LETTER

Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing

Volker Hovestadt[1]*, David T. W. Jones[2]*, Simone Picelli[1], Wei Wang[1], Marcel Kool[2], Paul A. Northcott[2], Marc Sultan[3], Katharina Stachurski[4], Marina Ryzhova[5], Hans-Jörg Warnatz[3], Meryem Ralser[3], Sonja Brun[6], Jens Bunt[7,8], Natalie Jäger[9], Kortine Kleinheinz[1,9], Serap Erkek[2,10], Ursula D. Weber[1], Cynthia C. Bartholomae[11,12], Christof von Kalle[11,12], Chris Lawerenz[13], Jürgen Eils[13], Jan Koster[8], Rogier Versteeg[8], Till Milde[14,15], Olaf Witt[14,15], Sabine Schmidt[16], Stephan Wolf[16], Torsten Pietsch[17], Stefan Rutkowski[18], Wolfram Scheurlen[19], Michael D. Taylor[20,21,22], Benedikt Brors[9], Jörg Felsberg[23,24], Guido Reifenberger[23,24], Arndt Borkhardt[4], Hans Lehrach[3], Robert J. Wechsler-Reya[6], Roland Eils[9,25,26,27], Marie-Laure Yaspo[3], Pablo Landgraf[4], Andrey Korshunov[28,29], Marc Zapatka[1], Bernhard Radlwimmer[1], Stefan M. Pfister[2,14] & Peter Lichter[1,27]

Epigenetic alterations, that is, disruption of DNA methylation and chromatin architecture, are now acknowledged as a universal feature of tumorigenesis[1]. Medulloblastoma, a clinically challenging, malignant childhood brain tumour, is no exception. Despite much progress from recent genomics studies, with recurrent changes identified in each of the four distinct tumour subgroups (WNT-pathway-activated, SHH-pathway-activated, and the less-well-characterized Group 3 and Group 4)[2–4], many cases still lack an obvious genetic driver. Here we present whole-genome bisulphite-sequencing data from thirty-four human and five murine tumours plus eight human and three murine normal controls, augmented with matched whole-genome, RNA and chromatin immunoprecipitation sequencing data. This comprehensive data set allowed us to decipher several features underlying the interplay between the genome, epigenome and transcriptome, and its effects on medulloblastoma pathophysiology. Most notable were highly prevalent regions of hypomethylation correlating with increased gene expression, extending tens of kilobases downstream of transcription start sites. Focal regions of low methylation linked to transcription-factor-binding sites shed light on differential transcriptional networks between subgroups, whereas increased methylation due to re-normalization of repressed chromatin in DNA methylation valleys was positively correlated with gene expression. Large, partially methylated domains affecting up to one-third of the genome showed increased mutation rates and gene silencing in a subgroup-specific fashion. Epigenetic alterations also affected novel medulloblastoma candidate genes (for example, LIN28B), resulting in alternative promoter usage and/or differential messenger RNA/ microRNA expression. Analysis of mouse medulloblastoma and precursor-cell methylation demonstrated a somatic origin for many alterations. Our data provide insights into the epigenetic regulation of transcription and genome organization in medulloblastoma pathogenesis, which are probably also of importance in a wider developmental and disease context.
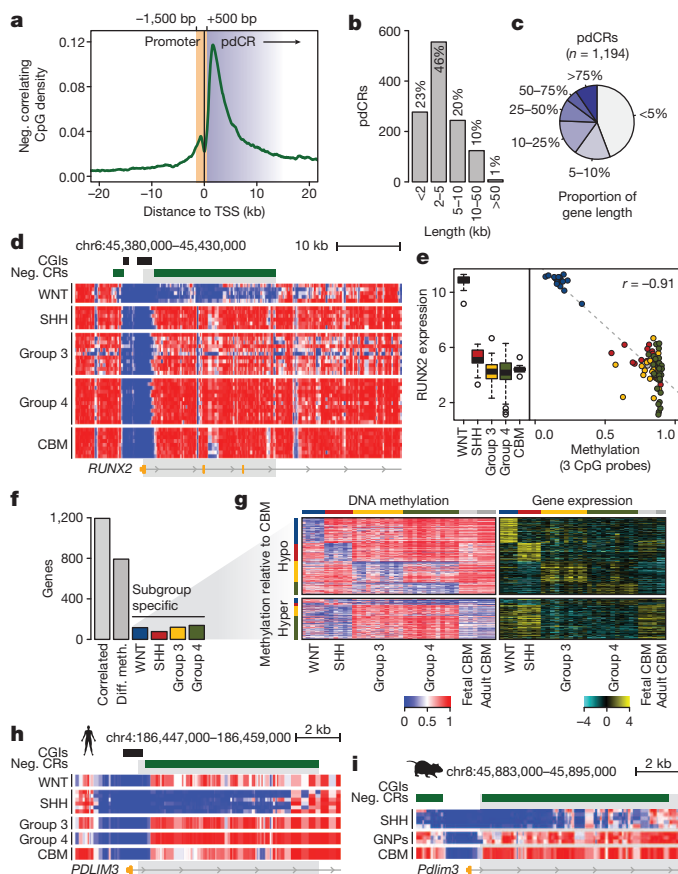
Medulloblastoma is an embryonal tumour with few differentiation markers and minimal normal-cell infiltration, making it well suited to epigenomic studies (where heterogeneity can be confounding). Thirty-four tumour and eight control samples were selected from a larger methylation array cohort for whole-genome bisulphite-sequencing analysis[5] (WGBS; average 28.5-fold average coverage and 91.4% of CpGs covered ≥10×; Supplementary Tables 1, 2 and Methods). Individual CpG sites showed a predominantly bimodal methylation pattern (either fully unmethylated or fully methylated), linked to CpG density (Extended Data Fig. 1a–c). Non-CpG methylation was low in tumours, but higher in adult cerebellum (Extended Data Fig. 1d, e), as recently reported for other brain regions[6].

Matching gene expression data enabled a comprehensive assessment of correlation with DNA methylation (see Methods), revealing negatively and positively correlating regions (CRs). Very few negative CRs were in promoter-associated CpG islands (CGIs). Of 13,380 genes with a promoter CGI, 333 (2.5%) were significantly differentially methylated, and only 79 showed negative correlation with expression (Extended Data Fig. 2a and Supplementary Table 3a). The WNK2 tumour suppressor[7], for example, was silenced by CGI hypermethylation in most WNT medulloblastomas and some SHH and Group 3 samples (Extended Data Fig. 2b, c). Overall, medulloblastomas showed slightly higher CGI methylation than normal cerebellum, but without evidence for a CGI methylator phenotype (Extended Data Fig. 2d). Thus, the classical notion of gene silencing through promoter hypermethylation was not a prominent feature, and many differentially methylated promoters were at genes that were not expressed even in unmethylated samples.

[1]Division of Molecular Genetics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [2]Division of Pediatric Neurooncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [3]Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, Berlin 14195, Germany. [4]Department of Pediatric Oncology, Hematology and Clinical Immunology, Heinrich Heine University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. [5]Department of Neuropathology, NN Burdenko Neurosurgical Institute, 4th Tverskaya-Yamskaya 16, Moscow 125047, Russia. [6]Tumor Initiation and Maintenance Program, National Cancer Institute (NCI)–Designated Cancer Center, Sanford-Burnham Medical Research Institute, 2880 Torrey Pines Scenic Drive, La Jolla, California 92037, USA. [7]Queensland Brain Institute, University of Queensland, QBI Building, St Lucia, Queensland 4072, Australia. [8]Department of Oncogenomics, AMC, University of Amsterdam, Meibergdreef 9, Amsterdam 1105 AZ, the Netherlands. [9]Division of Theoretical Bioinformatics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [10]European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, Heidelberg 69117, Germany. [11]Division of Translational Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [12]National Center for Tumor Diseases (NCT), Im Neuenheimer Feld 460, Heidelberg 69120, Germany. [13]Data Management Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [14]Department of Pediatric Oncology, Hematology & Immunology, Heidelberg University Hospital, Im Neuenheimer Feld 430, Heidelberg 69120, Germany. [15]Clinical Cooperation Unit Pediatric Oncology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [16]Genomics and Proteomics Core Facility, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [17]Department of Neuropathology, University of Bonn Medical Center, Sigmund-Freud-Strasse 25, Bonn 53105, Germany. [18]Department of Paediatric Haematology and Oncology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg 20246, Germany. [19]Cnopf'sche Kinderklinik, Nürnberg Children's Hospital, St.-Johannis-Mühlgasse 19, Nürnberg 90419, Germany. [20]Program in Developmental and Stem Cell Biology, The Arthur and Sonia Labatt Brain Tumour Research Centre, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [21]Division of Neurosurgery, Hospital for Sick Children, 555 University Avenue, Toronto, Ontario M5G 1X8, Canada. [22]Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, Ontario M5S 1A8, Canada. [23]Department of Neuropathology, Heinrich Heine University Düsseldorf, Moorenstrasse 5, Düsseldorf 40225, Germany. [24]German Consortium for Translational Cancer Research (DKTK), German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [25]Institute of Pharmacy and Molecular Biotechnology (IPMB), University of Heidelberg, Heidelberg 69120, Germany. [26]Bioquant Center, University of Heidelberg, Im Neuenheimer Feld 267, Heidelberg 69120, Germany. [27]Heidelberg Center for Personalised Oncology (DKFZ-HIPO), Im Neuenheimer Feld 280, Heidelberg 69120, Germany. [28]Department of Neuropathology, University of Heidelberg, Im Neuenheimer Feld 220, Heidelberg 69120, Germany. [29]Clinical Cooperation Unit Neuropathology, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 220-221, Heidelberg, 69120 Germany.
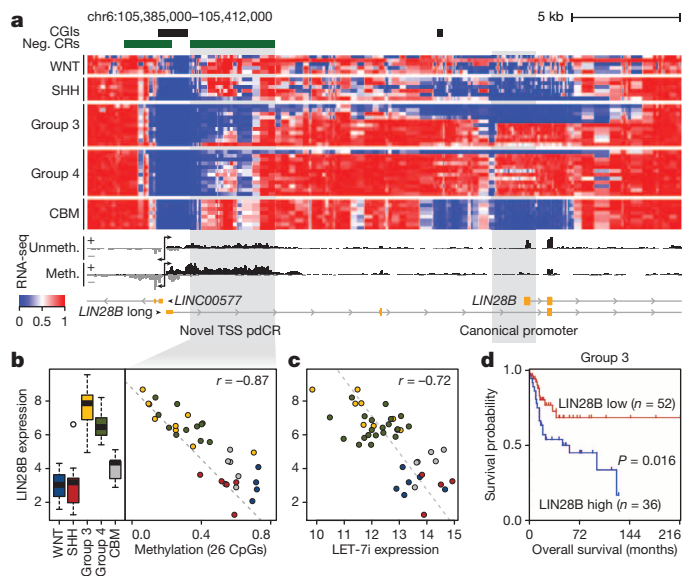*These authors contributed equally to this work.

Much more abundant were negative CRs extending several kilobases (kb) downstream of the promoter (promoter downstream (pd)CRs), affecting 1,194 genes (Supplementary Table 3b). Correlation density peaked at ~2 kb downstream of the transcription start site (TSS), with most pdCRs extending much farther (often tens of kilobases, and sometimes greater than 50 kb; Fig. 1a–c). For example, *RUNX2*, encoding a WNT-activated transcription factor[8], showed a hypomethylated pdCR and high expression exclusively in WNT medulloblastomas (Fig. 1d, e). Overall, pdCRs were found in 8.4% of expressed genes, with many showing tumour-subgroup-specific differential methylation (Fig. 1f, Extended Data Fig. 3a, b and Supplementary Table 3b). Methylation and expression levels of genes harbouring a pdCR clearly distinguished between the tumour subgroups (Fig. 1g). Almost 20% of genes specifically expressed by a subgroup contain a pdCR (636 out of 3,304), suggesting that this methylation pattern plays an important part in determining the distinct transcriptomes of molecular variants of medulloblastoma (Supplementary Table 3b, c).

Analysis of publicly available WGBS data from various tissues (Supplementary Table 2d) revealed pdCRs as a universal feature, more prominent in homogeneous cell populations (for example, sorted neuronal or haematopoietic lineages versus infiltrative tumours or whole organs). A similar pattern in CpG 'shores' and 'shelves' (up to 2 kb and 2–4 kb from CGIs, respectively) has been described in some tissues using array-based methods or small WGBS cohorts, and has been linked to tissue-specific expression patterns[9–11] and to the presence of H3K4me3 (ref. 12), an active chromatin mark. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) in two medulloblastoma cell lines (D425 and MEB-MED-8A) confirmed an overlap of pdCRs and elevated H3K4me3 (Extended Data Fig. 3c, d). Our results highlight the extent of pdCRs as an abundant, previously underappreciated pattern of correlation between methylation and expression. Possible explanations for this hypomethylation range from passive methylation loss during cell division at highly RNA-polymerase-occupied regions (inaccessible to DNA methyltransferases), to a regulated process that actively marks genes for increased transcription.

