecular technologies, DNA methylation profiling has had the most pronounced impact on brain tumour classification. Furthermore, using machine learning based-algorithms, DNA methylation profiling is changing diagnostic practice. This can be regarded as an exemplar for how molecular pathology can influence diagnostic practice and illustrates some of the unanticipated benefits and risks.

Introduction

Accurate diagnosis and classification are at the heart of all tumour pathology. However, in many tumour types, not only have we not known with certainty how to group them into biologically meaningful subtypes, but we have lacked the tools to provide objective sub-typing in clinical practice, relying heavily on the subjectivity of expert opinion. As an example, the diagnosis and classification of brain tumours has altered unrecognisably over the last few years through a range of molecular techniques, of which the most important, and at times controversial, has been DNA methylation profiling. In this review, we will outline the ways in which the field has changed because of this technology and we will attempt to draw some general implications that may be relevant as the technology impacts other tumour fields.

The clinical challenge of brain tumours

Brain tumours account for 3% of cancers but with only 14% of patients surviving over 10 years [1], there is a compelling need for better approaches to their management. The severity of this problem becomes particularly clear in children, teenagers and young adults. In developed countries, cancers are the commonest natural cause of death in children over 1 year [2]. Furthermore, brain tumours are the commonest cause of cancer-related death not only in children but also teenagers and young adults [1]. Therefore, as the major cancer-related cause of death in young people, brain tumours account for a disproportionate loss of years of life.

However, mortality is only part of the problem, as long-term life altering disability is prevalent amongst survivors [3]. This may be accounted for partly by the direct effects of the tumour on the nervous system but also probably reflects the impact of treatment, including radiotherapy, on the developing brain. Therefore, brain tumours in young people, particularly children, offer a paradigm for the challenges of accurate diagnosis that correctly predicts outcomes and response to treatment. If a diagnosis leads to insufficient treatment, the child may die. If a diagnosis leads to excessive treatment, the child may suffer life-long disability, potentially for decades.

While traditional histopathology has substantial power to stratify and remains the foundation of brain tumour diagnosis, it has had limitations. For example, there are tumour types where histology is poorly correlated with the underlying biology, such as low-grade epilepsy associated tumours [4-7]. Furthermore, histologically similar groups may contain biologically diverse sets of tumours that overlap with other morphologically distinct tumours, as with CNS Embryonal Tumours [8,9]. Moreover, there are tumours where histopathology has very poor inter-observer variability, suggesting that we do not have reliable tools to objectively subtype the tumours (e.g. low grade epilepsy associated tumours and ependymoma [6,7,10]).

As in other tumour types, much of this complexity is being resolved by 'omic molecular techniques but the most defining changes have come from DNA methylation profiling.

The biological role of DNA methylation

DNA methylation is one of the most well-studied epigenetic mechanisms and is involved in widespread biological processes, including but not limited to regulation of gene expression, imprinting, chromosomal stability, embryogenesis, X-inactivation, and tumorigenesis [11,12]. The commonest form of DNA methylation, the addition of a methyl group (CH_3) to cytosine, typically inhibits transcription and suppresses gene expression [13,14]. In non-embryonic human DNA, this occurs almost exclusively at specific sites termed CpGs, in which a cytosine and guanidine residue are situated next to one another. Under normal circumstances, most of these sites are heavily methylated, while specific CpG-dense regions (CpG islands), often located near gene promoters, remain unmethylated to allow gene expression. By modifying the methylation state of these CpG islands, gene expression can be 'switched' on or off allowing complex patterns of transcriptional activity.

As a regulator of gene expression, DNA methylation plays an extensive role in cancer biology [15]. Illustrating this, tumorigenesis is associated with significant changes in patterns of DNA methylation. In normal tissues, ~4-6% of cytosines in

genomic DNA are methylated. In contrast, cancer cells experience a large overall loss of DNA methylation and acquire distinct patterns of hypo- and hypermethylation [16]. In some instances, such as *MGMT* promoter hypermethylation, these patterns may also be directly clinically relevant [17]. As a result, DNA methylation has attracted widespread attention for investigation across cancers, something which has been facilitated by the advent of fast, cost effective methods for methylation profiling.

Methods for detecting DNA methylation in neuro-oncology

Several methods exist for assaying a given tissue's methylation profile. However, the most utilised tool for assaying human DNA methylation in the neuro-oncology field are Illumina BeadChip arrays. Originally arising from modified SNP genotyping chips, these arrays have been expanded over several iterations to cover 853,307 CpG sites across the human genome in the most recent chips [18]. Methylation arrays have a number of features which make them attractive for use in a clinical setting including comparatively low cost per sample, a quick turnaround time from sample acquisition to analysable data, high reproducibly due to in-built control probes, and scalability to accommodate large or small cohorts. Moreover, they are amenable to small amounts of starting material and suboptimal quality DNA, such as that derived from archival FFPE tissue. This latter point is particularly relevant, as FFPE material is significantly more prevalent in a clinical setting than fresh frozen tissue. For research, the widespread use of a unified platform facilitates comparative studies of discrete datasets and enables data integration in a previously unprecedented manner.

The disadvantages of reliance on a single proprietary platform are obvious. A major challenge for the field is to develop classification techniques that are platform independent and will work with array data, bisulphite sequencing, mass spectrometry or whichever technique is most suitable for the clinical context. While the field has relied heavily on array-based technology, algorithms have been developed that are agnostic to the platform and are compatible with alternative methods including mass spectrometry; this approach is perhaps best developed in medulloblastoma [19].

