Globally altered epigenetic landscape and delayed osteogenic differentiation in H3.3-G34W-mutant giant cell tumor of bone

Pavlo Lutsik, Annika Baude, Daniela Mancarella, Simin Öz, Alexander Kühn, Reka Toth, Joschka Hey, Umut H. Toprak, Jinyeong Lim, Viet Ha Nguyen, Chao Jiang, Anand Mayakonda, Mark Hartmann, Felix Rosemann, Kersten Breuer, Dominik Vonficht, Florian Grünschläger, Suman Lee, Maren Kirstin Schuhmacher, Denis Kusevic, Anna Jauch, Dieter Weichenhan, Jozef Zustin, Matthias Schlesner, Simon Haas, Joo Hyun Park, Yoon Jung Park, Udo Oppermann, Albert Jeltsch, Florian Haller, Jörg Fellenberg, Anders M. Lindroth, Christoph Plass

The neoplastic stromal cells of giant cell tumor of bone (GCTB) carry a mutation in H3F3A, leading to a mutant histone variant, H3.3-G34W, as a sole recurrent genetic alteration. We show that in patient-derived stromal cells H3.3-G34W is incorporated into the chromatin and associates with massive epigenetic alterations on the DNA methylation, chromatin accessibility and histone modification level, that can be partially recapitulated in an orthogonal cell line system by the introduction of H3.3-G34W. These epigenetic alterations affect mainly heterochromatic and bivalent regions and provide possible explanations for the genomic instability, as well as the osteolytic phenotype of GCTB. The mutation occurs in differentiating mesenchymal stem cells and associates with an impaired osteogenic differentiation. We propose that the observed epigenetic alterations reflect distinct differentiation stages of H3.3 WT and H3.3 MUT stromal cells and add to H3.3-G34W-associated changes.

1 Division of Cancer Epigenomics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 2 Faculty of Biosciences, Ruprecht Karl University of Heidelberg, Im Neuenheimer Feld 234, 69120 Heidelberg, Germany. 3 Bioinformatics and Omics Data Analytics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 4 Graduate School of Cancer Science and Policy, Cancer Biomedical Science, National Cancer Center, Goyang-si, Gyeonggi-do 10408 Republic of Korea, Republic of Korea. 5 Botnar Research Centre, Oxford NIHR BRC, Nuffield Department of Orthopaedics, Rheumatology and Musculoskeletal Sciences, University of Oxford, Oxford OX3 7LD, UK. 6 Section Translational Cancer Epigenomics, Division of Translational Medical Oncology, National Center for Tumor Diseases (NCT) & German Cancer Research Center (DKFZ), Im Neuenheimer Feld 460, 69120 Heidelberg, Germany. 7 Division of Stem Cells and Cancer, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 8 Heidelberg Institute for Stem Cell Technology and Experimental Medicine-HI-STEM gGmbH, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 9 Department of Biochemistry, Institute of Biochemistry and Technical Biochemistry, University of Stuttgart, Allmandring 31, 70569 Stuttgart, Germany. 10 Institute of Human Genetics, Ruprecht Karl University of Heidelberg, Im Neuenheimer Feld 366, 69120 Heidelberg, Germany. 11 Institute of Pathology, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, 20251 Hamburg, Germany. 12 Department of Nutritional Science and Food Management, Ewha Womans University, 52 Ewheyeoedae-gil, Daehyeon-dong, Seodaemun-gu, Seoul 03760, Republic of Korea. 13 FRIAS-Freiburg Institute of Advanced Studies, Albert Ludwig University of Freiburg, Alberstraasse 19, 79104 Freiburg, Germany. 14 Institute of Pathology, University Hospital Erlangen, Friedrich Alexander University Erlangen-Nürnberg, Krankenstrasse 8, 91054 Erlangen, Germany. 15 Department of Experimental Orthopaedics, Orthopaedic University Hospital Heidelberg, Ruprecht Karl University of Heidelberg, Schlierbacher Landstrasse 200a, 69118 Heidelberg, Germany. 16 German Consortium for Translational Cancer Research (DKTK), Heidelberg, Germany. 17 Present address: Division of Neuroblastoma Genomics, German Cancer Research Center (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. 18 These authors contributed equally: Pavlo Lutsik, Annika Baude, Daniela Mancarella, Simin Öz. Email: lindroth@ncc.re.kr; c.plass@dkfz.de
The discovery of mutated histone genes in aggressive cancers raised a lot of interest in the cancer research community due to their ability to globally alter the epigenomic landscape. A frequently mutated histone is the non-canonical histone variant H3.3. In contrast to canonical H3.1 and H3.2, the incorporation of histone variant H3.3 is replication-independent and its turnover occurs throughout the cell cycle. Deposition occurs either via the histone regulator A (HIRA) chaperone complex at sites of gene activation or through alpha-thalassemia/mental retardation X-linked-death domain associated protein (ATRX-DAXX) into heterochromatic regions and the silent allele of imprinted genes. Mouse embryonic stem cells require H3.3 for correct establishment of H3K27me3 patterns at bivalent promoters of developmentally regulated genes. While two human genes, H3 histone family member 3A (H3F3A) and 3B (H3F3B), encode for an identical H3.3 protein, oncogenic mutations occur gene-specifically in different tumor types. Lysine-27-to-methionine (H3.3-K27M) and glycine-to-arginine or valine substitution (H3.3-G34R/V) in pediatric gliomas, as well as glycine-34-to-tryptophan or leucine (H3.3-G34W/L) substitutions in giant cell tumor of bone (GCTB) have been described, all due to mutations in H3F3A. For H3F3B, mutations leading to a lysine-36-to-methionine (H3.3-K36M) substitution were reported in chondroblastomas. The molecular consequence of the H3.3-K27M mutation is a global loss of the repressive chromatin mark H3K27me3 through interference with histone methyltransferases nuclear receptor binding SET domain protein 2 (NSD2) and SET domain containing 2 (SETD2) of the PRC2 complex. Similarly, the H3.3-K36M mutation suppresses the deposition of the H3K36me3 mark by interference with histone methyltransferases nuclear receptor binding SET domain protein 2 (NSD2) and SET domain containing 2 (SETD2). Recent findings suggested reduced levels of H3K36me3 and increased levels of H3K27me3 in cis in HeLa cells overexpressing H3.3-G34W. However, the detailed effects of this mutant histone variant on the epigenome are yet to be determined and to be analyzed in patients. GCTB, where this mutation was shown as the sole alteration, offers a unique system to study these effects in primary patient material.

GCTB is a rare locally aggressive bone neoplasm that typically affects the meta-epiphyseal regions of long bones in young adults. These tumors consist of three major cell types: stromal cells originating from mesenchymal stem cells (MSC), multinuclear giant cells and mononuclear histiocytic cells. The GCTB stromal cells show high expression of H3.3-G34W in more than 90% of cases and display markers of both MSC and pre-osteoblast cell populations. The neoplastic stromal cell population secretes high levels of the receptor activator of NF-κB ligand (RANKL) and reduced levels of its decoy receptor, osteoprotegerin (OPG), thereby attracting and activating surrounding monocytes. Upon activation, the recruited monocytes fuse to form multinucleated giant cells, which resemble osteoclasts and lead to massive bone destruction.

Here we investigate the effects of H3.3-G34W on global epigenomic patterns in patient samples from four different centers. We find epigenetic distortions that contribute to the phenotypes of GCTB, stochastic genomic instability and increased osteolysis. Furthermore, we demonstrate that neoplastic and non-neoplastic GCTB stromal cells represent distinct stages of osteogenic differentiation. Differentiation-related epigenetic differences add to the overall picture of H3.3-G34W-associated global epigenetic alterations, whereas the differentiation delay is potentially driven by the direct effects of H3.3-G34W. Our findings collectively suggest that the single-residue alteration of H3.3 induces epigenomic changes with implications for the development of stromal cells and the tumorigenic process.

**Results**

H3K36me is unaltered in H3.3-G34W-expressing stromal cells. Recent biochemical studies have shown that G34 substitutions in H3.3, including G34W, inhibit the activity of the histone methyltransferase SETD2, which is responsible for H3K36me3, in cis (on the same histone tail). We verified this effect in HEK293T cells stably overexpressing H3.3-G34W or wild-type H3.3 as a control and confirmed in cis effects on H3.3-G34W K36 trimethylation levels. We did not observe any in trans effects on endogenous H3 modifications. To specifically test whether these biochemical findings apply to patient samples, we obtained access to GCTB biopsies from four different cohorts (Supplementary Data 1). For the initial characterization of 30 GCTB samples (Supplementary Data 1), we performed immunohistochemical analysis with a H3.3-G34W-specific antibody. Positive staining was observed and validated in 29 of 30 cases (Fig. 1a, b and Supplementary Fig. 1b). The H3.3-G34W-negative case (unidentified patient, UPI-13) carried a H3F3A (c.103_104GG>TT) mutation encoding a H3.3-G34L substitution that has already been described for GCTB4. We established both, neoplastic, H3.3-G34W-expressing (H3 MUT) and non-neoplastic, H3.3-G34W-negative (H3 WT) stromal cell lines from primary tumor tissue (Fig. 1b; Supplementary Data 1). Cell type differences were ruled out by flow cytometric analyses which revealed a high expression of MSC markers and low expression of hematopoietic markers in both H3.3 WT and H3.3 MUT cell lines, suggesting that both are of mesenchymal origin (Supplementary Fig. 1d). In total, we collected 96 tissue samples from 95 different GCTB patients from four different cohorts (Supplementary Data 1) and were able to establish 26 stromal cell lines from 24 different GCTB patients from two cohorts (Fig. 1c).