We also produced WGBS profiles for murine SHH medulloblastomas arising owing to inactivation of patched homologue 1 (*Ptch1*; *Math1-creER^{T2}:Ptch1^{fl/fl}* (ref. 13)) together with matched purified cerebellar granule neuron precursors (GNPs)—the precursor for SHH medulloblastoma and therefore the optimal normal control. When comparing mouse tumours and precursor cells, 161 pdCRs larger than 2 kb were detected (Supplementary Table 3d). For example, *Pdlim3* shows a clear pdCR and high expression in human and mouse SHH medulloblastoma, but not in GNPs (Fig. 1h, i, and Extended Data Fig. 3e, f). Other known SHH medulloblastoma genes, such as *Ptch1*, *Cdk6* and *Boc*, showed a similar pattern. Thus, although some DNA methylation differences between the subgroups are probably linked to an epigenetic 'fingerprint' or 'memory' of developmental lineages, alterations also arise somatically during tumour development.

Figure 1 | Negative correlation between methylation and expression is enriched in extensive regions downstream of promoters. a, Density plot of CpGs negatively (neg.) correlating with gene expression with respect to distance from the TSS (*n* = 36 samples). **b, c**, pdCR length in kb (**b**) and as a proportion of gene length (**c**). **d**, Methylation plot for *RUNX2* (*n* = 42), showing a large WNT-medulloblastoma hypomethylated pdCR extending ~19 kb downstream of the promoter. Grey box, region of interest; red, methylated; blue, unmethylated. **e**, Correlation of expression and methylation for the *RUNX2* pdCR in the extended validation cohort (*n* = 95). CBM, cerebellum; *r*, Pearson's correlation coefficient. **f**, Overview of genes associated with a pdCR (correlated), those that are significantly differentially methylated (Diff. meth.) across medulloblastoma subgroups and controls, and those that are specifically differentially methylated in a given tumour subgroup. **g**, Heat-map representation of methylation (*n* = 42) and gene expression (*n* = 36) of subgroup-specific pdCRs. **h, i**, pdCR at *PDLIM3* in human (*n* = 42) (**h**) and murine (*n* = 9) (**i**) SHH medulloblastoma, but not in matched GNPs, indicating a somatic origin.

Figure 2 | Differential methylation around *LIN28B* reveals a novel promoter, tightly correlated with expression in Group 3 and Group 4 medulloblastomas. a, Methylation plot for the *LIN28B* promoter region (*n* = 42), with a novel TSS giving rise to an alternative transcript (*LIN28B* long). Arrows indicate bidirectional enhancer RNA transcription. Representative RNA-seq data are shown for samples methylated (Meth.) or unmethylated (Unmeth.) at the canonical promoter. Red, methylated; blue, unmethylated. CBM, cerebellum. **b**, Correlation of expression and methylation of an upstream CR at the novel TSS. *r*, Pearson's correlation coefficient. **c**, Correlation of *LIN28B* and *LET-7i* (*n* = 59). **d**, Association of *LIN28B* expression with overall survival in Group 3 medulloblastoma (Kaplan–Meier analysis, *P*, log-rank test). See also Extended Data Fig. 5f.

We next sought to identify novel promoters regulated by pdCRs. RNA sequencing (RNA-seq) data for 43 medulloblastomas and 7 control samples were assessed for splice junctions linking to unannotated 5′ exons (that is, novel first exons; see Methods). This revealed 2,479 exons not annotated in RefSeq in 1,937 genes (Supplementary Table 3e). Of these exons, 1,714 were listed in GENCODE, providing support for our method. Around 40% were upstream of the annotated 5′ end of the gene, with the remainder being internal. For example, *C9orf3*, hosting the miRNA cluster *MIR-23b–27b–24-1* cluster, was associated with a novel gene-body start site that harboured a pdCR specifically in the WNT subgroup (Extended Data Fig. 4a). Upregulation of this microRNA (miRNA) cluster has previously been noted in WNT medulloblastomas[14], and correlation with methylation was observed in our miRNA data set (Extended Data Fig. 4b–e).

To avoid confounding signals from canonical promoters, we focused on novel exons located >15 kb upstream of the previously annotated start site (the farthest being >500 kb upstream). Of 262 such exons, 49 showed patterns indicative of a pdCR (Supplementary Table 3d). One notable example is the miRNA-processing gene *LIN28B*, which showed differential regulation at a novel alternative first exon in Group 3 and Group 4 medulloblastomas. A pdCR at this start site (representing a bidirectional promoter with *LINC00577*) is hypomethylated in almost all Group 3 and Group 4 medulloblastomas, with increased LIN28B expression, whereas the canonical promoter was mostly hypermethylated (Fig. 2a, b and Extended Data Fig. 5a–c). *LIN28B* regulates multiple oncogenic processes, in part by downregulating the tumour-suppressive *LET-7* miRNA family[15]. Group 3 and 4 medulloblastomas showed significantly lower expression of most *LET-7* miRNAs compared with WNT

and SHH medulloblastomas (Fig. 2c and Extended Data Fig. 5d, e). High *LIN28B* expression has been linked with poor prognosis in neuroblastoma[16], and this also held true for Group 3 and 4 medulloblastomas (Fig. 2d and Extended Data Fig. 5f).

Broader inspection of genome-wide methylation revealed megabase-scale partially methylated domains (PMDs)[10,17–19], affecting up to one-third of the tumour genome (Fig. 3a, b and Extended Data Fig. 6a, b). These regions account for the decreased global methylation observed in WNT and Group 3 tumours (Extended Data Fig. 1c). Although PMDs have been noted as a prominent feature of tumour methylomes[1], with global hypomethylation common in cancer, the affected regions showed clear subgroup specificity in WNT and Group 3 medulloblastomas, and they were virtually absent in SHH and Group 4 tumours (Fig. 3b and Extended Data Fig. 6c–e).

PMDs have been linked with inactive chromatin and large-scale gene silencing[17,18,20]. Genes inside subgroup-specific PMDs were significantly less expressed than non-PMD genes in that subgroup (positive correlation with methylation; Supplementary Table 4a), even (to a lesser extent) in tumours without clear hypomethylation (Extended Data Fig. 7a, b). Intra-subgroup heterogeneity, however, suggested some divergence in PMD formation (for example, for calmin (*CLMN*) in Group 3; Extended Data Fig. 7c, d). ChIP-seq in D425 cells confirmed an association of PMDs with long, organized domains of repressive marks (H3K9me3, H3K27me3)[17–19], and an absence of the active H3K36me3 mark (Fig. 3c, d and Extended Data Fig. 7e). Longer PMDs showed widespread H3K9me3 bordered by H3K27me3, whereas shorter PMDs showed H3K27me3 only (Fig. 3d). A significantly increased somatic mutation rate was observed
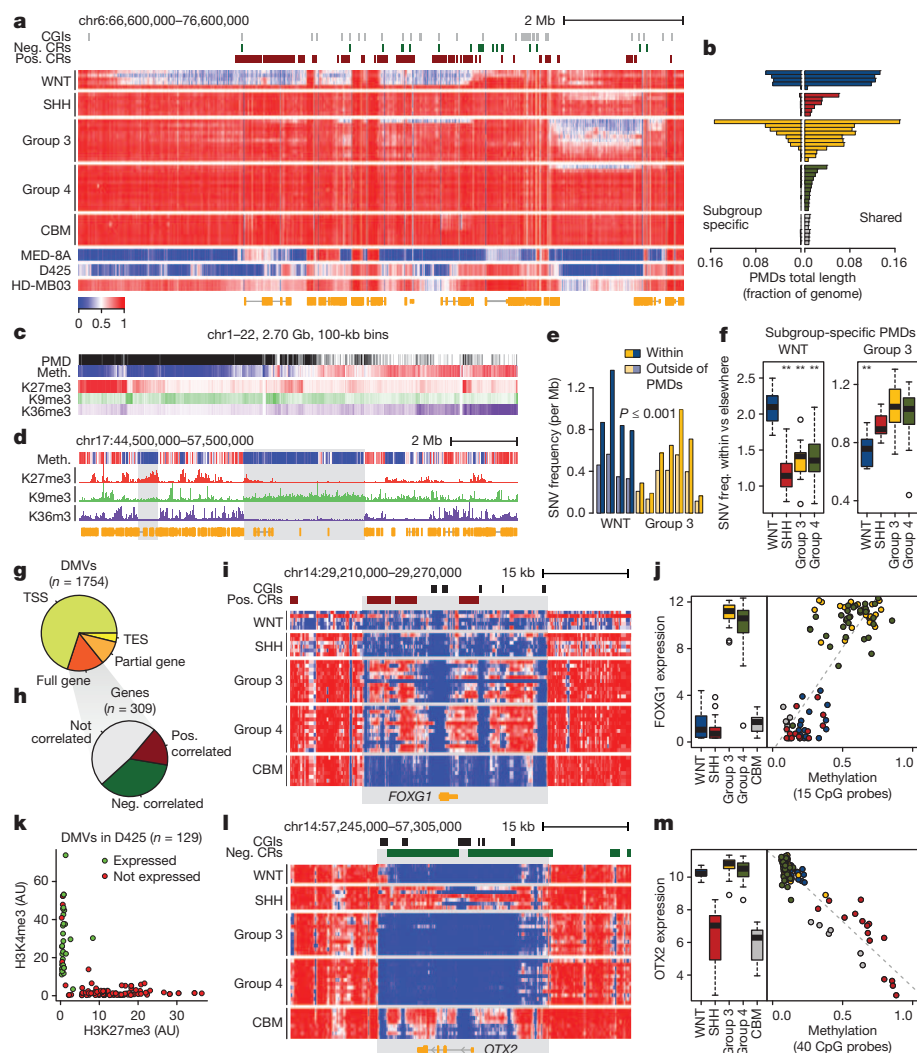
Figure 3 | Megabase-scale silenced domains and smaller DMVs are associated with distinct histone marks and positive correlation with expression. a, Example of a region on chromosome 6 showing large, subgroup-specific PMDs (*n* = 42). Negative (Neg.) and positive (Pos.) CRs are indicated. Partial methylation on chromosome 6 in WNT medulloblastomas (4/5 show monosomy 6) supports the notion that PMDs are stochastically acquired and not allele specific. Red, methylated; blue, unmethylated. CBM, cerebellum. **b**, Genome-wide quantification of PMDs. **c, d**, D425 cell line methylation (Meth.) and ChIP-seq data. PMDs are associated with increased H3K9me3 (for larger stretches) or H3K27me3 (typically smaller regions), and H3K36me3 depletion. **e**, Regions of the genome covered by a PMD show a higher somatic mutation frequency than non-PMD regions in a given sample (paired *t*-test). **f**, For WNT- or Group-3-specific PMDs, the ratio of single nucleotide variation (SNV) frequencies (freq.) within PMDs compared with elsewhere is higher in the PMD-containing subgroup. **P < 0.01 (*t*-test). **g, h**, Most DMVs overlap gene regions, including many that encompass entire genes. Both negative and positively correlating DMVs are observed. **i, j**, Methylation plot around *FOXG1* (*n* = 42), which is encompassed by a positively correlating DMV. Re-establishment of methylation in Group 3 and Group 4 tumours correlates with increased expression (*n* = 95). **k**, DMVs in the D425 cell line show either the active H3K4me3 mark (when expressed) or the inactive H3K27me3 mark (when silenced), but rarely both. AU, arbitrary units. **l, m**, Methylation plot around *OTX2*, with a negatively correlating DMV, showing extensive hypomethylation in WNT, Group 3 and Group 4 tumours (*n* = 42) linked with higher expression in these groups (*n* = 95).

in PMDs (Fig. 3e, f), consistent with increased mutagenesis in late-replicating heterochromatin[21]. Mutant allele fractions indicated that most of this additional mutational burden is probably occurring early in tumorigenesis (data not shown).

The subgroup specificity of PMDs implies a non-random heterochromatization. Expression of tumour PMD-related genes was lower in normal cerebellum than non-PMD genes (Extended Data Fig. 7a, b), indicating pre-existing silencing in certain normal precursors. Variable hypomethylation within tumours, however, suggests a stochastic loss among different cell populations (Extended Data Fig. 6c, d). WGBS data of medulloblastoma cell lines support a gradual acquisition model, as hypomethylation was much more pronounced in long-established lines (MEB-MED-8A, D425) compared with a recent model (HD-MB03; Extended Data Fig. 6a). Thus, PMDs probably arise progressively due to failed transmission of methylation in late replicating, repressive domains[22], and are probably secondary to, rather than causative of, heterochromatin formation.

More localized hypomethylated regions surrounding smaller genes, termed DNA methylation valleys (DMVs) or lowly methylated 'canyons', have recently been reported as regulating tissue-specific gene expression[23,24]. This pattern, associated with H3K27me3 and/or H3K4me3 histone marks, is reminiscent of previously described, evolutionarily conserved polycomb-binding sites[25]. Our WGBS data revealed several hundred valleys per sample, with 1,754 unique DMVs overlapping a gene (Fig. 3g). Almost 20% ($n = 309$) encompassed an entire gene, and many were correlated with expression (Fig. 3h, Extended Data Fig. 8a–c and Supplementary Table 4b). As previously noted[24], a large number of DMVs affect developmental transcription factors and neuronal lineage genes.

Interestingly, 16.8% of DMVs showed a positive correlation between methylation and expression. These regions are reminiscent of PMDs, but are present in all tumours and controls, are much shorter, and have a higher CpG density (Supplementary Table 4c). One example is the region around *FOXG1*, a transcription factor linked with stem cell self-renewal in Group 3 and Group 4 medulloblastomas[26]. This gene is hypomethylated and not expressed in normal cerebellum, whereas hypermethylation of the DMV in Group 3 and Group 4 tumours is coupled with markedly increased expression (Fig. 3i, j).