Advances in tumour classification and pathology

The first major impact of methylation profiling has been on the classification of CNS tumours. For some brain tumour types like medulloblastoma, methylation profiling has defined the molecular subtypes of the tumour to an exceedingly high level of granularity [20]. For others, such as low-grade epilepsy associated tumours, it has helped resolve the histological diagnostic uncertainties. In contrast, in the now defunct group of 'CNS PNETs', it has contributed to the disbanding of the diagnostic category [4,8]. Figure 1 shows a summary of the paediatric subtypes of CNS tumours recognisable by their methylation profile. We will illustrate the impact by considering two specific examples, covering low- and high-grade tumours respectively.

Embryonal tumours

The commonest embryonal tumour of the brain is medulloblastoma and this tumour type reveals complex underlying molecular architecture which has a pronounced effect on clinical management. Detailed advances in classification and stratification of medulloblastoma subgroups have been achieved thanks to extended periods of data collection and combination of multicentre data. The current iteration of the WHO classification recognises four histologically distinct entities (classic, desmoplastic/nodular, extensive nodularity, and large cell/anaplastic) and four genetically defined subgroups (WNT-activated, SHH-activated *TP53*-wildtype, SHH-activated *TP53*-mutant and non-WNT/non-SHH) [21]. Clustering of large methylation array datasets reproducibly identifies molecularly defined groups of medulloblastoma, separating non-WNT/non-SHH into Group 3 and Group 4 [22]. However, recent studies analysing larger tumour cohorts indicate much greater molecular granularity across Groups 3 and 4 [20,23-25], and a consensus of 8 separate groups has been reached; associated with different clinical outcomes [20] (Figure 1A).

Resolving the classification of the non-medulloblastoma embryonal tumours,

previously described by the now redundant term CNS-PNET, into multiple subtypes clearly highlights the importance of detailed cohort analysis and requirement of non-subjective diagnostic approaches. These tumours describe a mixed-bag of undifferentiated high-grade tumours [9]. Detailed methylation analysis reveals that many of these tumours can be re-diagnosed as distinct previously recognised embryonal entities, including embryonal tumour with multilayered rosettes (ETMR: formerly ETANTR, medulloepithelioma and ependymoblastoma) and atypical teratoid/rhabdoid tumours (AT/RT). Many can also be reclassified as non-embryonal tumours such as ependymoma and high-grade gliomas, or non-neuroepithelial tumours such as sarcomas (Figure 1D) [8]. Additionally, four novel tumour types defined by recurrent genetic features were identified amongst an archival cohort of CNS-PNET using methylation array data [8]. These included CNS neuroblastoma with *FOXR2* alteration, CNS Ewings tumour EFT with *CIC* alteration, High-grade neuroepithelial tumours (HGNET) with *BCOR* alteration and HNGET with *MN1* alterations. However, many of these genetically defined tumour groups are not exclusively embryonal tumours. For example, only neuroblastoma with *FOXR2* alteration could be considered to belong to the class of embryonal tumours, while the remaining cases more often either histologically or genetically resemble sarcomas (CNS Ewings tumour EFT with *CIC* alteration) or glial tumours (HGNET with *BCOR* alteration, HGNET with *MN1* alteration). This study is a prime example of how retrospective methylation studies can lead not only to discovery of new tumour types but can completely alter our approach to classification. A major ongoing challenge is to use clinical outcome data in these new groups to direct patient stratification and improve outcome prediction. However, optimal treatment strategies and appropriate trial designs for such rare tumours remain particularly challenging.

Low-grade tumours

In the last few years, methylation studies have resulted in the identification and characterisation of a spectrum of novel low-grade tumour subgroups, as well as facilitating the refinement of existing entities. Molecular data derived from methylation studies has helped define the underlying biology of tumours where conventional histological practice is inaccurate and poorly reproducible. An example of this are recent studies in low-grade glioneuronal tumours, a group of tumours that usually present with devastating epilepsy. Traditional histological descriptions are dominated by two histological variants – dysembryoplastic neuroepithelial tumour and ganglioglioma. However, in clinical practice, cases often lack distinctive histological features and consequently pathological diagnosis shows extremely poor inter-observer agreement [7]. We were able to use methylation profiling data to classify such glioneuronal tumours, including those with uninformative histology, by unsupervised clustering into two distinct molecular subgroups that only partially correlated with histology [4]. This clustering solution was also shown in a subsequent independent cohort of tumours [6]. Further inspection of our glioneuronal cohort revealed distinct features enriched within each group: *BRAF* V600E mutations and an astrocytic expression phenotype in one group, in contrast with *FGFR1* mutations and an oligodendrocyte precursor cell phenotype in the other. These data suggest that DNA methylation profiling represents a valuable tool in the classification of low-grade glioneuronal tumours, particularly those with poorly defined or indistinct histological features. More generally, this may also bear relevance in a wider range of cases where conventional histological diagnosis is complicated, for example in the case of small or sub-optimal biopsies or where numerous differential diagnoses are possible.

In addition to refining existing diagnoses, a handful of potential new low-grade entities have been identified with the help of methylation profiling data. One such example is the recent description of polymorphous low-grade neuroepithelial tumours of the young (PLNTY), a subgroup of oligodendroglioma-like tumours with a distinct DNA methylation profile that sets them apart from other low-grade neuroepithelial tumours [26]. PLNTY also harbour a variety of abnormalities affecting

BRAF, *FGFR1* and *FGFR1*. Interestingly, this group of tumours also displays marked histological variability which may indicate several distinct sub-entities or, like the glioneuronal tumours detailed previously, suggests that histological appearance does not always predict the underlying biology of the tumour.