In addition to H3.3 WT cells, we analyzed bone marrow-derived primary MSCs from non-GCTB patients from here on referred to as nontumoral stromal cells (nt-SC) (Supplementary Data 1). We verified the mutational status of the cells using Sanger resequencing, ultra-deep resequencing on the MiSeq platform or whole-genome sequencing (Supplementary Data 1, Fig. 1b, Supplementary Fig. 1c, e). In addition to the common mutation leading to the G34W alteration in H3.3, whole-genome sequencing of seven patient-derived primary H3.3 MUT cell lines (Supplementary Data 1) did not reveal any recurrent genetic alterations (Fig. 1d, Supplementary Fig. 1e, f). In contrast to other malignancies carrying H3F3A mutations, as for example pediatric glioblastoma with co-occurring mutations in tumor protein 53 (TP53) and ATRX4,5 (Fig. 1g, Supplementary Fig. 1g), and to other bone tumors, e.g., osteosarcoma (Fig. !f, Supplementary Fig. 1h, i), GCTB showed an extremely low overall mutation frequency for H3.3 MUT and H3.3 WT cells (Fig. 1f, Supplementary Fig. 1e, f). This places GCTB in a unique position to uncover epigenomic alterations linked to H3.3-G34W.

To understand how H3.3-G34W exerts its function in tumor cells, we analyzed protein fractions by Western Blot and found H3.3-G34W incorporated into chromatin (Supplementary Fig. 2a). To analyze in cis effects of H3.3-G34W in patient derived cells, we used a specific and verified antibody to map H3.3-G34W and identify sites of H3.3-G34W enrichment in two independent patient cell lines (Supplementary Data 1). We identified high-confidence genomic regions showing H3.3-G34W enrichment (Supplementary Fig. 2b, c, Supplementary Data 2) and profiled H3K36me3 along with several other histone modifications using chromatin immunoprecipitation (ChIPmentation). Changes of H3K36me3 or H3K27me3 between H3.3 WT and H3.3 MUT stromal cells at sites of H3.3-G34W enrichment were minute and did not recapitulate the strong loss of H3K36me3 in cis observed in HEK293T cells (Supplementary Fig. 2b, c). Nucleosomes enriched by H3.3-G34W-ChIP most probably contain H3.3-G34W, as well as wild-type H3.3
potentially masking in cis effects on the H3.3-G34W N-terminal tail. Using Western Blot analysis of whole cell lysates we did not reveal any changes in the total amount of K36 methylated histone H3 indicating that the G34W substitution does not affect methylation of K36 on other histones lacking the mutation, which is in line with the analysis in HEK293T cells in Supplementary Fig. 1a (Supplementary Fig. 2d).

The epigenome of H3.3 MUT cells is globally altered. As the effects of H3.3-G34W on the global level of H3.3K36me3 were insignificant, we tested whether H3.3-G34W might exert its tumorigenic effects through other epigenetic mechanisms. We analyzed several H3.3 MUT and H3.3 WT cells, as well as a large number of biopsies of the four GCTB cohorts (Supplementary Data 1), using HumanMethylation450 arrays (89 samples in total).
and observed full-range DNA methylation changes of a large number of CpG sites (Fig. 2a). Dimensionality reduction and clustering analysis of the methylation data showed the most pronounced changes between primary patient-derived H3.3 MUT and H3.3 WT cell lines as compared to GCTB biopsies. Most of the primary tumor samples had an intermediate methylome identifying them as mixtures of normal and tumor cells (Fig. 2a, b), as also confirmed by deep resequencing of the H3F3A gene locus of tissue-derived DNA showing a lower variant allele frequency than the theoretically expected 0.5 for a heterozygous
Supplementary Fig. 2e). In contrast, other histone modifications, we used a genome-wide approach to stratify the analysis. Type cells into a single control group (H3.3 WT) for further nt-SCs clustered with H3.3 WT cells, we combined all H3.3 wild substitution to be the major determinant of variability in DNA methylation pro-Data 1). Global hierarchical clustering proved the H3.3-G34W mutation (Fig. 1b). This observation prompted us to restrict all our subsequent analyses to H3.3 MUT and H3.3 WT cells in order to obtain a clean view of H3.3-G34W-associated epigenetic alterations in pure cell populations. To further a detailed analysis of the epigenome, we profiled DNA methylation at single CpG resolution using whole genome bisulfite sequencing (WGBS), analyzed chromatin accessibility with the assay for transposase-accessible chromatin using sequencing (ATAC-seq) and analyzed global distribution of several histone marks (H3K4me1, H3K4me3, H3K9me3, H3K27me3, H3K27ac, H3K36me3) to investigate their potential redistribution. Since the initial methylation analysis showed high homogeneity within the mutant and wild-type groups and culturing primary tumor cells did not yield sufficient material for a comprehensive omics profiling of each line, we analyzed H3.3 WT and H3.3 MUT cell lines from several patients to exclude patient-dependent alterations (Supplementary Data 1). Global hierarchical clustering proved the H3.3-G34W substitution to be the major determinant of variability in DNA methylation, chromatin accessibility and posttranslational histone modifications (H3K27me3, H3K27ac, H3K4me1) with H3.3 MUT and H3.3 WT groups forming separate clusters (Fig. 2c–e, Supplementary Fig. 2e). In contrast, other histone modifications (H3K36me3, H3K4me3) did not result in H3.3 MUT-separating and H3.3 WT-separating clusters (Supplementary Fig. 2e). Since nt-SCs clustered with H3.3 WT cells, we combined all H3.3 wild type cells into a single control group (H3.3 WT) for further analysis.

In order to identify loci with more pronounced epigenetic changes, we used a genome-wide approach to stratify the observed differences from all epigenetic layers using the ENCODE functional annotation of the MSC epigenome21. We observed profound differences in several groups of chromatin states including heterochromatic and Polycomb-repressed states, as well as bivalent domains (Fig. 2f). Heterochromatic regions were noticeably hypomethylated while bivalent domains accumulated a whole range of alterations, including gain of DNA methylation, loss of chromatin accessibility, as well as loss of several histone modifications (H3K27me3, H3K4me3 etc). Collectively, we conclude that the epigenome of H3.3 MUT cells shows significant and reproducible differences to that of H3.3 WT stromal cells.

**Hypomethylation in megabase domains and genomic instability.** We first followed up changes associated with loss of DNA methylation. In contrast to the homogeneity of DNA methylation profiles observed within each of the two groups (Supplementary Fig. 3a), DNA methylation was non-uniformly altered between H3.3 WT and H3.3 MUT cells across the entire genome (Fig. 3a). The majority of significantly changed loci showed profound reduction of DNA methylation in H3.3 MUT cells, while a minority was hyper-methylated (Fig. 3a, Supplementary Fig. 3b). Overall, H3.3 MUT cells exhibited a 20% genome-wide reduction of DNA methylation (Fig. 3b, Supplementary Fig. 3c). To systematically characterize the abundant DNA methylation changes, we segmented the genome into large methylation domains (LMD I to IV) with sizes >20 kb (Supplementary Fig. 3d, Supplementary Data 2) using a combination of breakpoint analysis and clustering (see Methods for details). The majority of regions (predominantly LMDs III and IV) matched the criteria for partially methylated domains i.e., megabase-scale domains of predominantly repressive chromatin with low gene density22. Each LMD had a distinct level of DNA methylation, a discrete pattern of histone marks and gene density, suggesting a non-uniform domain structure. It is noteworthy that most variable methylation probes clustered iso-H3.3-WT clones dispersed off, indicating profound differences in several groups of chromatin states including heterochromatic and Polycomb-repressed states, as well as bivalent domains (Fig. 2f, Supplementary Fig. 2e).