ChIP-seq data for DMVs in D425 cells that are also present in primary tumours and encompass an entire gene ($n = 129$) revealed a notable pattern of either H3K4me3 (when the gene was expressed) or H3K27me3 (no expression) marks, with few bivalent sites (Fig. 3k). The absence of DNA methylation in DMVs showing positive correlation between methylation and expression may reflect a particularly dense heterochromatin that is inaccessible to the methylation machinery. Increased methylation is associated with gain of the active mark H3K4me3, loss of H3K27me3, and increased expression (for example, at the *PITX2* locus; Extended Data Fig. 8d, e). Thus, reactivation may involve a 'normalization' of the chromatin state from densely packed heterochromatin to a more open conformation, with re-establishment of the DNA methylation pattern typical of other genomic regions. DMVs are generally thought to be hypomethylated in normal tissues, and have been reported to contain CGIs that are particularly prone to silencing by hypermethylation in cancer[24,27,28]. We demonstrate, however, that cancer-specific gain of DNA methylation in DMVs is also linked to reactivation of repressed genes. This relationship is similar to that observed previously for a subset of PMDs in embryonic stem cells, which showed high levels of H3K27me3 and low methylation presumably linked with polycomb-mediated repression. Decreased H3K27me3 and increased DNA methylation in several of these PMDs, indicating relieved repression, was noted in differentiated IMR90 fibroblasts[19]. Both shorter PMDs and positively correlating DMVs may therefore be linked to similar regulatory processes.

Negatively correlating DMVs ($n = 111$) were also identified, with hypomethylation and concomitant upregulation of expression (Supplementary Table 4d). One notable example was *OTX2*, a known medulloblastoma oncogene with roles in cell cycle regulation and differentiation[29]. It is strongly hypomethylated and markedly overexpressed in WNT, Group 3 and Group 4 tumours (Fig. 3l, m).

OTX2 was also highlighted by an analysis of additional focal regions that did not fit into the categories described earlier, but clearly showed differential methylation. Localized lowly methylated regions (LMRs) have previously been linked with transcription-factor-binding sites and gene regulation[19,30]. To prove a functional link between LMRs and transcription-factor binding in medulloblastoma, we performed ChIP-seq for OTX2. As expected, ChIP-seq peaks were associated with hypomethylation, which was strongest at the centre of the peak. The periodicity in flanking methylation levels matched the size of a nucleosome, suggesting that hypomethylation of transcription-factor-binding sites may spread to adjacent linker regions (Fig. 4a). Conversely, 75.9% of LMRs that contained an OTX2-binding motif, identified in a systematic analysis, overlapped with an OTX2 ChIP-seq peak (Extended Data Fig. 9a).

Approximately 70,000 LMRs were identified per sample, with >45,000 unique regions showing subgroup-specific patterns (Supplementary Table 5a). Clustering revealed six groups: four with specificity for each medulloblastoma subgroup, plus one non-SHH/non-cerebellum and one non-Group 3/non-Group 4 cluster (Fig. 4b and Extended Data Fig. 9b).
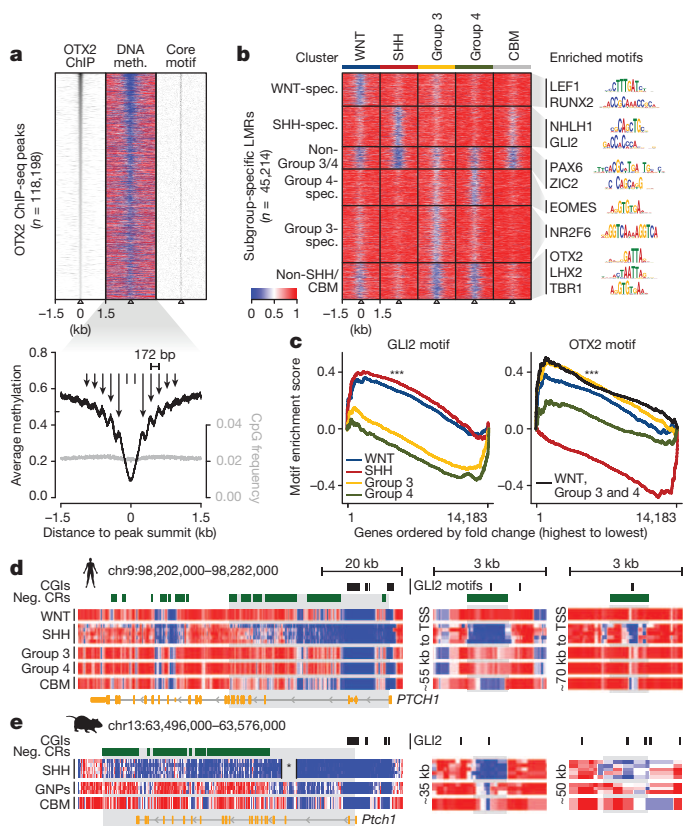
Figure 4 | Focal LMRs mark binding sites for key transcriptional regulators. a, Sites of OTX2 binding as defined by ChIP-seq in the D425 cell line are marked by hypomethylation and possess an OTX2-binding motif. Methylation (meth.) around the peak summit reveals a nucleosome periodicity. Red, methylated; blue, unmethylated. **b**, Heat-map representation of subgroup-specific (spec.) LMRs (±1.5 kb around centre) grouped according to *k*-means clustering. Motif analysis on these LMR clusters revealed enrichment for several key transcription factors, with representative examples shown (see Supplementary Table 5b). CBM, cerebellum. **c**, Examples of downstream regulation of gene expression. Binding motifs for the identified transcription factors are strongly enriched in the LMRs of the most highly overexpressed genes in the respective subgroup. ***$P < 0.001$ (motif enrichment score, see Methods). **d, e**, Example of upstream LMRs at the *PTCH1/Ptch1* locus in human/murine medulloblastomas with GLI2-binding sites. The same sites are not hypomethylated in mouse GNPs, indicating a somatic origin. There is also a pdCR at this site in both mouse ($n = 9$) and human ($n = 41$; sample ICGC_MB1, with homozygous *PTCH1* deletion, is not shown). Asterisk indicates deletion of *Ptch1* exon 3 (flanked by LoxP sites in the *Ptch1*[fl/fl] model).

In a systematic analysis, each cluster showed enrichment for certain transcription-factor-binding motifs, implicating known medulloblastoma-related genes and new candidates (for example, *NR2F6* in Group 3 medulloblastoma; Supplementary Table 5b). To test whether the transcription factors highlighted by LMR clustering might regulate downstream networks, we also looked at LMRs associated with the most highly expressed genes within the subgroups. Many transcription-factor-binding motifs identified in a particular LMR cluster were also highly enriched in these genes, demonstrating a key role for these transcription factors in determining subgroup identity (Fig. 4c and Supplementary Table 5b).

Analysis of differentially methylated LMRs therefore represents a powerful tool to provide a global, unbiased assessment of the key transcription factor regulatory networks acting in different tissues.

Analysis of our own and public mouse WGBS data again supported a somatic origin for many LMRs. *PTCH1/Ptch1*, for example, harboured several upstream LMRs with GLI2-binding sites in human and mouse SHH medulloblastoma that were not seen in normal cerebellum or GNPs (Fig. 4d, e and Extended Data Fig. 10). The GLI2-binding motif was enriched in mouse-tumour-specific LMRs, in keeping with the human data (Extended Data Fig. 9c and Supplementary Table 5c).

Our integrative analysis has uncovered a wealth of alterations in the methylome landscape of medulloblastoma. The regulatory effect of such epigenetic alterations in modulating the cancer phenotype allows for gradual global changes during tumour evolution, and their interplay with single time-point events such as genetic mutations will be an important future avenue of investigation. New targets that may be important for the pathophysiology of these tumours and for therapeutic intervention were also identified. These data, and the analysis methods applied, represent a valuable resource for further investigation into epigenetic regulatory mechanisms in cancer biology and in a wider context of other disease processes and organismal development.

METHODS SUMMARY

All patients contributing material to this study provided informed consent. Tumour samples were collected before adjuvant chemo- or radiotherapy. Control cerebellum samples were from a commercial supplier (BioChain). Cohort details are given in Supplementary Table 1. Tumour subgrouping was based on DNA methylation profiling, as previously described[5]. Genome-wide methylome profiles were derived by whole-genome bisulphite sequencing, as described in Methods. Sequencing metrics are given in Supplementary Table 2. CRs were derived by merging overlapping 5-CpG windows that showed correlation between DNA methylation and gene expression. Differential methylation and expression were assessed using analysis of variance and post-hoc testing. Further details are provided in Methods.

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

1. Shen, H. & Laird, P. W. Interplay between the cancer genome and epigenome. *Cell* **153,** 38–55 (2013).
2. Taylor, M. D. *et al.* Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* **123,** 465–472 (2012).
3. Jones, D. T. W. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488,** 100–105 (2012).
4. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nature Rev. Cancer* **12,** 818–834 (2012).
5. Hovestadt, V. *et al.* Robust molecular subgrouping and copy-number profiling of medulloblastoma from small amounts of archival tumour material using high-density DNA methylation arrays. *Acta Neuropathol.* **125,** 913–916 (2013).
6. Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain development. *Science* **341,** 1237905 (2013).
7. Hong, C. *et al.* Epigenome scans and cancer genome sequencing converge on WNK2, a kinase-independent suppressor of cell growth. *Proc. Natl Acad. Sci. USA* **104,** 10974–10979 (2007).
8. Gaur, T. *et al.* Canonical WNT signaling promotes osteogenesis by directly stimulating *Runx2* gene expression. *J. Biol. Chem.* **280,** 33132–33140 (2005).
9. Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.* **41,** 1350–1353 (2009).
10. Hansen, K. D. *et al.* Increased methylation variation in epigenetic domains across cancer types. *Nature Genet.* **43,** 768–775 (2011).
11. Irizarry, R. A. *et al.* The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nature Genet.* **41,** 178–186 (2009).
12. Hodges, E. *et al.* Directional DNA methylation changes and complex intermediate states accompany lineage specificity in the adult hematopoietic compartment. *Mol. Cell* **44,** 17–28 (2011).
13. Yang, Z. J. *et al.* Medulloblastoma can be initiated by deletion of Patched in lineage-restricted progenitors or stem cells. *Cancer Cell* **14,** 135–145 (2008).
14. Cho, Y. J. *et al.* Integrative genomic analysis of medulloblastoma identifies a molecular subgroup that drives poor clinical outcome. *J. Clin. Oncol.* **29,** 1424–1430 (2011).
15. Shyh-Chang, N. & Daley, G. Q. Lin28: primal regulator of growth and metabolism in stem cells. *Cell Stem Cell* **12,** 395–406 (2013).
16. Molenaar, J. J. *et al.* LIN28B induces neuroblastoma and enhances MYCN levels via *let-7* suppression. *Nature Genet.* **44,** 1199–1206 (2012).
17. Berman, B. P. *et al.* Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature Genet.* **44,** 40–46 (2012).
18. Hon, G. C. *et al.* Global DNA hypomethylation coupled to repressive chromatin domain formation and gene silencing in breast cancer. *Genome Res.* **22,** 246–258 (2012).
19. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462,** 315–322 (2009).
20. Wen, B., Wu, H., Shinkai, Y., Irizarry, R. A. & Feinberg, A. P. Large histone H3 lysine 9 dimethylated chromatin blocks distinguish differentiated from embryonic stem cells. *Nature Genet.* **41,** 246–250 (2009).
21. Schuster-Böckler, B. & Lehner, B. Chromatin organization is a major influence on regional mutation rates in human cancer cells. *Nature* **488,** 504–507 (2012).
22. Aran, D., Toperoff, G., Rosenberg, M. & Hellman, A. Replication timing-related and gene body-specific methylation of active human genes. *Hum. Mol. Genet.* **20,** 670–680 (2011).
23. Jeong, M. *et al.* Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nature Genet.* **46,** 17–23 (2014).
24. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153,** 1134–1148 (2013).
25. Tanay, A., O'Donnell, A. H., Damelin, M. & Bestor, T. H. Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl Acad. Sci. USA* **104,** 5521–5526 (2007).
26. Manoranjan, B. *et al.* FoxG1 interacts with Bmi1 to regulate self-renewal and tumorigenicity of medulloblastoma stem cells. *Stem Cells* **31,** 1266–1277 (2013).
27. Ohm, J. E. *et al.* A stem cell-like chromatin pattern may predispose tumor suppressor genes to DNA hypermethylation and heritable silencing. *Nature Genet.* **39,** 237–242 (2007).
28. Schlesinger, Y. *et al.* Polycomb-mediated methylation on Lys27 of histone H3 pre-marks genes for *de novo* methylation in cancer. *Nature Genet.* **39,** 232–236 (2007).
29. Bunt, J. *et al.* OTX2 directly activates cell cycle genes and inhibits differentiation in medulloblastoma cells. *Int. J. Cancer* **131,** E21–E32 (2012).
30. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480,** 490–495 (2011).