Outside of novel low-grade tumour entities and the refinement of existing subgroups, methylation has also assisted in the further characterisation of rare tumour entities, such as rosette forming glioneuronal tumours (RGNT). Recently, methylation array profiling on a large cohort of tumours diagnosed as RGNT, shows that they possessed a distinct methylation profile and clustered away from other low-grade CNS tumours [27]. This molecularly defined group of RGNTs all possessed *FGFR1* hotspot mutations, occurring alongside *PIK3CA* mutations in a majority, and in some cases mutations in *NF1*. As with the aforementioned glioneuronal tumour studies, the alignment of methylation profiling and clustering with recurrent mutations illustrates the ability of methylation data to pull out biologically distinct groups in order that they can be characterised further by complementary methods.

The unknowns

One of the more difficult questions is to determine how many unknown tumours types remain to be discovered. The recent proliferation of reports describing new entities would suggest that there are still significant numbers of tumours yet to be accurately defined. However, not only it is difficult to estimate how many undefined tumours exist, it also remains uncertain how many of these new entities will resolve into clinically significant tumour types.

There are a few studies that do offer an estimate and suggest that there are a significant number of unknown types. For example, in the Sturm *et al.* paper of tumours historically called CNS-PNETs, 15% of cases did not classify as distinct or known entities [8].

We collected a UK-based cohort of high-grade CNS tumours that had been

traditionally difficult to diagnose [28]. We were able to confidently diagnose 51% of these tumours using conventional neuropathology without advance molecular techniques. This suggests that advances in our understanding of tumour subtypes, along with conventional reagents such as mutation specific antibodies, was enough to solve many historical diagnostic conundrums. We were able to solve a further 17% of these cases with the assistance of a methylation array, leaving 32% of cases that defied diagnosis by any technique. Having made an estimate of the impact of technical factors, we estimated that 15% of the total cohort were novel entities, remarkably like the Sturm *et al.* paper [8]. Therefore, there remains a significant group of undefined tumours that will require clinical, molecular and pathological characterisation.

Implementation of diagnostic algorithms in clinical practice

These examples show how methylation profiling has redefined the spectrum of recognised tumours types. The challenge now is how to incorporate these findings into clinical practice. As early as 2015, models were designed to allow classification of CNS tumours according to their methylation profiles [29]. More recently, efforts have been made to develop diagnostic tools for the prospective classification of tumours by comparison against extensive training sets of archival tumours. The most prominent of these is the Molecular Neuropathology (MNP) platform (https://www.molecularneuropathology.org/mnp), a tool that utilises a random forest classification method to classify tumours against a reference training set of 2,801 CNS tumours [30,31]. By comparing the methylation status of a subset of 10,000 CpGs the MNP model generates a score representing how well a given sample matches each of the 91 diagnostic classes within the training set (Figure 2). Since its release, this platform has been incorporated into clinical pipelines at several centres across the world [28,30,32,33].

In addition to the MNP platform, other groups have designed and built classification models for the refinement of diagnosis within individual tumour groups. Two such examples are the minimal methylation classifier (MIMIC) and methylation array

classifier (MAC) for identification of medulloblastoma subtypes from MassARRAY and Illumina methylation array data, respectively (http://medulloblastomadiagnostics.ncl.ac.uk) [19].

Diagnostic impact

The implementation of methylation algorithms offers a standardised approach to molecular diagnosis based on the community's existing knowledge. In assessing the impact of these, the temptation has been to compare them directly to traditional approaches in well-defined curated cohorts. Indeed, using data from several centres, the MNP methylation classifier shows good correlation with standard neuropathology [28,30,33]. However, a more realistic assessment of its impact is to determine the value it adds above standard-of-care technology, such as histopathology, and to do so in a real-world setting. We found using the UK paediatric experience with conservative criteria that the MNP classifier had diagnostic value in real time and above standard practice in 35% of all cases [28].

All diagnostic testing is subject to error and misdiagnosis, and while an algorithm-based technique removes the subjectivity inherent to traditional diagnosis, it is likely to be associated with other biases and errors. The existing studies report low error rates. For example, in our study we found only 3 cases (0.6%) where the algorithm confidently classified the tumour but in a misleading way [28]. However, none of these cases would have led to a genuine clinical error if taken in the full clinical and pathological context. This emphasises the importance of integrating innovative technologies in the context of a reasonable clinical context and standard of care testing.

More concerning are the cases where the algorithm provides a lower confidence score. We found a much higher rate of misleading diagnoses when the confidence of the algorithm was low. What is more, we found that the confidence of the algorithm to classify tumours depends not only on the tumour type but on the clinical question being asked. This suggests that the assessment of errors and impact will be

dependent on the clinical context.

Diagnostic classification is important *per se* because it allows certainty and prognosis for patients, but it is more important when it directs a change in treatment. Few studies of molecular profiling have determined the impact on conventional treatment, instead relying necessarily on theoretical predictions of actionability [34]. However, methylation profiling differs to sequencing technology as it focuses on tumour subtyping rather than identifying individual targetable sequence variants. As such, one would predict that changes in diagnosis based on the methylation profile might lead to changes in conventional treatment. Indeed, in paediatric neuro-oncology, we estimated that methylation profiling would change conventional treatment in 4% of children [28].

Biologically informed classification methods like methylation profiling can be used to stratify patients into clinically relevant risk groups and subtypes that are of prognostic important. Moreover, these methods are able to highlight known and novel pathways, often targetable, that may drive tumour formation and progression. This additional granularity may in turn direct a change in treatment. We estimated that if potential trial targets are including then methylation profiling could change treatment in 10% of children with brain tumours [28].