Collectively, we conclude that the epigenome of H3.3 MUT cells shows significant and reproducible differences to that of H3.3 WT stromal cells. Isogenic cells recapitulate H3.3 MUT DNA hypomethylation. As H3.3 MUT stromal cells showed the strongest epigenetic differences to H3.3 WT stromal cells on the DNA methylation level, we expected DNA methylation to play a major role in stromal cell transformation leading to GCTB. To verify that the observed epigenetic changes were dependent on H3.3-G34W expression, we aimed to recapitulate the findings in an unrelated cell line system. To this end, we introduced the H3.3-G34W encoding mutation into HeLa cells by targeting the endogenous H3F3A locus as earlier described (Supplementary Fig. 4)24. Individual iso-H3.3-WT and iso-H3.3-G34W clones were isolated, of which four iso-H3.3-WT and four iso-H3.3-G34W monoclonal lines and one parental HeLa cell line were subjected to HumanMethylationEPIC DNA methylation analysis (Fig. 4a). The iso-H3.3-G34W samples showed similar alterations as those found when comparing H3.3 MUT and H3.3 WT stromal cells of GCTB (Fig. 4a–e). Principal component analysis of the most variable methylation probes clustered iso-H3.3-WT clones together while iso-H3.3-G34W clones dispersed off, indicating changes in DNA methylation in iso-H3.3-G34W but not WT isogenic cell lines (Fig. 4a). Similar to primary GCTB cells, the largest principal component (PC1, 31% of variance explained) captured widespread hypomethylation in iso-H3.3-G34W accompanied by focal hypermethylation events (Fig. 4a).
Clustering analysis further enforced the distinct difference between iso-H3.3-WT and G34W clones (Fig. 4b). A scatter plot of all probes indicated that iso-H3.3-G34W cells showed predominantly hypomethylation as also shown for H3.3 MUT cells from GCTB (Fig. 4c). We found 9047 differentially methylated probes (Δβ-value>0.2 and FDR < 0.05) of which 5688 (63%) were hypomethylated in iso-H3.3-G34W cells (Fig. 4d). These data recapitulate the bias towards hypomethylation found in H3.3 MUT stromal cells in comparison to H3.3 WT stromal cells (compare Fig. 3a, b), and confirm that the changes in DNA methylation are associated with H3.3-G34W. CpG sites that lost methylation were more frequently localized in intergenic regions and promoter-proximal exons, while hypermethylated ones were in addition moderately enriched at promoters and depleted at iso-H3.3-G34W.
exons (Fig. 4e). This was coherent with widespread hypomethylation of intergenic regions (LMDs III and IV in Fig. 3c, d, e, h) and overrepresentation of promoters among hypermethylated regions (Supplementary Fig. 3b) observed in H3.3 MUT cells. A specific example for a hypomethylated region in iso-H3.3-G34W cells is the RANKL locus showing hypomethylation at exon 3, a potential alternative promoter element (Fig. 4f). This is in line with the methylation differences observed between H3.3 WT and H3.3 MUT stromal cells at the same locus (Fig. 5a). RANKL signaling has been extensively studied in GCTB. The expression of RANKL, a master regulator of osteoclast differentiation, has been shown to be upregulated in GCTB stromal cells causing an osteolytic phenotype. We verified the increased expression of RANKL (TNFSF11) in H3.3 MUT cells (Fig. 5b, Supplementary Fig. 5a) and additionally observed decreased expression and secretion of its decoy receptor Osteoprotegerin (OPG, TNFRSF11B) (Fig. 5c, Supplementary Fig. 5a, b). Similar expression patterns were already described by us earlier. The OPG locus showed decreased levels of the active histone marks H3K4me3 and H3K27ac potentially indicating a missing activation by a transcription factor (Supplementary Fig. 5c). One known transcription factor of OPG is the early B-cell factor 2 (EBF2) which is a bivalent gene (Fig. 5f) and belongs to the key regulators of osteogenic differentiation in mice. OPG became downregulated after siRNA mediated EBF2 knockdown, confirming a role of EBF2 in OPG expression in stromal cells (Fig. 5d, Supplementary Fig. 5d). The EBF family is a conserved group of four transcription factors. Our RNA-seq analysis found EBF2 and EBF3 to be differentially expressed between H3.3 WT and H3.3 MUT stromal cells (Fig. 5e, Supplementary Fig. 5e). EBF3, previously reported as a tumor suppressor in glioblastoma, has shown to be differentially expressed between H3.3 WT and H3.3 MUT stromal cells whereas EBF2 expression was almost completely lost. We found the EBF2 locus to be hypomethylated with a focal hypermethylation around the promoter region (Fig. 5f). Increased H3K9me3 and H3K27me3 levels and decreased levels of H3K27ac supported a repressed state of EBF2. Lost ATAC signals in H3.3 MUT cells indicated differentially closed chromatin. These findings link the H3.3 G34W-associated epigenetic dysregulation of EBF2 expression to the osteolytic phenotype of GCTB.

Taken together, we showed that the stable introduction of H3.3-G34W into a GCTB unrelated cell line recapitulates the DNA hypomethylation trend seen in GCTB stromal cells. Furthermore, we could detect epigenetic alterations that directly and indirectly affected the expression of key regulators of bone metabolism shedding novel light on the emergence of the osteolytic phenotype in GCTB.

**Bivalent promoters and delayed differentiation in H3.3 MUT.** Besides dominant hypomethylation in heterochromatic regions, we additionally observed profound epigenetic changes at bivalent
Fig. 5 The epigenetic determinants of the osteolytic phenotype in GCTB. a Genome browser view of the gene TNFSF11 encoding for RANKL. Each lane (dash-separated) represents normalized averaged signals of DNA methylation (WGBS), the level of H3K4me3, H3K4me1, H3K9me3, H3K27ac, H3K27me3, H3K36me3, and chromatin accessibility (ATAC) in several replicates of H3.3 WT (lane-wise top) or H3.3 MUT (lane-wise bottom) cells. b, c Expression levels of RANKL (TNFSF11) and OPG (TNFRSF11B) in H3.3 WT (blue) H3.3 MUT (red) cells as analyzed by RNA-seq. TPMS, transcripts per million. Differences were borderline significant for both genes with p = 0.092 (TNFSF11) and p = 0.062 (TNFRSF11B), Wilcoxon rank sum test with n = 6 independent patient cell lines in each group. d Expression analysis of EBF2 and OPG (TNFRSF11B) by qRT-PCR 48 h after siRNA-mediated knockdown of EBF2 in H3.3 WT cells (UPI-13). Log fold change (LFC) relative to expression in cells transfected with control siRNAs. n = 4 independent experiments are shown. Results showed significance in a one-sample t-test with p-values 0.006 and 0.018. e Expression levels of all members of the EBF family in H3.3 WT (blue) H3.3 MUT (red) cells as analyzed by RNA-seq. TPMS, transcripts per million. Differences were significant for EBF2 (p = 4.8 × 10^-3) and EBF3 (p = 2.1 × 10^-3), Wilcoxon rank sum test with n = 6 independent patient cell lines in each group. f Genomic browser view of the EBF2 locus. Legend as in a. All analyzed stromal cells with their UPIs are listed in Supplementary Data 1. In all boxplots bar represents the mean, box the IQR, upper and lower whiskers extend from the smallest to the largest value within 1.5 IQR from the lower and upper edges of the box, respectively.

domains (Fig. 2f), which are well known sites of H3.3 deposition. Changes included gain of DNA methylation which coincide with decreased accessibility, decrease of H3K27me3 and increased levels of H3K36me3 (Fig. 2f). Most promoter-associated ATAC peaks that were lost in H3.3 MUT cells overlapped with bivalent promoters (Supplementary 6a). As a consequence of epigenetic disturbance, bivalent genes comprised a significant portion of differentially expressed genes both upregulated and downregulated in H3.3 MUT cells (Supplementary Fig. 6b). It was previously reported that comparable epigenetic regulation and downregulated in H3.3 MUT cells (Supplementary Fig. 6c, Supplementary Data 3). Gene ontology (GO) analysis of closed regions revealed many categories related to differentiation (Supplementary Fig. 6e). Impaired differentiation was already suggested for tumor entities harboring mutations in histones including GCTB. To investigate whether H3.3 WT and H3.3 MUT cells differ in their osteogenic differentiation state, we performed a comparison across all genes covered by RNA-seq with those obtained from an in vitro differentiation of MSCs. The expression changes during osteogenic differentiation largely correlated inversely with the expression changes observed between H3.3 WT and H3.3 MUT cells, with anticorrelation increasing from very early to late differentiation stages (Fig. 6a). Accordingly, an overlap of the differentially expressed genes from both experiments showed that the majority of matches were upregulated during differentiation and downregulated in H3.3 MUT cells (Supplementary Fig. 6f). Examples included insulin growth factor 2 (IGF2) and leptin (LEP) previously implicated in osteogenic differentiation (Fig. 6a). A global principal component analysis confirmed a less
differentiated state of H3.3 MUT cells resembling MSC in a very early stage of differentiation and H3.3 WT cells showing an expression profile more similar to MSC in early to middle stage of osteogenic differentiation (Fig. 6b). To confirm differences in the osteogenic differentiation state, we stained H3.3 WT and H3.3 MUT stromal cells for the activity of the osteogenic marker alkaline phosphatase (ALP). Most of the H3.3 WT cells exhibited ALP activity whereas only some H3.3 MUT cells showed ALP activity (Fig. 6c). Quantification of the ALP activity relative to viability as a surrogate for cell count confirmed reduced ALP activity for H3.3 MUT stromal cells (Fig. 6d). Impaired differentiation of GCTB stromal cells was suggested earlier based on the analysis of histological markers and transcriptomic profiling. We performed an osteogenic differentiation of H3.3 MUT