Author Contributions D.T.W.J., S.Pi., W.W., M.S., S.B., J.B., C.C.B., C.v.K., R.V., S.S., S.W., J.F. and P.La performed and/or coordinated experimental work. V.H., M.K., P.A.N., K.S., N.J., H.-J.W., M.Ra., K.K., S.E., C.L., J.E., J.K. and P.La. performed data analysis. M.Ry., T.M., O.W., T.P., S.R., W.S., M.D.T. and A.K. collected data and provided patient materials. V.H., D.T.W.J., M.K., P.A.N., M.Z., B.R., S.M.P. and P.Li. prepared the initial manuscript and figures. D.T.W.J., U.D.W., B.B., G.R., A.B., H.L., R.J.W.-R., R.E., M.-L.Y., A.K., P.La., M.Z., B.R., S.M.P. and P.Li. provided project leadership.

Author Information Short-read sequencing data have been deposited in the European Genome-phenome Archive (http://www.ebi.ac.uk/ega/) under accession number EGAS00001000561. Methylation array data have been deposited in the Gene Expression Omnibus under accession number GSE54880. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.R. (b.radlwimmer@dkfz-heidelberg.de), S.M.P. (s.pfister@dkfz-heidelberg.de), or P.Li. (peter.lichter@dkfz-heidelberg.de).

METHODS

Sample collection and preparation. An Institutional Review Board ethical vote (Ethics Committee of the Medical Faculty of Heidelberg) was obtained according to the International Cancer Genome Consortium (ICGC) guidelines (http://www.icgc.org), along with informed consent for all participants. No patient underwent chemotherapy or radiotherapy before surgical removal of the primary tumour. Tumour tissues were subjected to neuropathological review for confirmation of histology and for tumour cell content >80%. Analytes were isolated as previously described[3]. Normal cerebellum samples of eight individual donors were purchased from BioChain. Lot numbers are given in Supplementary Table 1.

Cells were cultured at 37 °C with 5% CO_2. D425_Med (D425; a gift from D. D. Bigner) and MEB-MED-8A cells (from the authors' own stocks; T. Pietsch) were cultured in DMEM with 10% FCS (Life Technologies). HD-MB03 cells[31] were grown in RPMI-1640 with 10% FCS (Life Technologies). All cells were regularly authenticated and tested for mycoplasma.

Details of the *Math1-creER^{T2}:Ptch1^{fl/fl}* mouse medulloblastoma model have been described previously[13]. Mice were bred and maintained at the Sanford Burnham Animal Facility. Experiments were performed in accordance with national regulations and procedures were approved by the Institutional Animal Care and Use Committee. GNPs were harvested at postnatal day 4 (three biological replicates of pooled precursor-cell populations), and the mean age of mice for tumour tissue collection was 131 days (range 81–260 days, five individual tumours from female mice).

Library preparation and sequence data generation

Whole-genome bisulphite sequencing. Whole-genome bisulphite library preparation was carried out as recently described[32], with modifications to a previously published protocol[33]. In brief, 5 μg of genomic DNA were sheared using a Covaris device. After adaptor ligation, DNA fragments with insert lengths of 200–250 bp were isolated using an E-Gel electrophoresis system (Life Technologies) and bisulphite converted using the EZ DNA Methylation kit (Zymo Research). PCR amplification of the fragments was performed in six parallel reactions per sample using the FastStart High Fidelity PCR kit (Roche). Library aliquots were then pooled per sample and sequenced on an Illumina HiSeq 2000 machine.

Whole-genome sequencing. Samples up to ICGC_MB51 were described previously[3]. Data for six additional samples were generated using identical conditions.

RNA-seq. RNA-seq libraries were prepared as described[3,34] using methods to preserve strand specificity. A subset of libraries was prepared using purified polyA+ RNA fractions, with the remainder using ribosomal-RNA-depleted fractions. Sequencing was carried out on the HiSeq 2000 platform using 2× 51 cycles according to the manufacturer's instructions. All samples profiled by WGBS sequencing, whole-genome sequencing, 450k array and RNA sequencing were *in silico* genotyped to exclude possible sample swaps.

miRNA sequencing. Small RNA libraries were prepared essentially as previously described[35], with minor adjustments. 3′-Adaptor ligation was performed on total RNA with barcoded pre-adenylated adapters and five samples were pooled for successive steps. 3′-Adaptor-ligated products corresponding to small RNAs of 19–35 nucleotides were fractionated by polyacrylamide gel electrophoresis for further processing. Sequencing of pooled, barcoded samples was performed on an Illumina HiSeq 2000 following standard procedures.

ChIP-seq. H3K4me3 (AM#39159, Active Motif), H3K9me3 (AM#39161, Active Motif), H3K27me3 (#07-449, Millipore) and H3K36me3 (AM#61101, Active Motif) ChIP library preparation of D425 and MEB-MED-8A cell lines was performed at Active Motif according to proprietary methods. Libraries were sequenced on the Illumina HiSeq 2000 platform using 2× 101 cycles according to the manufacturer's instructions.

OTX2 immunoprecipitation of the D425 cell line was performed as previously described[36]. Sequencing libraries were generated using standard Illumina adapters and sequenced on the Illumina GA IIx platform using 1× 33 cycles according to the manufacturer's instructions.

Expression array analysis. Affymetrix U133 Plus2.0 expression array data were extracted from publicly available data sets via the R2 software tool for analysis and visualization of genomic data (http://r2.amc.nl) or for additional cases on an early access basis through a collaboration between the DKFZ and the Microarray Department of the University of Amsterdam. Sample library preparation, hybridization and quality control were performed according to protocols recommended by the manufacturer. The MAS5.0 algorithm of the GCOS program (Affymetrix Inc.) was used for normalization and assignment of detection *P* values. Log2-transformed expression values were used for all analyses and illustrations. Gene expression values in Fig. 1g were centred by subtracting the gene-wise median. Genes were considered as expressed if the maximum log2 expression value in any sample of the WGBS cohort was larger than 6 ($n = 14,183$ genes).

Affymetrix Mouse Genome 430 2.0 expression array data of GNPs and SHH medulloblastoma tumours was generated according to the manufacturer's instructions at the DKFZ Genomics and Proteomics Core Facility.

Methylation array analysis. In addition to whole-genome bisulphite sequencing, a partially overlapping validation cohort was assessed using the Illumina Infinium HumanMethylation450 BeadChip array. Sample processing and hybridization was performed according to the manufacturer's instructions at the DKFZ Genomics and Proteomics Core Facility. Downstream array processing was performed as previously described[5], with exclusion of probes containing an annotated single nucleotide polymorphism (SNP) (dbSNP132 Common) within 5 bp of and including the targeted CpG site, or probes not mapping uniquely to the genome allowing for one mismatch. Except for normalization to internal controls, no additional normalization techniques were applied.

Next-generation-sequencing data analysis

Whole-genome bisulphite sequencing. WGBS sequencing data were analysed using methylCtools (V.H. *et al.*, manuscript in preparation). In brief, methylCtools builds upon BWA and adds functionality for aligning bisulphite-treated DNA to a reference genome in a similar manner to that described previously[37]. Sequencing reads were adaptor-trimmed using SeqPrep (https://github.com/jstjohn/SeqPrep) and translated to a fully C-to-T converted state. Alignments were performed against a single index of both *in silico* bisulphite-converted strands of the human reference genome (hg19, NCBI build 37.1) or mouse reference genome (mm10, NCBI build 38) using BWA version 0.6.1-r104 (ref. 38) and the non-default parameters -q 20 -s. Previously translated bases were translated back to their original state, and reads mapping antisense to the respective reference strand were removed. Putative PCR duplicates were filtered using Picard MarkDuplicates (http://picard.sourceforge.net). Single-base-pair methylation ratios (β-values) were determined by quantifying evidence for methylated (unconverted) and unmethylated (converted) cytosines at all CpG positions. Only properly paired or singleton reads with mapping quality of ≥1 and bases with a Phred-scaled quality score of ≥20 were considered. To account for population variability, we filtered CpGs for which more than 25% of reads at a given position (on either strand) were not supportive of this CpG being in fact a CpG in the sample being analysed. For mouse samples, which share the same genetic background, this filtering step was performed on the merged data set. Subsequently, information from both strands was combined and CpGs with a coverage of less than five reads were not considered in the respective sample. Non-conversion rates were estimated on the basis of lambda phage genome spike-ins. Concordance of methylation β-values between WGBS sequencing and DNA methylation array was extremely high ($r = 0.973 \pm 0.006$).

Methylation plots were generated using the Gviz Bioconductor package. Essentially, CpGs within the region being visualized were binned into 1,000 equally sized windows. The average methylation levels of all CpGs per window were then weighted by their distance to adjacent CpGs, and plotted using a heat-map scale.

Whole-genome sequencing. Whole-genome sequencing data were analysed as previously described[3]. For the analysis integrating SNVs and PMDs, only SNVs mapping to chromosome 1 to 22 were considered.

RNA-seq. Strand-specific RNA-seq data were analysed using STAR version 2.3.0e[39], supplying RefSeq gene annotations (obtained from the UCSC genome browser). Uniquely mapping reads were obtained by setting outFilterMultimapNmax to 1, otherwise default parameters were used. Reads per million (RPM) values were calculated relative to all sequencing reads mapping to annotated exons.

Novel first exons were detected by performing reference-guided (RefSeq gene annotations) transcript assembly using cufflinks[40] version 2.1.1 with the following non-default parameter: library-type fr-firststrand. Transcript assemblies were combined using cuffmerge version 1.0.0 supplying reference annotations. Novel first exons were filtered for being smaller than 2 kb in size and being linked directly or through additional novel exons to annotated transcripts by junctions covered by at least 50 sequencing reads.

miRNA sequencing. Small RNA sequencing data were analysed by aligning adaptor- and barcode-trimmed reads to a reference database containing precursor sequences of miRBase version 18[41]. Bowtie[42] was used for sequence alignment with the following non-default parameters: seedmms 1, maqerr 1000, seedlen 21, norc, best, strata, M 1. Expression of mature microRNAs was quantified relative to all aligned reads. For all analyses and illustrations, log2-transformed RPM values were used.

ChIP-seq. Histone ChIP-seq data were analysed using BWA version 0.5.10. Putative PCR duplicates were filtered using Picard MarkDuplicates. For downstream analyses, we generated whole-genome coverage tracks with reads normalized to all properly paired reads (RPM; paired-end reads/fragments per million). We used igvtools version 2.2.2 (http://www.broadinstitute.org/igv/igvtools) and the non-default parameter pairs and a window size of 25.

OTX2 ChIP-seq data were analysed using BWA. Peak calling was performed using MACS[43] version 1.4.1 using default parameters and enabling calling of subpeaks. A whole-genome coverage track was generated using the detected fragment size and a window size of 10 bp.

Integrative analysis of DNA methylation and gene expression. Downstream analyses were performed using R version 2.15.2 (ref. 44), making extensive use of

the GenomicRanges Bioconductor package. We developed a novel approach to define CRs between DNA methylation and gene expression. To this aim, methylation levels for every sample of the WGBS cohort were combined in windows of five adjacent CpGs for up to 100 kb surrounding and within annotated genes. These five CpG windows were then correlated to Affymetrix-derived expression levels of annotated genes. Five CpG windows were termed correlated if showing a Pearson's correlation coefficient smaller/larger than ± 0.5 and a maximum methylation difference between any two samples of at least 0.5. Additionally, we used a permutation test (permutation of sample labels) to estimate the probability of obtaining a more extreme correlation coefficient by chance. Only those windows associated with a P value smaller than 0.001 (0.01 for analysis of mouse pdCRs) were taken as being correlated. Overlapping correlated five-CpG windows were then merged into CRs. Both negative and positive CRs were required to contain at least ten CpGs to be further considered. Average methylation levels of CRs were calculated by taking the mean of all CpGs per sample. If there were multiple gene expression probesets associated with a single gene, the one containing the most CpGs within CRs was used (ties were resolved by higher mean expression).

pdCRs. pdCRs were defined using a novel approach. Every CpG starting from 500 bp downstream of the TSS (to not overlap with gene promoter annotations) was given a score based on being located within a negative CR ($+1$) or not (-2). CpGs unmethylated in controls (average methylation level <0.15 in 1 kb windows; see later) were not scored. The cumulative sum from the first to every downstream CpG was calculated. The outer pdCR boundary was defined as the last CpG within a CR for which the cumulative sum was larger than 0.

DMVs. DMVs for each sample were detected in a similar fashion to that previously described[24]. Average methylation levels within windows of 1 kb were calculated (individual CpGs weighted by the distance of both adjacent CpGs), in steps of 1 bp. Overlapping 1 kb windows with average methylation levels smaller than 0.15 were merged and resulting regions larger than 5 kb were termed DMVs. DMVs of individual samples were merged if present in at least two tumour samples or any control sample.

PMDs. PMDs for each sample were detected in a similar fashion to that previously described[17]. Average methylation levels within windows of 10 kb were calculated (individual CpGs weighted by the distance of both second-next CpGs), in steps of 1 bp. Overlapping 10 kb windows with an average methylation level <0.6 were merged, and resulting regions larger than 100 kb were termed PMDs.

LMRs. LMRs for each sample were detected using the MethylSeekR Bioconductor package[45] with a methylation level threshold of 0.5 and a minimal number of four CpGs. An additional filtering step of LMRs probably located within PMDs was performed by filtering LMRs associated with an average methylation level <0.75 of the 30 adjacent CpGs on both ends (<0.6 of 10 adjacent CpGs in mouse). LMRs of individual samples were merged if present in at least two tumour samples or any control sample.