Implementation of methylation profiling also informs neuropathologists because it acts as a learning tool. Rare, difficult to diagnose variants, and novel groups identified by their molecular basis become more readily recognised histologically as a result of feedback from profiling, potentially speeding up the diagnostic process. Also, as more is discovered about specific tumour subgroups, relevant diagnostic markers are also being identified. For example, the RELA subgroup of ependymoma can be identified by p65 nuclear staining and L1CAM staining [35]; while the posterior fossa A and B subgroups are distinguished by H3K27me3 staining [36]. This means that the diagnostic impact of the array technology will vary as other techniques develop in parallel.

Inference of copy number data

In addition to profiling the methylation status of a given sample, BeadChip methylation arrays can be utilised to derive broad copy number (CN) data, for which several tools, such as *conumee,* are available [37] (Figure 3). This additional data allows the arrays to be used for the identification of large-scale chromosomal gains and losses, focal gene amplifications and deletions and may be suggestive of gene fusion events [38]. The main caveats are that CN data is restricted to changes that are covered by probes targeted by the array and that analysis of subtle or focal changes on frequently 'noisy' plots introduces some subjectivity compared to other CN detection methods. Nevertheless, the availability of this data allows for broad detection of CN alterations with little extra effort. In our experience detection of CN events by methylation arrays results in comparable specificity, although lower sensitivity compared to the current gold standard methods [28]. DNA extracted for methylation arrays is usually derived from bulk tissue, which may mask low-frequency events in a heterogeneous cell population compared to cytogenetic data such as fluorescent *in-situ* hybridisation.

Epigenetic biomarkers

Aside from changes in overall methylation profile, specific features and patterns of methylation can provide diagnostic and prognostic utility, allowing subtyping of tumours and risk stratification at a more granular level. The most well-recognised example is hypermethylation of *MGMT*, which predicts response to alkylating agents in adult glioblastoma. Specifically, hypermethylation of the *MGMT* promoter is a positive predictor of drug-responsiveness and survival [17]. More generally, the CpG island methylator phenotype (CIMP) is an important prognosticator across gliomas [39]. This phenotype refers to a recurrent pattern of genome-wide hypermethylation of CpG islands, first described in colorectal cancer, but later seen in a range of tumours including those affecting the CNS. In the context of astrocytomas, CIMP was first described in glioblastoma, often occurring alongside mutations affecting the gene *IDH*, but has also been observed in low- and high-grade glioma[40,41].

Future perspectives

Methylation profiling has had a significant impact on central nervous system tumour pathology by driving discovery of novel molecular subgroups, characterisation of existing entities alongside their underlying biology, and modification of diagnostic approaches. The distinctive methylation profiles of other non-CNS tumours suggest that a similar approach is likely to be helpful in diagnosing other tumours and algorithms are being developed in a wide range of tumour types including soft tissue tumours and uterine tumours [42-44].

The use of the methylation profiling for clinical diagnosis raises challenges for how we practice diagnostic pathology beyond the implementation of this specific technology. The first is that it introduces for the first time a truly algorithm-based technique into routine clinical histopathology. This is important for several reasons. First, used wisely an algorithm has the potential to reduce errors because it is by its nature an objective process and eliminates the subjectivity of opinion-based diagnosis. However, this puts an onus on the user to fully understand the limitations of the technique and to understand when the objective predictions are incorrect. The second change will be how we use advanced techniques in diagnosis. The traditional approach to diagnosis is to first generate an opinion-based hypothesis on the morphology and clinical features and then to test this hypothesis by more objective molecular techniques. The introduction of algorithm-based techniques suggests an alternative model where the objective algorithm-based technique is the first step. In this context, the role of the expert pathologist is to check that the diagnosis made by the algorithm fits the clinical and pathological findings, to do additional tests to support or refute the diagnosis, and to solve cases that cannot be solved by the algorithm. How this change happens in practice and what impact it has is uncertain and needs to be tested. It also raises issues relating to turnaround times and health economics that have not been thoroughly examined. Furthermore, the impact in different parts of the world with access to different specialist expertise and different health care models is completely untested.

Finally, the emergence of algorithm-based classification will ultimately change the way that we develop as expert pathologists, learning from, refining, and feeding back into these new diagnostic tools as they develop (Figure 4). As such, it is important that going forwards diagnostic pathologists are proficient in the application of these approaches and aware of their limitations, to accurately interpret the additional data they provide and flag discrepancies where necessary.

In a practice where one undertakes a methylation array on every case, the pathologist receives an ‘objective’ answer for all cases a week or two after making the initial diagnosis. This is in contrast with the traditional model where one obtained a subjective expert second opinion only on the small proportion of cases that were challenging. The array effectively provides rapid objective feedback in an unselected manner across the whole practice. How this improves diagnostic practice is, as of yet, untested, but it is potentially a very powerful form of both quality assurance and continuing professional development.

Acknowledgements

We are grateful for funding from the Brain Tumour Charity, Children with Cancer UK, Great Ormond Street Hospital (GOSH) Children's Charity, Olivia Hodson Cancer Fund, Cancer Research UK and the National Institute of Health Research (NIHR). Our work is supported by the INSTINCT network funded by The Brain Tumour Charity, GOSH Children's Charity and Children with Cancer UK, and the EVEREST centre funded by The Brain Tumour Charity. All research at GOSH NHS Foundation Trust and UCL Great Ormond Street Institute of Child Health is made possible by the NIHR GOSH Biomedical Research Centre. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health.

Author contribution statement

All authors contributed to writing this review and reviewing the final manuscript.

Ethical statement

As a review article, there are no specific ethical issues.

References

1. Cancer Statistics for the UK. 2015 May 13. Available from: https://www.cancerresearchuk.org/health-professional/cancer-statistics-for-the-uk

2. Cunningham RM, Walton MA, Carter PM. The Major Causes of Death in Children and Adolescents in the United States. N Engl J Med. 2018 Dec 20;379(25):2468–75.