Fig. 6 Delay of osteogenic differentiation in H3.3 MUT stromal cells. a Correlation between differentiation-associated and H3.3-G34W-associated gene expression changes. For the differentiation log-fold changes (LFC) were calculated between average of profiles from each differentiation timepoint group (very early: 0.5 to 2 h; early: 4 to 554 16 h; mid: 24 to 168 h; and late: 336 to 504 h) and the profiles of mesenchymal stem cells (MSCs) sampled prior to the start of the experiment. Each hexagon reflects the density of the underlying data points. Genes with most pronounced changes in both data sets are marked. b Principal component analysis of RNA-Seq data from osteogenic differentiation of MSC (greyscale rectangles) and expression profiles of GCTB samples, H3.3 WT (blue triangles) and H3.3 MUT (red circles), only PC2 and PC3 are shown, while PC1 capturing batch effects between two experiments is not shown. The sampled differentiation timepoints were grouped as in a. For H3.3 WT and H3.3 MUT each data point represents one patient. The isochores and background shading give a surface fit-based estimate of differentiation time for each point of the (PC2, PC3) plane. c Alkaline phosphatase staining of nt-SC, H3.3 WT, and H3.3 MUT cells during osteogenic differentiation. d Quantification of alkaline phosphatase (ALP) activity relative to the viability as a surrogate for cell count over the time course of 15 days of in vitro osteogenic differentiation. Each data point represents one patient. Middle bars represent means, error bars show standard deviations within each group. H3.3 WT in blue, n = 8 and H3.3 MUT in red, n = 7 biologically independent cell lines from different patients. P-values are from unpaired (for within-group comparisons) and paired (for between-group comparisons) two-tailed t-tests. e Model for GCTB tumorigenesis: The H3F3A mutation encoding H3.3-G34W occurs in osteoblastic precursor cells and leads to alterations in osteogenic differentiation. Epigenetic differences described between H3.3 WT and H3.3 MUT stromal cells are the result of measuring cells at different differentiation stages and direct effects of the mutated histone.
and H3.3 WT cells from several patients in vitro to analyze the potential of H3.3 MUT and H3.3 WT cells to differentiate. While both H3.3 WT and H3.3 MUT stromal cells showed an increase of ALP activity indicating the potential for osteogenic differentiation (Fig. 6c, d), we noticed that H3.3 MUT stromal cells lagged behind and did not achieve the level of ALP activity reached by H3.3 WT stromal cells (Fig. 6c, d). We concluded that H3.3-G34W associates with changes in bivalent regions and that their deregulation potentially has an effect on osteogenic differentiation. H3.3 WT and H3.3 MUT stromal cells therefore differ in their state of osteogenic differentiation in GCTB patients and this difference contributes to the herein found epigenetic alterations (Fig. 6e).

Discussion
Mutant epigenetic regulator proteins, including histones, have been described in multiple cancer types. Despite the fact that oncogenic mutations in histones influence the epigenetic landscape, deep insight into the mechanistic ramifications to cancer initiation is still incomplete. Several studies investigated the mechanism of how the G34W substitution in histone variant H3.3 affects the epigenetic machinery responsible for its post-translational modifications. To this end, reduced H3K36 methylation and increased H3K27me3 in cis were described in model systems and verified by us in this work. However, whether these effects can be found in patients and whether epigenetic changes are involved in tumorigenesis of GCTB was not known so far. In this paper, we therefore analyzed GCTB tissue, as well as patient-derived primary stromal cells from four different centers to investigate H3.3-G34W-associated epigenetic changes and their contribution to the GCTB neoplasia. The high incidence of H3.3-G34W in GCTB in a largely unaltered genomic context leaves GCTB as a suitable system to study histone mutation-driven tumorigenesis.

We found that, despite that H3.3-G34W is incorporated into chromatin, its presence did not lead to changes in the global amount of H3K36me3 as shown for other substitutions affecting H3.3 lysine 27 and lysine 36. This ruled out a possible in trans effects of H3.3-G34W on histone posttranslational modifications. Using ChIP sequencing, we identified high confidence H3.3-G34W enrichment sites in patient cell lines and analyzed common histone marks at these regions. No pronounced differences were found that could recapitulate the in cis effects on H3.3 lysine methylation observed in HEK293T cells. The absence of clear-cut in cis effects upon chromatin modifications could be explained by an inability to distinguish wild-type and mutant H3.3 in GCTB samples by western blot analysis. In ChIP analysis, nucleosomes of mutant cells could include both, wild-type, as well as mutated H3.3 obliterating possible in cis effects. Moreover, antibodies against histone modifications cannot distinguish between H3 variants and the overall small fraction of the mutant H3.3 (25% of overall H3.3) in the total H3 pool makes analysis of H3.3 modifications challenging.

We show that H3.3-G34W, the sole recurrent alteration of GCTB, associates with large-scale differences in multiple epigenetic marks. H3.3 MUT stromal cells show heterochromatic defects, specifically reduced genome-wide levels of DNA methylation and gained accessibility at heterochromatic regions which is in line with previous reports on pediatric glioblastomas with the G34R substitution in H3.3.39 These changes could potentially contribute to the genomic instability described for GCTB.23 We furthermore observed localized changes at bivalent regions, some of the same sites previously described as targets of H3.3 deposition.67 DNA methylation changes could be recapitulated using an isogenic system verifying effects on the methylome to be H3.3-G34W-specific. Moreover, we show H3.3-G34W-associated methylation changes to directly or indirectly affect the expression of the main players of bone metabolism, RANKL and OPG. These results connect the H3.3-G34W-associated epigenetic dysregulation to the osteolytic phenotype, a hallmark of GCTB.

While the epigenomic differences between H3.3 WT and H3.3 MUT stromal cells can be partially ascribed to direct effects of H3.3-G34W, other alterations observed can be explained by the fact that H3.3 MUT and H3.3 WT stromal cells represent distinct differentiation stages. Assuming epigenetic reprogramming during differentiation as demonstrated in recent studies,60,61 this could also contribute to the epigenetic differences seen between H3.3 WT and H3.3 MUT cells. While expression profiles from H3.3 MUT cells resemble precursor states of osteoblasts, the H3.3 WT cells resemble more mature osteoblasts, suggesting impaired or delayed differentiation of H3.3 MUT cells as already suggested by42. A weak differentiation signature is in line with previous reports suggesting that neoplastic GCTB stromal cells are expressing markers of an early osteoblastic differentiation.63 Genes such as IGF2, LEP, and stathmin-like 2 (STMN2), found to be downregulated in H3.3 MUT stromal cells, were described to play a role in mesenchymal stem cell differentiation and osteoclastogenesis in mice, in promoting bone differentiation and as a marker for osteogenesis, respectively. Consistently, H3.3 was found to be closely associated with differentiation processes.64,65 The impairment of differentiation of H3.3 MUT stromal cells could be mediated by H3.3-G34W-associated downregulation or prevention of induction of genes upregulated during osteogenic differentiation through a yet unknown mechanism. Epigenetic repression of bivalent genes and dysregulation of PRC2 targets described in this study could potentially contribute to the observed phenotype. H3.3 MUT stromal cells are still able to undergo osteogenic differentiation, as it was also found in studies not separating H3.3-mutant neoplastic from H3.3-wild-type stromal cells of GCTB tissue.

Direct effects of H3.3-G34W upon the epigenome could be confirmed in isogenic HeLa cells. The effects on DNA methylation were less pronounced in HeLa cells compared to the alterations observed comparing H3.3 WT and H3.3 MUT stromal cells, potentially reflecting the relatively small passage number after the introduction of the H3F3A mutation (approximately 20 passages) in the HeLa cells compared to primary cells. Another explanation may be that DNA methylation differences between H3.3 WT and MUT stromal cells largely reflect differentiation-associated epigenomic alterations shown to occur during osteogenesis as discussed above. A detailed analysis of H3.3-G34W-induced effects on the epigenome and their connection to impeded osteogenic differentiation will require systems capable of reproducing the differentiation context in vitro or in vivo, such as MSC-derived immortalized cell lines or animal models. Alternatively, detailed transcriptomic and epigenomic maps of the osteogenic differentiation can be instrumental for delineating H3.3-G34W-induced and differentiation-related alterations. Furthermore, meticulous computational approaches, e.g.,59 may help disentangle differentiation-related from cancer-specific epigenetic alterations.

Taken together, our results suggest that the molecular mechanism behind H3.3-G34W-induced epigenomic alterations and neoplastic transformation in GCTB is different to the action of other mutant histone variants such as H3.3-K27M and H3.3-K36M, where the mutated amino acid is a direct target for histone modification.11,13,15 This suggests that the H3.3-G34W substitution alters the N-terminal interactome as previously described, recruiting novel binders or disrupting interactions with binders. The described DNA methylation changes can be indicative of H3.3-G34W affecting either the binding or the activity of...
enzymes writing (DNA methyltransferases, DNMTs) or erasing (ten eleven translocators, TETs) this modification. On the structural side it is known that DNMT3a/b recognize H3K36me3 with their PWWP-domains. The G34W substitution might interfere with binding of DNMT3a/b PWWP preventing proper establishment of DNA methylation patterns. The thereby induced remodeling of the epigenome associated with H3.3-G34W, in particular changes at heterochromatic and bivalent regions, could contribute to an impaired osteogenic differentiation and the phenotypes of GCTB: stochastic chromosomal rearrangements and increased RANKL signaling. Precise sequences of molecular events leading to these phenotypes will be a matter of further studies.

