ARTICLE

Active medulloblastoma enhancers reveal subgroup–specific cellular origins

Charles Y. Lin[1]†*, Serap Erkek[2,3]*, Yiai Tong[4], Linlin Yin[5], Alexander J. Federation[1], Marc Zapatka[6], Parthiv Haldipur[7], Daisuke Kawauchi[3], Thomas Risch[8], Hans–Jörg Warnatz[8], Barbara C. Worst[3], Bensheng Ju[9], Brent A. Orr[10], Rhamy Zeid[1], Donald R. Polaski[1], Maia Segura-Wang[2], Sebastian M. Waszak[2], David T. W. Jones[3,11], Marcel Kool[3,11], Volker Hovestadt[6], Ivo Buchhalter[12], Laura Sieber[3], Pascal Johann[3], Lukas Chavez[3], Stefan Gröschel[13], Marina Ryzhova[14], Andrey Korshunov[11,15], Wenbiao Chen[5], Victor V. Chizhikov[16], Kathleen J. Millen[7,17], Vyacheslav Amstislavskiy[8], Hans Lehrach[8], Marie–Laure Yaspo[8], Roland Eils[12,18], Peter Lichter[6,11], Jan O. Korbel[2], Stefan M. Pfister[3,11,19]§, James E. Bradner[1]§ & Paul A. Northcott[3,4]§

Medulloblastoma is a highly malignant paediatric brain tumour, often inflicting devastating consequences on the developing child. Genomic studies have revealed four distinct molecular subgroups with divergent biology and clinical behaviour. An understanding of the regulatory circuitry governing the transcriptional landscapes of medulloblastoma subgroups, and how this relates to their respective developmental origins, is lacking. Here, using H3K27ac and BRD4 chromatin immunoprecipitation followed by sequencing (ChIP–seq) coupled with tissue–matched DNA methylation and transcriptome data, we describe the active *cis*-regulatory landscape across 28 primary medulloblastoma specimens. Analysis of differentially regulated enhancers and super–enhancers reinforced inter–subgroup heterogeneity and revealed novel, clinically relevant insights into medulloblastoma biology. Computational reconstruction of core regulatory circuitry identified a master set of transcription factors, validated by ChIP–seq, that is responsible for subgroup divergence, and implicates candidate cells of origin for Group 4. Our integrated analysis of enhancer elements in a large series of primary tumour samples reveals insights into *cis*-regulatory architecture, unrecognized dependencies, and cellular origins.

Medulloblastoma is a highly malignant paediatric brain tumour classified into four biologically and clinically distinct molecular subgroups[1]. The present clinical approach to medulloblastoma involves maximal safe surgical resection, cytotoxic chemotherapy, and cranio-spinal radiation, which together are associated with profound morbidity in the developing child, underscoring the need for new subgroup-specific therapeutic insights.
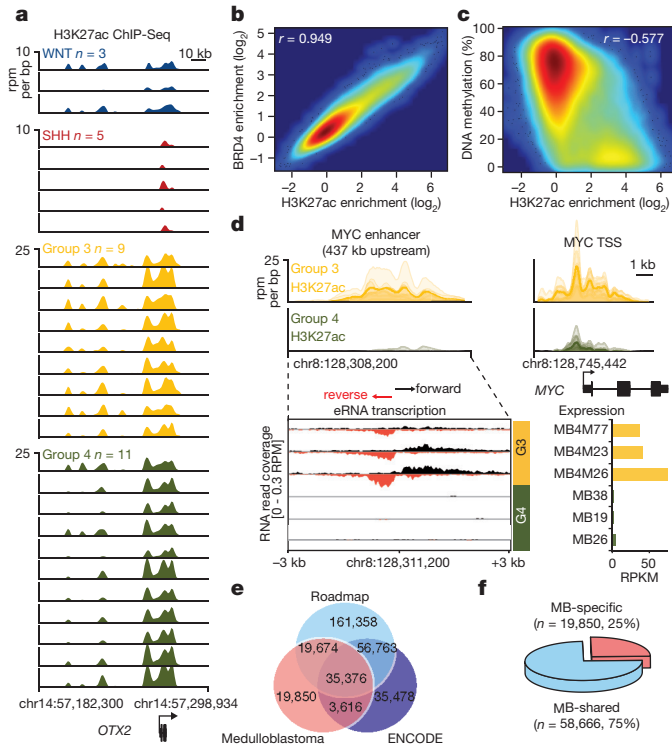
Transcriptional diversity amongst WNT, SHH, Group 3, and Group 4 subgroup medulloblastomas is partially explained by active and discriminatory signalling pathways, such as the Wingless/WNT and Sonic hedgehog/SHH developmental cascades inherent to WNT and SHH medulloblastomas, respectively. Somatically altered driver genes including *MYC* (Group 3), *KDM6A* (Group 4), *GFI1* and/or *GFI1B* (Group 3 and Group 4), and others contribute further to subgroup divergence[2–4]. Recurrent targeting of genes involved in chromatin modification has been the most consistent theme to emerge from recent next-generation sequencing (NGS) studies[5], strongly suggesting deregulation of the epigenome as a critical step during medulloblastoma pathogenesis. However, this hypothesis has yet to be experimentally substantiated and knowledge pertaining to how the medulloblastoma epigenome influences subgroup-specific transcriptional programs remains in its infancy[6].

Enhancers are *cis*-acting regulatory elements that recruit transcription factors (TFs) and chromatin-associated regulatory complexes, which together signal to RNA polymerase to regulate target gene expression[7]. Consortia such as ENCODE[8,9] and the Roadmap Epigenomics Consortium[10] have extensively mapped enhancers, advancing our understanding of enhancer/gene regulation across a comprehensive spectrum of cell lines and tissues. These resources empower our understanding of the complex cartography of the human regulatory landscape, provide testable hypotheses regarding disease–risk association, contribute evolutionary inferences, and establish robust analytical techniques. To deeply characterize the active *cis*-regulatory circuitry of a single disease entity, here medulloblastoma, we performed high-resolution chromatin immunoprecipitation with sequencing (ChIP-seq) for active enhancers (H3K27ac) in 28 primary tumour specimens and three established cell lines. Our approach to studying enhancers genome-wide in a large set of primary tissue samples led to a regulatory explanation for subgroup transcriptional diversity, previously unrecognized subgroup-specific dependencies, and firm insights into medulloblastoma cellular origins, in particular for the poorly characterized Group 3 and Group 4 subgroups.
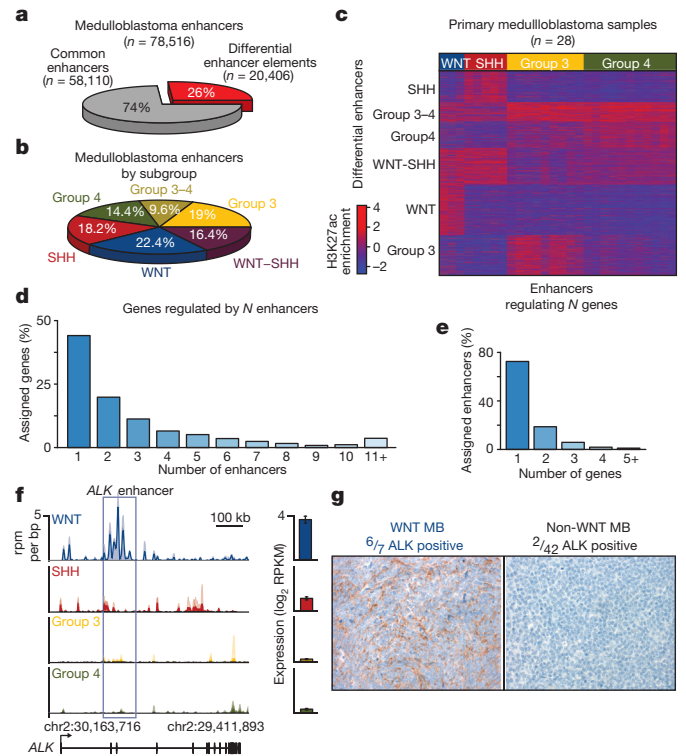
Figure 1 | The enhancer landscape of primary medulloblastoma.
a, Highly active enhancers at the *OTX2* locus across 28 primary
medulloblastomas. **b**, H3K27ac versus BRD4 ChIP-seq signals at
medulloblastoma enhancers ($n = 78,516$). **c**, H3K27ac ChIP-seq signal
versus DNA methylation (WGBS) at medulloblastoma enhancers
($n = 78,516$). **d**, Group 3-specific eRNA expression (lower left)
overlapping a Group 3-specific *MYC* enhancer (upper left) in a subset of
medulloblastomas ($n = 6$). *MYC* gene expression (reads per kilobase of
transcript per million, RPKM) is also shown for the same cases (lower
right). **e, f**, Overlap of medulloblastoma enhancers with ENCODE and
Roadmap enhancers.

**Figure 2 | Differentially regulated enhancers in medulloblastoma
subgroups. a**, ANOVA classification of medulloblastoma enhancers.
b, Distribution of differentially regulated enhancers among
medulloblastoma enhancer classes. **c**, *K*-means clustering of differentially
regulated medulloblastoma enhancers ($n = 20,406$). **d**, Proportion of
enhancer/gene assignments to *N* enhancers. **e**, Proportion of enhancer/
gene assignments to *N* genes. **f**, WNT-specific enhancer activity (reads per
million mapped reads per bp, rpm per bp) and expression (log$_2$ RPKM,
$n = 140$) of *ALK*. Error bars represent standard deviation (s.d.) of the
mean. **g**, Immunohistochemical validation of ALK expression in WNT
medulloblastoma patients ($n = 49$).

The medulloblastoma enhancer landscape

Recent large-scale efforts annotating active regulatory elements
genome-wide in human tissues (for example, through DNase I hyper-
sensitivity, H3K27ac and BRD4 ChIP-seq), have catalogued enhancers
in immortalized or malignant cell lines and normal human tissues,
often under-representing discrete disease entities[8,10]. For medulloblas-
toma, only a single long-term culture cell line (D721; first reported in
1997) is included amongst 125 cell types initially studied by ENCODE[9].
Further, cancer cell lines often exhibit drastic genomic and transcrip-
tional divergence from their corresponding primary tumour tissues as
exemplified in non-Hodgkin's lymphoma where our prior epigenomic
analyses identified greater likeness between primary tumour samples
and normal lymphoid tissues than between tumours and cell lines[11].
Given the apparent limitations of using cell lines to faithfully study the
tumour epigenome, and the recognized subgroup-dependent hetero-
geneity of medulloblastoma, we collected a series of 28 treatment-
naive, fresh-frozen medulloblastoma specimens and profiled the active
enhancer landscape by H3K27ac ChIP-seq (Fig. 1a; Extended Data
Fig. 1a–c).

This cohort is inclusive of all four medulloblastoma subgroups
(Supplementary Table 1; WNT, $n = 3$; SHH, $n = 5$; Group 3, $n = 9$;
Group 4, $n = 11$) and includes three additional Group 3 cell lines
(MED8A, D425, and HD-MB03). Using MACS[12] to identify signifi-
cantly enriched H3K27ac peaks, we inferred 78,516 enhancers, effec-
tively saturating the medulloblastoma enhancer landscape (Extended
Data Fig. 1d). These regions of promoter distal H3K27ac enrichment
mainly (~80%) covered introns and intergenic regions (Extended
Data Fig. 1e). Parallel ChIP-seq was performed for bromodomain

containing 4 (BRD4), an enhancer-associated transcriptional coacti-
vator[11,13], in 27/31 cases. Enrichment of H3K27ac and BRD4 ChIP-seq
signals strongly correlated at putative enhancer loci (Pearson
correlation, $r = 0.949$), further enforcing their active enhancer
classification (Fig. 1b)[11,13]. Likewise, H3K27ac peaks were strongly anti-
correlated with DNA methylation (Pearson correlation, $r = -0.577$;
Fig. 1c) and showed a high degree of overlap with the active/poised
enhancer H3K4me1 but not the repressive H3K27me3 histone marks
(Extended Data Fig. 1f). Finally, strand-specific RNA-seq data gen-
erated from the same cohort detected short, unspliced, bidirectional
RNA transcripts overlapping H3K27ac peaks (Fig. 1d), in accordance
with recently described enhancer RNAs (eRNAs)[14]. Active enhanc-
ers exhibited a modest statistical enrichment for overlap with focal
amplifications and deletions identified in Group 3 and Group 4
(ref. 4; $P = 0.028$ for amplifications, $P = 0.016$ for deletions; Extended
Data Fig. 1g). Comparison of predicted medulloblastoma enhancers
with those reported using analogous methods employed by the
ENCODE and Roadmap Epigenomics projects revealed 19,850
novel regulatory regions, indicative of potentially hindbrain- or
medulloblastoma-specific enhancers in our data set (Fig. 1e, f). Primary
medulloblastoma enhancer landscapes exhibited poor overlap and
correlation with those generated from medulloblastoma cell lines
(Extended Data Fig. 1h, i), further emphasizing the importance of
studying the epigenome in primary tumours.

Analysis of variance (ANOVA) identified sets of enhancers dif-
fering according to known molecular subgroup, revealing 20,406
differentially active enhancers (26% of all inferred enhancers; Fig. 2a, b).
The remaining 74% ($n = 58,110$) displayed varied activity across

subgroups, suggesting either ubiquitous activity of for example, 'housekeeping' genes or a general role in medulloblastoma or cerebellar identity (Fig. 2a; Supplementary Table 2). *K*-means clustering of differentially regulated enhancers delineated six distinct medulloblastoma enhancer classes, including one for each subgroup as well as WNT-SHH and Group 3-Group 4 shared classes (Fig. 2b, c). Group 3 and Group 4 subgroups are known to exhibit some degree of transcriptional similarity[15,16], consistent with the enhancer clustering results, whereas a common subset of shared enhancers between WNT and SHH subgroups was unexpected.

Medulloblastoma enhancer/gene assignment

We next sought to assign enhancer elements to target genes, a process typically hindered by the majority of enhancer–promoter interactions occurring over extensive and highly variable genomic distances[17]. To overcome these challenges, we leveraged sample-matched RNA-seq gene expression data to identify putative enhancer/gene interactions that are (1) contained within the same topologically associated domain (TAD[18]) and (2) exhibit significant positive correlations between enhancer H3K27ac signal and gene expression (false discovery rate (FDR) < 0.05, Extended Data Fig. 2a–i). This approach assigned 8,775 enhancers (43% of all differential enhancers) to at least one protein-coding target gene (Supplementary Table 3). The majority (44%) of inferred target genes were assigned to a single enhancer, but in many cases, several enhancers were predicted to converge on the regulation of a single gene (Fig. 2d). Likewise, 73% of enhancers were assigned to only a single gene target (Fig. 2e). To validate the robustness of our methods, we used 4C-seq[19] to query Group 3-specific enhancer–promoter interactions for enhancers showing conserved activity in both primary Group 3 tumours and cell lines. This approach confirmed enhancer–promoter interactions for both *TGFBR1* and *SMAD9* in the Group 3 cell line HD-MB03, a low-passage line more faithful to primary Group 3 tumours than older models[6,20] (Extended Data Fig. 2j, k).

Medulloblastoma subgroup 'signature' genes have been extensively documented using various expression-profiling methods[15,16]. Enhancer/gene assignments derived from coupling H3K27ac ChIP-seq with RNA-seq produced a refined 'lens' for investigating subgroup-related diversity in medulloblastoma, implicating themes previously undisclosed through expression data alone. For example, enhancers

regulating *ALK*, a receptor tyrosine kinase frequently altered in a variety human cancers[21], were found to be highly active in the WNT subgroup and explained the largely WNT-specific expression pattern detected by RNA-seq and confirmed by immunohistochemical staining of primary patient samples ($n = 49$; $P = 1.35 \times 10^{-5}$, Fisher's exact test; Fig. 2f, g). Further investigations into the potential oncogenic role of ALK in WNT subgroup medulloblastoma are essential but rational given that ALK inhibitors are currently FDA-approved for the treatment of non-small-cell lung carcinoma (that is, crizotinib and ceritinib).

Rational target-based treatment options remain scarce for Group 3 and Group 4 subgroup patients, necessitating additional biological insights to direct future mechanistic and translational research. Functional pathway analysis (see Supplementary Methods) performed on differential enhancer/gene target assignments identified enrichment of neuronal transcriptional regulators in Group 4 and thematic pathways associated with TGF-β signalling in Group 3 (Extended Data Fig. 3a–c). Notably, we uncovered a ~450 kilobase focal amplification at the *ACVR2A* locus in one Group 3 sample that encompassed both the gene and the upstream enhancer regions (Extended Data Fig. 3d). In this sample, enhancers regulating TGF-β pathway components exhibited increased H3K27ac versus other Group 3 tumours (Extended Data Fig. 3e). These data, combined with our prior observations that TGF-β receptor genes are recurrently amplified in Group 3 (ref. 4), further suggest TGF-β signalling as a putative oncogenic driver in this subgroup.

