LETTER

doi:10.1038/nature11329

Medulloblastoma exome sequencing uncovers subtype-specific somatic mutations

Trevor J. Pugh[1,2,3], Shyamal Dilhan Weeraratne[3,4], Tenley C. Archer[3,4], Daniel A. Pomeranz Krummel[5], Daniel Auclair[1], James Bochicchio[1], Mauricio O. Carneiro[1], Scott L. Carter[1], Kristian Cibulskis[1], Rachel L. Erlich[1], Heidi Greulich[1,2,3], Michael S. Lawrence[1], Niall J. Lennon[1], Aaron McKenna[1], James Meldrim[1], Alex H. Ramos[1,2,3], Michael G. Ross[1], Carsten Russ[1], Erica Shefler[1], Andrey Sivachenko[1], Brian Sogoloff[1], Petar Stojanov[1], Pablo Tamayo[1], Jill P. Mesirov[1], Vladimir Amani[3,4], Natalia Teider[3,4], Soma Sengupta[3,4], Jessica Pierre Francois[3,4], Paul A. Northcott[6], Michael D. Taylor[6], Furong Yu[7], Gerald R. Crabtree[7,8], Amanda G. Kautzman[7], Stacey B. Gabriel[1], Gad Getz[1], Natalie Jäger[9], David T. W. Jones[9], Peter Lichter[9], Stefan M. Pfister[9], Thomas M. Roberts[2,3], Matthew Meyerson[1,2,3,10], Scott L. Pomeroy[1,3,4] & Yoon-Jae Cho[1,3,4,7]

Medulloblastomas are the most common malignant brain tumours in children[1]. Identifying and understanding the genetic events that drive these tumours is critical for the development of more effective diagnostic, prognostic and therapeutic strategies. Recently, our group and others described distinct molecular subtypes of medulloblastoma on the basis of transcriptional and copy number profiles[2–5]. Here we use whole-exome hybrid capture and deep sequencing to identify somatic mutations across the coding regions of 92 primary medulloblastoma/normal pairs. Overall, medulloblastomas have low mutation rates consistent with other paediatric tumours, with a median of 0.35 non-silent mutations per megabase. We identified twelve genes mutated at statistically significant frequencies, including previously known mutated genes in medulloblastoma such as *CTNNB1*, *PTCH1*, *MLL2*, *SMARCA4* and *TP53*. Recurrent somatic mutations were newly identified in an RNA helicase gene, *DDX3X*, often concurrent with *CTNNB1* mutations, and in the nuclear co-repressor (N-CoR) complex genes *GPS2*, *BCOR* and *LDB1*. We show that mutant DDX3X potentiates transactivation of a TCF promoter and enhances cell viability in combination with mutant, but not wild-type, β-catenin. Together, our study reveals the alteration of WNT, hedgehog, histone methyltransferase and now N-CoR pathways across medulloblastomas and within specific subtypes of this disease, and nominates the RNA helicase DDX3X as a component of pathogenic β-catenin signalling in medulloblastoma.

Medulloblastomas are aggressive tumours of primitive neuroectodermal origin. More than one third of patients diagnosed with medulloblastoma succumb to their disease within 5 years[6] and surviving patients often have significant long-term adverse effects from current therapies. Identifying the underlying genetic events responsible for medulloblastomas can help guide the development of more effective therapies and refine the selection of currently available chemotherapy and radiotherapy. Recent efforts profiling transcriptional and DNA copy number changes in medulloblastoma have provided insights into the biological processes involved in these tumours and have underscored the molecular heterogeneity of this disease[2–4]. Based on these data, four broad subgroups have been established, known according to a consensus nomenclature as SHH, WNT, Group 3 and Group 4 (ref. 5).

The first genome-scale sequencing of protein coding regions in medulloblastoma was reported recently[7]. Altered genes encoding for histone modification proteins were identified in 20% of cases, most notably *MLL2* and *MLL3* (ref. 7). This initial survey was limited by a small discovery sample size (22 patients), lack of subtype-specific analysis, and use of Sanger sequencing technology insensitive to variants present at low allelic fraction. Here we survey coding somatic mutations at deeper coverage in a larger cohort of 92 medulloblastoma/normal pairs and assess these mutations in the context of specific molecular subtypes (Supplementary Table 1).

In total, 1,908 mutations were detected within 1,671 out of 18,863 genes sequenced to a median of 106× coverage (Supplementary Table 2). Confirmation of 20 candidate mutations in selected genes (*CTNNB1*, *DDX3X*, *SMARCA4*, *TP53* and *CTDNEP1*) was performed by amplification of 48 exons using a microfluidic PCR device (Fluidigm) followed by single-molecule real-time sequencing (SMRT, Pacific Biosciences) (Supplementary Information). Sequence data was unavailable for one *DDX3X* mutation because of poor PCR amplification from the sample. All remaining 19 mutations were confirmed by this orthogonal method (median 73 redundant sub-reads, range 3–287, Supplementary Fig. 1).