3. Ness KK, Hudson MM, Ginsberg JP, Nagarajan R, Kaste SC, Marina N, et al. Physical Performance Limitations in the Childhood Cancer Survivor Study Cohort. JCO. 2009 May 10;27(14):2382–9.

4. Stone TJ, Keeley A, Virasami A, Harkness W, Tisdall M, Izquierdo Delgado E, et al. Comprehensive molecular characterisation of epilepsy-associated glioneuronal tumours. Acta Neuropathol. 2018 Jan;135(1):115–29.

5. Stone TJ, Rowell R, Jayasekera BAP, Cunningham MO, Jacques TS. Review: Molecular characteristics of long-term epilepsy-associated tumours (LEATs) and mechanisms for tumour-related epilepsy (TRE). Neuropathol Appl Neurobiol. 2018 Feb;44(1):56–69.

6. Blümcke I, Coras R, Wefers AK, Capper D, Aronica E, Becker A, et al. Review: Challenges in the histopathological classification of ganglioglioma and DNT: microscopic agreement studies and a preliminary genotype-phenotype analysis. Neuropathol Appl Neurobiol. 2019 Feb;45(2):95–107.

7. Blümcke I, Aronica E, Becker A, Capper D, Coras R, Honavar M, et al. Low-grade epilepsy-associated neuroepithelial tumours - the 2016 WHO classification. Nat Rev Neurol. 2016 Dec;12(12):732–40.

8. Sturm D, Orr BA, Toprak UH, Hovestadt V, Jones DTW, Capper D, et al. New Brain Tumor Entities Emerge from Molecular Classification of CNS-PNETs. Cell. 2016 Feb 25;164(5):1060–72.

9. Schwalbe EC, Hayden JT, Rogers HA, Miller S, Lindsey JC, Hill RM, et al. Histologically defined central nervous system primitive neuro-ectodermal tumours (CNS-PNETs) display heterogeneous DNA methylation profiles and show relationships to other paediatric brain tumour types. Acta Neuropathol. 2013 Dec;126(6):943–6.

10. Ellison DW, Kocak M, Figarella-Branger D, Felice G, Catherine G, Pietsch T, et al. Histopathological grading of pediatric ependymoma: reproducibility and clinical relevance in European trial cohorts. J Negat Results Biomed. 2011 May 31;10(1):7–13.

11. Greenberg MVC, Bourc'his D. The diverse roles of DNA methylation in mammalian development and disease. Nat Rev Mol Cell Biol. 2019 Oct;20(10):590–607.

12. Robertson KD. DNA methylation and human disease. Nat Rev Genet. 2005 Aug;6(8):597–610.

13. Siegfried Z, Eden S, Mendelsohn M, Feng X, Tsuberi BZ, Cedar H. DNA methylation represses transcription in vivo. Nat Genet. 1999 Jun;22(2):203–6.

14. Deaton AM, Bird A. CpG islands and the regulation of transcription. Genes & Development. 2011 May 16;25(10):1010–22.

15. Kulis M, Esteller M. Chapter 2 - DNA Methylation and Cancer. 1st ed. Vol. 70, Epigenetics and Cancer, Part A. Elsevier Inc; 2010. 30 p.

16. Esteller M. Aberrant DNA methylation as a cancer-inducing mechanism. Annu Rev Pharmacol Toxicol. 2005;45:629–56.

17. Olson RA, Brastianos PK, Palma DA. Prognostic and predictive value of epigenetic silencing of MGMT in patients with high grade gliomas: a systematic review and meta-analysis. J Neurooncol. 2011 Apr 27;105(2):325–35.

18. Pidsley R, Zotenko E, Peters TJ, Lawrence MG, Risbridger GP, Molloy P, et al. Critical evaluation of the Illumina MethylationEPIC BeadChip microarray for whole-genome DNA methylation profiling. Genome Biology. 2016 Oct 6;:1–17.

19. Schwalbe EC, Hicks D, Rafiee G, Bashton M, Gohlke H, Enshaei A, et al. Minimal methylation classifier (MIMIC): A novel method for derivation and rapid diagnostic detection of disease-associated DNA methylation signatures. Sci Rep. 2017 Oct 18;7(1):13421–8.

20. Sharma T, Schwalbe EC, Williamson D, Sill M, Hovestadt V, Mynarek M, et al. Second-generation molecular subgrouping of medulloblastoma: an international meta-analysis of Group 3 and Group 4 subtypes. Acta Neuropathol. 2019 May 10;:1–18.

21. Pickles JC, Hawkins C, Pietsch T, Jacques TS. CNS embryonal tumours: WHO 2016 and beyond. Neuropathol Appl Neurobiol. 2018 Feb;44(2):151–62.

22. Taylor MD, Northcott PA, Korshunov A, Remke M, Cho YJ, Clifford SC, et al. Molecular subgroups of medulloblastoma: the current consensus. Acta Neuropathol. 2011 Dec 2;123(4):465–72.

23. Schwalbe EC, Lindsey JC, Nakjang S, Crosier S, Smith AJ, Hicks D, et al. Novel molecular subgroups for clinical classification and outcome prediction in childhood medulloblastoma: a cohort study. Lancet Oncol. 2017 Jul;18(7):958–71.

24. Northcott PA, Buchhalter I, Morrissy AS, Hovestadt V, Weischenfeldt J, Ehrenberger T, et al. The whole-genome landscape of medulloblastoma subtypes. Nature. 2017 Jul 19;547(7663):311–7.