Medulloblastoma subgroup super-enhancers

In multiple tumour types, super-enhancers (SEs), broad spatially co-localized enhancer domains[22,23], have recently been shown to drive oncogenes, genes required for maintenance of tumour cell identity, and genes associated with cell-type-specific functions. To determine whether SEs might play a role in characterizing subgroup-specific identity, we undertook a systematic mapping of SEs across all 28 medulloblastoma samples (see Supplementary Methods; Extended Data Fig. 4a). Massive (>50 kb) SE domains were identified at the cerebellar-specific TFs, *ZIC1* and *ZIC4*[24] (Extended Data Fig. 4b, c), and at ~70% of a queried set of established medulloblastoma driver genes and chromatin modifiers implicated in cancer, including *GLI2*, *MYC*, *OTX2* and others[4] (Extended Data Fig. 4d).

To identify subgroup-specific SEs, we took the union of all enhancer regions in a given subgroup and ranked them by average H3K27ac
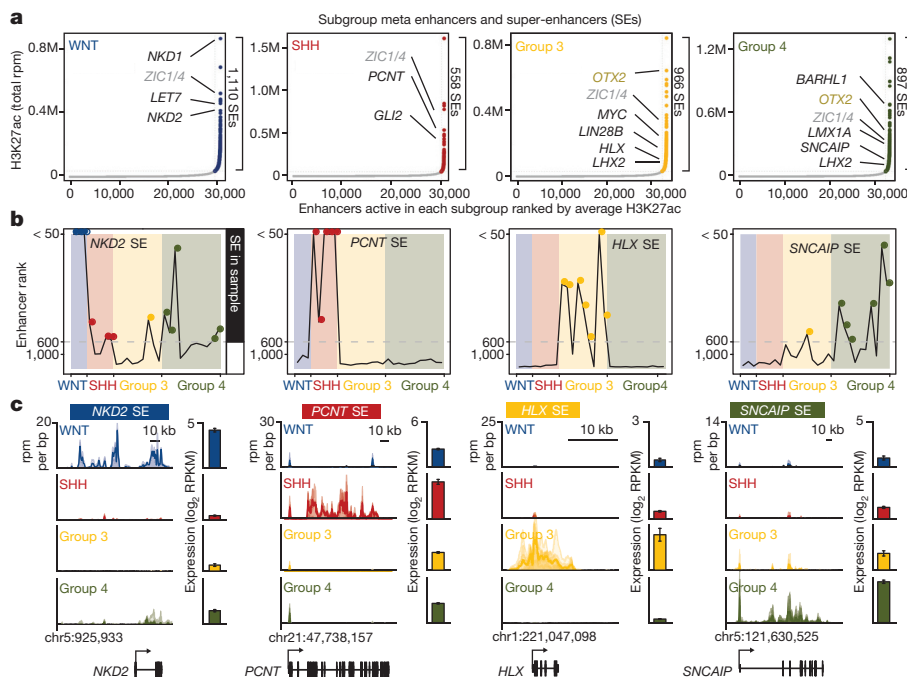
Figure 3 | Medulloblastoma super-enhancers characterize subgroup-specific identity. a, Ranked enhancer plots defined across composite H3K27ac landscapes of WNT, SHH, Group 3, and Group 4 medulloblastomas. Select genes associated with SEs in each subgroup are highlighted and shaded according to enhancer class specificity. **b**, Enhancer rankings for candidate subgroup-specific SEs across all samples according to subgroup. **c**, Meta tracks of H3K27ac ChIP-seq signal (rpm/bp) across medulloblastoma subgroups for the loci shown in **b**. Candidate gene expression (mean RPKM) is shown to the right of each H3K27ac track ($n = 140$). Error bars represent standard deviation (s.d.) of the mean.

enrichment across all samples in that subgroup[23], resulting in ~3,000 distinct SE-containing loci (~600–1,100 SEs per subgroup; Fig. 3a; Extended Data Fig. 4e; Supplementary Table 4). Compared to typical enhancers, SEs showed higher occupancy of BRD4 and greater enhancer signal dynamic range between subgroups (Extended Data Fig. 4f–h). Targets of differential enhancers contained within SEs (that is, SE target genes) included a large fraction of established medulloblastoma signature genes (32%; Supplementary Table 3), as well as novel candidates (Fig. 3a–c). Medulloblastoma SEs were inferred to regulate known Cancer Gene Census genes, including the aforementioned *ALK* in WNT, *SMO* and *NTRK3* in SHH, *LMO1*, *LMO2*, and *MYC* in Group 3, and *ETV4* and *PAX5* in Group 4, among others (Supplementary Table 3). Furthermore, several actionable, SE-regulated genes were revealed in our analysis (Supplementary Table 5).

Unbiased hierarchical clustering of SEs across samples was sufficient to recapitulate transcriptional subgroupings using no prior knowledge of subgroup status, suggesting that SEs might play a pivotal role in characterizing subgroup identity (Extended Data Fig. 4a). SEs from established Group 3 medulloblastoma cell lines clustered with one another, but failed to show similarity to primary Group 3 samples or samples from any other subgroup.

To experimentally validate the activity of medulloblastoma subgroup-specific SEs, we synthesized twenty-two unique SE loci (size range, 1.1–2.1 kb) and evaluated them using Tol2 transposon-mediated zebrafish transgenesis (see Supplementary Methods)[25]. These *in vivo* reporter assays resulted in a validation rate of 45% (10/22), with all reproducibly active enhancer constructs showing specific activity in the zebrafish central nervous system (Fig. 4a, b; Extended Data Fig. 5a–l). We used TF ChIP-seq data for HLX, LHX2, and LMX1A—all highly expressed and SE-regulated in Group 3 and/or Group 4 (Fig. 3a and data not shown)—to enable precise definition of enhancer coordinates (based on TF occupancy) for testing in zebrafish (Fig. 4e), potentially explaining the remarkably high *in vivo* validation rate we observed. These experiments confirmed zebrafish hindbrain-specific activity for an SE (active in WNT and Group 3) mapping approximately 90 kb upstream of *MYC* inferred to regulate MYC expression (Fig. 4b–e). This SE was not found in other common human cancers (Fig. 4d), and in only 4/77 different primary tissues included in Roadmap, suggesting that this validated *MYC* SE is highly specific to the developing hindbrain and/or medulloblastoma (Extended Data Fig. 5m). Importantly, identified *MYC* SEs clearly demarcate a focal amplification hotspot in published Group 3 medulloblastoma copy-number data[4] (Fig. 4c, d), strongly implicating these SEs in the oncogenic regulation of MYC. Collectively, these *in vivo* validation data further substantiate our highly integrative approach for the identification of enhancers and SEs, and inference of their target genes.

SE-regulated TFs reveal cellular origins

Among subgroup-specific SE target genes, we observed an enrichment of TFs involved in neuronal development ($P \approx 0.0001$, Fisher's exact test; Extended Data Fig. 6a). Overall, subgroup-specific TFs displayed similar patterns of expression, enhancer motif enrichment, and overlap of target genes (Extended Data Figs 6b and 7). TFs were also enriched in subgroup-specific SE targets as compared to subgroup-specific non-SE targets ($P \approx 0.002$, Fisher's exact test), consistent with prior observations that SEs regulate key TFs required for tumour cell identity and maintenance[11,13,22]. Given evidence in embryonic stem cells that pluripotency master regulator TFs (OCT4, SOX2, and NANOG) are driven by SEs and themselves bind to and establish SEs[23], we hypothesized that a reverse analysis of SEs in medulloblastoma might enable a *de novo* reconstruction of tumour identity-defining TFs and their associated regulatory circuitry, thereby providing novel insights into medulloblastoma origins.

Pursuant to this idea, we proposed a definition of core regulatory circuitry TFs in which the TFs are SE-regulated and the TFs themselves bind to the SEs of one another (Fig. 5a, see Supplementary Methods). For each SE-regulated TF, these criteria are quantified by measuring the
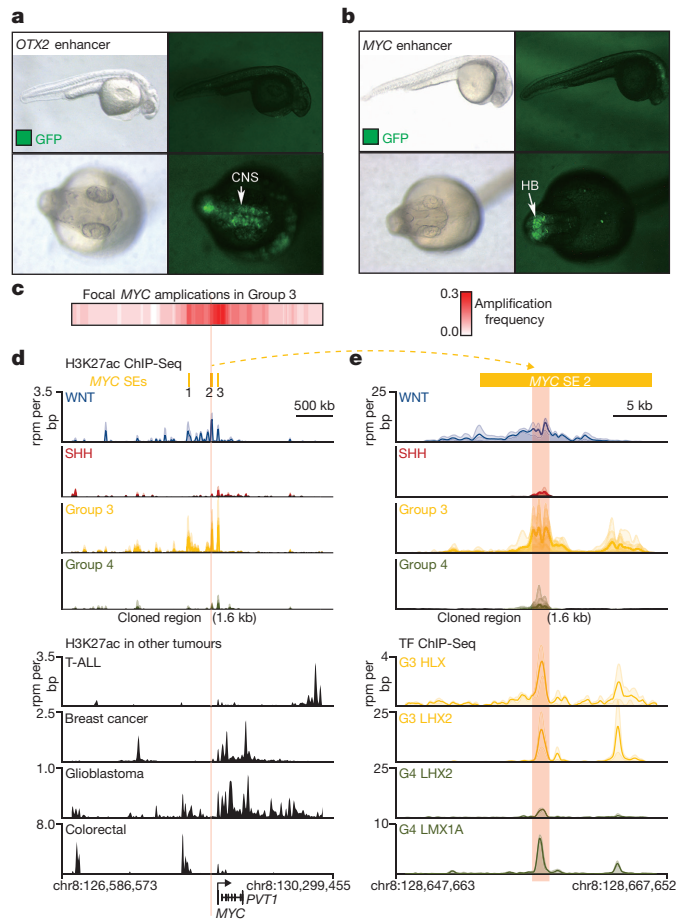

Figure 4 | I*n vivo* validation of medulloblastoma super-enhancers.
a, **b**, Zebrafish reporter assays for *OTX2* (**a**) and *MYC* (**b**) enhancers observed in medulloblastoma. Arrows indicate the locations of GFP signal. CNS, central nervous system; HB, hindbrain. **c**, Heat map summarizing *MYC* copy-number data derived from a published series of Group 3 medulloblastomas ($n = 168$). **d**, **e**, H3K27ac ChIP-seq (upper panels) data showing a shared WNT/Group 3 *MYC* enhancer not found in other human cancers (**d**, lower panel) and occupied by SE-regulated TFs HLX, LHX2, and LMX1A as determined by TF ChIP-seq (**e**, lower panel).

inward binding of other SE associated TFs (in degree) and the outward binding of the TF to other SEs (out degree) (Fig. 5a, b). Regulatory circuitry reconstruction across all SE-associated TFs in medulloblastoma identified cliques of TFs with similar patterns of in/out degree, strong interconnectivity via motif binding, and higher likelihoods of pairwise protein–protein interaction and motif co-occurrence at enhancers (see Supplementary Methods, Extended Data Fig. 8). This reconstruction creates for the first time a core regulatory circuitry blueprint for each subgroup, and implicates specific sets of TFs in establishing medulloblastoma subgroup identity (Extended Data Fig. 9). Importantly, ChIP-seq for the homeodomain TFs HLX (Group 3 network), LMX1A (Group 4 network), and LHX2 (shared Group 3/Group 4 network) performed on select Group 3 and Group 4 primary samples ($n = 4$) largely validated the computationally derived regulatory networks constructed for these subgroups (Fig. 5c, d; Extended Data Figs 8 and 9).

Distinct cellular origins for WNT and SHH medulloblastomas have been experimentally established using a variety of genetically engineered mouse models[26–28]. The origins of Group 3 and Group 4 medulloblastoma, however, are unknown and yet essential to define, as these tumours account for ~60% of all diagnoses, lack targeted therapies, and are frequently associated with poor clinical outcomes[1].

Cell identity is most essentially defined by the activity of master regulator TFs. As such, we hypothesized that the regulatory SE
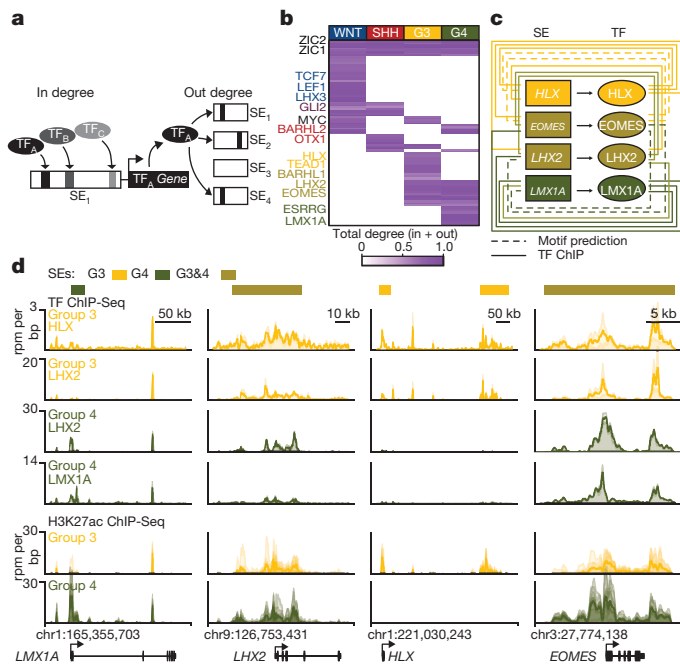
Figure 5 | Super-enhancers characterize medulloblastoma regulatory circuitry. a, Methodology for inferring medulloblastoma core regulatory circuitry. **b**, Heat map of all SE-associated TFs in medulloblastoma (rows) clustered by similarity of regulatory degree. Selected TFs with similar subgroup-specific patterns of regulatory degree are annotated. **c**, Subgroup-specific regulatory circuitry in Group 3 and Group 4 medulloblastoma. **d**, TF and H3K27ac ChIP-seq meta tracks for the SE-regulated TFs *LMX1A*, *LHX2*, *HLX*, and *EOMES*.

regions governing endogenous expression of candidate master TFs and embedded in the core regulatory circuitry of medulloblastoma subgroups might inform cellular origins of the disease via their cell-type-specific activity. During early cerebellar development, LMX1A, EOMES, and LHX2—master regulator Group 4 TFs deduced from our core regulatory circuitry analysis (Fig. 5c)—exhibit overlapping spatiotemporal restricted expression in the nuclear transitory zone (NTZ; Fig. 6a), an assembly point for immature deep cerebellar nuclei (DCN). DCN residing in the NTZ at this time point are predominantly glutamatergic projection neurons that originate from earlier progenitors of the upper rhombic lip (uRL), a transient germinal zone producing progenitors with distinct cellular fates, including DCN and cerebellar granule neurons[29]. Immunofluorescence microscopy confirmed compartmentalized expression of LMX1A, EOMES, and

LHX2 that was notably distinct from ATOH1 expression, the latter marking the early external granule layer (EGL) at this developmental stage (Fig. 6b).

Both LMX1A enhancer activity and expression are highly discriminatory for Group 4, nominating this TF as a master regulator of the Group 4 transcriptional program (Figs 3a and 5c; Extended Data Figs 8 and 9). Indeed, LMX1A ChIP-seq performed on Group 4 primary samples verified >90% of predicted target genes inferred through motif-driven computational analyses (Extended Data Figs 8 and 9). LMX1A is a LIM-homeodomain TF previously shown to function as a critical regulator of cell-fate decisions in the uRL and essential for normal cerebellar development[30]. Spontaneous *Lmx1a* loss-of-function null mutations are causative in *dreher* mice, resulting in profound cerebellar phenotypes typified by premature regression of the RL, reduced choroid plexus, and cerebellar hypoplasia predominantly affecting the posterior vermis (Fig. 6c)[31]. To further investigate the molecular targets associated with *dreher* cerebellar phenotypes, we microdissected uRL from wild-type and *dreher* (dr^J/dr^J) mice at embryonic day 13.5 and delineated transcriptional differences through expression profiling. Strikingly, SE-regulated TFs contained in Group 3/Group 4 regulatory circuitry (Extended Data Fig. 9) were among the most differentially expressed genes in *dreher* uRL compared to controls (Fig. 6d, e). Collectively, these phenotypic and molecular data further support Lmx1a as a master regulator TF in both the cerebellar uRL and in Group 4 medulloblastoma, implicating the uRL compartment and its derivate precursors as putative cells-of-origin for Group 4.