A median of 16 somatic mutations (12 non-silent, 4 silent) per tumour was identified, corresponding to a mutation rate of 0.35 non-silent mutations per megabase of callable sequence, less than most adult solid tumours and consistent with results from ref. 7. Six of the twelve most frequently mutated tumours were from the oldest patients (16–31 years at diagnosis), consistent with increased mutation frequency in adult versus childhood medulloblastomas ($P = 7.7 \times 10^{-5}$, Wilcoxon rank-sum test, Supplementary Fig. 2).

To identify genes mutated at statistically significant frequencies across our cohort, we used the MutSig algorithm[8] which takes into account gene size, sample-specific mutation rate, non-silent to silent mutation ratios, clustering within genes, and base conservation across species. In our cohort of 92 samples, we identified 12 significantly mutated genes ($q < 0.1$, Table 1 and Supplementary Table 3). Strikingly, these genes were not mutated in c5 (Group 3) and c4 (Group 4) tumours with extensive somatic copy number alteration (Fig. 1), indicating that these subtypes are driven primarily by structural variation, rather than base mutation. Not unexpectedly, *CTNNB1* (β-catenin) and *PTCH1* were the two most significantly mutated genes (see Table 1 and Fig. 1). Point mutations of *CTNNB1* in combination with loss of chromosome 6 were found in all WNT subgroup tumours and were concurrent with several other recurrently mutated genes, namely *CSNK2B*, *DDX3X*, *TP53* and *SMARCA4*. Mutations involving *PTCH1* occurred exclusively in SHH subgroup tumours and mutations of genes associated with the hedgehog pathway were also restricted to this subgroup ($P < 0.0001$, Fisher's exact test). All but one of the tumours with *PTCH1* mutations had somatic loss of 9q, resulting in hemizygosity for the mutant allele. The remaining tumour had apparent copy neutral loss-of-heterozygosity of 9q22. Other

[1]Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. [2]Center for Cancer Genome Discovery, Departments of Medical Oncology and of Biological Chemistry and Molecular Pharmacology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. [3]Harvard Medical School, Boston, Massachusetts 02115, USA. [4]Department of Neurology, Children's Hospital Boston, Boston, Massachusetts 02115, USA. [5]Brandeis University, Waltham, Massachusetts 02453, USA. [6]The Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [7]Departments of Neurology and Neurosurgery, Stanford University School of Medicine, Stanford, California 94305, USA. [8]Howard Hughes Medical Institute at Stanford University, Stanford, California 94305, USA. [9]German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. [10]Department of Pathology, Brigham and Women's Hospital, Boston, Massachusetts 02115, USA.

Table 1 | Genes mutated at a statistically significant frequency in 92 medulloblastomas.

Gene	Description	Mutations	Patients	Unique sites	Silent	Missense	Indel or null	Double null	*q*
CTNNB1	β-catenin	6	6	4	0	6	0	0	$<1.8\times10^{-11}$
PTCH1	Patched homologue 1 (*Drosophila*)	7	7	7	0	0	7	0	4.0×10^{-9}
MLL2	Myeloid/lymphoid or mixed-lineage leukaemia 2	10	8	10	0	2	4	4	4.0×10^{-9}
DDX3X	DEAD box polypeptide 3, X-linked	7	7	7	0	7	0	0	2.3×10^{-8}
GPS2	G protein pathway suppressor 2	3	3	3	0	1	2	0	1.2×10^{-4}
TP53	Tumour protein p53	3	3	3	0	3	0	0	0.039
KDM6A	UTX, lysine (K)-specific demethylase 6A	3	3	3	0	2	1	0	0.042
BCOR	BCL6 co-repressor	3	3	3	0	0	3	0	0.046
SMARCA4	ATP-dependent helicase	4	4	3	0	4	0	0	0.046
LDB1	LIM domain binding 1	2	2	2	0	1	1	0	0.047
CTDNEP1	CTD nuclear envelope phosphatase 1	2	2	2	0	0	2	0	0.047
CSNK2B	Casein kinase 2, β polypeptide	2	2	2	0	2	0	0	0.071

Null, nonsense, frameshift or splice-site mutations; double null, null mutations co-occurring in a single tumour; *q*, *q*-value, false discovery rate (Benjamini–Hochberg procedure). See Supplementary Table 3 for further statistics and subtype analysis.

somatic mutations of hedgehog pathway members include a splice site mutation in *SUFU*, an in-frame deletion in *WNT6*, and missense mutations in *GLI2*, *SMO*, *PRKACA*, *WNT2* and *WNT2B*.

Two patients with SHH subgroup tumours had germline variants in *PTCH1*, one with somatic loss of 9q resulting in hemizygosity for a loss-of-function germline allele (MD-085, c.3030delC, p.Asn1011Thrfs*38), and the other with a substitution previously reported in patients with holoprosencephaly (MD-286, p.T1052M, ref. 9). Two additional cases (MD-097 and MD-335) had loss-of-function variants in *SUFU* (1 frameshift deletion and 1 nonsense) that began as heterozygotes in the germline and became hemizygous in the tumour, due to somatic loss of chromosome 10 in one case and copy neutral loss-of-heterozygosity in the other.