Methods

Patient samples. Biopsies and derived primary cell lines were obtained from patients at the Orthopaedic University Hospital, Heidelberg (UHK), the Korean Cancer Center (KCC), the University Clinic of Leipzig (UKL) and the University Medical Center Hamburg-Eppendorf (UKHE). The complete list of samples used in the study is available in Supplementary Data 1. The use of patient samples and the experiments performed in this study was approved by and in accordance with guidelines by the Ethics Committees of the University of Heidelberg, University Clinic of Leipzig, University Medical Center Hamburg-Eppendorf, and the National Cancer Center of Korea (IRB NCC2015-0070). Informed consent has been obtained from all patients. All H3.3 WT and H3.3 MUT cells analyzed in the manuscript were obtained from different patients or indicated as relapse samples by addition of ‘2’ to the UPI.

Isolation of GCTB stromal cells from patient tissue. Tumor tissue from surgical resections was mechanically cut into small pieces and digested with 1.5 mg/ml collagenase B (Roche Diagnostics, Mannheim, Germany) at 37 °C for 3 h in Dulbecco’s Modified Eagle Medium (DMEM) (Lonza GmbH, Wuppertal, Germany). High glucose supplemented with 10% fetal calf serum (FCS) (Biochrom, Berlin, Germany) and 100 U/ml penicillin/streptomycin (Lonza GmbH), Cells were collected by centrifugation, washed twice in PBS and cultured in DMEM as described above. Twenty-four hours after plating, cells were carefully treated with Trypsin/EDTA (Lonza GmbH) leaving the giant cells attached in the culture flask. Detached cells were cultured for further 2 passages to eliminate histiocytes and remaining giant cells.

Isolation of nt-SC cell lines from the bone marrow. Nt-SCs were isolated from fresh bone marrow samples derived from the iliac crest under the approval of the Ethics Committee of the University of Heidelberg. Bone marrow cells were purified on a Ficoll-Paque™ Plus density gradient (GE Healthcare, Munich, Germany). Bone marrow cells were digested with 1.5 mg/ml collagenase B (Roche Diagnostics, Mannheim, Germany) at 37 °C for 3 h in Dulbecco’s Modified Eagle Medium (DMEM) (Lonza GmbH, Wuppertal, Germany). Cells were then washed in PBS and treated with erythrocyte lysis buffer (0.154 M NaCl, 10 mM KHCO3, 0.1 mM EDTA) to remove erythrocytes. The nt-SC-enriched fraction was seeded and cultured in DMEM high-glucose supplemented with 12.5% FCS (Biochrom), 2 mM L-glutamine, 100 U/ml penicillin, 100 µg/ml streptomycin (Lonza), 4 ng/ml basic fibroblast growth factor (Merck Chemicals GmbH, Darmstadt, Germany), 50 µM 2-mercaptoethanol and 1% non-essential amino acids (Invitrogen, Karlsruhe, Germany). After 48 h, cultures were washed with PBS to remove non-adherent material. During expansion, medium was replaced twice a week.

Cell line maintenance. The GCTB stromal and nt-SC cell lines were cultivated in DMEM/F12 no phenol red (Gibco, Invitrogen, Life Technologies, Paislay, UK) supplemented with 10% FCS (Biochrom, Berlin, Germany) and 2 ng/ml basic FGF (Bionlegend, San Diego, USA) at 37 °C and 5% CO2. To avoid contact inhibition, cells were split after reaching maximum 90% confluency.

Commercially available cell lines. All commercially available cell lines used within this work were authenticated: Hs1143B (DSMZ 09/2018), matching STR-based assay 100%; HEK293T (COA Multiplexion 06/2020), identity based on a SNP-based assay 100%; HeLa: Korean Cell Line Bank, KOA 03/2018, full matching STR-based assay 100%; HeLa: Korean Cell Line Bank, KOA 03/2018, full matching STR-based assay 100%. Biopsies and derived primary cell lines were obtained from patient tissue. Genomic DNA was isolated using the Quick DNA Miniprep kit (Zymo research, Freiburg, Germany) according to the manufacturer’s protocol. PCR amplification was performed using H3F3A wild-type and H3F3A-G34W specific primer, respectively. The reaction consisted of 2 U HotStar Taq polymerase (New England BioLabs, USA), 0.6 µg MgCl2 (50 mM), 0.4 µl dNTPs (10 mM each), 0.5 µl of each primer (10 µM) and 100 ng genomic DNA as template in a total volume of 20 µl. Samples were incubated at 94 °C for 3 min followed by 34 cycles of denaturation at 94 °C for 15 s, annealing at 66 °C for 20 s and extension at 72 °C for 30 s and a final extension step at 72 °C for 7 min. PCR products were separated on a 1.6% agarose gel, visualized by Midori Green (Biozym, Hessisch Oldendorf, Germany) and imaged. All primers are listed in Supplementary Data 5.

Detection of the H3F3A-G34W mutation by mutation-specific PCR. Genomic DNA was isolated using the Quick DNA Miniprep kit (Zymo research, Freiburg, Germany) according to the manufacturer’s protocol. PCR amplification was performed using H3F3A wild-type and H3F3A-G34W specific primer, respectively. The reaction consisted of 2 U HotStar Taq polymerase (New England BioLabs, USA), 0.6 µg MgCl2 (50 mM), 0.4 µl dNTPs (10 mM each), 0.5 µl of each primer (10 µM) and 100 ng genomic DNA as template in a total volume of 20 µl. Samples were incubated at 94 °C for 3 min followed by 34 cycles of denaturation at 94 °C for 15 s, annealing at 66 °C for 20 s and extension at 72 °C for 30 s and a final extension step at 72 °C for 7 min. PCR products were separated on a 1.6% agarose gel, visualized by Midori Green (Biozym, Hessisch Oldendorf, Germany) and imaged. All primers are listed in Supplementary Data 5.
100 ng genomic DNA as template in a total volume of 40 μl. Samples were incubated at 95 °C for 15 min followed by 10 cycles of denaturation at 94 °C for 45 s, annealing at 61–66 °C for 20 s (touchdown – 0.5 °C/cycle) and extension at 72 °C for 60 s followed by 25 cycles of denaturation at 94 °C for 45 s, annealing at 56 °C for 30 s and extension at 72 °C for 60 s and a final extension step at 72 °C for 10 min. Sequencing was performed by GATC Biotech AG, Konstanz, Germany. All primers are listed in Supplementary Data 5.

**Deep targeted resequencing using MiSeq.** Deep resequencing of H3F3A amplicons was performed according to32. DNA was extracted using the Qiang Mini Kit (Qiagen, Hilden, Germany), according to the manufacturer’s instructions. The PCR reaction to amplify the mutation spanning region of the H3F3A gene consisted of 0.35U HotStart Q5 polymerase (NEB, Ipswich, USA), 0.6 μl dNTPs (10 mM each, Fermentas, St. Leon-Rot, Germany), 0.3 μl of each primer (10 μM, Sigma, St. Louis, USA) and 25 ng of DNA in a total volume of 25 μl. Samples were incubated at 98 °C for 1 min followed by 33 cycles of denaturation at 98 °C for 10 s, annealing at 61 °C for 30 s and extension at 72 °C for 20 s, and a final extension step at 72 °C for 2 min. Primer sequences are listed in Supplementary Data 5. Primers included a sequence complementary to the primers used for library preparation. Samples were separated on a 1.2% agarose gel and DNA was visualized by Ethidium Bromide. Gel extraction was performed using the Gel extraction Kit (Qiagen, Hilden, Germany). Libraries were prepared using 12.5 μl NEB NexT 2x PCR mix (New England Biolabs, USA) in combination with 0.75 μl of 10 μM IDT primers with Nextera handles and TrueSeq Unique Indexes (6x, Illumina, USA), 0.3 μl 100X Sybr Green and 11 μl water to a final volume of 50 μl. Amplification was performed for 6 cycles with the following progression: 98 °C 30 s, 98 °C 10 s, 62 °C 30 s, and 72 °C 15 s. Libraries were pooled and sequenced on a Illumina MiSeq in 300 bp paired-end run (Illumina, USA). The Illumina FastQ files were demultiplexed using custom scripts, aligned to the GRCh37 human genome (release 37.3). We excluded reads with zero mapping quality and reads with at least one N base pair in the alignment. We used minimum base quality threshold set to zero [-T 0] and remaining settings left at default values, followed by coordinate-sorting with bamsort (with compression option set to best) and merging duplicate reads with hammarduplicates (with compression option set to best)53; both are part of bowtiebam package (v.0.14.16)52. Somatic SNVs were identified with the DFKZ SNV-calling workflow51. Somatic SNVs and indels in matched tumor normal pairs were identified using the DKEF core variant calling workflows of the ICCG Pan-cancer Analysis of Whole Genomes (PCAGW) project (https://dkfzccrcc.de/pcagw) and biobambam workflow. Tumor and matched control samples were analyzed by Platypus (v. 1.0)26 to identify indel events. SNVs and indels from all samples were annotated using ANNOVAR (v.2015Jul1626) according to GENCODE gene annotation (v. 19) and overlapped with variants from dbsNP111 (build 141) and 1000 Genomes Project release 20150302. Genomic structural rearrangements were detected using SOPHIA (v.3.40) (https://bitbucket.org/utoprak/sophia/src) as described in ref. 66. Briefly, SOPHIA uses supplementary alignments as produced by bwa mem as indicators of a possible underlying SV. SV candidates are filtered by comparing them to a background control set of sequencing data obtained using tumor-normal samples from a published Pan-Cancer Pan-Genome dataset55, which is composed of 3261 patients from published TCGA and icgc studies and both published and unpublished DKFZ studies, sequenced using Illumina HiSeq 2000 (100 bp), 2500 (100 bp), and HiSeq X (151 bp) platforms and aligned uniformly using the same workflow as in this study. An SV candidate is discarded if (i) it has more than 85% read support from low quality reads; (ii) the second breakpoint of the SV was unapplicable in the sample and the first breakpoint was detected in 10 or more background control samples; (iii) an SV with two identified breakpoints had one breakpoint present at least in 98 control samples (3% of the control samples); or (iv) both breakpoints have less than 5% read support. Statistics over SVs for 9 samples with matched control and integrated variant analysis over all samples were based on SOAPHA calls. Allele-specific copy-number aberrations were detected using ACESeq (v.1.0)26. SVs called by SOPHIA were incorporated to improve genome segmentation.