25. Cavalli FMG, Remke M, Rampasek L, Peacock J, Shih DJH, Luu B, et al. Intertumoral Heterogeneity within Medulloblastoma Subgroups. Cancer Cell. 2017 Jun 12;31(6):737–754.e6.

26. Huse JT, Snuderl M, Jones DTW, Brathwaite CD, Altman N, Lavi E, et al. 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. 2017 Mar;133(3):417–29.

27. Sievers P, Appay R, Schrimpf D, Stichel D, Reuss DE, Wefers AK, et al. Rosette-forming glioneuronal tumors share a distinct DNA methylation profile and mutations in FGFR1, with recurrent co-mutation of PIK3CA and NF1. Acta Neuropathol. 2019 Sep;138(3):497–504.

28. Pickles JC, Fairchild AR, Stone TJ, Brownlee L, Merve A, Yasin SA, et al. DNA methylation-based profiling for paediatric CNS tumour diagnosis and treatment: a population-based study. The Lancet Child and Adolescent Health. 2020 Feb;4(2):121–30.

29. Danielsson A, Nemes S, Tisell M, Lannering B, Nordborg C, Sabel M, et al. MethPed: a DNA methylation classifier tool for the identification of pediatric brain tumor subtypes. Clin Epigenetics.2015;7(1):62–9.

30. Capper D, Jones DTW, Sill M, Hovestadt V, Schrimpf D, Sturm D, et al. DNA methylation-based classification of central nervous system tumours. Nature. 2018 Mar 22;555(7697):469–74.

31. Perez E, Capper D. DNA-Methylation-based Classification of Paediatric Brain Tumours. Neuropathol Appl Neurobiol. in press.

32. Karimi S, Zuccato JA, Mamatjan Y, Mansouri S, Suppiah S, Nassiri F, et al. The central nervous system tumor methylation classifier changes neuro-oncology practice for challenging brain tumor diagnoses and directly impacts patient care. Clinical Epigenetics; 2019 Dec 4;:1–10.

33. Jaunmuktane Z, Capper D, Jones DTW, Schrimpf D, Sill M, Dutt M, et al. Methylation array profiling of adult brain tumours: diagnostic outcomes in a large, single centre. Acta Neuropathologica Communications; 2019 Feb 18;:1–18.

34. George SL, Izquierdo E, Campbell J, Koutroumanidou E, Proszek P, Jamal S, et al. A tailored molecular profiling programme for children with cancer to identify clinically actionable genetic alterations. Eur J Cancer. 2019 Nov;121:224–35.

35. Gessi M, Giagnacovo M, Modena P, Elefante G, Gianno F, Buttarelli FR, et al. Role of Immunohistochemistry in the Identification of Supratentorial C11ORF95-RELA Fused Ependymoma in Routine Neuropathology. The American Journal of Surgical Pathology. 2019 Jan;43(1):56–63.

36. Panwalkar P, Clark J, Ramaswamy V, Hawes D, Yang F, Dunham C, et al. Immunohistochemical analysis of H3K27me3 demonstrates global reduction in group-A childhood posterior fossa ependymoma and is a powerful predictor of outcome. Acta Neuropathol. 2017 Jul 21;134(5):705–14.

37. conumee. Available from: http://bioconductor.org/packages/conumee/.

38. Capper D, Stichel D, Sahm F, Jones DTW, Schrimpf D, Sill M, et al. Practical implementation of DNA methylation and copy-number-based CNS tumor diagnostics: the Heidelberg experience. Acta Neuropathol. 2018 Aug;136(2):181–210.

39. Malta TM, de Souza CF, Sabedot TS, Silva TC, Mosella MS, Kalkanis SN, et al. Glioma CpG island methylator phenotype (G-CIMP): biological and clinical implications. Neuro-oncology. 2018 Apr 9;20(5):608–20.

40. Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010 May 18;17(5):510–22.

41. Turcan S, Rohle D, Goenka A, Walsh LA, Fang F, Yilmaz E, et al. IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. Nature. 2012 Feb 15;483(7390):479–83.

42. Koelsche C, Hartmann W, Schrimpf D, Stichel D, Jabar S, Ranft A, et al. Array-based DNA-methylation profiling in sarcomas with small blue round cell histology provides valuable diagnostic information. Modern Pathology. 2018 Aug;31(8):1246–56.

43. Kommoss FKF, Stichel D, Schrimpf D, Kriegsmann M, Tessier-Cloutier B, Talhouk A, et al. DNA methylation-based profiling of uterine neoplasms: a novel tool to improve gynecologic cancer diagnostics. J Cancer Res Clin Oncol. 2020 Jan;146(1):97–104.

44. Röhrich M, Koelsche C, Schrimpf D, Capper D, Sahm F, Kratz A, et al. Methylation-based classification of benign and malignant peripheral nerve sheath tumors. Acta Neuropathol. 2016 Jun;131(6):877–87.

45. Pajtler KW, Witt H, Sill M, Jones DTW, Hovestadt V, Kratochwil F, et al. Molecular Classification of Ependymal Tumors across All CNS Compartments, Histopathological Grades, and Age Groups. Cancer Cell. 2015 May;27(5):728–43.