MLL2 was also subject to recurrent inactivating mutations, consistent with findings from ref. 7 and providing further evidence for dysregulated histone modification in medulloblastoma. Indeed, six

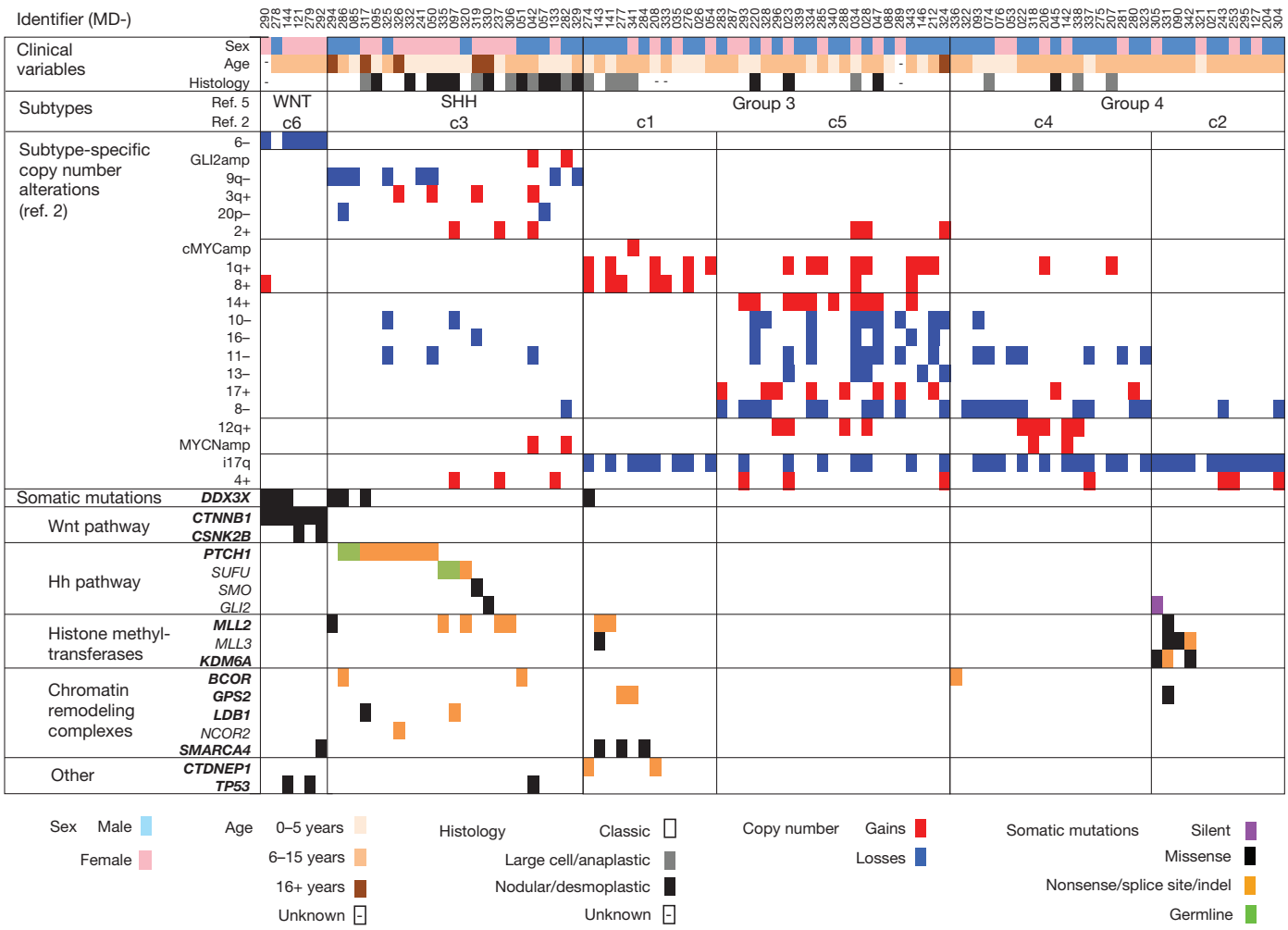

Figure 1 | Demographic characteristics, molecular subtypes and selected copy number alterations and somatic mutations across 92 medulloblastoma cases. Data tracks describing 92 medulloblastoma cases. Identifier, unique name used to denote each case. Identifiers also link samples to those analysed in ref. 2. Sex, males in blue, females in pink. Age, years of age at diagnosis binned as infants, children or adults. Histology, pathology review of primary tissue specimen. Subtypes, based on copy number profiles derived from sequence or microarray data. Consensus subtypes from refs 2 and 5 as published. Copy number alterations, selected copy number alterations used to assign tumours to subtypes. Blue, losses; red, gains. Somatic mutations, gene names (HUGO symbols) grouped by functional category. MutSig gene names are in bold. Black, missense mutations; orange, nonsense/splice site/indel mutations; purple, silent mutations; green, germline variants.

of the twelve most significantly mutated genes are involved in histone modification and/or related chromatin remodelling complexes (*MLL2*, *GPS2*, *KDM6A*, *BCOR*, *SMARCA4* and *LDB1*; see Table 1). As a gene set, histone methyltransferases (HMTs) were enriched for somatic mutation with 21 tumours having apparent, predominantly loss-of-function, HMT mutations ($q = 5.8 \times 10^{-9}$; Fig. 2 and Supplementary Table 4).