Subgroup-specific differential methylation. Subgroup-specific differential methylation was determined independently for the different methylation patterns described. For pdCR differential methylation analysis, negative CRs were intersected with pdCRs. For DMV differential methylation analysis, negative and positive CRs were intersected with DMVs. For global positive CR differential methylation analysis, all positive CRs were used. However, CRs could not be located farther than the adjacent annotated expressed gene, and at least 50 CpGs falling into a CR were required per gene. For promoter CGI differential methylation analysis (obtained from the UCSC genome browser), CGIs were linked to genes if located within $-1,500/+500$ of annotated TSSs and all overlapping CpGs were used. Similarly, for LMR differential methylation analysis, all overlapping CpGs were used.

For all patterns described earlier, the average methylation level per gene and sample was calculated for the entire WGBS cohort. Analysis of variance was used to determine genes differentially methylated between all four medulloblastoma subgroups and combined controls (adjusted P value < 0.001). The Benjamini–Hochberg procedure was used to adjust for multiple testing. Subgroup-specific differentially methylated genes were determined by applying a post-hoc test on genes previously determined as being differentially methylated. (R package: multcomp, functions: mcp and glht). Individual medulloblastoma subgroups were required to be significant against all other subgroups combined, and against control samples separately (P value < 0.001; in essence eight comparisons were made). This testing procedure was deemed appropriate because of large sample sizes and continuous methylation levels resulting of averaging of multiple CpGs. Identical testing was performed on the array-based methylation verification cohort ($n = 284$ samples, using all CpG probes that were located within CRs), and the array-based gene expression cohort ($n = 95$) to provide further proof of correlation in regions of interest.

Additional statistical analyses. Kaplan–Meier analysis was performed for samples with matching Affymetrix U133 Plus2.0 expression array data and patient survival data. Samples were stratified into high or low expression of *LIN28B* using a threshold of 250 expression units (probeset 229349_at).

An unpaired t-test was used to test for differences between continuous variables, with the exception of Fig. 3e and Extended Data Fig. 7a, b in which paired t-tests were used. Unequal variances within groups were assumed for all tests.

Boxplots displayed in Figs 1e, 2b, 3f, j, m and Extended Data Figs 1c, d, 2c, 3b, d–f, 4b–e, 5b–d, 7a, b, 8a, e and 10a, b, d were generated in R using default settings (function: boxplot).

The Jaccard index as a measure of PMD similarity between any two samples is defined as the ratio of the combined length of the genomic regions within PMDs in both samples (intersect) and in at least one sample (union).

Position weight matrices of transcription factor binding motifs (as defined by the TRANSFAC database version 2013 or a recent study[46]) were mapped against the genome using the FIMO tool (part of the MEME suite, version 4.9.0) and the following parameters: thresh 1e-5, max-seq-length 4e9, max-stored-scores 1e7. Enrichment of transcription-factor-binding motifs overlapping LMRs was tested using the chi-squared test over the clusters generated by k-means clustering (R function: kmeans, 10 starting sets, maximum of 10,000 iterations, using mean methylation levels per sample and LMR). The Benjamini–Hochberg procedure was used to adjust for multiple testing. The ratio of observed over expected motif occurrences was taken as a measure for determining motifs enriched in a particular cluster. As transcription factors of the same family often share similar binding motifs, the transcription factors indicated in Fig. 4b and Extended Data Fig. 9b, c were further refined by taking into account available gene expression data (see also Supplementary Table 5b, c). For heat-map representations of OTX2 ChIP-seq peaks and LMRs 2,000 sites were randomly sampled for illustration purposes.

Motif enrichment of downstream targets was analysed adapting the gene set enrichment analysis (GSEA) approach[47]. Genes were ordered by the $\log_2$ fold change in expression between indicated subgroups and remaining samples, and the running motif enrichment score was calculated by motif occurrences within LMRs associated with the particular gene (up to 500 kb upstream, within gene body), allowing for multiple LMRs per gene. 'Hits' (motif overlapping LMRs) were weighted by the absolute $\log_2$ fold change. The maximum running motif enrichment value was used as the motif enrichment score. A P value was calculated by performing a permutation test (randomizing association of motifs and LMRs, in 200,000 iterations) to obtain a more extreme motif enrichment score by chance.

31. Milde, T. *et al.* HD-MB03 is a novel Group 3 medulloblastoma model demonstrating sensitivity to histone deacetylase inhibitor treatment. *J. Neurooncol.* **110,** 335–348 (2012).
32. Richter, J. *et al.* Recurrent mutation of the *ID3* gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. *Nature Genet.* **44,** 1316–1320 (2012).
33. Lister, R. *et al.* Hotspots of aberrant epigenomic reprogramming in human induced pluripotent stem cells. *Nature* **471,** 68–73 (2011).
34. Jones, D. T. W. *et al.* Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nature Genet.* **45,** 927–932 (2013).
35. Hafner, M. *et al.* Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods* **44,** 3–12 (2008).
36. Bunt, J. *et al.* Joint binding of OTX2 and MYC in promotor regions is associated with high gene expression in medulloblastoma. *PLoS ONE* **6,** e26058 (2011).
37. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27,** 1571–1572 (2011).
38. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
39. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).
40. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28,** 511–515 (2010).
41. Kozomara, A. & Griffiths-Jones, S. miRBase: integrating microRNA annotation and deep-sequencing data. *Nucleic Acids Res.* **39,** D152–D157 (2011).
42. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).
43. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
44. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2010).
45. Burger, L., Gaidatzis, D., Schubeler, D. & Stadler, M. B. Identification of active regulatory regions from DNA methylation data. *Nucleic Acids Res.* **41,** e155 (2013).
46. Jolma, A. *et al.* DNA-binding specificities of human transcription factors. *Cell* **152,** 327–339 (2013).
47. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* **102,** 15545–15550 (2005).
48. Li, H. *et al.* Transcription factor *MEF2C* influences neural stem/progenitor cell differentiation and maturation *in vivo. Proc. Natl Acad. Sci. USA* **105,** 9397–9402 (2008).
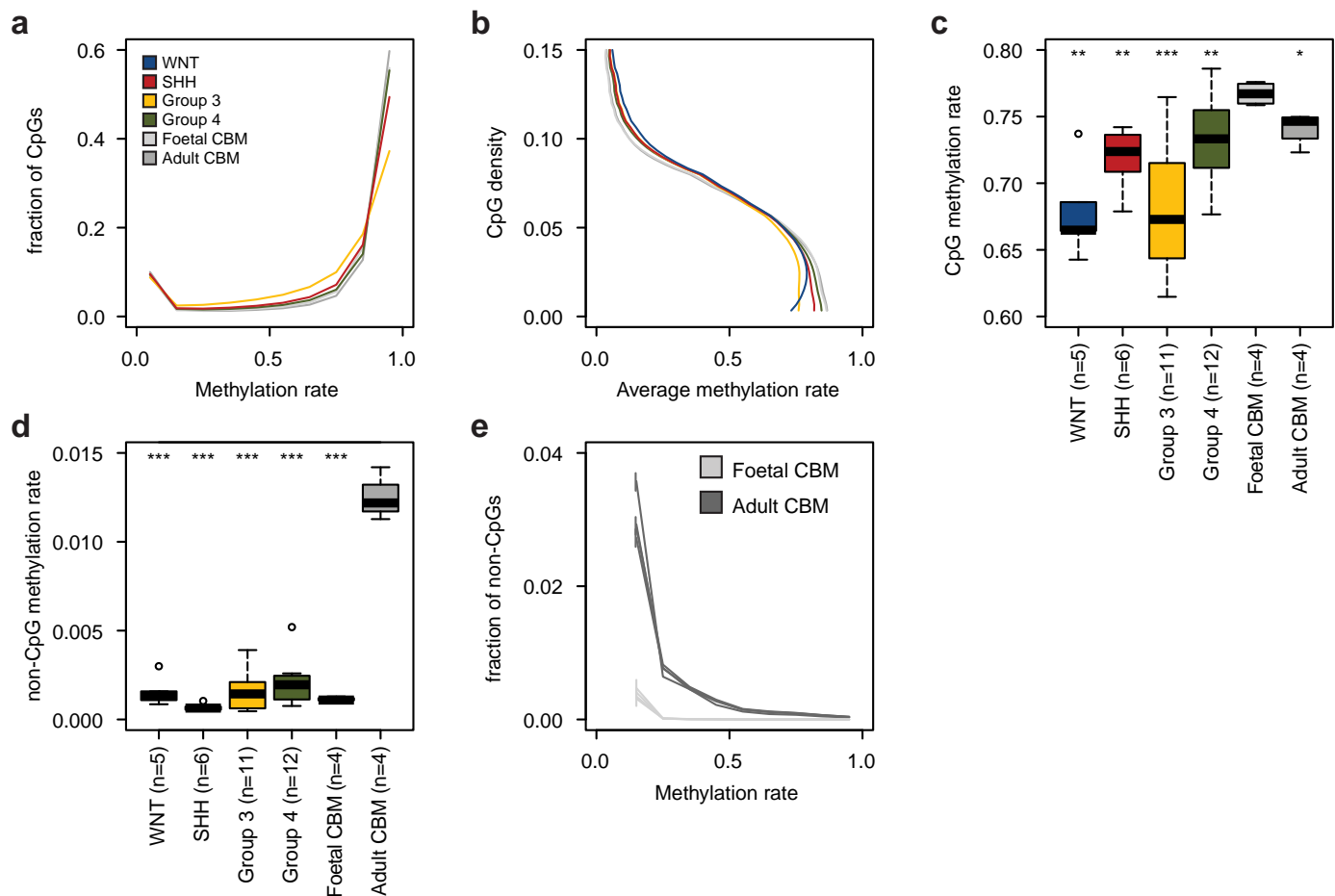
Extended Data Figure 1 | Global properties of DNA methylation.
a, Fraction of genome-wide CpGs showing different methylation states, binned into ten windows. The majority of CpGs show a bimodal pattern close to either fully methylated or unmethylated, with few intermediate values. **b**, Distribution of average methylation rate with varying CpG density, showing that CpG-dense regions are typically unmethylated, with high methylation in CpG-poor regions. **c**, Genome-wide methylation rates by tumour subgroup or age (for controls). All tumour subgroups show significantly reduced methylation compared with fetal cerebellum (CBM), with the largest reduction

in WNT and Group 3 tumours. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$.
d, Minimal non-CpG (CH) methylation is seen in tumours or fetal cerebellum, with levels significantly below that of adult normal cerebellum. ***$P < 0.001$.
**e**, The distribution of methylation states for non-CpG methylation does not show a similar bimodal pattern to CpG methylation, with a marked shift towards low percentages (indicating heterogeneity of methylation levels within a tissue). The curves are truncated at a lower limit of 0.1 to increase resolution in the higher range, as the vast majority of sites show methylation rates <0.1.
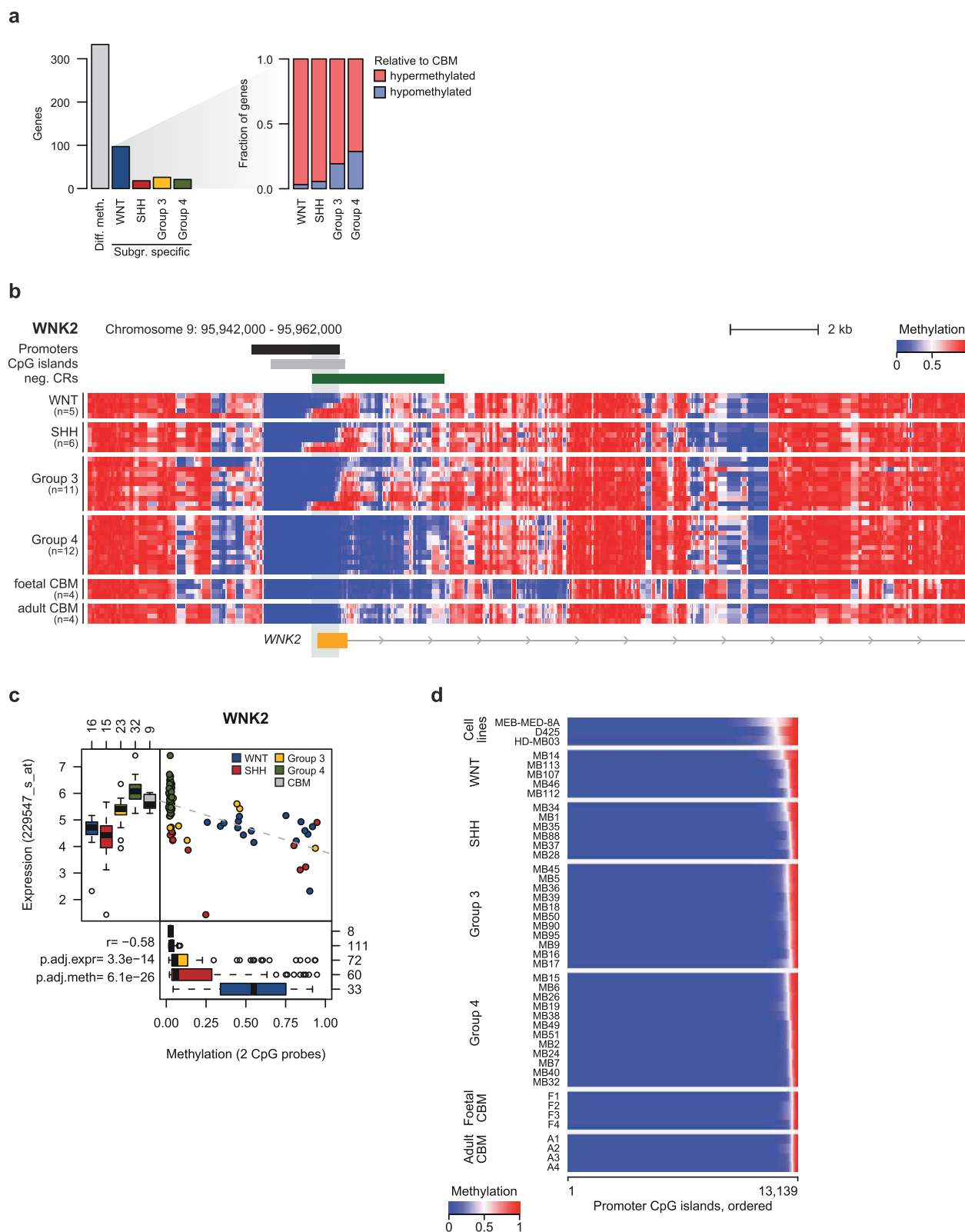
Extended Data Figure 2 | Hypermethylation of CGIs is rare in medulloblastoma. a, Overview of number of genes associated with significantly differentially methylated CGIs across the four medulloblastoma subgroups and control cerebellum, and those that are specifically differentially methylated in a given tumour subgroup. The fraction of differentially methylated promoter CGIs per subgroup that are hypo- (blue) or hypermethylated (red) relative to control cerebellum is also shown. **b**, Methylation plot for the 5′ end of *WNK2*, showing a region of methylation extending into the promoter CGI in some WNT, SHH and Group 3