Figure legends

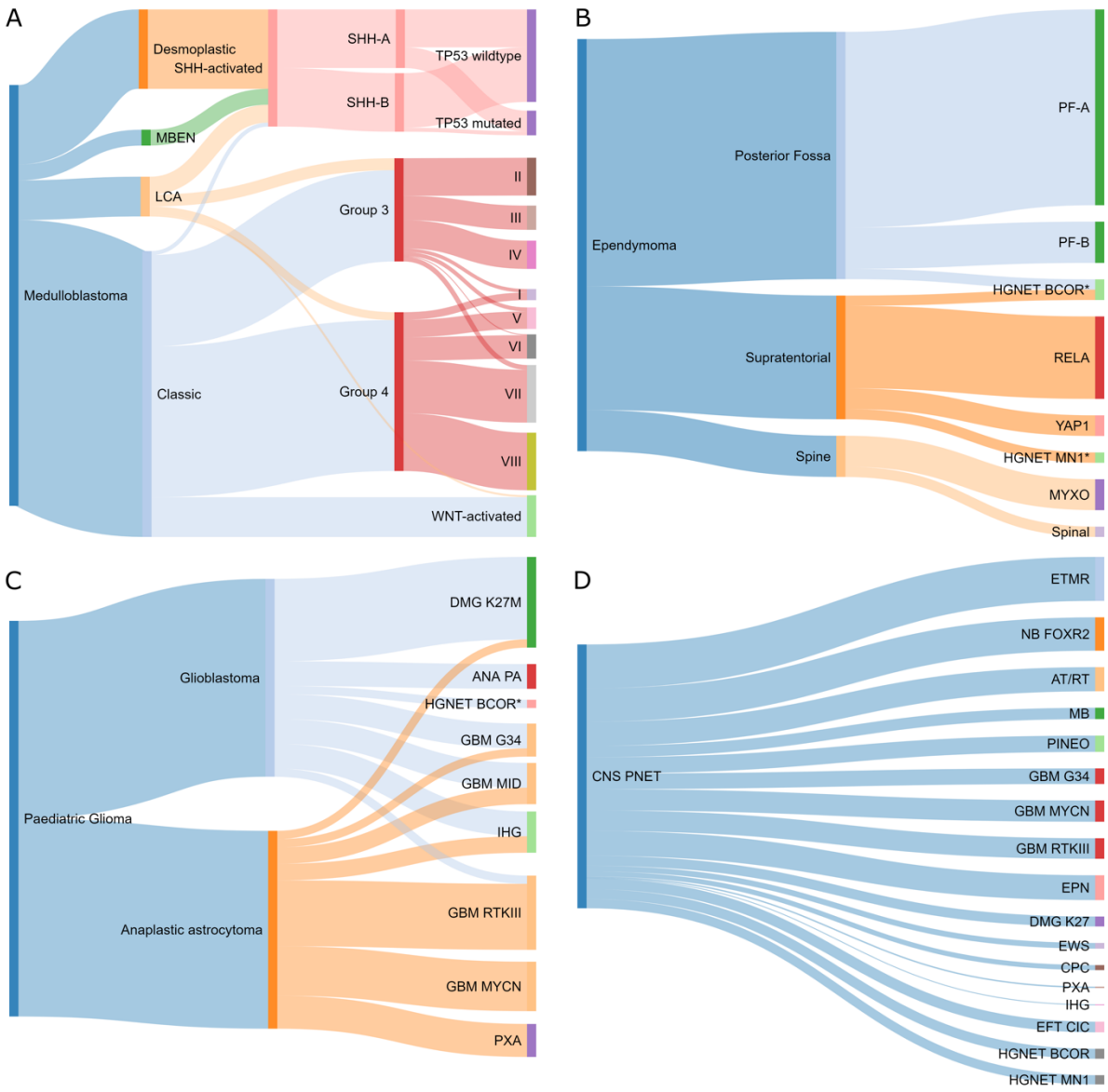

Figure 1. Advances in molecular classification of paediatric CNS tumours. Flow diagram describing the relationship between histological and molecular diagnosis. **(A)** There are four main histological subtypes of medulloblastoma, while the current MNP classifier can distinguish between five methylation classes (SHH-A, SHH-B, Group 3, Group 4 and WNT-activated), further molecular grouping has been described (e.g. I-VIII) along with specific prognostic factors (e.g. *TP53* status) associated with clinical outcomes [20]. **(B)** Ependymoma has multiple histological subtypes (clear cell, anaplastic, myxopapillary) and classification is based on location [45]. (C) Glial tumours can be divided histologically between astrocytoma

and glioblastoma, with multiple molecular subtypes now recognised ([28] & unpublished data) (D) Re-diagnosis of mixed tumours formerly known as CNS-PNET into known and novel subgroups [8,28]. * = Reclassification, do not belong to the broad tumour grouping. *Abbreviations:* MBEN: medulloblastoma with extensive nodularity, LCA: large cell/anaplastic, SHH-A: Infant group, SHH-B: child and adolescent group, DMG K27M: diffuse midline glioma with K27 mutation, ANA PA: anaplastic pilocytic astrocytoma, GBM G34: Glioblastoma with G34 mutation, GBM MID: glioblastoma midline, IHC: infantile hemispheric glioma, GBM RTKIII: glioblastoma IDH wildtype subclass RTKIII, GBM MYCN: glioblastoma IDH wildtype subclass MYCN, PXA: pleomorphic xanthoastrocytoma, PF-A: posterior fossa subgroup A, PF-B: posterior fossa subgroup B, HGNET: high grade neuroepithelial tumour, EFT: Ewing sarcoma family tumour, MB: medulloblastoma, EPN: Ependymoma, PINEO: pineoblastoma, EWS: Ewings sarcoma, CPC: Choroid plexus carcinoma, MYXO: myxopapillary.