Subtype-specific MutSig analysis identified additional significant mutations of histone-modifying genes, *MLL3* and *HDAC2*, in Group 4 tumours along with *KDM6A* mutations ($q = 0.039$ and 0.066, see Supplementary Table 3). Mutations in KDM6A, interestingly, occurred exclusively in tumours with an i17q as the sole autosomal alteration ($P = 0.0023$, Fisher's exact test) with the one female case with *KDM6A* mutation also having loss of a chromosome X. Notably, the two 'i17q only' tumours without *KDM6A* mutations had other histone-modifying enzymes mutated, namely *THUMPD3*, *ZMYM3* and *MLL3*, perhaps suggesting a distinct biology for tumours with this karyotype.

Mutations in several genes encoding components of the nuclear co-repressor (N-CoR) complex were observed at a statistically significant frequency: *BCOR* in 3 tumours, *GPS2* in 3 tumours, and *LDB1* in 2 tumours. *BCOR* mutations have recently been reported at high frequency in retinoblastoma[10] and in 'copy-neutral' acute myelogenous leukaemia[11]. *BCOR* is located on the X-chromosome and two hemizygous frameshift mutants were found in tumours from males (allele fractions 0.90 and 0.92). A third nonsense mutation was also found in a male but at low allelic fraction (0.12), indicating a subclonal event. Two out of three *BCOR* mutations occurred in SHH subgroup tumours. *LDB1* missense and nonsense mutations were found in two additional SHH tumours, both appearing hemizygous due to loss of 10q and complete chromosome 10 loss, respectively (allele fractions 0.81 and 0.78). Both *BCOR* and *LDB1* promote assembly of the repressive N-CoR complex[12] and harbour apparent loss of function mutations. *GPS2*, which encodes a critical subunit of the N-CoR complex, a repressor of JNK/MAPK signalling through partnership with histone deacetylases[12], was mutated in two Group 3 tumours. The *GPS2* mutations cluster within amino acids 53–90, the domain critical for heterodimerization with *NCOR2* (also known as SMRT) and interacting with a TBL1 amino-terminal domain tetramer to assemble the N-CoR repression complex[12]. Finally, an additional nonsense mutation in *NCOR2* was identified in a single SHH subgroup tumour, underscoring the central role of N-CoR dysregulation in medulloblastoma development and particularly within the SHH subgroup.

Several genes encoding subunits of the SWI/SNF-like chromatin-remodelling complex were also mutated in our cohort, including significant recurrent mutations of *SMARCA4(Brg/BAF190)*, which encodes a DNA helicase with ATPase activity[13] and has been reported to be mutated in lung, ovarian, and pancreatic cancers[14] as well as medulloblastoma[15]. In our cohort, *SMARCA4(Brg/BAF190)* mutations clustered in helicase domains and occurred in three Group 3 tumours (significant within the c1 subtype, $q = 0.019$), and one WNT tumour. In addition, mutations were found in the alternative ATPase subunit *SMARCD2(Brm)* (missense at a highly conserved residue) and two other members of the SWI/SNF complex, *ARID1B(BAF250b)* (2 base pairs (bp) frameshift deletion) and *SMARCC2(BAF170)* (splice site). These were all apparent loss-of-function mutations and occurred in SHH tumours. Thus, it seems that disruption of this complex is frequent across medulloblastomas.

New and hemizygous mutations were found in *CTDNEP1* (previously known as *DULLARD*), a phosphatase with roles in *Xenopus* neural development through regulation of BMP receptors[16], and as a direct regulator of LIPIN, an integral component of the mTOR complex[17]. *CTDNEP1* mutations were found in two Group 3 tumours (significant within the subtype, $q = 0.0087$), a 2-bp frameshift deletion and a substitution disruptive of a splice site. Both tumours have i17q chromosomes, resulting in loss of the wild-type allele at 17p13.