Discussion

We describe the active medulloblastoma enhancer landscape across a series of 28 fresh-frozen, treatment-naive tissue samples and three cultured cell lines, to our knowledge representing the largest such data set for any single cancer entity. Our data reveal drastic divergence between primary tumour and tumour cell line material and uncover considerable *cis*-regulatory element heterogeneity between subgroups of the disease that would be unsubstantiated in series limited to just a few cases.

Clinically relevant medulloblastoma subgroups are principally defined based on their underlying transcriptional profiles. Differentially regulated medulloblastoma enhancers and SEs are here shown to recapitulate these subgroups, and importantly extend our understanding of this disease to inferences regarding cell specification and actionable tumour dependencies. Biological themes and signalling networks extracted from transcriptional data have served as the primary source of annotation for medulloblastoma subgroups, with WNT and SHH subgroups characterized by activation of their respective signalling pathways, and Group 3 and Group 4 recognized for their GABAergic and glutamatergic expression phenotypes, respectively. Although these
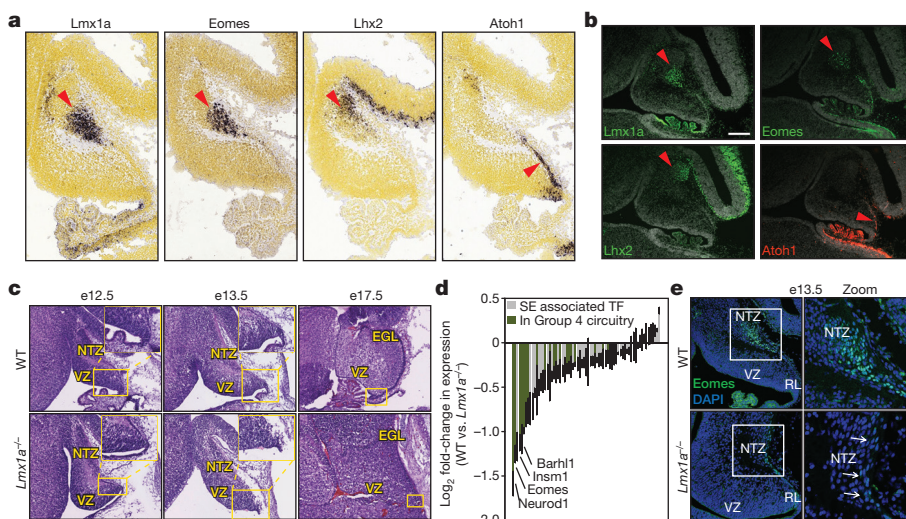
Figure 6 | Master transcription factors implicate Group 4 cellular origins. a, Expression (*in situ* hybridization) of *Lmx1a*, *Eomes*, *Lhx2*, and *Atoh1* in the embryonic cerebellum (e13.5). **b**, Immunofluorescence microscopy for the TFs shown in **a** performed on sagittal sections of the e13.5 murine cerebellum. **c**, Haematoxylin and eosin-stained cerebellar sections (sagittal) of wild-type and dr^J/dr^J (*Lmx1a$^{-/-}$*) embryos. The RL is demarcated by a yellow box in each panel. **d**, Differentially expressed TFs in the e13.5 RL of *Lmx1a$^{-/-}$* embryos versus wild-type controls. Error bars represent standard error of the mean (*n* = 3). **e**, Immunofluorescence microscopy confirming Eomes downregulation in *Lmx1a$^{-/-}$* embryos (e13.5).

data provide a functional and phenotypic annotation of medulloblastoma, they fail to articulate the developmental identity of individual subgroups. Using a reverse analysis of the medulloblastoma chromatin landscape starting at the level of differentially regulated enhancers and SEs, we have reconstructed and experimentally validated the core regulatory circuitry inherent to medulloblastoma subgroups, inferring master transcriptional regulators responsible for subgroup-specific divergence. The majority of these master regulator TFs were not previously implicated in medulloblastoma development, nor were they visible amongst transcriptionally-derived gene sets dominated by overwhelming phenotypic signatures. Through tracing the spatiotemporal activity of a subset of Group 4 master TFs, these studies identified DCN of the cerebellar NTZ, or plausibly their earlier precursors originating from the uRL, as putative cells-of-origin for this large subgroup of patients. Together these approaches establish a framework for the inference of tumour cell-of-origin through enhancer core regulatory circuitry mapping.

Identifying the cellular origins of cancer has broad implications for the understanding and treatment of malignancy[32]. Although tumour cells deviate from their developmental origins during transformation, numerous cancers, especially those of the immune compartment, still maintain developmental TF activity and as such are treatable through targeting of the lineage (for example, anti-B cell therapies for leukaemia)[33,34]. As medulloblastoma is believed to originate from cell populations that normally exist ephemerally during development, targeting the aberrant persistence of tumour cells from these lineages may represent a novel therapeutic strategy with minimal effect on the normal tissue compartment. Moreover, elucidation of master TFs of medulloblastoma implicates upstream signalling pathways, transcriptional co-activators, and downstream effectors as potential subgroup-specific targets for rational therapeutic intervention. These insights demonstrate the critical importance of epigenetic analyses of primary tumours as opposed to cell line model systems and highlight the broad utility of core regulatory circuitry mapping, especially in poorly characterized and clinically heterogeneous malignancies.

1. Northcott, P. A., Korshunov, A., Pfister, S. M. & Taylor, M. D. The clinical implications of medulloblastoma subgroups. *Nat. Rev. Neurol.* **8,** 340–351 (2012).
2. Jones, D. T. *et al.* Dissecting the genomic complexity underlying medulloblastoma. *Nature* **488,** 100–105 (2012).
3. Northcott, P. A. *et al.* Enhancer hijacking activates GFI1 family oncogenes in medulloblastoma. *Nature* **511,** 428–434 (2014).
4. Northcott, P. A. *et al.* Subgroup-specific structural variation across 1,000 medulloblastoma genomes. *Nature* **488,** 49–56 (2012).
5. Northcott, P. A. *et al.* Medulloblastomics: the end of the beginning. *Nature Rev. Cancer* **12,** 818–834 (2012).
6. Hovestadt, V. *et al.* Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature* **510,** 537–541 (2014).
7. Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nature Rev. Genet.* **15,** 272–286 (2014).
8. Consortium, E. P.; ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489,** 57–74 (2012).
9. Thurman, R. E. *et al.* The accessible chromatin landscape of the human genome. *Nature* **489,** 75–82 (2012).
10. Kundaje, A. *et al.*; Roadmap Epigenomics Consortium. Integrative analysis of 111 reference human epigenomes. *Nature* **518,** 317–330 (2015).
11. Chapuy, B. *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24,** 777–790 (2013).
12. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9,** R137 (2008).
13. Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153,** 320–334 (2013).
14. Kim, T. K. *et al.* Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465,** 182–187 (2010).
15. 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).
16. Northcott, P. A. *et al.* Medulloblastoma comprises four distinct molecular variants. *J. Clin. Oncol.* **29,** 1408–1414 (2011).
17. Jin, F. *et al.* A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503,** 290–294 (2013).
18. Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515,** 402–405 (2014).
19. Gröschel, S. *et al.* A single oncogenic enhancer rearrangement causes concomitant EVI1 and GATA2 deregulation in leukemia. *Cell* **157,** 369–381 (2014).
20. 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).
21. Hallberg, B. & Palmer, R. H. Mechanistic insight into ALK receptor tyrosine kinase in human cancer biology. *Nature Rev. Cancer* **13,** 685–700 (2013).
22. Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155,** 934–947 (2013).
23. Whyte, W. A. *et al.* Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153,** 307–319 (2013).
24. Aruga, J. *et al.* Mouse Zic1 is involved in cerebellar development. *J. Neurosci.* **18,** 284–293 (1998).
25. Rada-Iglesias, A. *et al.* A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470,** 279–283 (2011).
26. Gibson, P. *et al.* Subtypes of medulloblastoma have distinct developmental origins. *Nature* **468,** 1095–1099 (2010).
27. Schüller, U. *et al.* Acquisition of granule neuron precursor identity is a critical determinant of progenitor cell competence to form Shh-induced medulloblastoma. *Cancer Cell* **14,** 123–134 (2008).
28. 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).
29. Fink, A. J. *et al.* Development of the deep cerebellar nuclei: transcription factors and cell migration from the rhombic lip. *J. Neurosci.* **26,** 3066–3076 (2006).
30. Chizhikov, V. V. *et al.* Lmx1a regulates fates and location of cells originating from the cerebellar rhombic lip and telencephalic cortical hem. *Proc. Natl Acad. Sci. USA* **107,** 10725–10730 (2010).
31. Millonig, J. H., Millen, K. J. & Hatten, M. E. The mouse *Dreher* gene *Lmx1a* controls formation of the roof plate in the vertebrate CNS. *Nature* **403,** 764–769 (2000).
32. Gilbertson, R. J. Mapping cancer origins. *Cell* **145,** 25–29 (2011).
33. Byrd, J. C. *et al.* Targeting BTK with ibrutinib in relapsed chronic lymphocytic leukemia. *N. Engl. J. Med.* **369,** 32–42 (2013).
34. Hale, G. *et al.* Remission induction in non-Hodgkin lymphoma with reshaped human monoclonal antibody CAMPATH-1H. *Lancet* **332,** 1394–1399 (1988).

Author Contributions P.A.N., J.E.B., and S.M.P. conceived and co-led the study. C.Y.L. and S.E. performed all bioinformatics related to the analysis of medulloblastoma enhancers and super-enhancers. Y.T., L.Y., D.K., B.C.W., B.J., and W.C. validated subgroup-specific enhancers *in vivo*. C.Y.L. and A.J.F. constructed medulloblastoma regulatory circuitry networks. M.Z., S.M.W., R.Z., D.R.P., M.S.-W., D.T.W.J., M.K., V.H., I.B., and L.C. provided informatics and general scientific support. P.H., V.V.C., and K.J.M. performed the developmental studies with *dreher* and WT mouse embryonic cerebella. T.R., H.-J.W., V.A., H.L., and M.-L.Y. conducted RNA-seq data generation and enhancer RNA analysis. B.A.O. performed ALK staining on medulloblastoma TMAs. L.S., P.J., and S.G. performed 4C-seq experiments. M.R. and A.K. provided medulloblastoma tissue samples. R.E., P.L., J.O.K., S.M.P., J.E.B., and P.A.N. provided institutional support and project supervision. C.Y.L., S.E., S.M.P., J.E.B., and P.A.N. prepared the figures and wrote the manuscript.

METHODS

No statistical methods were used to predetermine sample size.

Identifying super-enhancer constituents for reporter assays. We sought to identify candidate Group 3 and Group 4 super-enhancer constituents for validation by reporter assays. We identified candidate Group 3 and Group 4 super-enhancer constituents by first locating nucleosome free "valleys" in the H3K27ac data using an algorithm adapted from ref. 35. Valleys that showed strong evidence of TF ChIP-seq binding for respective Group 3 (HLX and LHX2) and Group 4 (LHX2 and LMX1A) TFs were selected and manually curated for validation in reporter assays. Based on restrictions for DNA synthesis and cloning, candidate reporter regions of roughly ±1 kb flanking the valley centre were used (Fig. 4 and Extended Data Fig. 5).

Zebrafish _in vivo_ enhancer assays. All experiments involving zebrafish (_Danio rerio_, AB strain) were approved by the Vanderbilt Institutional Animal Care and Use Committee. For _in vivo_ zebrafish reporter assays, a minimum, ~150–200 embryos (male and female) were injected per reporter construct and assays were repeated 2–3 times per construct to confirm reproducibility. No randomization of enhancer assays was performed. The scientist who performed the injections had no prior knowledge related to the enhancer constructs and was therefore blinded to the experiment. Microinjection was done as described previously[36]. In brief, a mixture of individual enhancer-containing vector DNA (25 μg ml^{-1}) and transposase RNA (25 μg ml^{-1}) was injected into zebrafish zygotes (1 nl per zygote). The injected embryos were cultured in 0.3× Danieau's solution at 28.5 °C. After 24 hours, the embryos were examined for eGFP expression under a fluorescent dissecting microscope (Zeiss Discovery V12) to determine the stereotypic expression pattern conferred by the enhancer. The total number of embryos injected with the construct and the number of embryos with the stereotypical eGFP pattern were determined to calculate the frequency of the pattern. Embryos were dechorionated and imaged using a Zeiss AxioCam HRc digital camera.

Immunofluorescence microscopy. Spatial protein expression of medulloblastoma$_{G4}$-specific transcription factors in e13.5 cerebella was determined by IHC. PFA-fixed frozen tissues were sectioned (12 μm thickness) and processed without antigen retrieval steps. The antibodies used here are Tbr2 (1:100, Abcam, ab23345), Lmx1a (1:100, Novus Biologicals, NBP1-81303), Atoh1 (1:500, Abcam, ab105497) and appropriate secondary antibodies conjugated with Alexa fluorophores (1:400, Invitrogen). The images were captured by an epifluorescence microscopy.

Analysis of Allen Brain Atlas data portal. Endogenous expression of candidate TFs was determined by querying the Allen Brain Atlas Data Portal (http://developingmouse.brain-map.org) at various developmental time points.

Medulloblastoma tissue microarrays (TMAs). The molecular subgroup of 49 medulloblastoma samples on tissue microarrays were determined as previously described[37]. Immunohistochemistry was performed using clone ALK01 (#790-2918, Ventana) with appropriate secondary reagents. Individual tumours were scored positive in the presence of cytoplasmic immunoreactivity for ALK1, whereas the tumour was considered negative in the absence of immunoreactivity.

Phenotypic analysis of Dreher (_Lmx1a_$^{-/-}$) embryos. All mouse (_Mus musculus_, B6C3HFe background) experiments were done in accordance with the guidelines laid down by the Institutional Animal Care and Use Committee (IACUC), of Seattle Children's Research Institute. No randomization or experimental blinding related to mouse experiments was performed. _Lmx1a_$^{+/-}$ mice were crossed and the day of plug was taken as e0.5. WT and _Lmx1a_$^{-/-}$ embryos (male and female) were dissected out between e12.5 and e17.5 and subsequently fixed in 4% paraformaldehyde (PFA) for 2–6 hours. The fixed embryos were washed in PBS and incubated in 30% sucrose overnight. The following day, embryos were frozen in optimum cutting temperature (OCT) compound. Mid-sagittal cryo-sections of the cerebellum at 11 μm were taken. Haematoxylin and eosin staining and immunohistochemistry were performed as described previously[38]. Briefly, cryosections were incubated at room temperature for 1 hour after which they were subjected to heat-mediated antigen retrieval. All sections were blocked using 5% serum containing 0.35% Triton X, and then incubated with the primary antibody (Eomes (Tbr2); #14-4875, ebioscience, mouse, 1:200), overnight. The following day fluorescent-dye-labelled secondary antibodies (Alexa Fluor 488, 1:1000, Molecular Probes) were used. Sections were counter stained using DAPI (4′,6-diamidino-2-phenylindole) (Vector Laboratories). All images were captured at room temperature. Haematoxylin and eosin-stained sections were imaged a using Hamamatsu Nanozoomer whole slide scanner. All confocal images were captured using Zeiss LSM Meta and Zen 2009 software.

Collection of patient material and cell lines. An Institutional Review Board ethical vote (Ethics Committee of the Medical Faculty of Heidelberg) was obtained according to 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%. The ChIP-seq cohort was established based on tissue availability and availability of orthogonal data types (for example, WGS, RNA-seq) and patient metadata (for example, molecular subgroup). Subgroup assignments were made using the Illumina 450K DNA methylation array as described[39]. Medulloblastoma cell lines were cultured at 37 °C with 5% CO$_2$. D425_Med (D425; a gift from D. D. Bigner) and MED8A cells (from the authors' own stocks; T. Pietsch) were cultured in DMEM with 10% FCS (Life Technologies). HD-MB03 cells[20] were grown in RPMI-1640 with 10% FCS (Life Technologies). All cells were regularly authenticated and tested for mycoplasma (Multiplexion, Heidelberg, Germany).