medulloblastomas. **c**, Methylation of the *WNK2* promoter region is negatively correlated with gene expression in the extended, array-based validation cohort. Sample numbers are indicated next to the boxplots (*n* = 95 overlap). CBM, cerebellum; *r*, Pearson's correlation coefficient; p.adj.expr, expression-adjusted *P* value (analysis of variance (ANOVA), Benjamini–Hochberg adjustment); p.adj.meth, methylation-adjusted *P* value. **d**, A summary of methylation levels at all promoter CGIs across medulloblastoma cell lines and tumours, and normal tissues.
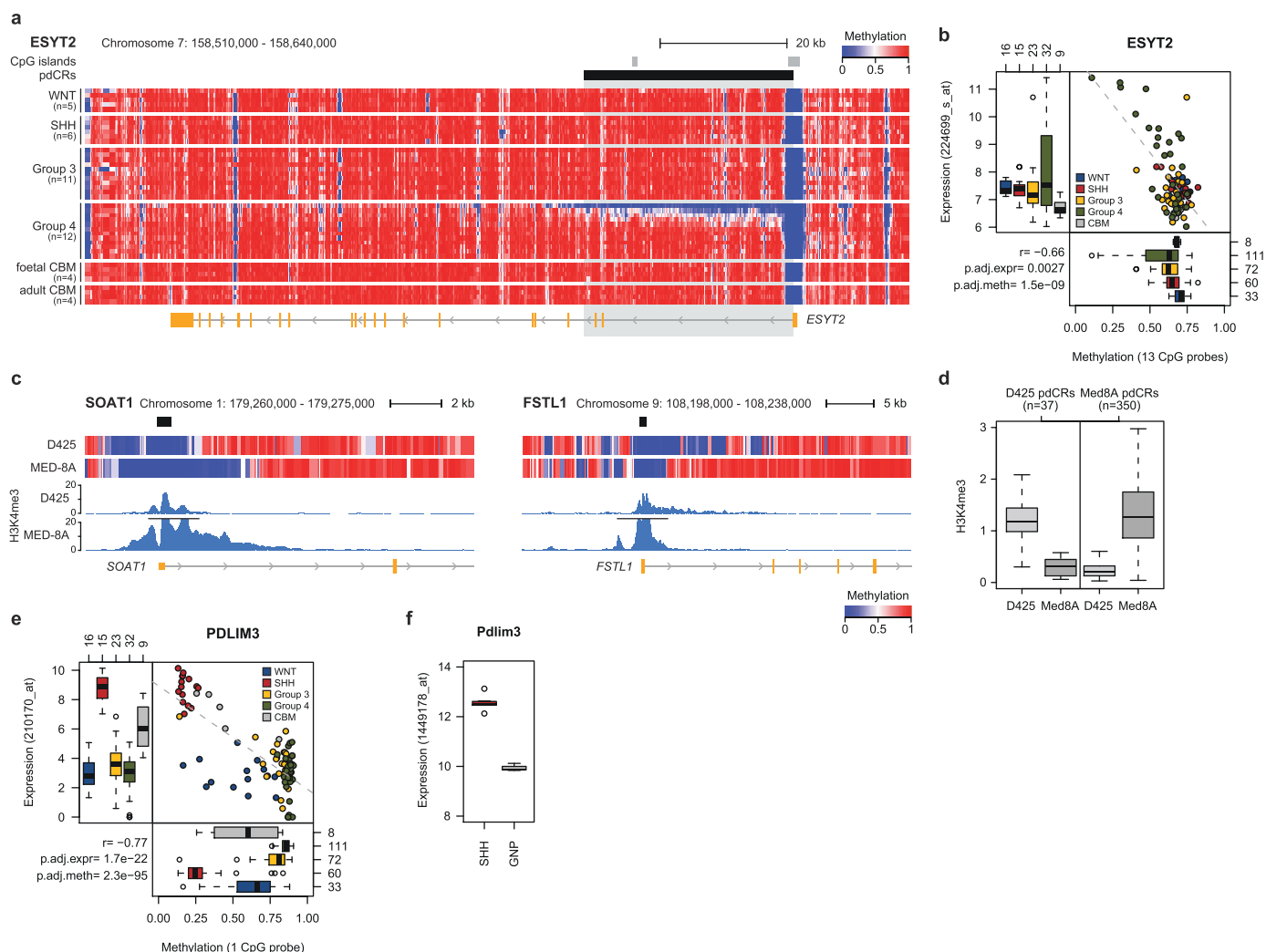
Extended Data Figure 3 | pdCRs overlap with the H3K4me3 histone mark.
a, Methylation plot for *ESYT2*, showing hypomethylation of a pdCR in some Group 4 medulloblastomas. The variable methylation states at the boundaries of the pdCR suggest subpopulations of differing values within a tumour. Additionally, the heterogeneity across the subgroup indicates a potential somatic acquisition of this demethylation. **b**, Methylation of the *ESYT2* pdCR is negatively correlated with gene expression in the extended, array-based validation cohort. **c**, Methylation plot including H3K4me3 ChIP-seq data for the *SOAT1* and *FSTL1* loci, showing differential pdCRs methylation in two medulloblastoma cell lines. Notably, the pdCR overlaps with the presence of H3K4me3. **d**, When looking at segments of pdCRs unique to either medulloblastoma cell line, H3K4me3 levels are significantly higher in the cell line showing the extended pdCR. **e**, Methylation levels at the pdCR of *PDLIM3* are negatively correlated with expression in the extended array cohort (see also Fig. 1h). **f**, Mouse SHH medulloblastomas also show hypomethylation of a pdCR in *Pdlim3* that is associated with increased expression, which is not seen in GNPs (that is, somatic in origin, see also Fig. 1i).
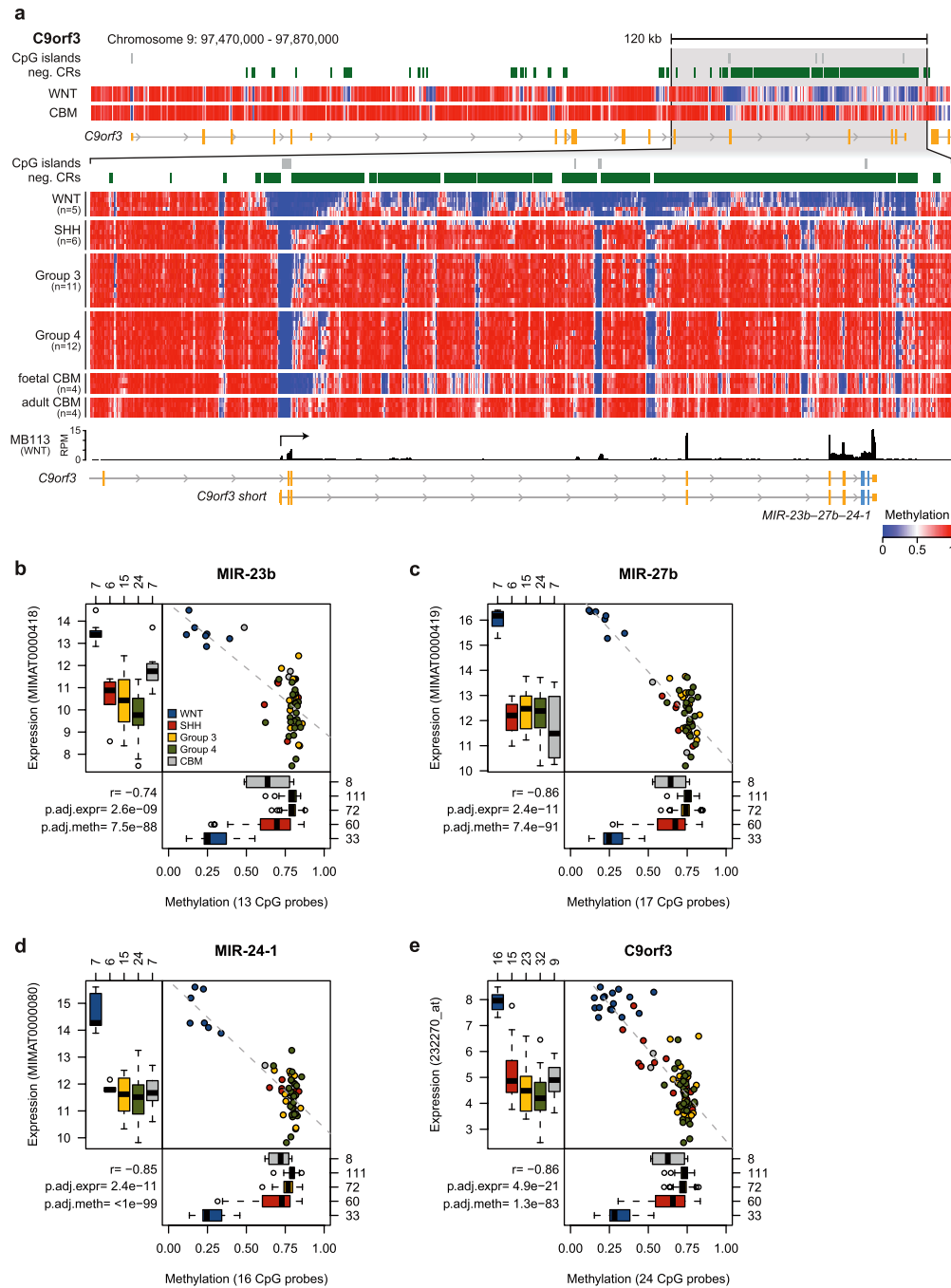
Extended Data Figure 4 | The *MIR-23b–27b–24-1* cluster is epigenetically regulated in WNT medulloblastoma. a, An example of a novel first exon within a gene body is seen for *C9orf3*, which is the primary transcript hosting the *MIR-23b–27b–24-1* cluster.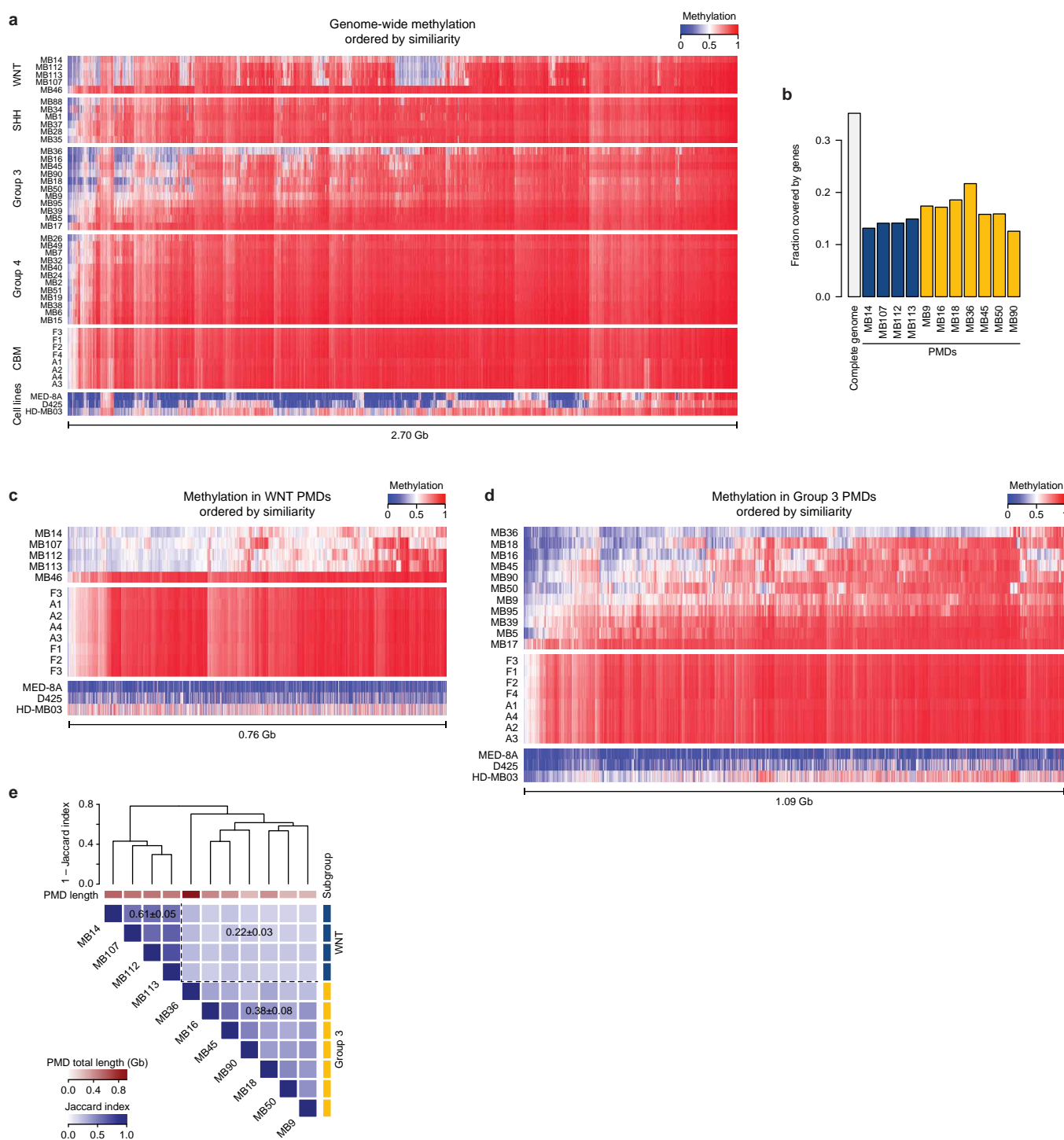 **b–d,** Methylation around *MIR-23b*, *MIR-27b* and *MIR-24-1* is negatively correlated with expression of these miRNAs in an extended validation cohort. **e,** Negative correlation of methylation and expression is also observed for *C9orf3* itself.
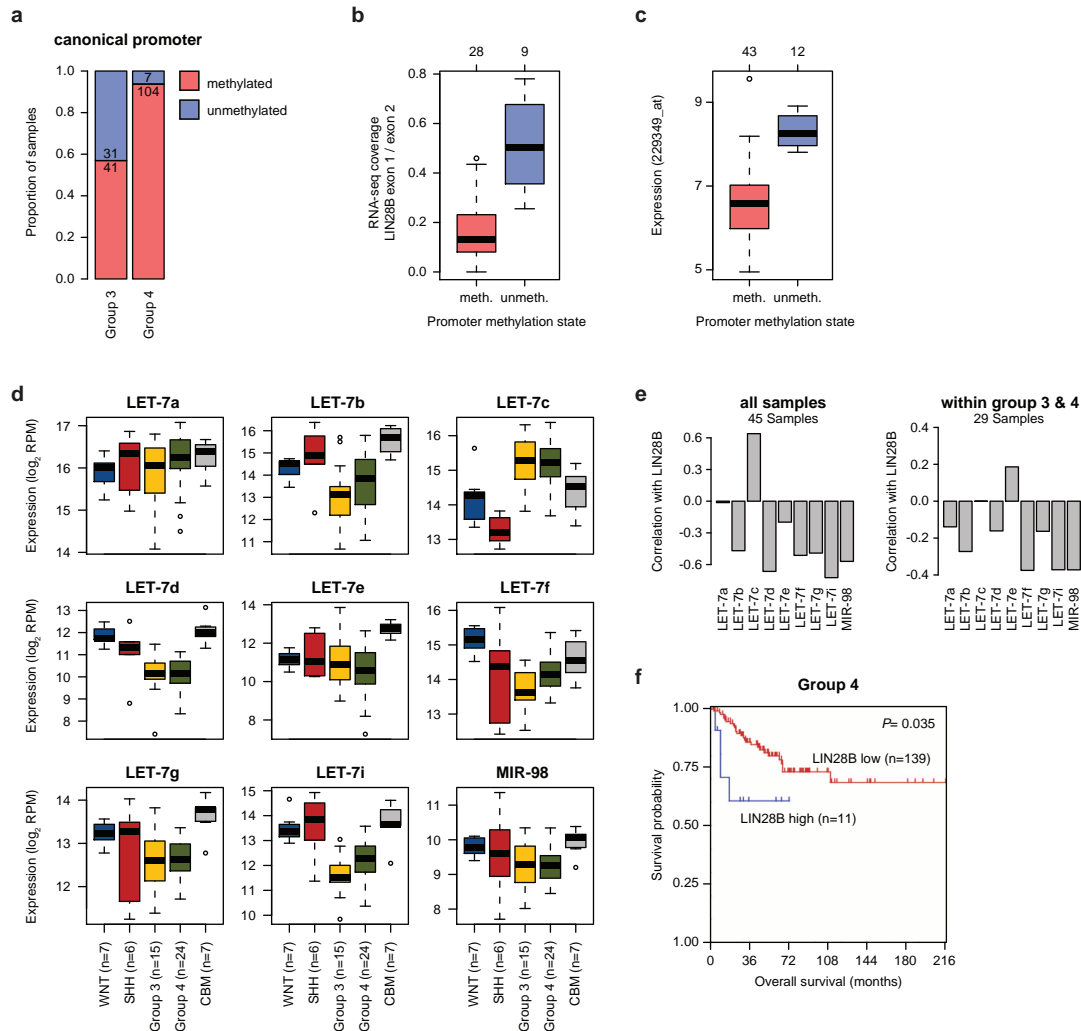
Extended Data Figure 5 | Members of the LET-7 family negatively correlate with *LIN28B* expression. **a**, Proportion of Group 3 and Group 4 medulloblastomas showing methylation of the canonical *LIN28B* promoter in the extended validation cohort. Sample numbers are indicated. **b**, Coverage of RNA-seq reads for exon 1 and 2 of the canonical *LIN28B* transcript separated by promoter methylation state for Group 3 and 4 medulloblastomas, showing a loss of exon 1 expression in those samples with methylation of the canonical promoter. **c**, Group 3 and 4 medulloblastomas that are unmethylated at the canonical promoter show even higher *LIN28B* expression than other samples in these subgroups. **d**, Expression levels of *LET-7* family miRNAs for medulloblastoma subgroups and normal cerebellum. RPM, reads per million. **e**, Correlation of *LIN28B* expression with *LET-7* family miRNAs within Group 3 and Group 4 medulloblastomas and across all medulloblastoma subgroups. **f**, Association of *LIN28B* expression with overall survival in Group 4 medulloblastomas (Kaplan–Meier analysis; *P*, log-rank test).

