SUPPLEMENTARY METHODS

Experimental pipeline for genome-wide DNA methylation profiling

After bisulfite conversion, extracted DNA was profiled on the Infinium MethylationEPIC BeadChip array (Illumina, San Diego, CA) in accordance with the manufacturer's instructions. DNA methylation data and statistical analysis were performed in R version 3.6.0 (www.R-project.org). Raw data files generated by the iScan array scanner were read and preprocessed using minfi Bioconductor package according to steps as implemented in Illumina GenomeStudio software.[1] In addition, probes with the following characteristics were removed: targeting of the X and Y chromosomes, a nucleotide polymorphism (dbSNP132 Common) within five base pairs of and including the targeted CpG-site, and probes not mapping uniquely to the human reference genome (hg19) after allowing for one mismatch. We used 786,425 probes for downstream analysis. Methylation profiles were examined by unsupervised clustering within the study cohort, and then with relevant reference classes from a published cohort of CNS tumors (2801 samples from 91 tumor and control classes, Gene Expression Omnibus GSE90496) (Supplementary Fig S4).[2] Results were visualized by using hierarchical clustering and t-SNE analysis as previously described.[2] In brief, the 10,000 most variably methylated probes measured by standard deviation across combined samples were selected. One-Pearson correlation was calculated as distance measure between samples. The raw data were also analyzed via the German Cancer Research Center (DKFZ) MolecularNeuropathology (MNP) brain tumor classifier (www.molecularneuropathology.org, version 11b4) by random forest algorithm, with

methylation class assignment carrying a calibrated score ≥0.70 considered as potentially relevant and those <0.70 considered as “not matching”.

Copy-number variation analysis based on Illumina methylation arrays was performed by using conumee Bioconductor package with default settings.[3] The combined intensities of all available CpG probes were normalized against control samples from normal brain tissue by using a linear regression approach. Twenty-six normal brain tissues profiled on MethylationEPIC array were used as controls. Copy-number plots were manually examined for copy-number alterations.

Whole-exome and -genome sequencing

Tumor and germline DNA exomes were captured by using the SureSelect Human All Exon V5 (Agilent Technologies, Santa Clara, CA, USA) platform and sequenced paired-end. FASTQ files derived from WES/WGS were aligned to hg38 by using BWA (v0.7.17) and then sorted and marked for duplicates by using the MarkDuplicates tool from Picard (http://broadinstitute.github.io/picard).[4] To account for systematic errors, base quality score recalibration was done using the BaseRecalibrator tool from GATK (v4.1.0) to produce BAM files ready for SNP and indel detection.

Variant Calling

Somatic variant calling was done by following the GATK best practices workflow using the tools within GATK (v.4.1.0).[5] Briefly, a panel of normals (PON) was constructed by using all available germline samples to account for technical artifacts for the tumor-only cohort; for each patient who also had germline material, a PON was similarly built, omitting that patient’s germline sample. All samples were then called for SNPs and indels

by using Mutect2 with a PON and matched germline when available. Germline variant calling was performed on all germline samples by using the tool HaplotypeCaller followed by hard filtering using the tool VariantFiltration with default parameters. All variant calls were filtered for sequencing artifacts by using the CollectSequencingArtifacts tool and required to have an alternate allele depth greater than 3, total sequencing coverage greater than 10, allele frequency greater than 0.2, and an ExAC allele frequency less than 0.0001.[6] Variants were then annotated by using ANNOVAR and required to reside within exonic regions and be functionally annotated as one of the following types of coding mutations: stop-gain, stop-loss, frameshift indel, splice-site, or inframe indel.[7] Variants were then filtered based on a list of causally implicated cancer genes from the COSMIC cancer gene census (Supplementary Table S4).[8]

RNA-seq analysis

FASTQ files from paired-end, strand-specific RNA sequencing were aligned by using STAR (v.1.6.0) to hg38 with gene annotations from Ensembl (v93).[9,10] Raw gene counts were input into DESeq2 (v1.24.0) to generate a normalized count matrix; expression boxplots were created by taking the $\log_2$(Normalized counts+1).[11] The likelihood ratio test was used to test for differences in gene expression across all PB subgroups and non-PB entities. Differential expression analysis was performed by using DESeq2 to compare each PB subgroup to the other subgroups; all p-values were then adjusted for multiple comparisons (|L2FC|>1; P-value <0.01). Gene-set enrichment analysis (GSEA) was done by using the R packages fgsea (v1.10.0) and gage (v.2.34.0). Only Gene Ontology (GO) terms from MsigDB that were significant (Normalized enrichment score >1 & adjusted P<0.05) in both methods were retained.[12-14] All significant GO terms are listed

(Supplementary Table S5). Unsupervised clustering was performed to group GO terms into functional modules based on a GO term distance matrix using 1-Jaccard's distance between two GO terms' gene lists. The resulting GO network was then plotted in Cytoscape (v.3.6.1).

SUPPLEMENTARY REFERENCES

1. Aryee MJ, Jaffe AE, Corrada-Bravo H, et al: Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30:1363-9, 2014
2. Hwang EI, Kool M, Burger PC, et al: Extensive Molecular and Clinical Heterogeneity in Patients With Histologically Diagnosed CNS-PNET Treated as a Single Entity: A Report From the Children's Oncology Group Randomized ACNS0332 Trial. Journal of Clinical Oncology 36:3388-3395, 2018
3. Hovestadt V, Zapatka M: conumee: Enhanced copy-number variation analysis using Illumina 450k methylation arrays. R package version 0.99 4, 2015
4. Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754-60, 2009
5. Van der Auwera GA, Carneiro MO, Hartl C, et al: From FastQ data to high confidence variant calls: the Genome Analysis Toolkit best practices pipeline. Curr Protoc Bioinformatics 43:11 10 1-33, 2013
6. Lek M, Karczewski KJ, Minikel EV, et al: Analysis of protein-coding genetic variation in 60,706 humans. Nature 536:285-91, 2016
7. Wang K, Li M, Hakonarson H: ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38:e164, 2010

8. Tate JG, Bamford S, Jubb HC, et al: COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res 47:D941-D947, 2019

9. Dobin A, Davis CA, Schlesinger F, et al: STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29:15-21, 2013

10. Aken BL, Ayling S, Barrell D, et al: The Ensembl gene annotation system. Database (Oxford) 2016, 2016

11. Love MI, Huber W, Anders S: Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol 15:550, 2014

12. Sergushichev A: An algorithm for fast preranked gene set enrichment. bioRxiv, 2016

13. Luo W, Friedman MS, Shedden K, et al: GAGE: generally applicable gene set enrichment for pathway analysis. BMC Bioinformatics 10:161, 2009

14. Liberzon A, Subramanian A, Pinchback R, et al: Molecular signatures database (MSigDB) 3.0. Bioinformatics 27:1739-40, 2011