**Whole-genome sequencing and analysis.** Read pairs were mapped to the human reference genome (build 37, version hs37d5), using bwa mem (v. 0.7.8)53 with minimum base-quality threshold set to zero [-T 0] and remaining settings left at default values, followed by coordinate-sorting with bamsort (with compression option set to best) and merging duplicate reads with hammarduplicates (with compression option set to best); both are part of bowtiebam package (v.0.14.16)52. Somatic SNVs were identified with the DFKZ SNV-calling workflow51. Somatic SNVs and indels in matched tumor normal pairs were identified using the DKEF core variant calling workflows of the ICCG Pan-cancer Analysis of Whole Genomes (PCAGW) project (https://dkfzccrcc.de/pcagw). Sequencing data was analyzed by the 1000 Genomes Project (PGCBHM group) were obtained from PedBrain Tumor Project of the International Cancer Genome Consortium20. The processing and variant calling (SNVs, SVs, CNVs) was performed consistently with the GCTB data as described above. To balance sample sizes for the analysis of recurrent genetic events, in addition to analyzing the complete BOCA cohort (Supplementary Fig. 1i), we draw several random 10-sample subsets of all WGS samples are given in Supplementary Data 1. WGS data of 73 bone cancer patients (BOCA cohort) were obtained via the Pan-cancer Analysis of Whole Genomes Project27. Pediatric glioblastoma data (PGCBHM group) were obtained from the DFCI Tumor Project of the International Cancer Genome Consortium20. The sequencing and variant calling (SNVs, SVs, CNVs) was performed consistently with the 1000 Genomes Project (PGCBHM group) were obtained from PedBrain Tumor Project of the International Cancer Genome Consortium20. The processing and variant calling (SNVs, SVs, CNVs) was performed consistently with the GCTB data as described above.
NCBI build 37.1) using Bowtie2 (v. 2.3.0)65. The non-default parameters -q 20 -s were used. PCR duplicates were removed by Picard MarkDuplicates (v. 1.125). Signal tracks were calculated using deepTools. CLE-based ATAC-seq data processing workflow is available at https://github.com/CompEpigen/ATACSeq_workflows. Peaks were called using Macs2 (v. 2.1.1)67 with the parameters --nomodel -shift -50 -extsize 100 -gvalue 0.01. All peaks were split to create a common bed file with read counts before differential analysis using edgeR (v. 2.3.9)68. Gene annotations were made using CePReKanno (v. 3.18.0)69. Transcription binding motif analysis was performed using HOMER (v. 4.9)70. Motifs with a P-value <0.01 and a ratio of motif to background above 1.1 were defined as significantly enriched.

**ChiP-sequencing and analysis.** We used the ChiP-mentation protocol71 to map the genomic distribution and analysis of WT and G3H3.3 total, H3.3 as well as histone modifications (H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K36me3) in a subset of samples (see details in Supplementary Data 1). To confirm the specificity of the H3.3 and H3.3 G43W antibody used for ChiP-seq analysis, we performed a validation experiment with a Histone code peptide array (JPT, Berlin, Germany) containing short peptides that densely cover most of the known H3.3 and H3.3 modifications. We did not observe any significant binding of the G34W-specific antibody to H3.3 peptides and a highly specific binding of the H3.3 antibody. Cells of a control transfection were used. PCR duplicates were removed by Picard MarkDuplicates (v. 2.3.3)66. A compatible CWL-based data processing workflow for ChiP-seq data processing work was performed for 10 cycles of PCR using KAPA HiFi Uracil- DNA Polymerase with customized primer (Supplementary Data 5) and an annealing temperature of 69 °C according to the settings for the PE libraries in the technical Data Sheet (KAPA HiFi HotStart Uracil – Ready Mix, KR0413–version 1.12, pglab). Amplified libraries were purified with 1x Agencourt AMPure XP beads (Beckman Coulter Inc.) and quantified using Qubit fluorometer. The final libraries were pooled and clustered on the cBOT (illumina) according to the manufacturer’s instructions with a final concentration of 250 pm, spiked with 5% PhiX control v3. Paired-end bp sequencing on HiSeq X was performed using standard Illumina protocols. Biological samples having at least 3 rpm, a minimum length of 50 bp and a Benjamin-Hochberg corrected P value >0.05 were selected. All DMRs were filtered requiring a minimal mean methylation-value difference of 0.1.

**HumanMethylation450 and HumanMethylationEPIC analysis.** Methylation analysis using HumanMethylation450 arrays was performed by the Genomics and Proteomics Core Facility according to the manufacturer’s instructions. Profiling of isogenic HeLa cell lines with HumanMethylation450 arrays was conducted at Korean Cancer Center according to the manufacturer’s instructions. Unnormalized signals (IDAT files) were loaded into R using Bioconductor software (v. 2.2.0)72 and subjected to preprocessing with default optimization settings. 10,000 sites most variable across the sequencing results is given in Supplementary Data 1. Raw reads were processed using Trimmomatic (v. 0.36)73 and aligned against reference sequence of the Genome Research Consortium (v. 37) using bwa mem (https://github.com/brentp/bwa-meth) wrapping bwa mem (v. 0.7.8)74 with default parameters, except for the maximum separation of fragments allowed (v. 0.7). After alignment, peaks were marked using EpiCypher (v. 1.125). Methylation calling was performed with MethyDackel (v. 3.0.0). A compatible CWL-based data processing workflow for bisulite sequencing is available at https://github.com/CompEpigen/WGBS_workflows (BWA-meth_start_with_trimmed.cwl). Subsequently and prior to DMR calling, Bsmoother was used (v. 1.4)0 with default parameters to smooth the processed methylation profiles in all samples74. We then used we DSS (v. 2.27.0)75 to call DMRs for pairwise comparison between H3.3 WT and H3.3 M7C cells. Regions with at least 3 Cpgs, a minimum length of 50 bp and a Benjamin-Hochberg corrected P value >0.05 were selected. All DMRs were filtered requiring a minimal mean methylation-value difference of 0.1.

**RNA-sequencing and analysis.** Poly-A RNA sequencing of GCTB samples was performed according to the standard protocol77. Total RNA was prepared for each cell line by using the RNeasy Mini Kit (Qiagen, Hilden, Germany) and library preparation was done using TruSeq Stranded mRNA Kit (Illumina), according the manufacturer’s instructions. Paired-end 125 bp sequencing runs were performed on Illumina HiSeq2000 v4 machines. Raw sequence reads were preprocessed using cutadapt (v. 1.10)78 to remove sequence primers and adapters. Reads were aligned to the GRCh37 human reference genome with HISAT2 (v. 2.0.4)79 with additional non-default parameters –max-intronlen 20000 –no-unal –ds. Transcripts were assembled and quantified with StringTie (v. 1.3.3)79 with the GRCH37 transcript database. The RNA-seq processing workflow is available from https://github.com/CompEpigen/RNASeq_GCTB. Differential expression analysis was performed using DESeq2 (v. 1.8.1)80. Genes were called differentially expressed at FDR 0.05 and the absolute log-fold difference of greater or equal to one.

**RNA-Seq of the differentiation samples.** RNA-seq of the MSC differentiation samples was performed according to the following protocol. 50,000 cells were plated in a 24 well plate in medium: MesenPDS™ (Thermo Fisher Scientific, Massachusetts, USA). After 3 days, medium was replaced with osteogenic differentiation medium: DMEM high glucose supplemented with 10% FBS, 1x non-essential amino acids (NEAA), 2 mM L-glutamine, 0.28 mM ascorbic acid, 10 mM β glycerophosphate, and 10 mM dexamethasone (Sigma-Aldrich, USA). At each time point: 0, 0.5, 1, 2, 4, 8, 12, 16, 24, 48, 72, 125, 250, 500, 1000, and 3004 h during osteogenisis, total RNA was isolated using TRIzol (Invitrogen) Life Technologies, Carlsbad, USA) with the Direct-zol RNA kit (ZymoResearch, USA) according to the manufacturer’s instruction. RNA-seq library preparation was carried out using the NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Multiplex Ultra Directional RNA Library Prep Kit (New England Biolabs, USA) according to the manufacturer’s instruction. The quantity and quality of the cDNA library were assessed using the Agilent 2200 Tapestation

**Calling of H3.3-G34W enriched regions.** We used the Poisson test-based binarization module of the ChromHMM software (v. 1.18)72 to generate 200 bp windows with statistically significant enrichment of the wild-type H3.3 or H3.3-G34W compared to the wild-type background. Adjacent windows were merged to generate primary enrichment regions. The regions were filtered against a union of the ENCODE ChiP-seq blacklists.