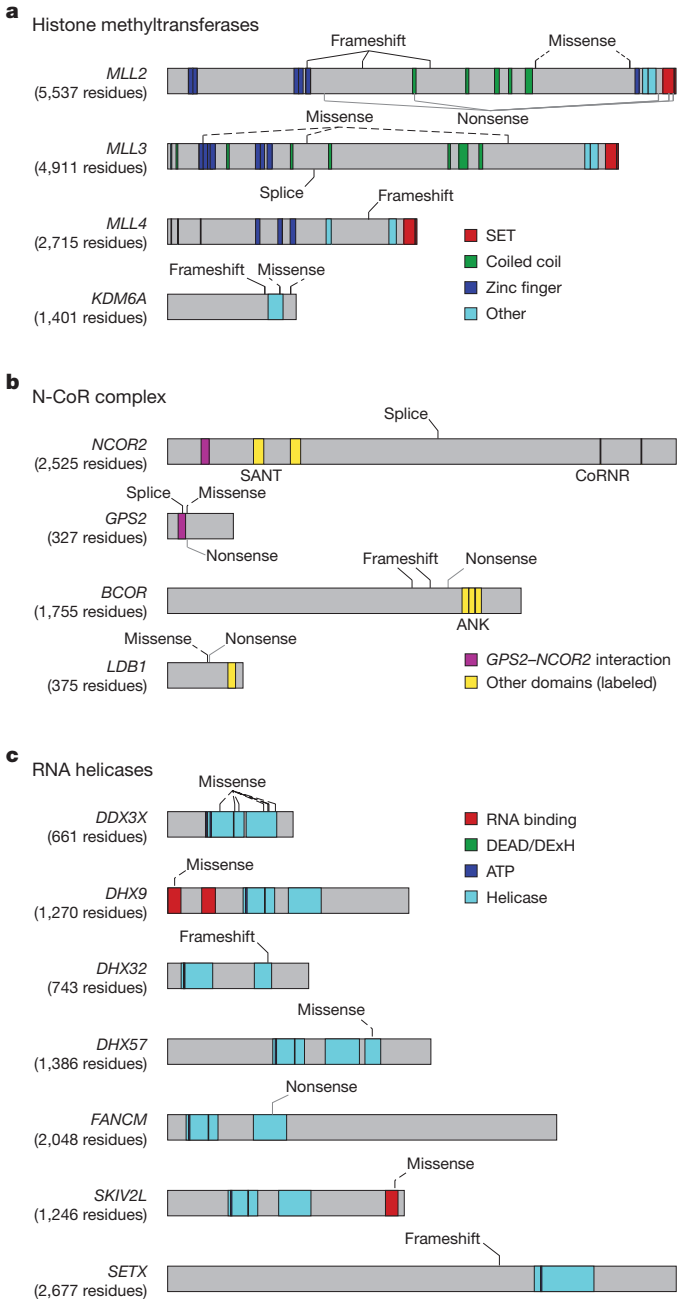

Figure 2 | Location of mutations in histone methyltransferases, RNA helicases and N-CoR complex-associated genes. Location of somatic mutations on linear protein domain models of genes from sets frequently mutated in medulloblastoma. All domain annotations are from UniProt and InterPro annotations. Diagrams were constructed using Domain Graph (DOG)[28], version 2.0. **a**, Histone methyltransferase domains: red, SET; green, coiled-coil; blue, zinc-finger; cyan, other. **b**, N-CoR complex-associated domains: purple, anti-parallel coiled-coil domains required for GPS2–NCOR2 (SMRT) interaction[12]; yellow, other interaction domains as labelled (SANT domains binds DNA, CoRNR domains binds nuclear receptors, ANK repeats mediate a diversity of protein–protein interactions, and LIM-binding domains bind a common protein structural motif). **c**, RNA helicase domains: cyan, helicase and helicase-associated (InterPro); red, RNA-binding and RNA polymerase sigma factor (InterPro); blue, ATP binding site; green, DEAD or DExH box motif. See Supplementary Table 1 for UniProt protein model identifiers.

Mutations in *DDX3X*, an ATP-dependent RNA helicase with functions in transcription, splicing, RNA transport and translation[18], were found in seven tumours, including half of the WNT pathway tumours ($P = 0.005$, Fisher's exact test) and several SHH subgroup tumours.

DDX3X mutations have recently been reported at low frequency in five other tumour types (Catalogue of Somatic Mutations in Cancer, COSMIC[15]) but the significance of these mutations for DDX3X function remains unclear. To understand the consequence of observed point mutations on the physical structure of DDX3X, we mapped the mutations onto the previously reported crystal structure of DDX3X[19] and its orthologue DDX4 (also known as VASA, ref. 20) (Fig. 3a; Supplementary Fig. 3 and Supplementary Table 5). The mutations seem to cluster in two structural domains, a helicase ATP-binding domain (residues 211–403) and a helicase carboxy-terminal domain (residues 414–575). The location of these mutations indicates that they may alter DDX3X–RNA interaction (Fig. 3a and Supplementary Table 5).

As half of the β-catenin mutated tumours contained concurrent *DDX3X* mutations, we investigated whether DDX3X could enhance the of ability β-catenin to transactivate a TCF4-luciferase reporter (TOPflash) and if DDX3X/β-catenin co-expression had a measureable effect on cell viability/proliferation. In combination with wild-type β-catenin, neither wild-type nor mutant DDX3X alone significantly transactivated the TOPflash reporter. However, in combination with mutant β-catenin (S33Y substitution), the majority of DDX3X point mutants in our cohort potentiated reporter activity ($P < 0.05$, Fig. 3b). This potentiation was also apparent in cell viability assays in both HeLa (data not shown) and D425 medulloblastoma cell lines ($P < 0.05$, Fig. 3c).