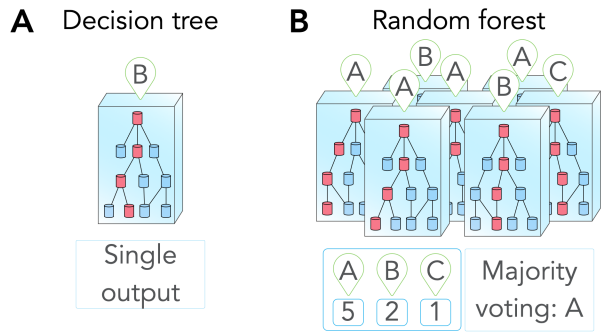

Figure 2 Random forests are collections of randomised decision trees. (A) A single decision tree is built from all variables in a dataset, and as a result classification can be vulnerable to the order in which variables appear in the tree. **(B)** Random forests avoid this by randomly selecting variables from the dataset to build many trees (e.g. a forest) with combinations of variables. In the case of the DKFZ classification model these variables are a subset of 10,000 CpG sites. These individual decision trees 'vote' for a class based on their own individual criteria and variable set, then the classification with the most 'votes' is considered the consensus classification.

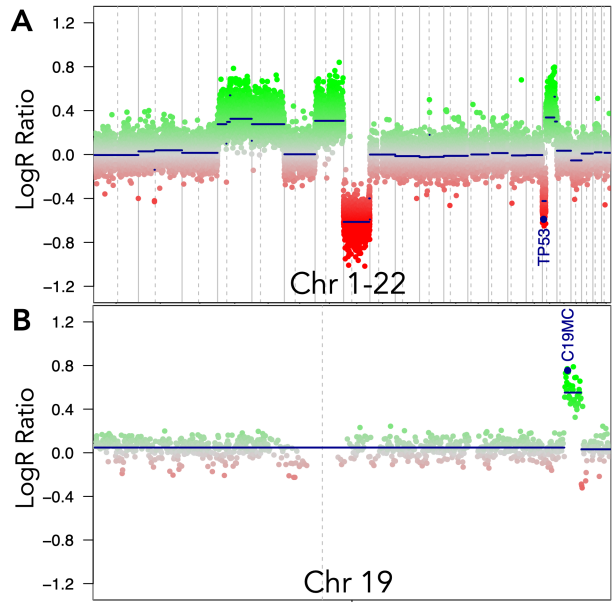

Figure 3 Large and small-scale copy number changes are observed in methylation array data. Methylation array data can be used for the detection of large- and small-scale copy number changes, including those characteristic of specific diagnostic groups. **(A)** Isochromosome 17q in a medulloblastoma. **(B)** C19MC miRNA cluster amplification in an embryonal tumour with multilayered rosettes (ETMR).

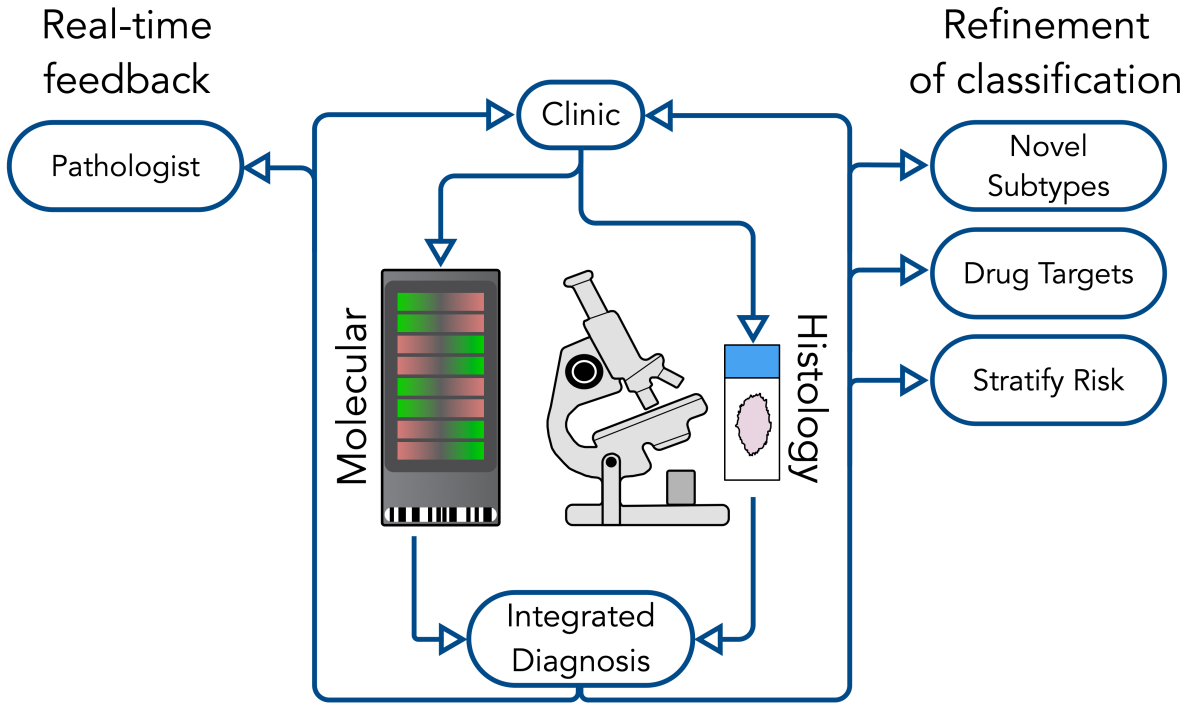

Figure 4 Integration of histological and molecular diagnoses as a positive feedback system. The integration of molecular data alongside traditional histology promotes continuous improvement of tumour diagnosis and treatment. Rapid molecular feedback acts as a process of continuous development for the diagnostic pathologist. Also understanding the underlying biology of the tumour allows for the identification of novel subtypes, novel treatment targets, and improved risk stratification, which can be fed back into the clinic to improve patient survival and quality of life.