ChIP-sequencing. H3K27ac, BRD4, H3K27me3, H3K4me1, LMX1A, LHX2, and HLX ChIP was performed at ActiveMotif (Carlsbad, CA) using antibodies against H3K27ac (AM#39133, Active Motif), BRD4 (#A301-985A, Bethyl Laboratories), H3K27me3 (#07-449, Millipore), H3K4me1 (AM#39298, ActiveMotif), LMX1A (#AB10533, Millipore), LHX2 (#sc-19344, Santa Cruz), and HLX (#HPA005968, Sigma). Fresh-frozen medulloblastoma tissues (or cell lines) were submersed in PBS + 1% formaldehyde, cut into small pieces and incubated at room temperature for 15 min. Fixation was stopped by the addition of 0.125 M glycine (final concentration). The tissue pieces were then treated with a TissueTearer and finally spun down and washed 2× in PBS. Chromatin was isolated by the addition of lysis buffer, followed by disruption with a Dounce homogenizer. Lysates were sonicated and the DNA sheared to an average length of ~300–500 bp. Genomic DNA (input) was prepared by treating aliquots of chromatin with RNase, proteinase K and heat for de-crosslinking, followed by ethanol precipitation. Pellets were resuspended and the resulting DNA was quantified on a NanoDrop spectrophotometer. Extrapolation to the original chromatin volume allowed quantitation of the total chromatin yield. An aliquot of chromatin (30 μg) was precleared with protein A (G – for goat pc or monoclonal antibodies) agarose beads (Invitrogen). Genomic DNA regions of interest were isolated using 4 μg of antibody. ChIP complexes were washed, eluted from the beads with SDS buffer, and subjected to RNase and proteinase K treatment. Crosslinks were reversed by incubation overnight at 65 °C, and ChIP DNA was purified by phenol-chloroform extraction and ethanol precipitation. Quantitative PCR (qPCR) reactions were carried out in triplicate on specific genomic regions using SYBR Green Supermix (Bio-Rad). The resulting signals were normalized for primer efficiency by carrying out qPCR for each primer pair using Input DNA.

Illumina sequencing libraries were prepared from the ChIP and Input DNAs by the standard consecutive enzymatic steps of end-polishing, dA-addition, and adaptor ligation. After a final PCR amplification step, the resulting DNA libraries were quantified and sequenced on the Illumina HiSeq 2000 platform using 2 × 101 cycles according to the manufacturer's instructions. Alignment, and downstream processing of ChIP-seq data was performed as described[6].

RNA-sequencing and transcriptome read alignment. RNA was extracted from fresh frozen tissue samples using the AllPrep DNA/RNA/Protein Mini kit (Qiagen) including DNase I treatment on column. All samples were subjected to quality control on a Bioanalyzer instrument. RNA sequencing libraries were prepared from 10 μg of total RNA. Strand-specific RNA sequencing was performed following a protocol described previously[40,41]. Sequencing was carried out with 2×51 cycles on a HiSeq 2000 instrument (Illumina). All reads were aligned to the human reference genome (1000 genomes version of human reference genome hg19/GRCh37) using BWA (v 0.5.9-r16). Aligned reads were converted to the SAM/BAM format using SAMtools. Gene annotation was based on Ensembl v70 (_Homo sapiens_).

4C-seq. 4C samples were prepared from Group 3 medulloblastoma cell line HD-MB03 using the method as described[19,42]. DpnII was used as the primary restriction enzyme and Csp6I as the secondary restriction enzyme in template generation. Sample libraries for SMAD9 and TGFBR1 were amplified using the primers, SMAD9_F: TTATCCAGGCAAGGAAGATC, SMAD9_R: ATTACCTCATCTGCAAAACC, TGFBR1_F: CATTCTTTCTCCCCATGATC, and TGFBR1_R: ACACAATCTTGGGTGTTTTT, respectively. Amplified libraries were multiplexed, spiked with 40% PhiX viral genome and sequenced on Hiseq 2000. Reads were mapped to human genome (hg19) using Bowtie (v 1.0.0)[43].

Identification of enhancer RNA candidates. Forward and reverse RNA transcription based on directional RNA sequencing data was quantified in 3 kb windows upstream and downstream of enhancer peaks that were based on H3K27ac ChIP-seq data, resulting in four RNA expression values for each enhancer region: (L_fwd) forward transcription left of enhancer peak, (R_fwd) forward transcription right of enhancer peak, (L_rev) reverse transcription left of enhancer peak, and (R_rev) reverse transcription right of enhancer peak. We calculated the "directionality index" D, a measure of the directionality of transcription in an enhancer region, with D ranging from 0 to 1, by $D = |R\_fwd - L\_rev|/(R\_fwd + L\_rev)$ as described before[14], with low D values representing bidirectional eRNA transcription.

For correlation of eRNA transcription values with corresponding gene expression values, we calculated eRNA transcription values in 3 kb windows upstream and downstream of enhancer peaks by eRNA_transcription $= (R\_fwd + L\_rev)/2$.

Genomic coordinates and gene annotation. All coordinates in this study were based on human reference genome assembly hg19, GRCh37 (http://www.ncbi.nlm.nih.gov/assembly/2758/). Gene annotations were based on gencode annotation release 19 (http://www.gencodegenes.org/releases/19.html).

Calculating read density. We calculated the normalized read density of a ChIP-seq data set in any genomic region using the Bamliquidator (version 1.0) read density calculator (https://github.com/BradnerLab/pipeline/wiki/bamliquidator). Briefly, ChIP-seq reads aligning to the region were extended by 200 bp and the density of reads per base pair (bp) was calculated. The density of reads in each region was normalized to the total number of million mapped reads producing read density in units of reads per million mapped reads per bp (rpm per bp).

Plotting meta representations of ChIP-seq signal. To compactly display medulloblastoma H3K27ac ChIP-seq signal at individual genomic loci and across subgroups, we developed a simple meta representation (Fig. 1d and others). For all samples within a group, ChIP-seq signal is smoothed using a simple spline function and plotted as a translucent shape in units of rpm per bp. Darker regions indicate regions with signal in more samples. An opaque line is plotted and gives the average signal across all samples in a group.

Peak finding and classification. H3K27ac peak finding was performed using MACS[12] with a P-value threshold of 1×10^{-9}, and with other settings as default parameters. Peak finding for each medulloblastoma was performed separately and as a control background for each H3K27ac ChIP-seq sample, its matched genomic DNA was used. The SPOT statistic[44], a measure of read fraction found in enriched regions developed by the ENCODE consortium, was used to quantify H3K27ac enrichment quality. Primary medulloblastoma data sets had a median SPOT score of 0.62 which was equivalent to cell line data and on par with primary human data generated in the Epigenome ROADMAP. Afterwards, H3K27ac peaks were merged into a single coordinate file. Peaks which can not be identified in at least two primary medulloblastomas and contained completely within the region surrounding ± 1 kb TSS were excluded from any further analysis. This resulted in final combined and filtered peak set ($n = 78516$). H3K27ac enrichments were calculated on the final peak set using the following formula: $\log2(((Cnt_{ChIP}/LSize_{ChIP}*\min(LSize_{ChIP}, LSize_{cnt}))+pscnt)/ ((Cnt_{cnt}/LSize_{cnt}*\min(LSize_{ChIP}, LSize_{cnt}))+pscnt))$, where Cnt_{ChIP} denotes the total number of reads mapping to the enhancer coordinate in ChIP sample, $LSize_{ChIP}$ is the total library size for the ChIP sample, Cnt_{cnt} is the total number of reads mapping to the enhancer coordinate in the control genomic DNA, $LSize_{cnt}$ is the total library size for the control sample, and pscnt is a constant number (pscnt $= 8$), which was used to stabilize enrichments based on low read counts. (Peaks showing statistically significant differential H3K27ac enrichment across medulloblastoma subgroups were determined using ANOVA and the ones with FDR < 0.01 were preserved after multiple testing correction. From the resulting peak-set, peaks having 1.5 ($\log_2$) fold change difference across any medulloblastoma subgroup comparison were called as "subgroup specific" enhancers ($n = 20,406$). Peaks that do not fulfil these criteria were referred as "common" enhancers ($n = 58,110$). Subgroup-specific enhancers were further clustered using k means, with $k = 6$ into 6 groups as "SHH", "WNT", "Group4", "WNT-SHH", "Group3-Group4", and "Group3" (Fig. 2).

Coverage of medulloblastoma enhancers in the genome. Genome was classified into regions as exon, intron, intergenic and promoter (region surrounding ± 1 kb transcriptional start sites) by following the hierarchy promoter $>$ exon $>$ intron $>$ intergenic. Then, medulloblastoma enhancers were intersected with these defined elements and fraction covered by each element was calculated.

Enhancer saturation analysis. To better understand whether our enhancer profiling adequately captured the primary medulloblastoma enhancer landscape, we performed a saturation analysis. We measured the total number of discreet regions and the fraction of novel regions gained by increasing sample number. This was performed across 1,000 permutations of the 28 medulloblastoma samples to establish 95% confidence intervals (Extended Data Fig. 1d).

Comparison of H3K27ac with BRD4 occupancy and DNA methylation at enhancers. Enrichment values for H3K27ac at enhancers were calculated as the ratio between library size normalized read counts for H3K27ac ChIP and its sample matched genomic DNA control. The formula used for the enrichment calculation is as follows: $\log2(((Cnt_{ChIP}/LSize_{ChIP}*\min(LSize_{ChIP}, LSize_{cnt}))+pscnt)/ ((Cnt_{cnt}/LSize_{cnt}*\min(LSize_{ChIP}, LSize_{cnt}))+pscnt))$, where Cnt_{ChIP} denotes the total number of reads mapping to the enhancer coordinate in ChIP sample, $LSize_{ChIP}$ is the total library size for the ChIP sample, Cnt_{cnt} is the total number of reads mapping to the enhancer coordinate in the control genomic DNA, $LSize_{cnt}$ is the total library size for the control sample, and pscnt is a constant number (pscnt $= 8$), which was used

to stabilize enrichments based on low read counts. To compare BRD4 enrichment with H3K27ac enrichment at the enhancers, BRD4 enrichments were calculated in the same way as H3K27ac enrichments. DNA methylation values at enhancers were determined by calculating the average DNA methylation of all medulloblastoma samples where DNA methylation data are available[6].

Comparison of H3K27ac occupancy with H3K4me1, H3K27me3 and BRD4 occupancy. We generated ChIP-seq data for H3K4me1 and H3K27me3 for only three Group 3 medulloblastomas (MB-1M21,MB-4M23, and MB-4M26). Therefore, comparison of H3K27ac occupancy with H3K4me1, H3K27me3 and BRD4 (Extended Data Fig. 1f) was done using the data from only these three Group 3 samples. To analyse the occupancy of the marks at H3K27me3 enriched regions, we called H3K27me3 peaks using MACS. ChIP-seq reads covering each base pair either in the region ± 5 kb around Group 3-specific enhancer midpoints (Extended Data Fig. 1f top panel) or in the region ± 5 kb around H3K27me3 peak midpoints (Extended Data Fig. 1f bottom panel) were quantified. Read coverage was averaged in 100-bp windows along the regions and the values were scaled to arrange between 0–1. Resulting values were represented as heat maps.

Comparison of H3K27ac peak calling using whole genome sequencing or whole cell extract backgrounds. We repeated H3K27ac peak finding (running MACS with a P-value threshold of 1×10^{-9}, and with other settings as default parameters) for the two medulloblastomas (MB12 and MB200) using their input chromatin as the backgrounds instead of using their matched whole genome sequencing. Resulting set of peaks identified using whole chromatin extract were compared to the ones identified using whole genome sequencing in scatter plots in Extended Data Fig. 1c.

Comparison of medulloblastoma H3K27ac enhancers with published H3K27ac data. ENCODE[8] H3K27ac peaks were downloaded from http://ftp.ebi.ac.uk/pub/databases/ensembl/encode/integration_data_jan2011/byDataType/peaks/jan2011/histone_macs/optimal/hub/ and all peaks were merged into a single coordinate file. Regarding ROADMAP data[45,46], all available H3K27ac alignment files were downloaded and peak finding on individual samples was performed using MACS[12]. All ROADMAP H3K27ac peaks were as well merged into a single coordinate file. Resulting peaks from both ENCODE and ROADMAP were intersected with medulloblastoma H3K27ac peaks (with a minimum 50% overlap criteria; Fig. 1e, f).

Comparison of medulloblastoma H3K27ac enhancers with CNV data. To determine the overlap of enhancer loci with CNVs, medulloblastoma enhancer loci were intersected with focal amplifications and deletions obtained from[4]. To determine the statistical significance of the overlap, we performed 10,000 random simulations whereby CNV locations were randomly permuted across the genome without overlap using the bedtools shuffle utility (http://bedtools.readthedocs.org) and excluding regions found in the ENCODE[8] blacklist (https://sites.google.com/site/anshulkundaje/projects/blacklists). This distribution of random overlaps was used to calculate an empirical P-value of the observed overlap significance (Extended Data Fig. 1g).

Quantification of gene expression and assignment of subgroup specific expression. Expression values in RPKM were calculated using "qCount" function of Bioconductor package "quasR" (http://www.bioconductor.org/packages/release/bioc/html/QuasR.html). Genes showing differential gene expression across four medulloblastoma subgroups were determined using ANOVA (FDR less than 1%). Then, subgroup specific assignment of gene expression was done by performing a post-hoc test (using "glht" function of R package "multicomp")[56].

Identification of enhancer target genes. Target gene identification of enhancers was performed as described[47]. For each enhancer, topology-associated domain (TAD)[18] which it belongs to was identified. Then, genes with transcriptional start sites falling into the same TAD were determined. We filtered nearby genes for protein coding status, as eRNAs and other enhancer associated ncRNAs are likely to emanate from enhancers and obfuscate distal target genes. Correlation tests (Spearman's rank correlation coefficient) for H3K27ac enrichment of the enhancer and expression level of genes which are in the same TAD were performed. After repeating this procedure for each enhancer, all P-values obtained via correlation tests were combined and corrected for multiple testing globally using Bioconductor package "qvalue" (http://www.bioconductor.org/packages/release/bioc/html/qvalue.html). Correlations with a FDR less than 5% were preserved. For each enhancer, gene whose expression best correlates with the H3K27ac enrichment of the enhancer was selected as the potential target gene. For the cases where the difference between spearman correlation coefficients for the best and second best correlating genes were less than 0.1, the second best correlating gene was also selected as another potential target gene. Identification of enhancer target genes was performed for subgroup specific and common enhancers separately. After getting final gene lists for targets of subgroup specific and common enhancers, genes which are identified as targets both for subgroup specific and common enhancers were removed from common enhancer target gene list.

Classification of enhancer targets according to enhancer regulation. Genes regulated by differential enhancers were classified into categories depending on the number of differential enhancers they are targeted by (Fig. 2d). As mentioned in "identification of enhancer targets" part, to assign the enhancers to their targets with highest probability, in the final list of enhancer target genes, number of genes per enhancer was restricted to 2 genes having the highest correlation coefficient. However, to evaluate the number of genes targeted by each enhancer overall, enhancers were classified into categories depending on the number of genes they target by including all the genes targeted by enhancers (satisfying FDR < 0.05 criteria) (Fig. 2e).

Overlap of target genes with regulatory information from literature. Medulloblastoma signature genes were defined to be the genes regulated differentially in 4 medulloblastoma subgroups[16]. To be conservative on the signature genes, for each medulloblastoma subgroup, top 100 genes differentially regulated in the respective subgroups were included in the analysis. Resulting gene list were compared to the genes regulated by medulloblastoma subgroup specific enhancers and super-enhancers. Comparison to cancer genes was performed using the gene list provide in cancer gene census (http://cancer.sanger.ac.uk/cancergenome/projects/census). Target genes were overlapped with consensus TFs provided[48]. Inference whether the target genes we identified was druggable was done by intersecting target genes with the genes provided in the drug gene interaction database (http://dgidb.genome.wustl.edu/) by using "Expert curated" option in the source trust level category of the interactions. All information showing the overlap of target genes with gene lists from literature can be found in Supplementary Table 3.

Pathway analysis. Functional characterization of enhancer/gene assignments was conducted using the ClueGO plugin for cytoscape[49]. Subgroup-specific enhancer gene targets or SE-regulated TFs were queried against a compendium of gene sets from GO (Biological Process), KEGG, and REACTOME to identify processes/pathways that were significantly enriched in tested gene lists from our data set. Analyses were performed using the GO Term Fusion option in ClueGO and only processes/pathways with a P-value < 0.05 (right-sided hypergeometric test) following P-value correction (Bonferroni step down) were visualized. Manual trimming of ClueGO output was performed to remove processes/pathways affiliated with only a single gene set.

Functional comparison of Group 3 and Group 4 enhancers. To identify subgroup specific enhancers and their associated functional pathways, we performed a differential enhancer analysis[50] on Group 3 and Group 4 enhancers. We first took the union of the top 1,000 enhancer in Group 3 and Group 4 as defined by total H3K27ac signal (area under the curve). We next ranked all enhancer regions by the $\log_2$ fold change in H3K27ac (Extended Data Fig. 3b). Differential enhancer target genes as previously defined were depicted under associated enhancers. Visual inspection revealed a number of TGF-β pathway components associated with Group 3 specific enhancers. We visualized this by identifying all enhancer regulated TGF-β pathway components (obtained from KEGG, REACTOME, and GO Biological Process databases) and depicting their specific regulation by Group 3, Group 4, or Group 3-4 differential enhancers (Extended Data Fig. 3c).