a, Genome-wide methylation ordered by similarity

b

c, Methylation in WNT PMDs ordered by similarity

d, Methylation in Group 3 PMDs ordered by similarity

e

Extended Data Figure 6 | PMDs are restricted to WNT and Group 3 tumours and are subgroup specific. a, A global view of genome-wide methylation values, shown in 100 kb bins and sorted according to similarity, gives a clear picture of the PMDs in WNT and Group 3 tumours. Also notable is the extensive hypomethylation in the cell lines. **b**, Indication of the fraction of total PMD length in each tumour that is covered by genes, showing that PMDs are typically gene-poor regions compared to the genome average (left-most bar). **c**, Overview of methylation values in WNT PMDs, showing a high similarity between samples (with the exception of ICGC_MB46, which does not show the same pattern). **d**, An overview of methylation values in Group 3 PMDs, showing some similarity between samples, but with more variation in extent and sites of demethylation compared with WNT tumours. **e**, Similarity of PMD regions in WNT and Group 3 medulloblastoma samples that show elevated levels of PMDs (samples with a total of >0.2 Gb covered by a PMD). Correlation is higher within rather than between subgroups.
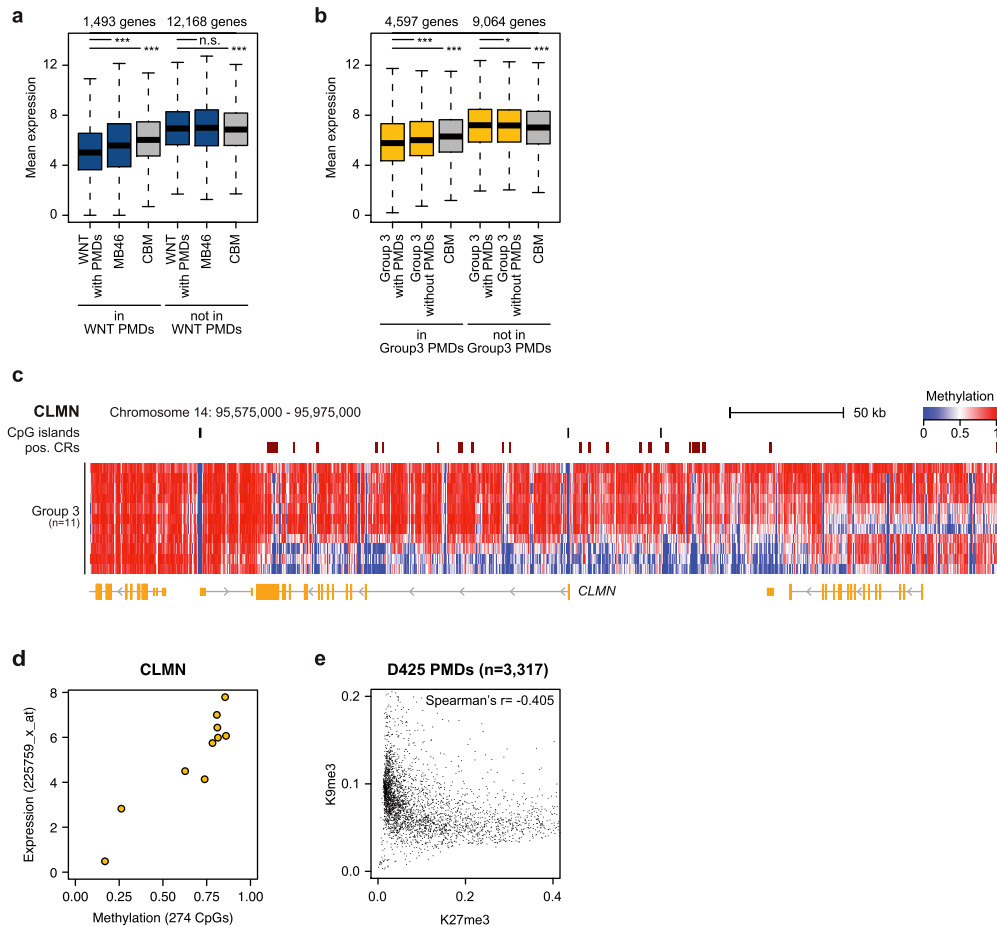
Extended Data Figure 7 | PMDs are associated with decreased gene expression. a, Genes located within WNT PMDs are expressed at a significantly lower level in those WNT tumours that have PMDs compared with normal cerebellum. WNT medulloblastomas lacking clear PMDs (that is, ICGC_MB46) show intermediate expression. Genes not in PMDs show a slightly elevated expression in the tumours. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. **b**, Genes located within Group 3 PMDs are expressed at a significantly lower level in those Group 3 tumours that have PMDs compared with normal cerebellum. Group 3 medulloblastomas lacking clear PMDs show intermediate expression. Genes not in PMDs show a slightly elevated expression in the tumours. *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$. **c**, An example of a PMD in Group 3 medulloblastoma around the *CLMN* gene, showing heterogeneous methylation levels within the subgroup. **d**, Methylation levels of the *CLMN* PMD are positively correlated with gene expression in Group 3 medulloblastoma ($n = 11$). **e**, PMDs are typically associated with H3K9me3 or H3K27me3 histone marks, but less commonly with both.
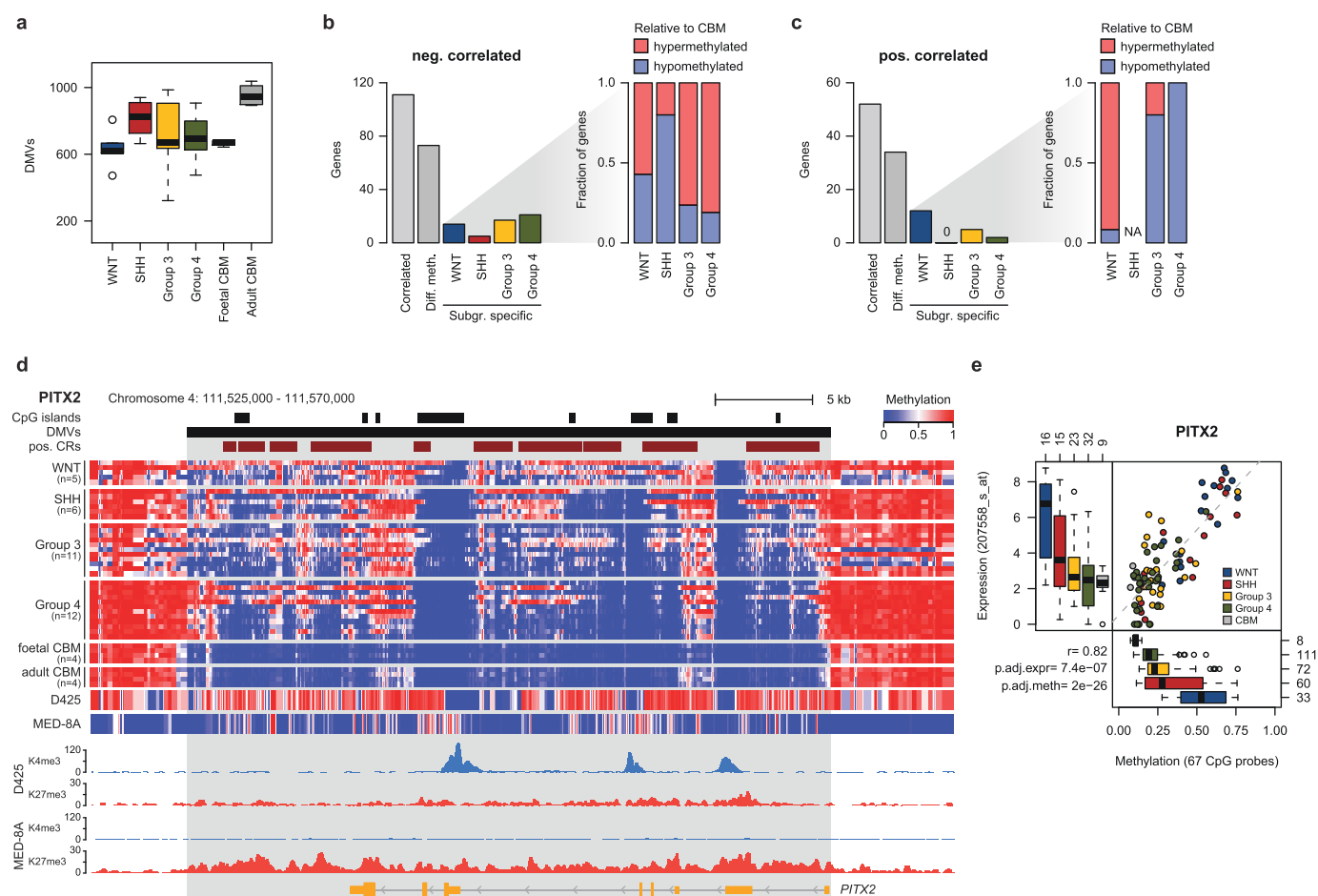
Extended Data Figure 8 | Positively correlating DMVs overlap regions of differential H3K27 trimethylation. a, Boxplots of the number of DMVs identified per sample, split by tumour subgroup (or age group for controls, *n* = 42 samples). **b, c**, Overview of genes encompassed by a negatively or positively correlating DMV, those that are significantly differentially methylated across the four medulloblastoma subgroups and control cerebellum, and those that are