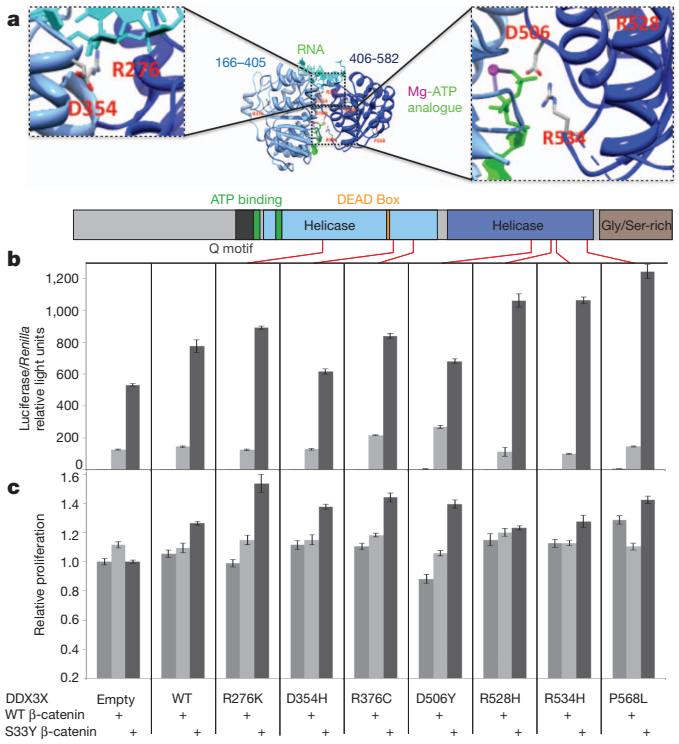

Figure 3 | Functional consequence of DDX3X point mutations. **a**, Three-dimensional model of the two recA-like domains of human DDX3X in complex with single-stranded RNA and a Mg-ATP analogue. Displayed are the residues mutated in the amino-terminal recA-like domain (R276K, D354H, R376C) and C-terminal recA-like domain (D506Y, R528H, R534H, P568L). Colouring: light blue, DDX3X residues 166–405; dark blue, DDX3X residues 406–582; cyan, single-stranded RNA; magenta and green, Mg-ATP analogue. Molecular graphics images were produced using the University of San Francisco Chimera package[29] (http://www.cgl.ucsf.edu/chimera). **b**, Mutant DDX3X potentiates mutant β-catenin transactivation of TOPflash promoter. Represented is relative luciferase activity in 293T cells co-transfected with TOPflash reporter, FOPflash control, and either wild-type or mutant DDX3X in combination with wild-type or mutant β-catenin. One-dimensional model of DDX3X displayed above bar graphs to illustrates the position of the mutations. WT, wild type. **c**, Cell viability assays of medulloblastoma D425 cells stably transduced with either wild-type or mutant DDX3X lentivirus in combination with either wild-type or mutant β-catenin lentivirus. For **b** and **c**, error bars depict the standard deviation of the mean from five replicate experiments performed for each condition. Student's *t*-tests were performed to evaluate significance of differences in TOPflash intensity or cell proliferation value distributions as follows: increases with DDX3X alone versus empty vector, increases with wild-type β-catenin versus DDX3X alone, increases with mutant β-catenin versus DDX3X alone, and increases with mutant β-catenin versus wild-type β-catenin.

Given the apparent importance of *DDX3X* mutations in medulloblastoma, we searched the genes listed in the RNA Helicase Database (http://www.rnahelicase.org/) for low frequency mutations in medulloblastoma. We found five tumours with mutations within RNA helicase or RNA binding domains of *DHX9*, *DHX32*, *DHX57*, *FANCM* and *SKIV2L* (Fig. 2 and Supplementary Table 6). The missense mutations were at conserved residues and predicted to be deleterious by the software packages SIFT, AlignGVGD and PolyPhen2. In addition, a frameshift insertion in *SETX* occurs upstream of, and probably disrupts, its RNA helicase domain. Overall, 15% of medulloblastomas seem to have some disruption of RNA helicase activity.

In summary, we report a next-generation sequencing analysis of medulloblastoma, the most common malignant brain tumour in children. Our results reveal mutations in several known pathways such as histone methylation (*MLL2* and others), sonic hedgehog (*PTCH1*, *SUFU* and others) and Wnt (*CTNNB1* and others), and also previously unrecorded mutations in genes including *DDX3X*, *BCOR*, *LDB1* and *GPS2*. Our preliminary functional studies implicate DDX3X as a candidate component of pathogenic WNT/β-catenin signalling. In a broader sense, *DDX3X* mutations have recently been reported in chronic lymphocytic leukaemia[21] and head and neck cancers[22], both of which have subsets of tumours with dysregulated WNT signalling. Studies investigating whether mutant DDX3X functions together with β-catenin in these contexts should provide additional insights into this multifaceted molecule and open potential avenues for novel therapies. Finally, the delineation of nuclear receptor co-repressor complex molecules as altered in medulloblastoma provides new insight into the pathogenesis of this deadly childhood disease.