Comparing enhancer acetylation at TGF-β pathway components in ACVR2A amplified vs non-amplified Group 3 tumours. We identified a focal amplification of the TGF-β pathway receptor gene *ACVR2A* in the Group 3 medulloblastoma sample MB-4M23. Whole genome sequencing $\log_2$ read depth ratio is plotted in Extended Data Fig. 3d. We hypothesized that in MB-4M23, amplification of ACVR2A leads to increased TGF-β pathway activity, including the increased H3K27ac at enhancers regulating TGF-β pathway components. We identified all Group 3 enhancers regulating TGF-β pathway components and compared the median enhancer normalized H3K27ac signal in MB-4M23 vs all other Group 3 medulloblastomas. Extended Data Fig. 3e shows all enhancers ranked by their $\log_2$ fold change in H3K27ac for MB-4M23 vs other Group 3 samples. The standard error of the mean was calculated for the fold change and is displayed as error bars in Extended Data Fig. 3e.

Nucleosome free region (NFR) identification. H3K27ac data for the samples within the same subgroup was combined. Nucleosome free regions per subgroup were identified by feeding these combine data sets to HOMER software (http://homer.salk.edu/homer/ngs/index.html) using "findPeaks" function with the option "-nfr".

Enrichment of TFs at subgroup-specific enhancers. TF binding sites obtained from TRANSFAC[51] and detected at NFRs using FIMO[52] were overlapped with NFRs located within each class of differentially regulated enhancers. For each TF, contingency tables showing the number of NFRs overlapping and non-overlapping with the respective TF were constructed. Significance of enrichment of TFs in NFRs of differentially regulated enhancers was determined using Chi-squared test. Resulting P-values were corrected for multiple testing (FDR < 0.01). TF enrichments were calculated as the ratio between observed counts over expected counts. To represent TF enrichments as a heat map (Extended Data Fig. 6b), for each class of enhancers, 4–5 TFs showing the highest enrichments were selected.

Linking subgroup-specific enhancers with TFs. For each of differentially regulated enhancers in the classes of WNT, SHH, Group 3 and Group 4, NFRs belonging to each subgroup were overlapped with the respective subgroup-specific enhancers targeting at least one gene. Overlapping NFRs were intersected with TF binding sites having top 20th percentile enrichment scores in the respective subgroup-specific enhancers and differentially expressed in the same subgroup. For each TF, NFRs having the top 10th percentile number of binding sites were identified as sites occupied by the respective TF. Then, resulting NFRs were linked back to enhancers they are located, which enabled the linking of TFs having binding sites in the respective enhancers with the target genes of the enhancers. TF regulatory networks for each subgroup (Extended Data Fig. 7), where TFs represented as "sources" and enhancer target genes represented as "targets" were constructed using visualization platform Gephi (http://gephi.github.io/). To connect LMX1A, LHX2 and EOMES with their targets (Extended Data Fig. 9b), same strategy was applied by restricting the initial set of TFs to only those three.

4C-seq data analysis. Aligned 4C data was further processed, filtered and visualized using Bioconductor package "Basic4Cseq"[53].

Mapping typical enhancers and super-enhancers using H3K27ac enhancer definitions. H3K27ac super-enhancers (SEs) and typical enhancers (TEs) in individual medulloblastoma samples were mapped using the ROSE2 software package described[13,23] and available at https://github.com/BradnerLab/pipeline. A 12.5 kb stitching window was used to connect proximal clusters of H3K27ac peaks into contiguous enhancer regions. These mappings identified on average ~600 SEs per sample.

Clustering medulloblastoma samples by SE patterns. Relationships between SE landscapes between samples were determined as in ref. 11. First, we defined the union of all regions considered to be an SE in any individual primary sample and in three Group 3 cell lines. Next H3K27ac signal was calculated at each region and median normalized for each sample. Samples were hierarchically clustered based on similarity of patterns of median normalized H3K27ac enhancer signal as determined using pairwise Pearson correlations.

Mapping SEs and typical enhancers across medulloblastoma subgroups (subgroup enhancer mapping). In order to map and quantify enhancer regions for each medulloblastoma subgroup, we first mapped all enhancers in each individual sample within the group. Across a group, we used the union of all enhancer regions within group samples as the landscape of enhancers. Within this landscape, enhancers were ranked by average H3K27ac signal (area under curve) and classified as SEs or TEs as previously described. This produced SE and TE meta enhancer landscapes for WNT, SHH, Group 3, and Group 4 medulloblastoma with between 558 and 1,110 SEs called per group (Fig. 3a). Locations for all SEs and TEs in each subgroup are provided in Supplementary Table 4.

Quantifying enhancer signal variance across samples at meta enhancer regions. To compare the dynamic range of SEs and TEs defined in each medulloblastoma subgroup, we quantified H3K27ac signal variance across samples. For SE and TE enhancer constituents (individual peaks of H3K27ac enrichment within broader enhancer domains) defined in each group, H3K27ac signal variance across samples as a fraction of the mean sample was calculated. The average H3K27ac signal variance across all SEs or TEs within a group is plotted in Extended Data Fig. 4f.

Quantifying average H3K27ac signal across samples at subgroup SEs and typical enhancers. We sought to examine trends in H3K27ac signal across medulloblastoma samples at regions defined as SEs or TEs in each group. First we mapped H3K27ac across all samples to enhancer constituents defined in each group. For each medulloblastoma sample, the average median normalized H3K27ac signal was plotted for SE and TE constituents respectively. For SEs and TEs defined in each group, the average sample H3K27ac signal is plotted with the mean and standard deviation shown as lines. This visualization enables a rapid assessment of H3K27ac variance within a group and of trends in H3K27ac signal for SEs and TEs defined in each group (Extended Data Fig. 4h). For instance, enhancer constituents in Group 3 SEs tend to have high signal in Group 4.

Quantifying group ChIP-seq signal at subgroup SEs and typical enhancers within and between groups. SEs have been shown to have higher H3K27ac and BRD4 signal density at constituents when compared to typical enhancers[13,23]. To determine if these trends were observed at medulloblastoma enhancers, we calculated H3K27ac and BRD4 ChIP-seq signal density across all samples at all regions defined as enhancers across groups (meta enhancers). In order to properly compare ChIP-seq signal density between SEs and TEs, for each enhancer constituent, we first determined if it was considered part of an SE in one or more groups, and if so, these groups defined the "active group context" for that particular enhancer constituent. Groups in which the enhancer constituent showed no evidence of enhancer activity (SE or TE) were considered the inactive group context. For enhancer constituents considered only part of a TE in one or more groups, groups in which the enhancer constituent was classified as a TE were considered the active

group context and all other groups were considered the inactive group context. For each SE or TE constituent, average H3K27ac or BRD4 signal density was calculated at all samples in the active group context or in the inactive group context. The distributions of H3K27ac or BRD4 signal for enhancer constituents classified by SE or TE status were plotted and the statistical significance of the difference in the mean was tested in the active or inactive group context using a Welch's two-tailed *t*-test (Extended Data Fig. 4g).

Identifying group specific and conserved SEs. We developed a method to identify SEs that were conserved across all medulloblastoma subgroups as well as SEs that showed highly group specific patterns of enhancer activity. We first took as the SE landscape all regions identified as SEs in the meta subgroup enhancer mapping. To account for sample-to-sample variability in H3K27ac ChIP-seq dynamic range, H3K27ac signal at enhancers in each medulloblastoma sample was rank transformed (Fig. 3b). As each medulloblastoma sample contained on average ~600 SEs, enhancer regions with an average rank of 600 or better in each subgroup were considered conserved. To identify enhancers with group specific patterns of activity, we calculated a "group rank *Z*-score" that compared average signal in one group to average signal in other groups. Here we considered whether enhancers might show group specific patterns for WNT, SHH, Group 3, Group 4, and as well for groupings of WNT/SHH, and Group 3/4. For each enhancer, this group rank *Z*-score was calculated for each group vs other combination. Enhancers with a group rank *Z*-score >1 (that is, those whose mean rank within a group was >1 standard deviation above the mean rank of all other samples) were considered group specific. To account for variability in enhancer ranks, only enhancers with a statistically significant difference in ranks (within group vs all other samples, Welch's two-tailed *t* test, *P*-value <0.01) were considered. Supplementary Table 4 contains all SE regions identified in medulloblastoma subgroups and their corresponding max group rank Z-score, *P*-value, and classification.

Mapping H3K27ac enrichment at the MYC gene desert. To provide a developmental context for medulloblastoma MYC SEs, we mapped H3K27ac enrichment at the *MYC* locus. H3K27ac data was obtained from the Epigenome ROADMAP as in Fig. 1e. The 500 kb region flanking the MYC SE No. 2 was divided into 5 kb bins and each bin was tested for overlap with a H3K27ac peak in each ROADMAP sample. ROADMAP samples were hierarchically clustered by similarity of H3K27ac peak pattern at the *MYC* locus (Extended Data Fig. 5m). Overlap with MYC SE No. 2 was found in 4/77 ROADMAP samples.

Calculating regulatory IN and OUT degree for all SE associated TFs. Medulloblastoma core regulatory circuitry analysis was performed using the COLTRON (https://pypi.python.org/pypi/coltron) that calculated inward and outward degree regulation of SE-regulated TFs. To quantify the interaction network of TF regulation, we calculated the IN and OUT degree of all SE associated TFs. The 92 SE associated TFs were those defined as either proximal to an SE (within 50 kb) or the target of a differential SE enhancer element. For any given TF (TF_i), the IN degree was defined as the number of TFs with an enriched binding motif at the proximal SE of TF_i (Fig. 5a). The OUT degree was defined as the number of TF associated SEs containing an enriched binding site for TF_i. Within any given SE, enriched TF binding sites were determined at putative nucleosome free regions (valleys) flanked by high levels of H3K27ac. Valleys were calculated using an algorithm adapted from ref. 35. In these regions, we searched for enriched TF binding sites using the FIMO[52] algorithm with TF position weight matrices defined in the TRANSFAC database[51]. An FDR cutoff of 0.01 was used to identify enriched TF binding sites. Using this approach, we calculated IN and OUT degree for all SE associated TFs within the meta H3K27ac landscape (average of all samples) of each medulloblastoma subgroup. This approach resulted in an IN and OUT degree estimate for each SE associated TF in each medulloblastoma subgroup (Extended Data Fig. 8a–d).

Identifying TF binding motifs for LMX1A, LHX2, and HLX. We sought to identify TF binding motifs for each TF in each subgroup. For each TF, we defined binding regions as the ±1,000 bp flanking the enriched region summit (as defined using MACS 1.4.2 with a *P*-value cutoff of 1×10^{-9}). We took the union of all regions bound in a given subgroup (for example, HLX bound regions in Group 3 samples) that overlapped an enhancer in that subgroup and did not overlap any ENCODE[8] blacklist regions. We next took the top 10,000 discreet regions as ranked by average TF ChIP-seq signal and used the ±100 bp region flanking the region centre as the input for *de novo* motif finding. *De novo* motif finding was performed using the MEME[54] suite using a 1st order background model and searching for motifs between 6 and 30 bp in length. The top motif for each TF is displayed as a position weight matrix in Extended Data Fig. 8i–l.

Visualizing TF regulatory networks. To visualize SE associated TF interactions in each subgroup, we ranked all SE associated TF by TOTAL degree (IN + OUT).

We visualized the top 50% of SE associated TFs in each subgroup as a network diagram with each node representing a SE associated TF, and with nodes coloured and ordered by increasing TOTAL degree (Extended Data Fig. 8e–h). Interactions between SE associated TF nodes were defined as a TF motif identified in the SE of a TF and are depicted as edges. For Group 3 and Group 4, edges validated by the presence of a TF ChIP-seq peak are coloured.

Clustering TFs by regulatory degree to identify and infer subgroup specific regulatory circuitry. To identify SE associated TFs with similar regulatory patterns likely to influence subgroup identity, we first normalized the TOTAL degree for each SE associated TF in each subgroup from 0 to 1. We then calculated the normalized TOTAL degree for each SE associated TF in each subgroup. We filtered out all TFs with a max TOTAL degree across medulloblastomas of less than 0.7. We next clustered all remaining TFs by their TOTAL degree pattern. Hierarchical clustering was performed using a Euclidian distance metric and the resulting clustergram tree was cut at a distance of 0.5 to produce 26 individual clusters. Of these 26 clusters, 12 showed a median TOTAL degree >0.7 in 1, 2, or all 4 subgroups. Clusters with >0.7 TOTAL degree in 3 subgroups were omitted for simplicity. TOTAL degree patterns of TFs in these 12 clusters are shown in Extended Data Fig. 9a. This filtering produced a list of 102 SE associated TFs, of which 71 had predicted interactions with one another. These 71 TFs fall into either conserved, subgroup specific, or dual subgroup clusters and together they comprise the inferred core regulatory circuitry of medulloblastoma subgroups. As in Extended Data Fig. 8e–h, regulatory interactions between these core regulatory circuitry TFs are depicted in Extended Data Fig. 9a with Group 3 and Group 4 validated edges coloured. A subset of this larger network containing the TFs HLX, LHX2, EOMES, and LMX1A is depicted in Fig. 5c with ChIP-seq validated edges drawn as solid lines and motif prediction edges drawn in dotted lines.

Quantifying protein–protein interactions of co-regulating SEs. We used the STRING interaction database[55] to quantify protein–protein interaction frequencies of SE associated TFs with similar regulatory patterns. TF pairs were considered co-regulatory if they shared 50% of the same OUT degree edges. Interaction frequencies for co-regulatory pairs were compared to those from 10,000 randomly assigned pairs of TFs expressed in that subgroup (Extended Data Fig. 8o).

Integration of TF ChIP-seq occupancy into enhancer landscape and TF regulatory network. To determine the fraction of motif predicted edges with evidence of actual TF ChIP-seq binding, we first took all predicted edges for HLX, LHX2, and LMX1A interacting SE associated with other TFs in Group 3 and Group 4. We validated all edges that contained a ChIP-seq peak within the same enhancer as the predicted TF motif. The fraction of validated edges for each TF in each subgroup is shown in Extended Data Fig. 8g, h, m.

Quantification of TF binding at Group 3 and 4 enhancers. To determine how Group 3 and Group 4 TF ChIP-seq levels varied at Group 3 and Group 4 specific enhancers, we quantified TF ChIP-seq signal at Group 3 and Group 4 enhancers. We first took the union of the top 1,000 enhancer regions as defined by H3K27ac signal in Group 3 and Group 4 (as in Extended Data Fig. 3b). We identified as Group 3 and Group 4 specific enhancer regions with a >1.0 $\log_2$ absolute fold change between Group 3 and Group 4. We identified as conserved enhancer regions with a <0.05 $\log_2$ absolute fold change between Group 3 and Group 4. We next identified all enhancer regions bound by LHX2 and HLX in Group 3 (G3 HLX and LHX2) or by LHX2 and LMX1A in Group 4 (G4 LMX1A and LHX2). TF ChIP-seq occupancy in units of average area under the curve (AUC) were quantified at TF bound regions overlapping Group 3 specific, Group 4 specific, and conserved enhancer region (Extended Data Fig. 8n). Statistical differences in the means of the distributions of TF ChIP-seq signal at different enhancer populations was determined using a Welch's two tailed *t*-test (Extended Data Fig. 8n).

Quantifying Group 4 TF gene expression changes in Dreher RL. To identify genes transcriptionally regulated by Lmx1a in the developing cerebellum, we isolated cerebellar uRL from WT and $Lmx1a^{-/-}$ embryos by laser capture microdissection. uRL was isolated from WT ($n = 3$) and $Lmx1a^{-/-}$ ($n = 3$) embryos (~3,000 cells per embryo) at e13.5, just before abnormal RL regression in $Lmx1a^{-/-}$ embryos. RNA was extracted using PicoPure RNA Isolation Kit (Arcturus) and hybridized to Illumina MouseRef8 v2 Expression BeadChips at the Johns Hopkins Array Core Facility. Next we identified all human TF genes with unambiguous mouse homologues that were detectably expressed in the WT mouse cerebellum (cut off of 100 arbitrary units). We subsequently quantified median normalized expression in WT or $Lmx1a^{-/-}$ samples and calculated the $\log_2$ fold-change for all TFs. We ranked the expression fold-change of all SE-associated TFs in medulloblastoma and plotted their $\log_2$ fold change in $Lmx1a^{-/-}$ vs WT (Fig. 6d). SE-associated TFs present in the Group 4 TF network (Extended Data Fig. 8h) were coloured in green.