Whole-genome bisulfite sequencing and analysis. WGBS Libraries were prepared using the TruSeq DNA PCR-Free LT Library Preparation Kits with partially modified steps in fragmentation and bead clean up size selection. Briefly, 2 µg genomic DNA (diluted in nuclease-free water) was fragmented to 150–200 bp using a Covaris ultrasonicator (Covaris, Inc.) and quality checked using Agilent TapeStation (Agilent Technologies). The fragmented DNA sample was diluted with water and split into two aliquots of ~ 1 µg in 50 µl, end-repaired and purified using 1X salmon DNA fragmentation buffer with 600 ng of ends and ligated. WGBS LT adapters were performed as described in manufacturers protocol. Then, adapter-ligated fragment libraries were treated with bisulfitie using the EpiTect Bisulfite Kit (Qagen) following the instructions in the Illumina WGBS for Methylation Analysis Guide. After bisulfitie conversion, fragment libraries were sequenced with 8 cycles of PCR using KAPA HiFi Uracil– DNA Polymerase with customized primer (Supplementary Data 5) and an annealing temperature of 69 °C according to the settings for PE libraries in the technical Data Sheet (KAPA HiFi HotStart Uracil – Ready Mix, KR0413–version 1.12, pglab). Amplified libraries were purified with 1X Agencourt AMPure XP beads (Beckman Coulter Inc.) and quantified using Qubit fluorometer. The final libraries were pooled and clustered on the cBOT (illumina) according to the manufacturer’s instructions with a final concentration of 250 pm, spiked with 5% PhiX control v3. Paired-end bp sequencing on HiSeq X was performed using standard Illumina protocols. Biological samples having at least 3 rpm, a minimum length of 50 bp and a Benjamin-Hochberg corrected P value >0.05 were selected. All DMRs were filtered requiring a minimal mean methylation-value difference of 0.1.

**Sample preparation for osteogenisis.** After osteogenesis, total RNA was isolated using TRIzol (Invitrogen) Life Technologies, Carlsbad, USA) with the Direct-zol RNA kit (ZymoResearch, USA) according to the manufacturer’s instruction. RNA-seq library preparation was carried out using the NEBNext Poly(A) mRNA Magnetic Isolation Module and NEBNext Multiplex Ultra Directional RNA Library Prep Kit (New England Biolabs, USA) according to the manufacturer’s instruction. The quantity and quality of the cDNA library were assessed using the Agilent 2200 Tapestation
Gene and genomic feature annotations. Unless specified otherwise, Ensembl transcript and gene annotations were used for the GRCh37 assembly (build 87). List of bivalent genes in H3.3 WT was compiled by overlapping the consensus H3.3 WT peak regions of H3K27me3 and H3K4me3 within 2 kb from a RefSeq TSS. A consensus list of bivalent genes in human ESCs was obtained from ref. 81. A list of PRC2 target genes was found in ref. 82. A comprehensive list of imprinted genes was obtained from ref. 83. Replication timing domains for the annotation of LMDs originates from Repli-Seq data of ref. 84, processed and publicly deposited by RepliScan package85. MSCI-specific chromatin states were taken from the 15-state ChromHMM model21 for bone-marrow derived MSCs generated by the Roadmap Epigenomics consortium21 (sample E026).

Identification of large-scale methylation domains. DNA methylation data was summarized in 20 kb tiling windows to eliminate the small-scale variability (e.g., related to CpG islands). The changepoints were then called using R package changepoint (v. 2.2.2). DNA methylation data was summarized in the obtained segments and the latter were clustered using standard hierarchical clustering resulting in 6 stable clusters. After clustering adjacent segments that belonged to the same cluster were merged. Two smallest clusters were removed since one contained less than 10 segments and the other one exclusively Y-chromosome segments. R code for LMD identification is available from https://github.com/CompEpigen/lutsik/CP-LMDs. For gene density estimation UCSC hg19 human gene annotation (R package RDAhapiens.UCSC.hg19.knownGene) were used and for each LMD segment the number of overlapping genes was divided by its length in megabases (Mb).

Genomic overlap enrichment analysis. We tested the significance of overlap of differential ATAC-seq peaks, H3.3 G34W incorporation regions and other regions of interest with publicly available genomic annotations using LOLA (v. 1.8.0)86. In the case of ATAC peaks a union of all called peaks in H3.3 WT and H3.3 MUT was used as background to test enrichments at H3.3 G34W gained and lost peaks relative to each other. For H3.3 WT and H3.4 G4W enrichment regions the background set consisted of regions called for wild-type H3.3 in H3.3 WT and H3.3 MUT groups, as well as the H3.3-G34W regions. We used the LOLA core database for most of the analyses. Enrichment of repeat elements was based on a custom LOLA database created using the UCSC Repeat Masker track (http://www.repeatmasker.org).

Genome set overrepresentation and enrichment analysis. We used gene sets from the Molecular Signatures Database v. 6.2 (MsigDB)87 to test for the overrepresentation of DEGs or genes associated with differential ATAC peaks, and the gene set enrichment analysis (GSEA) of DEGs. Overrepresentation analysis was performed with the help of R package GeneOverlap (v. 1.14.0). GSEA was performed with gsea package (v. 1.4.1)88 using 10,000 permutations.

References
1. Mohammad, F. & Helin, K. Oncohistones: drivers of pediatric cancers. Genes Dev. 31, 2313–2324 (2017).
2. Nacev, B. A. et al. The expanding landscape of ‘oncohistones’ mutational load in human cancers. Nature 567, 473–478 (2019).
3. Bennett, R. L. et al. A mutation in histone H2B represents a new class of oncogenic driver. Cancer Discov. 9, 1438–1451 (2019).
4. Szenker, E., Ray-Gallet, D. & Almouzni, G. The double face of the histone H3.3–H4.1 core database. https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentation/ChromModels/coreMarks/jointModel_filed/E026_15_coreMarks_segments.bed. Replication timing segments by Replicase project (Hansen et al., 2010; https://de.cyverse.org/anon/anon/home/plантеs/gundhals/pinsahan/2010_replicate/repliscan_50khg.txt).

Received: 9 April 2020; Accepted: 16 September 2020; Published online: 27 October 2020

Data availability
All raw patient-derived sequencing data from WGS, WGBS, ATAC-seq, ChiP-seq, RNA-seq and deep targeted resequencing have been deposited in the European Genome-phenome Archive (EGA) under restricted access with the accession code: EGAS00001083730. Processed sequencing data and microarray data have been deposited in ArrayExpress with the accession codes: E-MTAB-7184 (ChiP-Seq), E-MTAB-9512 (ATAC-Seq), E-MTAB-9513 (WGBS), E-MTAB-9515 (RNA-Seq). RNA-seq data of the MSC osteogenic differentiation experiment are deposited at the Gene Expression Omnibus ( GEO) with the accession code: GSE129036. We furthermore used gene sets from the Molecular Signatures Database v. 6.2 (MsigDB): https://www.gsea-msigdb.org/gsea/msigdb/download_files.jsp?filePath=mgdbs/releases/6.2/migdb_v6.2_files_to_download_locally.zip, ChromHMM states of human bone-marrow by ENCODE project (sample E026): https://egg2.west.wustl.edu/roadmap/data/byFileType/chromhmmSegmentation/ChromModels/coreMarks/jointModel_filed/E026_15_coreMarks_segments.bed, Repeat classification segments by Replicase project (Hansen et al., 2010; https://de.cyverse.org/anon/anon/home/plантеs/gundhals/pinsahan/2010_replicate/repliscan_50khg.txt).

Consensus list of ESC-specific bivalent genes41: http://www.oncotarget.com/index.php?journal=oncotarget&tpage=article&option=download&SuppFilePath=%20137146pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D%20137146%20pdf%3D/ftp.1000genomes.ebi.ac.uk/vol1/ftp/1410502/GRCh37 transcript database ftp://ftp.ncbi.nlm.nih.gov/refsSeq/H_sapiens/annotation/GRCh37_37/refsSeq_identifiers/GRCh37_37_latest_genomic.gff.gz. Repeat masker database: http://www.repeatmasker.org. Source data are provided with this paper.
13. Lewis, P. W. et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science 340, 857–861 (2013).

14. Bhandari A. et al. H3K27me3 and DNA hypomethylation are major drivers of gene expression in K27M mutant pediatric high-grade gliomas. Cancer Cell 24, 660–672 (2013).