METHODS SUMMARY

Informed consent was provided by families of medulloblastoma patients treated at Children's Hospital Boston, The Hospital for Sick Children, Toronto, Canada, and institutions contributing to the Children's Oncology Group/Cooperative Human Tissue Network, under approval and oversight by their respective Internal Review Boards. All tumours were obtained at the initial surgical resection and recurrent tumours were excluded from our analysis. Haematoxylin- and eosin-stained slides of tumour samples were reviewed by a pathologist to confirm the diagnosis of medulloblastoma, determine histological subtype when possible, and assess tumour purity. DNA was isolated from tumour specimens and matched peripheral blood as previously described[2]. Exome sequencing of DNA from 92 tumour/normal pairs was performed using in-solution hybrid-capture of 193,094 exons from 18,863 micro RNA (miRNA)- and protein-coding genes, followed by sequencing of 76 bp paired-end reads using Illumina sequencing-by-synthesis technology[23]. Reads were aligned to human genome build GRCh37[24] using a Burrows–Wheeler aligner (BWA)[25]. The ~33-megabase target region was sequenced to 106× mean coverage in each sample (range 73–234). Gene expression data and copy number profiles (derived from SNP microarrays or sequence data) were used to assign each tumour to a subgroup using published criteria[2]. Our cohort consisted of 6 WNT (c6), 23 SHH (c3), 33 Group 3 (12 c1, 21 c5), and 30 Group 4 (12 c2, 18 c4) tumours (see Supplementary Table 1 for case annotations). Mutations were detected using muTect, annotated using Oncotator[26], and manually reviewed using the Integrated Genomics Viewer (IGV)[27]. For validation, PCR on Access Array microfluidic chips (Fluidigm) was followed by single-molecule real-time sequencing (Pacific Biosciences) as per manufacturer's instructions. Sub-reads were extracted and assigned to samples using manufacturer's and custom software, and aligned to the hg19 (GRCh37) build of the human reference genome sequence using BWA-SW[25]. Candidate mutations were confirmed by manual review using IGV[27] (Supplementary Fig. 1). See Supplementary Information and http://www.broadinstitute.org/cancer/cga/ for complete descriptions of materials and methods.

Received 3 February; accepted 15 June 2012.
Published online 22 July 2012.