35. Ramsey, S. A. *et al.* Genome-wide histone acetylation data improve prediction of mammalian transcription factor binding sites. *Bioinformatics* **26,** 2071–2075 (2010).

36. Ni, T. T. *et al.* Conditional control of gene function by an invertible gene trap in zebrafish. *Proc. Natl Acad. Sci. USA* **109,** 15389–15394 (2012).

37. Robinson, G. *et al.* Novel mutations target distinct subgroups of medulloblastoma. *Nature* **488,** 43–48 (2012).

38. Haldipur, P. *et al.* Expression of Sonic hedgehog during cell proliferation in the human cerebellum. *Stem Cells Dev.* **21,** 1059–1068 (2012).

39. 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).

40. Borodina, T. & Adjaye, J. & Sultan, M. A strand-specific library preparation protocol for RNA sequencing. *Methods Enzymol.* **500,** 79–98 (2011).

41. Sultan, M. *et al.* A simple strand-specific RNA-Seq library preparation protocol combining the Illumina TruSeq RNA and the dUTP methods. *Biochem. Biophys. Res. Commun.* **422,** 643–646 (2012).

42. van de Werken, H. J. *et al.* 4C technology: protocols and data analysis. *Methods Enzymol.* **513,** 89–112 (2012).

43. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10,** R25 (2009).

44. John, S. *et al.* Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nature Genet.* **43,** 264–268 (2011).

45. Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L. & Almouzni, G. Epigenomics: Roadmap for regulation. *Nature* **518,** 314–316 (2015).

46. Skipper, M. *et al.* Presenting the epigenome roadmap. *Nature* **518,** 313 (2015).

47. Waszak, S. M. *et al.* Population variation and genetic control of modular chromatin architecture in humans. *Cell* **162,** 1039–1050 (2015).

48. Vaquerizas, J. M., Kummerfeld, S. K., Teichmann, S. A. & Luscombe, N. M. A census of human transcription factors: function, expression and evolution. *Nature Rev. Genet.* **10,** 252–263 (2009).

49. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25,** 1091–1093 (2009).

50. Hothorn, T., Bretz, F. & Westfall, P. Simultaneous inference in general parametric models. *Biom. J.* **50,** 346–363 (2008).

51. Matys, V. *et al.* TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. *Nucleic Acids Res.* **34,** D108–D110 (2006).

52. Grant, C. E., Bailey, T. L. & Noble, W. S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* **27,** 1017–1018 (2011).

53. Walter, C., Schuetzmann, D., Rosenbauer, F. & Dugas, M. Basic4Cseq: an R/Bioconductor package for analyzing 4C-seq data. *Bioinformatics* **30,** 3268–3269 (2014).

54. Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37,** W202–W208 (2009).

55. Franceschini, A. *et al.* STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res.* **41,** D808–D815 (2013).
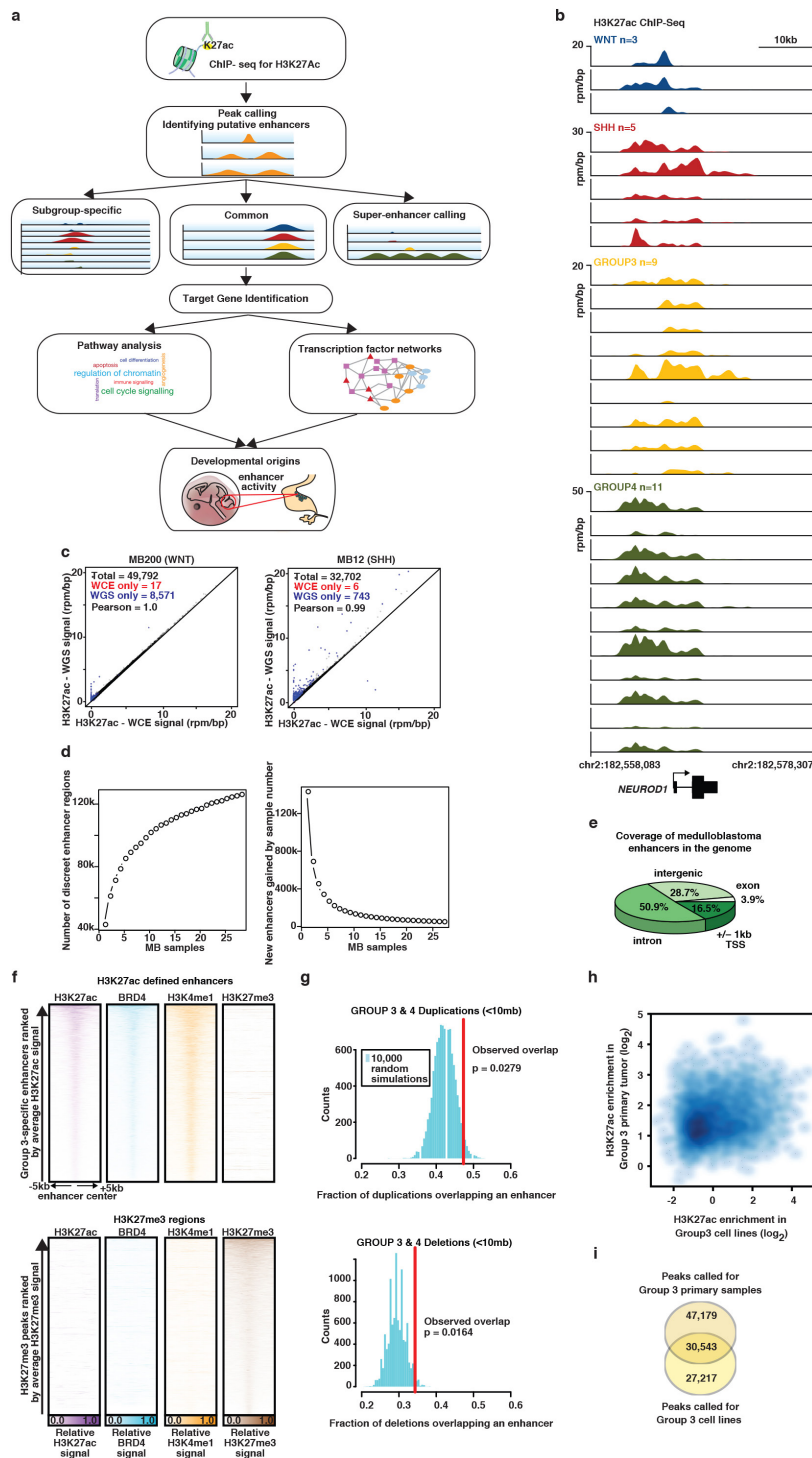
Extended Data Figure 1 | Enhancer landscape of primary medulloblastoma. This figure accompanies Fig. 1. **a**, Experimental workflow for studying enhancers and super-enhancers in primary medulloblastomas. **b**, H3K27ac ChIP-seq data showing a highly active enhancer at the *NEUROD1* locus across all 28 primary medulloblastoma samples from our series. **c**, Scatter plots showing Pearson correlation of H3K27ac peaks called using either sample-matched WGS or whole-cell extract (WCE) sequences as background for two samples from our series. **d**, Saturation analysis showing the number of discreet enhancer regions identified as a function of increasing sample number (top), or the fraction of newly gained discreet enhancer regions as a function of increasing sample number (bottom). Error bars represent 95% confidence intervals obtained from 1,000 permutations of sample order. **e**, Pie chart showing the genomic distribution of enhancer elements in medulloblastoma.
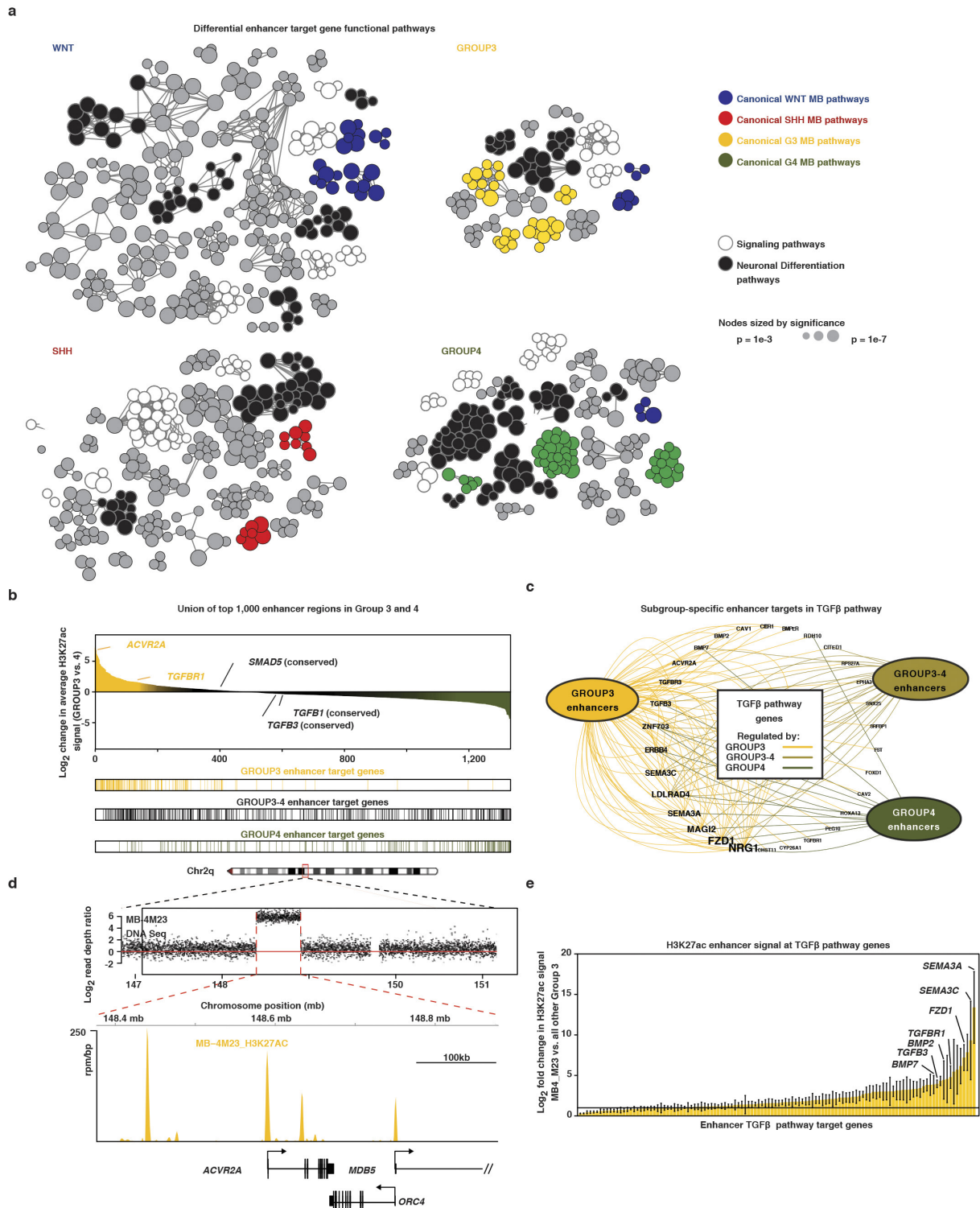
f, Heat maps of ChIP-seq data showing the scaled read densities for H3K27ac, BRD4, H3K4me1, and H3K27me3 in regions located $\pm 5\,\mathrm{kb}$ from Group 3-specific H3K27ac (top panel) and H3K27me3 peak midpoints (bottom panel). **g**, Histograms showing the fractional overlap of enhancers with focal amplifications (top) or focal deletions (bottom) in Group 3 and Group 4 medulloblastoma samples. The blue distributions represent expected fractional overlap generated from 10,000 random simulations. The red line depicts the actual observed fractional overlap with empirical *P*-value noted. **h**, Scatter plot correlating average H3K27ac enrichment in Group 3 cell lines with average H3K27ac enrichment in Group 3 primary medulloblastomas. Enrichments are calculated for peaks called in primary Group 3 samples. **i**, Venn diagram showing the overlap between H3K27ac peaks called for primary Group 3 medulloblastomas and Group 3 medulloblastoma cell lines.
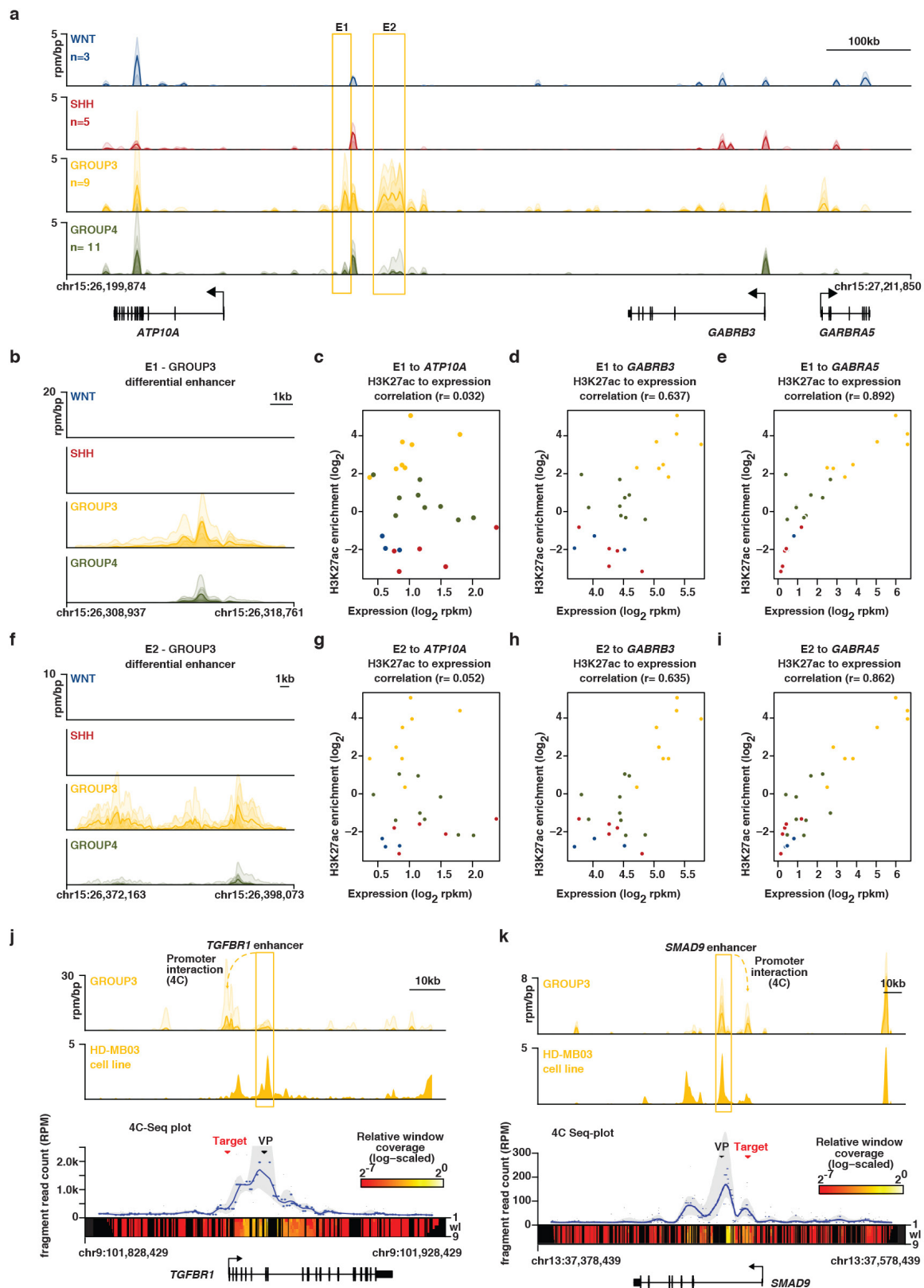
Extended Data Figure 2 | Enhancer/gene assignments in medulloblastoma. This figure accompanies Fig. 2. **a**, Meta H3K27ac ChIP-seq tracks of the Group 3-specific enhancers (E1 and E2) in the TAD containing *ATP10A*, *GABRB3*, and *GABRA5*. **b**, Zoom in meta H3K27ac ChIP-seq tracks of enhancer E1 from **a**. **c–e**, Scatter plots correlating sample-matched gene expression (log$_2$ RPKM, *x*-axis) of *ATP10A* (**c**), *GABRB3* (**d**), and *GABRA5* (**e**) with H3K27ac enrichment (log$_2$; *y*-axis)

for the Group 3-specific enhancer shown in **b**. **f**, Zoom in meta H3K27ac ChIP-seq tracks of enhancers E2 from **a**. **g–i**, Scatter plots correlating sample-matched gene expression (log$_2$ RPKM, *x*-axis) of *ATP10A* (**g**), *GABRB3* (**h**), and *GABRA5* (**i**) with H3K27ac enrichment (log$_2$; *y*-axis) for the Group 3-specific enhancer shown in **f**. **j**, **k**, 4C-seq validation of *TGFBR1* (**j**) and *SMAD9* (**k**) enhancer/promoter interactions in a Group 3 cell line (HD-MB03).