15. Lu, C. et al. Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. Science 352, 844–849 (2016).

16. Shi, L., Shi, J., Xi, L., Li, W. & Wen, H. Histone H3.3 G34 mutations alter histone H3K36 and H3K27 methylation in cis. J. Mol. Biol. 430, 1562–1565 (2018).

17. Amanatullah, D. F., Clark, T. R., Lopez, M. S., Borys, D. & Tamurian, R. M. Histone H3K36 mutations promote sarcomagenesis through the histone reader DPF3 and the BAF chromatin remodeling complex. J. Mol. Biol. 430, 1562–1565 (2018).

18. Mrozowsky, L. G. Genomic instability in giant cell tumor of bone. A study of 52 cases using DNA ploidy, relocation FISH, and array-CGH analysis. Genes Chromosomes Cancer 48, 468–479 (2009).

19. Lim, J. et al. The histone variant H3.3 G34W substitution in giant cell tumor of the bone link chromatin and RNA processing. Sci. Rep. 7, 13459 (2017).

20. Roadmap Epigenomics C. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330 (2015).

21. Zhou, W. et al. DNA methylation loss in late-replicating domains is linked to tumorigenesis. Proc. Natl. Acad. Sci. USA 112, 12727–12732 (2015).

22. Kieslinger, M. et al. EBF2 regulates osteoblast-dependent differentiation of Interstitial Heterochromatic Loci and Imprinted Genes. Cell Rep. 21, 316–327 (2017).

23. Huang, L., Teng, X. Y., Cheng, Y. Y., Lee, K. M. & Kumta, S. M. Expression of Hoxa-9 and Hoxa-10 in osteosarcoma. J. Orthop. Sci. 20, 134–140 (2015).

24. Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010).

25. Ziegler, K. et al. EEF1D and EEF1D2 are enriched in pediatric glioblastoma and are associated with survival. Acta Neuropathol. 129, 95–104 (2015).

26. Boyle, W. J., Simonet, W. S. & Lacey, D. L. Osteoclast differentiation and mitochondrial DNA variation and heteroplasmy in monozygotic twins clinically discordant for multiple sclerosis. Hum. Mutat. 37, 765–775 (2016).

27. Liu, L. et al. Enrichment of c-Met– ErbB2– tumorogenic stromal cells of giant cell tumor of bone and targeting by cabozantinib. Cell Death Dis. 5, e1471 (2014).

28. Kieslinger, M. et al. DNA methylation and histone H3K36 and H3K27 methylation in cis. J. Mol. Biol. 430, 1562–1565 (2018).

29. Dhayalan, A. et al. The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. J. Biol. Chem. 285, 26141–26142 (2010).

30. Souren, N. Y. et al. Mitochondrial DNA profiling of DNA methyltransferases reveals a role for DNMT3B in genic methylation. Nature 520, 243–247 (2015).

31. Hardouin, S. N., Guo, R., Romeo, P. H., Nagy, A. & Aubin, J. E. Impaired B-cell development and B-cell lymphoma in mice deficient for Igf2-P2 transcripts. Nat. Genet. 46, 912–918 (2014).

32. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 377–378 (2012).

33. McKee, M. Menosubmalignant treatment of giant cell tumour of bone. Lancet Oncol. 14, 801–802 (2013).

34. Boyle, W. J., Simonet, W. S. & Lacey, D. L. Osteoclast differentiation and activation. Nature 423, 337–342 (2003).

35. Kieslinger, M. et al. EEF2 regulates osteoblast-dependent differentiation of osteoclasts. Dev. Cell 9, 757–767 (2005).

36. Boyce, R. F., Yang, L. & Chen, D. Osteoprotegerin, the bone protector, is a surprising target for beta-catenin signaling. Cell Metab. 2, 344–345 (2005).

37. Wolock, S. L. et al. Mapping distinct bone marrow niche populations and their altered histone methylation landscape. Science 352, 844–849 (2016).

38. Amanatullah, D. F., Clark, T. R., Lopez, M. S., Borys, D. & Tamurian, R. M. Histone H3K36 mutations promote sarcomagenesis through the histone reader DPF3 and the BAF chromatin remodeling complex. Genes Dev. 31, 660–673 (2017).

39. Voon, H. et al. ATRX Plays a Key Role in Maintaining Silencing at Imprintable Heterochromatic Loci and Imprinted Genes. Cell Rep. 11, 405–418 (2015).

40. Hardouin, S. N., Guo, R., Romeo, P. H., Nagy, A. & Aubin, J. E. Impaired mesenchymal stem cell differentiation and osteogenesis in mice deficient for Igf2-P2 transcripts. Development 138, 203–213 (2011).

41. Xu, J. C., Wu, G. H., Zhou, L. L., Yang, X. J. & Liu, J. T. Leptin improves osteoblast differentiation of human bone marrow stroma cells. Eur. Rev. Med. Pharmacol. Sci. 20, 3507–3516 (2016).

42. Huang, L., Teng, X. Y., Cheng, Y. Y., Lee, K. M. & Kumta, S. M. Expression of proosteoblast markers and Cbfa-1 and Osterix gene transcripts in stromal tumour cells of giant cell tumour of bone. Bone 34, 393–401 (2004).

43. Lau, C. P. et al. Genome-wide transcriptome profiling of the neoplastic giant cell tumor of bone by RNA sequencing. J. Cell Biochem 118, 1349–1360 (2017).

44. Zheng, Y. et al. Molecular basis for the role of oncogenic histone mutations in modulating H3K36 methylation. Sci. Rep. 7, 43906 (2017).

45. Sturm, D. et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell 22, 425–437 (2012).

46. de la Rica, L. et al. PU.1 target genes undergo Fe2-coupled demethylation and DNMT3b-mediated methylation in monocyte-to-osteoclast differentiation. Genome Biol. 14, R99 (2013).

47. Heinz, S. et al. Osteogenesis depends on commissioning of a network of stem cell transcription factors that act as reproasers of adipogenesis. Nat. Genet. 51, 716–727 (2019).

48. Zhu, S. et al. Sabath, A., Gibbons, C. L., Giele, H. & Athanasou, N. Phenotypic and molecular studies of giant-cell tumours of bone and soft tissue. Hum. Pathol. 36, 945–954 (2005).

49. Murata, A., Fujita, T., Kawahara, N., Tsuhiya, H. & Tomita, K. Osteoblast lineage properties in giant cell tumours of bone. J. Orthop. Sci. 10, 581–588 (2005).

50. Chellini, C. et al. Stathmin-like 2, a developmentally-associated neuronal marker, is expressed and modulated during osteogenesis of human mesenchymal stem cells. Biochem. Biophys. Res. Commun. 374, 64–68 (2008).

51. Dong, H. et al. Global H3.K36 methyltransferase is necessary for murine osteoblast differentiation. J. Bone Miner. Res. 26, 1537–1548 (2011).
Acknowledgements

We acknowledge the High-throughput Sequencing Unit (head Stephen Wolf) of the Genomics and Proteomics Core Facility of the German Cancer Research Center (head Stephan Wiestmann) for their assistance with the high-throughput sequencing, and the Omics and Data Management Core Facility (heads Jürgen Edwards, Christian Lawerenz and Ivo Buchhalter) for the help with processing and management of the sequencing data. We are thankful to Christa Winkelmann (Institute of Pathology, Erlangen) for excellent preparation of immunostainings, Cansu Cirzi for the help with ChIP-mentation and Ana Banito for a thoughtful feedback about the manuscript. This work was in part supported by Funds from the Helmholtz Association and the Korean National Cancer Center (Grant No. 1810050-1 to A.M.L.), the Korean National Research Fund (Grant No. 2017R1E1A1A01073670 to A.M.L.), Oxford NIHR BRC (U.O.), Arthritis Research UK (program grant 20522 to U.O.), P.L. and A.B. are DKFZ Postdoctoral Fellows. P.L. was supported by the Helmholtz Association AMPro Project (ZT00026 to C.P.). A.K., A.M., T., and D.M. are fellows of the Helmholtz International Graduate School for Cancer Research. J.H. is fellow of the German-Israeli Helmholtz Research School in Cancer Biology, C.J. is a recipient of a PhD fellowship from the Royal Commission for the Exhibition 1851, China-Oxford Scholarship Fund and UCB-Oxford Fellowship.

Author contributions
C.P. and A.L. supervised the study. F.H., J.Z., and J.F. provided GCTB samples and performed their initial characterization. A.B., S.O., D.M., J.F., A. Jau., A.K., M.H., J.L., M. Schl., M.Schuh., A.J., D.V., F.G., S.H., V.H.N., S.L., J.H.P., Y.J.P., D.K., C.J., F.R., and D.W. performed experimental work and interpretation of experimental results. P.L., S.O., A.M., U.T., R.T., J.H., K.B., C.J., U.O. performed bioinformatical data analysis and interpretation. P.L., A.B., S.O., D.M., A.K., C.P., and A.L. wrote the manuscript.

Funding
Open Access funding enabled and organized by Projekt DEAL.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-18955-y.

Correspondence and requests for materials should be addressed to A.M.L. or C.P.

Peer review information Nature Communications thanks Xiaobing Shi, Eleni Tomazou and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s) 2020