1. Central Brain Tumor Registry of the United States. *Statistical Report: Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2004–2007* http://www.cbtrus.org/2011-NPCR-SEER/WEB-0407-Report-3-3-2011.pdf (CBTRUS, 2011).
2. 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).
3. Kool, M. *et al.* Integrated genomics identifies five medulloblastoma subtypes with distinct genetic profiles, pathway signatures and clinicopathological features. *PLoS ONE* **3,** e3088 (2008).
4. Remke, M. *et al.* Adult medulloblastoma comprises three major molecular variants. *J. Clin. Oncol.* **29,** 2717–2723 (2011).
5. Taylor, M. D. e. t. a. l. Molecular subgroups of medulloblastoma: the current consensus. *Acta Neuropathol.* **123,** 465–472 (2012).
6. Smoll, N. R. Relative survival of childhood and adult medulloblastomas and primitive neuroectodermal tumors (PNETs). *Cancer* **118,** 1313–1322 (2012).
7. Parsons, D. W. *et al.* The genetic landscape of the childhood cancer medulloblastoma. *Science* **331,** 435–439 (2011).
8. Getz, G. *et al.* Comment on ''The consensus coding sequences of human breast and colorectal cancers.''. *Science* **317,** 1500 (2007).
9. Ming, J. E. *et al.* Mutations in PATCHED-1, the receptor for SONIC HEDGEHOG, are associated with holoprosencephaly. *Hum. Genet.* **110,** 297–301 (2002).
10. Zhang, J. *et al.* A novel retinoblastoma therapy from genomic and epigenetic analyses. *Nature* **481,** 329–334 (2012).
11. Grossmann, V. *et al.* Whole-exome sequencing identifies somatic mutations of *BCOR* in acute myeloid leukemia with normal karyotype. *Blood* **118,** 6153–6163 (2011).
12. Oberoi, J. *et al.* Structural basis for the assembly of the SMRT/NCoR core transcriptional repression machinery. *Nature Struct. Mol. Biol.* **18,** 177–184 (2011).
13. Baek, S. H. *et al.* Regulated subset of G1 growth-control genes in response to derepression by the Wnt pathway. *Proc. Natl Acad. Sci. USA* **100,** 3245–3250 (2003).
14. Wilson, B. G. & Roberts, C. W. M. SWI/SNF nucleosome remodellers and cancer. *Nature Rev. Cancer* **11,** 481–492 (2011).
15. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4,** 177–183 (2004).
16. Satow, R., Kurisaki, A., Chan, T.-c., Hamazaki, T. S. & Asashima, M. Dullard promotes degradation and dephosphorylation of BMP receptors and is required for neural induction. *Dev. Cell* **11,** 763–774 (2006).
17. Peterson, T. R. *et al.* mTOR complex 1 regulates lipin 1 localization to control the SREBP pathway. *Cell* **146,** 408–420 (2011).
18. Garbelli, A., Beermann, S., Di Cicco, G., Dietrich, U. & Maga, G. A motif unique to the human dead-box protein DDX3 is important for nucleic acid binding, ATP hydrolysis, RNA/DNA unwinding and HIV-1 replication. *PLoS ONE* **6,** e19810 (2011).
19. Högbom, M. *et al.* Crystal structure of conserved domains 1 and 2 of the human DEAD-box helicase DDX3X in complex with the mononucleotide AMP. *J. Mol. Biol.* **372,** 150–159 (2007).
20. Sengoku, T., Nureki, O., Nakamura, A., Kobayashi, S. & Yokoyama, S. Structural basis for RNA unwinding by the DEAD-box protein *Drosophila* Vasa. *Cell* **125,** 287–300 (2006).
21. Wang, L. *et al. SF3B1* and other novel cancer genes in chronic lymphocytic leukemia. *N. Engl. J. Med.* **365,** 2497–2506 (2011).
22. Stransky, N. *et al.* The mutational landscape of head and neck squamous cell carcinoma. *Science* **333,** 1157–1160 (2011).
23. Bentley, D. R. *et al.* Accurate whole human genome sequencing using reversible terminator chemistry. *Nature* **456,** 53–59 (2008).
24. Genome Reference Consortium. Human Genome Overview http://www.ncbi.nlm.nih.gov/projects/genome/assembly/grc/human/ (2012).
25. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25,** 1754–1760 (2009).
26. Ramos, A. *et al.* Oncotator http://www.broadinstitute.org/oncotator/ (2012).
27. Robinson, J. T. *et al.* Integrative genomics viewer. *Nature Biotechnol.* **29,** 24–26 (2011).
28. Ren, J. *et al.* DOG 1.0: illustrator of protein domain structures. *Cell Res.* **19,** 271–273 (2009).
29. Pettersen, E. F. *et al.* UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25,** 1605–1612 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by NIH grants NHGRI U54HG003067 to E. S. Lander (E.S., D.A., S.B.G., G.G., M.M.); R01CA109467 (S.L.P., J.P.M.); R01CA105607 (H.G., T.M.R., M.M., S.L.P.); P30 HD18655 (S.L.P.); R01 CA030002 and CA050661 (T.M.R.); R01 NS046789 (G.R.C.); R01 CA154480 (P.T.); R25NS070682 (S.S.) and R01CA148699 (M.D.T.); St. Baldrick's Foundation Scholar Award and the Beirne Faculty Scholar endowment and Center for Children's Brain Tumors at Stanford University (Y.-J.C.); German Cancer Aid (109252) and the BMBF ICGC-PedBrain project (N.J., D.T.W.J., P.L., S.M.P.); HHMI (G.R.C.); the Pediatric Brain Tumor Foundation (M.D.T.); Canadian Institutes of Health Research Fellowship (T.J.P.); Restracomp funding from the Hospital for Sick Children (P.A.N.); and the Mullarkey Research Fund (S.L.P.). We thank Children's Oncology Group and the Cooperative Human Tissue Network for providing tumour samples, the staff of the Broad Institute Biological Samples, Genome Sequencing and Genetic Analysis Platforms for their assistance in genomic processing of samples and generating the sequencing data used in this analysis, K. Keho and M. Brown at Pacific Biosciences for technical support with sample barcoding methods, and L. Gaffney of Broad Institute Communications for assistance with figure layout and design.

Author Contributions Y.-J.C., M.M. and S.L.P. conceived the project. Y.-J.C., T.J.P., M.M. and S.L.P. wrote the manuscript with input from co-authors. S.D.W., T.C.A., J.P.F., S.S., N.T., Y.-J.C., A.G.K. and F.Y. performed functional characterization studies. D.A.P.K. generated *in silico* structural modelling of DDX3X mutations. T.J.P. conducted the bioinformatic analysis, supported by S.L.C., P.S., K.C., M.S.L., A.M., A.H.R., A.S., H.G., P.T., J.P.M., N.J. and D.T.W.J.; D.A., E.S., S.B.G., and G.G. facilitated transfer, sequencing and analysis of samples. P.A.N. and M.D.T. provided tissues for analysis. Y.-J.C., J.P.F. and V.A. processed tumour and blood samples for study. G.R.C. generated reagents used in functional characterization studies. P.L., S.M.P. and T.M.R. assisted with interpretation of results. J.B., M.O.C., R.L.E., N.J.L., J.M., M.G.R., C.R. and B.S. performed microfluidic PCR and single-molecule real-time sequencing for validation analysis.

Author Information Sequence data used for this analysis are available in dbGaP under accession phs000504.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.-J.C. (yjcho1@stanford.edu), M.M. (matthew_meyerson@dfci.harvard.edu) or S.L.P. (scott.pomeroy@childrens.harvard.edu).