a

Differential enhancer target gene functional pathways

WNT

GROUP3

SHH

GROUP4

Canonical WNT MB pathways
Canonical SHH MB pathways
Canonical G3 MB pathways
Canonical G4 MB pathways

Signaling pathways

Neuronal Differentiation pathways

Nodes sized by significance
$p = 1e\text{-}3$ ●●● $p = 1e\text{-}7$

b

Union of top 1,000 enhancer regions in Group 3 and 4

ACVR2A
TGFBR1
SMAD5 (conserved)
TGFB1 (conserved)
TGFB3 (conserved)

GROUP3 enhancer target genes

GROUP3-4 enhancer target genes

GROUP4 enhancer target genes

c

Subgroup-specific enhancer targets in TGFβ pathway

GROUP3 enhancers
GROUP3-4 enhancers
GROUP4 enhancers

TGFβ pathway genes
Regulated by:
GROUP3
GROUP3-4
GROUP4

d

Chr2q

MB-4M23 DNA Seq

Log₂ read depth ratio

Chromosome position (mb)

MB–4M23_H3K27AC

100kb

ACVR2A   MDB5   ORC4

e

H3K27ac enhancer signal at TGFβ pathway genes

SEMA3A
SEMA3C
FZD1
TGFBR1
BMP2
TGFB3
BMP7

Enhancer TGFβ pathway target genes

Extended Data Figure 3 | Enhancer-driven TGF-β activity in Group 3 medulloblastoma. This figure accompanies Fig. 2. **a**, Functional annotation of target genes assigned to subgroup-specific enhancers based on their significant overlap with gene sets annotated in Gene Ontology (GO Biological Process) and pathway databases (KEGG, Reactome). **b**, Waterfall plot discriminating the top 1,000 Group 3 and Group 4 subgroup-specific enhancers as defined by total H3K27ac signal. The distribution of assigned targets in Group 3, Group 4, and shared Group 3-4 targets are shown below the waterfall. **c**, Convergence of Group 3-specific enhancers on TGF-β pathway genes. Subgroup-specific enhancers are summarized as nodes according to their respective medulloblastoma enhancer class—Group 3, Group 4, and shared Group 3/Group 4—with

edges representing individual enhancer/TGF-β pathway gene assignments. **d**, Amplification of the TGF-β type II receptor, *ACVR2A*, in a Group 3 medulloblastoma from the ChIP-seq cohort (MB-4M23). Log₂ read depth data (tumour versus matched germline) derived from WGS data for this case is shown (upper panel). Highly active H3K27ac enhancer peaks overlapping the amplified *ACVR2A* locus are shown for the same case (lower panel). **e**, Bar plot showing the difference in H3K27ac enhancer signal between MB-4M23 (*ACVR2A*-amplified Group 3 sample) and all other Group 3 samples. Bar plot shows H3K27ac log₂ fold change at all enhancers regulating TGF-β component genes. Enhancers are ranked by increasing change in H3K27ac. Error bars represent standard error of the mean fold change.
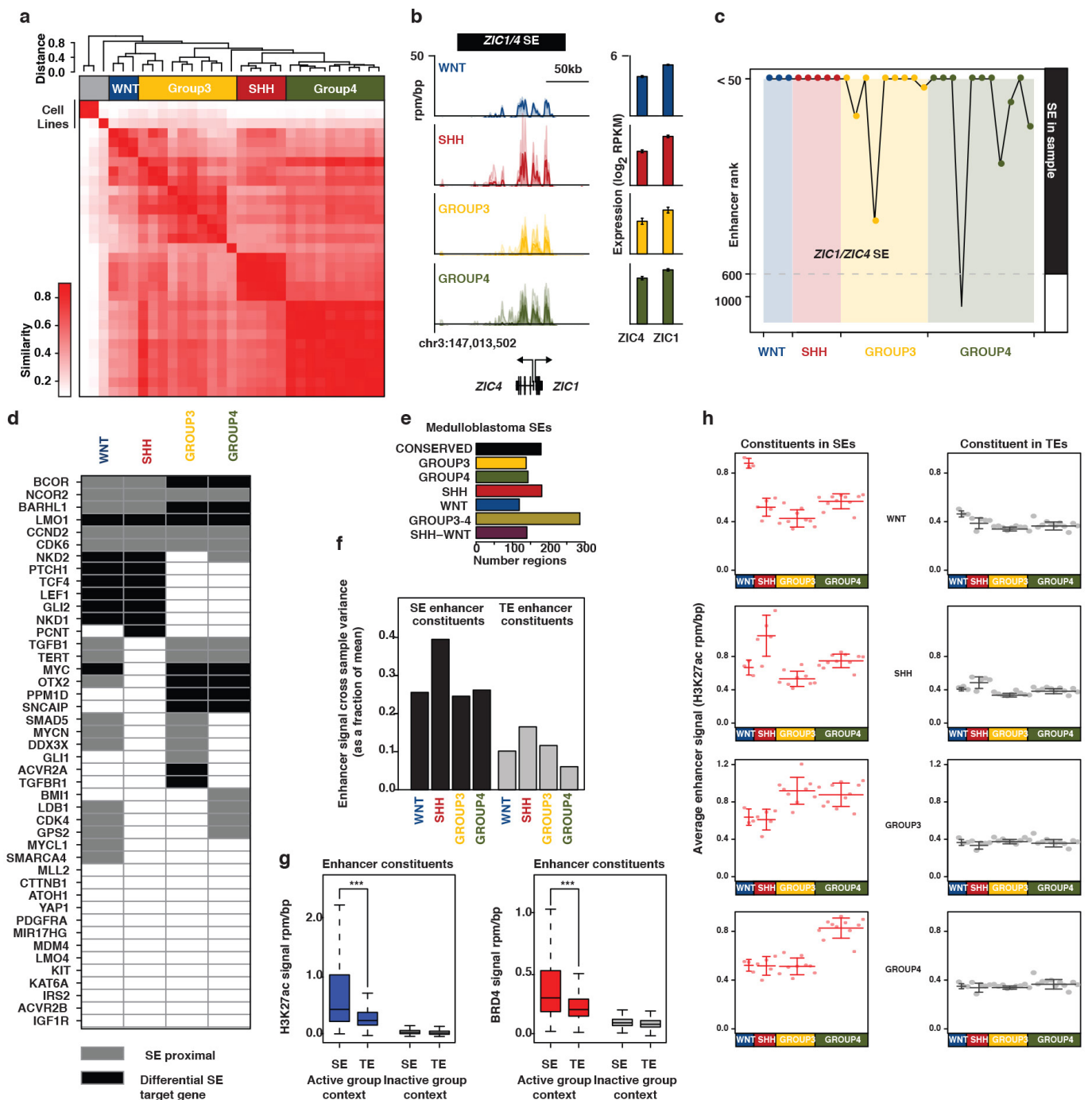
Extended Data Figure 4 | Features of medulloblastoma super-enhancers.
This figure accompanies Fig. 3. **a**, Unsupervised hierarchical clustering of
primary medulloblastomas and cell lines using H3K27ac signal calculated
at all SEs identified in each individual sample. **b**, Meta tracks of H3K27ac
ChIP-seq signal for the *ZIC1/ZIC4* SE locus. Expression (mean RPKM) for
both *ZIC4* (left) and *ZIC1* (right) is displayed as bar graphs to the right of
each H3K27ac track with error bars representing s.d. of the mean (*n* = 140
samples). **c**, Line plot showing the enhancer rank for the *ZIC1/ZIC4* SE
locus across all samples according to subgroup. **d**, Heat map showing the
SE association of known medulloblastoma driver genes and chromatin
modifiers. Genes with called differential SEs are shaded black, whereas
genes with proximal SEs (within 100 kb of TSS) are shaded grey, according
to their respective subgroup. **e**, Bar plot showing the number of SE regions
assigned to individual enhancer classes in medulloblastoma. **f**, Bar plot

of enhancer signal cross sample variance (*y*-axis) displayed as a fraction
of the mean for SE enhancer constituents (left, black) or TE enhancer
constituents (right, grey) identified in each medulloblastoma subgroup.
g, Box plots of H3K27ac (left, blue) or BRD4 (right, red) enhancer signal
at SEs or typical enhancers (TE) in their active group-specific context or
in their inactive group context (for example, for SEs or TEs present in
Group 3, active group context includes all Group 3 samples and inactive
group context includes all other samples). Differences in the means of the
distributions is quantified by a Welch's two-tailed *t*-test (***$P < 1 \times 10^{-9}$).
h, Dot plots of average H3K27ac enhancer signal in the constituents of SEs
(left) or TEs (right) for enhancer constituents identified in WNT, SHH,
Group 3, or Group 4 samples, respectively. Error bars represent standard
deviation of the mean across all samples in a subgroup.
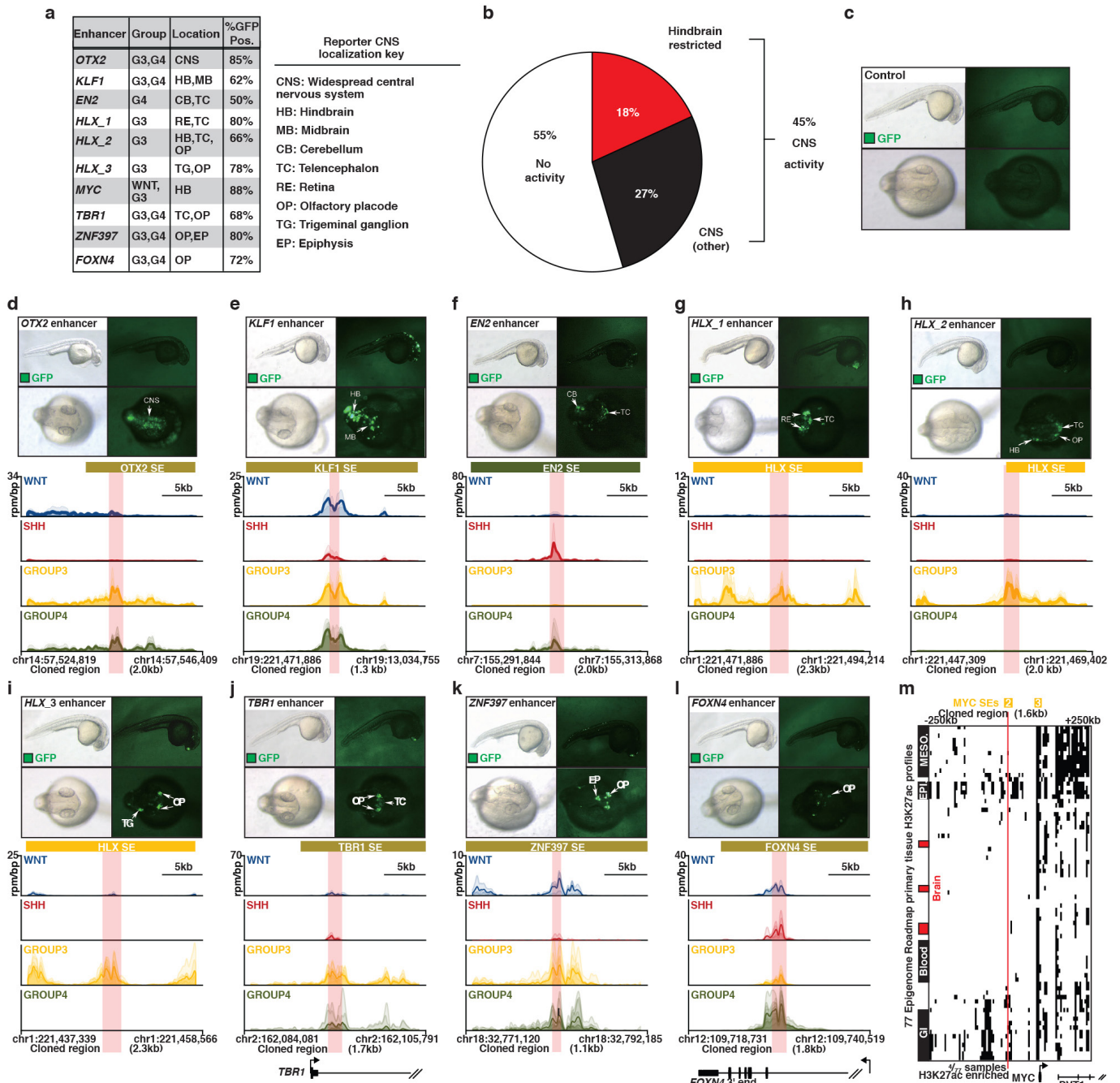
Extended Data Figure 5 | *In vivo* validation of Group 3 and Group 4 medulloblastoma super-enhancers. This figure accompanies Fig. 4. **a**, Summary of zebrafish reporter assays. **b**, Pie chart showing the fraction of all tested medulloblastoma enhancer regions that demonstrate any central nervous system localized reporter activity. **c–l**, Representative bright-field and fluorescence images of embryos (1 dpf) injected with individual enhancer-containing Tol2 vectors. Lateral views (60×) show GFP reporter expression in the whole body and dorsal views show GFP expression in the central nervous system (120×). White arrows indicate the locations of GFP signal. CNS, central nervous system; HB, hindbrain; MB, midbrain; CB, cerebellum; TC, telencephalon; RE, retina; OP, olfactory placode; TG, trigeminal ganglion. For each tested enhancer,

meta tracks of H3K27ac ChIP-seq signal across medulloblastoma subgroups for the cloned regulatory element are shown. **m**, Heat map showing H3K27ac enrichment at the ±250 kb region flanking the medulloblastoma *MYC* SE described in Fig. 4 (SE #2; panels **f**, **h–j**) across 77 Epigenome Roadmap tissues. Each row represents a single tissue. Each column represents a region of the *MYC* gene desert locus. Black shaded regions indicate the presence of H3K27ac enrichment. The samples are ordered by similarity of H3K27ac enrichment pattern. Notable clusters of mesoderm (MESO.), epithelial (EPI.), blood, brain, or GI lineage derived samples are noted. The cloned enhancer reporter region described in Fig. 4 (panels **f**, **h–j**) is depicted as a vertical line and shows overlap with only 4/77 H3K27ac Epigenome Roadmap samples.