specifically differentially methylated in a given tumour subgroup. The fraction of subgroup-specific, gene-encompassing DMVs that are hypo- (blue) or hypermethylated (red) relative to control cerebellum is also shown. **d**, Methylation plot for *PITX2*, located in a DMV that shows differential methylation in a variety of subgroups. This DMV is hypomethylated in control cerebellum, but methylation is re-established in several tumours. In the MED-8A cell line, this locus is hypomethylated and covered by the inactivating H3K27me3 mark, and is not expressed. In D425, in which *PITX2* is expressed, re-establishment of methylation is accompanied by loss of H3K27me3 and gain of the active H3K4me3 modification. **e**, This re-establishment of methylation, representing normalization of chromatin, is positively correlated with increased gene expression in the extended, array-based validation cohort.
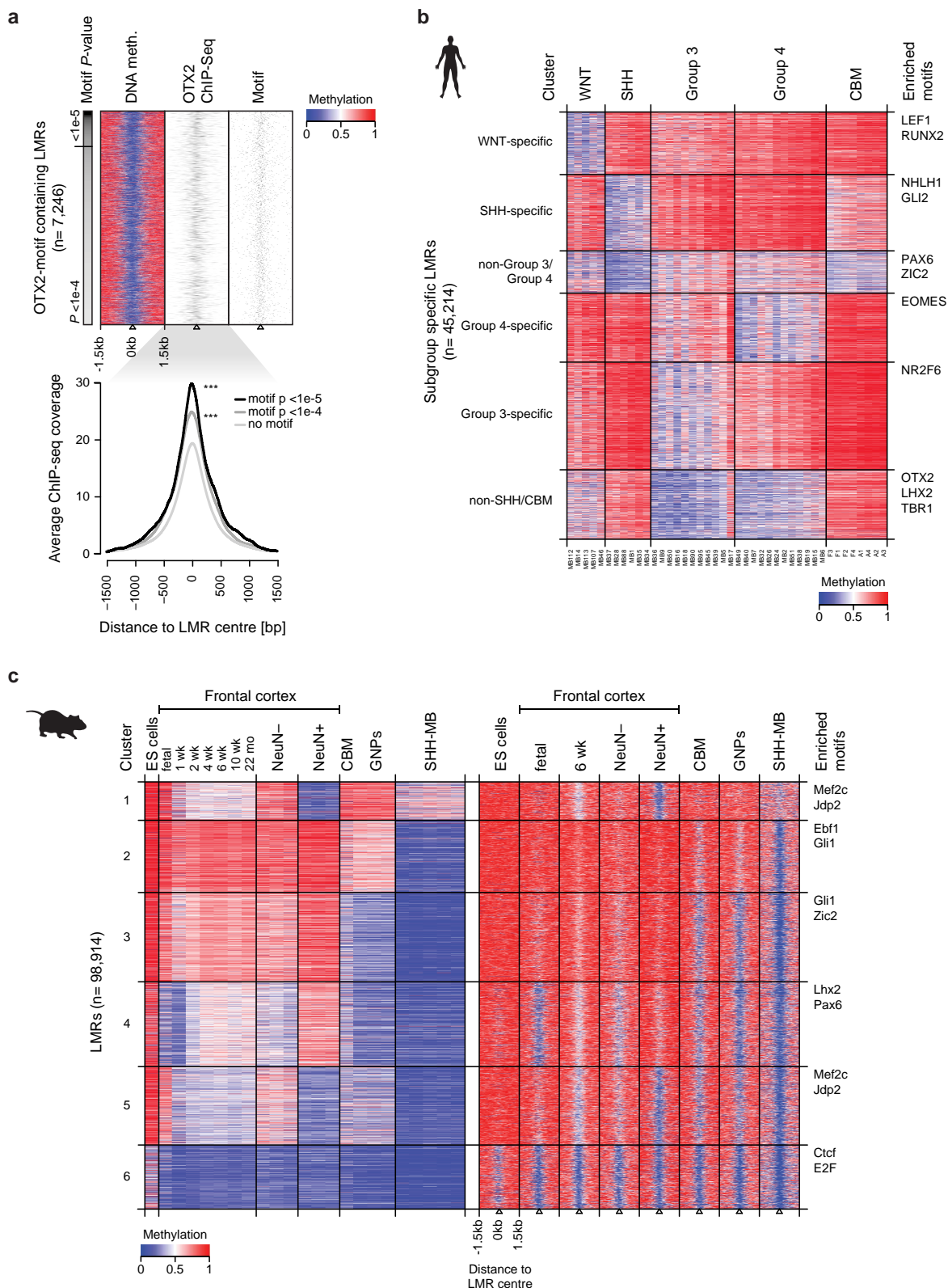
Extended Data Figure 9 | Systematic analysis of LMRs identifies transcriptional regulators in human and mouse medulloblastoma.
a, Heat-map representation of genomic regions of 3-kb-centred LMRs predicted to contain an OTX2-binding motif. More than 75% of LMRs overlap with a ChIP peak, with strongest binding at those sites where the motif more closely matches the consensus sequence. **b**, Heat-map representation of a *k*-means clustering of subgroup-specifically methylated LMRs. A single methylation value per sample and LMR is shown, as used for the clustering. **c**, Heat-map representation of a *k*-means clustering of LMRs in GNPs, SHH

medulloblastoma mouse model and external data sets. Left, a single methylation level per sample and LMR is shown, as used for the clustering. Right, genomic regions of 3 kb centred around the LMR of selected samples are shown. Selected transcription-factor-binding motifs enriched within specific clusters are indicated. The relevance of this analysis for highlighting transcriptional regulators is further supported by an LMR cluster with specificity for sorted neuronal cells (NeuN[+]), which showed clear enrichment for *Mef2c* binding (a key regulator of neuronal cell fate[48]).
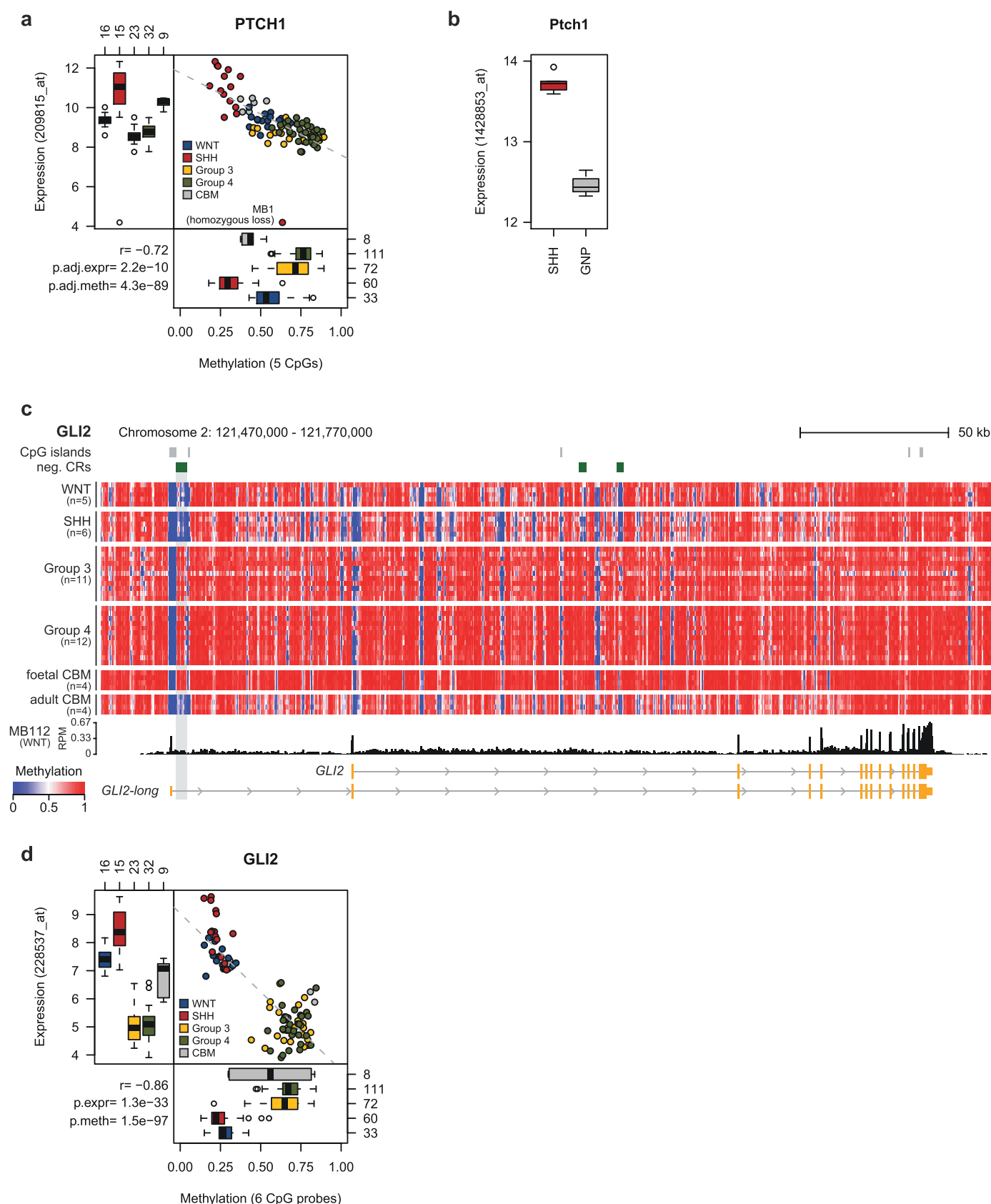
Extended Data Figure 10 | Epigenetic regulation of a novel GLI2 transcript variant. a, Methylation levels at a pdCR in *PTCH1* are negatively correlated with expression in the extended array cohort (see also Fig. 4d). **b**, Mouse SHH medulloblastomas also show hypomethylation of a pdCR in *Ptch1* that is associated with increased expression, which is not seen in GNPs (that is,

somatic in origin; see also Fig. 4e). **c**, Methylation plot for *GLI2*, showing an exon upstream of the annotated transcript in WNT and SHH medulloblastoma and adult cerebellum. RNA-seq data are shown in reads per million (RPM) below the heat map. **d**, Methylation at the *GLI2* upstream exon is negatively correlated with gene expression in the extended, array-based validation cohort.