a

b
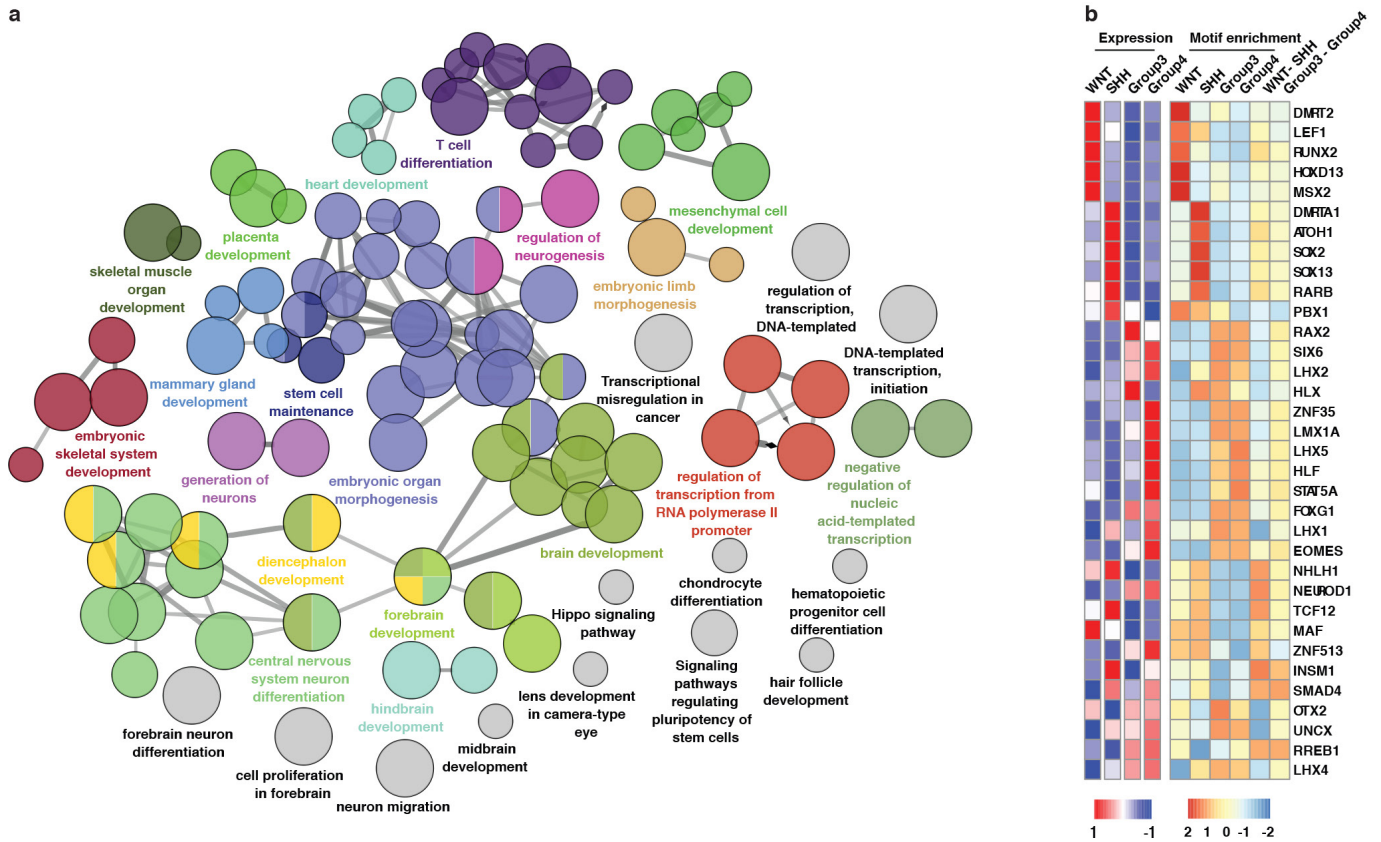
Extended Data Figure 6 | Pathways regulated by super-enhancer associated transcription factors in medulloblastoma. This figure accompanies Fig. 5. **a**, Functional pathways regulated by SE-associated TFs in medulloblastoma. **b**, Heat map of select subgroup-specific TFs showing their expression (left columns) and enhancer motif enrichment (right columns). Enhancer motif enrichment was calculated at differential enhancer elements in the respective enhancer classes.
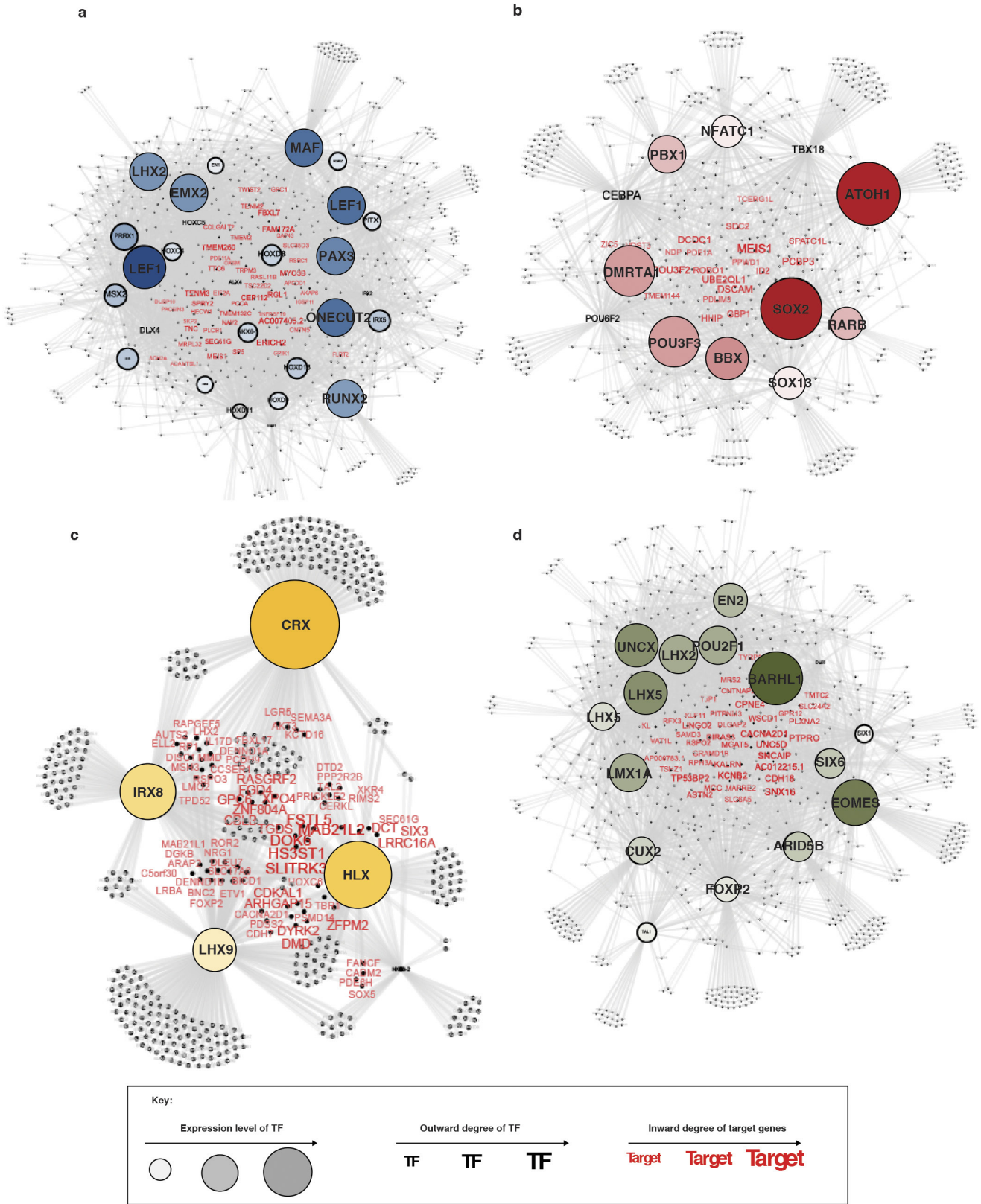
Extended Data Figure 7 | Medulloblastoma subgroup-specific transcription factors and their associated target genes. This figure accompanies Fig. 5. **a–d**, Network of subgroup-specific TFs and their predicted target genes for WNT (**a**), SHH (**b**), Group 3 (**c**) and Group 4 (**d**) subgroups. Nodes represent subgroup-specific TFs. In each subgroup, node size is scaled and shaded according to the expression level of the TF and node font is scaled and shaded according to the number of inferred target genes (that is, OUT degree). TF target genes are shown in red font scaled according to the number of TFs predicted to target that gene (that is, IN degree).
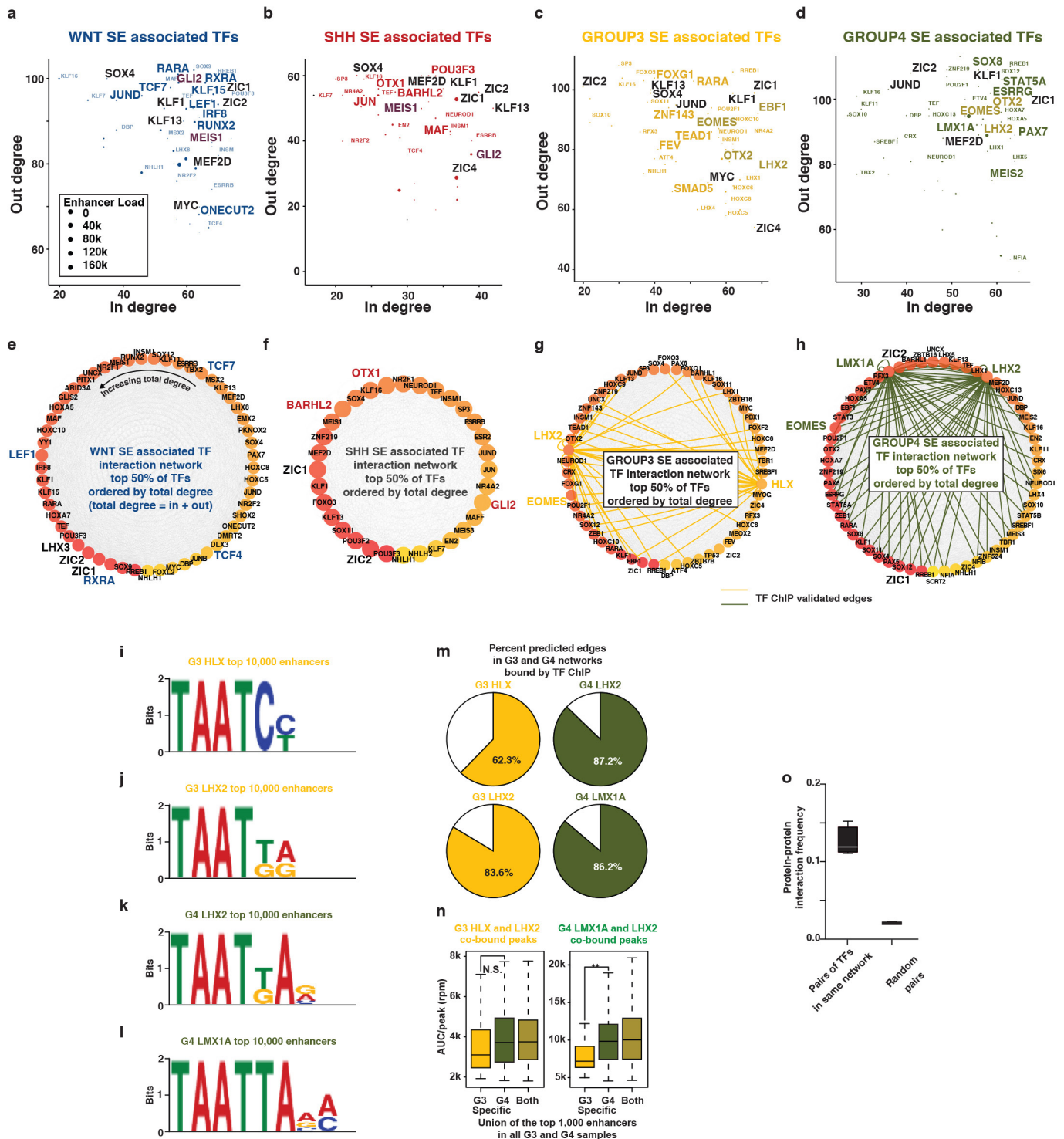
Extended Data Figure 8 | Super-enhancers define medulloblastoma regulatory circuitry. This figure accompanies Fig. 5. **a–d**, Scatter plots of IN (*x*-axis) and OUT (*y*-axis) regulatory degree for SE-associated TFs in each medulloblastoma subgroup. **e–h**, TF interaction networks for each medulloblastoma subgroup. Nodes represent the top 50% of SE-associated TFs in each subgroup as ranked by total degree (counter clockwise). Each node is coloured by total degree and predicted binding interactions with other TF SEs are shown as edges. For Group 3 and Group 4 networks, edges validated by TF ChIP-seq binding are coloured. **i–l**, Position weight matrices showing the top statistically enriched motif identified for each transcription factor at the top 10,000 bound enhancers in each subgroup.

m, Pie charts showing the fraction of predicted edges in each Group 3 and Group 4 TF networks that are validated by the presence of the respective TF ChIP-seq binding at the enhancer. **n**, Medulloblastoma subgroup distribution of shared, co-bound peaks for master regulatory TFs analysed by ChIP-seq. TF binding is quantified as area under curve per peak (AUC/peak) in units of rpm. Differences in the means of the distributions is quantified by a Welch's two-tailed *t*-test (N.S. $P > 0.1$, **$P < 1 \times 10^{-6}$). **o**, Box plot of protein–protein interaction frequency (*y*-axis) calculated from STRING database for pairs of SE-associated TFs showing patterns of subgroup-specific SE co-regulation (left) or randomized pairs (right).

a

b
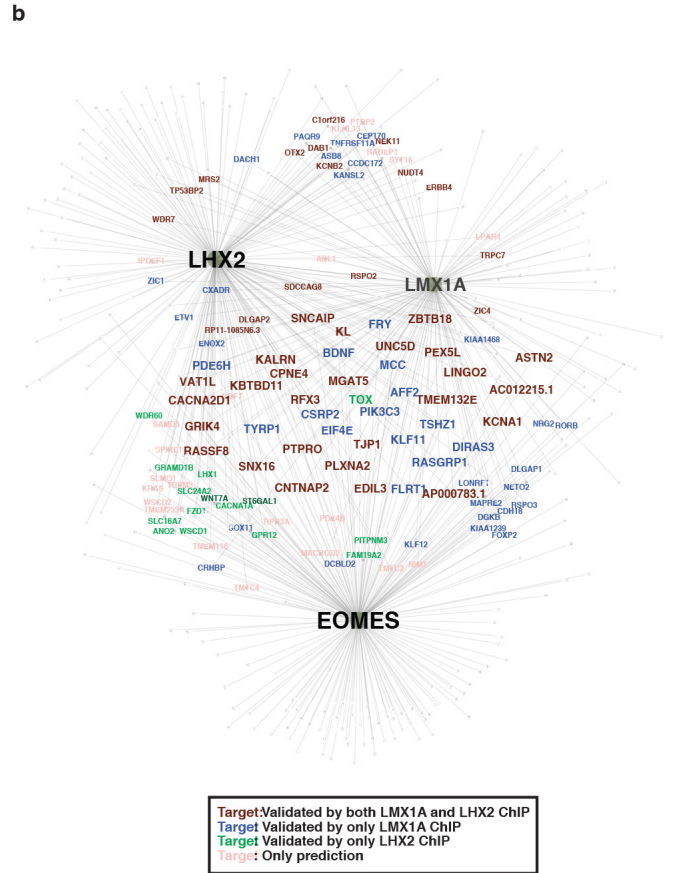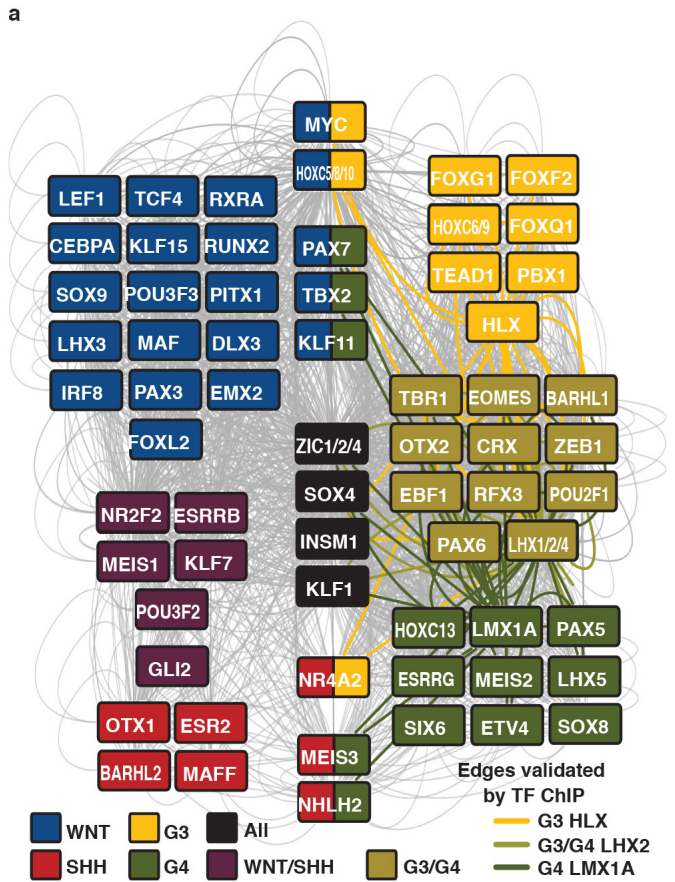
Extended Data Figure 9 | LMX1A, EOMES, and LHX2 are master transcriptional regulators of Group 4 medulloblastoma. This figure accompanies Fig. 5. **a**, Subgroup-specific regulatory circuitry. Nodes are TFs associated with an SE in a subgroup-specific context. Edges indicate co-regulating TFs as defined by enrichment of TF binding motifs in respective regulatory regions. Edges validated by TF ChIP-seq are coloured according to their respective subgroup association. **b**, Network involving

LHX2, LMX1A, and EOMES TFs and target genes inferred based on the presence of the respective TF motifs in Group 4-specific enhancers. Target genes are coloured according to their validation status based on LMX1A and LHX2 ChIP-seq, with genes arranged in the centre of the network inferred to be targeted by all three master TFs. For visualization purposes, these common targets are displayed with a larger font size compared to the genes in the surrounding network.