PGBD5 promotes site-specific oncogenic mutations in human tumors

Anton G Henssen,1,2,14 Richard Koche,2,22 Jiali Zhuang,3,22 Eileen Jiang,1 Casie Reed,1 Amy Eisenberg,1 Eric Still,1 Ian C MacArthur,1 Elias Rodríguez-Fos,4 Santiago Gonzalez,4 Montserrat Puiggròs,4 Andrew N Blackford,5 Christopher E Mason,6 Elisa de Stanchina,7 Mihat Gönen,8 Anne-Katrin Emde,9 Minita Shah,9 Kanika Arora,9 Catherine Reeves,9 Nicholas D Socci,10 Elizabeth Perlman,11 Cristina R Antonescu,12 Charles W M Roberts,13 Hanno Steen,14 Elizabeth Mullen,15 Stephen P Jackson,5,16,17,21 David Torrents,1,18 Zhiping Weng,3 Scott A Armstrong,2,19,20 & Alex Kentsis1,19,20

Genomic rearrangements are a hallmark of human cancers. Here, we identify the piggyBac transposable element derived 5 (PGBD5) gene as encoding an active DNA transposase expressed in the majority of childhood solid tumors, including lethal rhabdoid tumors. Using assembly-based whole-genome DNA sequencing, we found previously undefined genomic rearrangements in human rhabdoid tumors. These rearrangements involved PGBD5-specific signal (PSS) sequences at their breakpoints and recurrently inactivated tumor-suppressor genes. PGBD5 was physically associated with genomic PSS sequences that were also sufficient to mediate PGBD5-induced DNA rearrangements in rhabdoid tumor cells. Ectopic expression of PGBD5 in primary immortalized human cells was sufficient to promote cell transformation in vivo. This activity required specific catalytic residues in the PGBD5 transposase domain as well as end-joining DNA repair and induced structural rearrangements with PSS breakpoints. These results define PGBD5 as an oncogenic mutator and provide a plausible mechanism for site-specific DNA rearrangements in childhood and adult solid tumors.

Whole-genome analyses have now produced near-comprehensive topographies of coding mutations for certain human cancers, thus enabling detailed molecular studies of cancerogenesis and providing potential for precisely targeted therapies1–5. For certain childhood cancers, recent studies have begun to identify the essential functions of complex noncoding structural variants that induce aberrant expression of cellular proto-oncogenes6–7. However, for many aggressive childhood cancers, including solid tumors, such studies have identified distinct cancer subtypes that have no discernible coding mutations8–11. In addition, whereas defects in DNA-damage repair have been suggested to explain the increased incidence of some cancers in relatively young people, the causes of complex genomic rearrangements in other cancers in young children without apparent widespread genomic instability remain largely unknown.

Rhabdoid tumors are a prototypical example of this phenomenon. These tumors occur in the developing tissues of infants and children, and exhibit neuroectodermal, epithelial, and mesenchymal components in the brain, liver, kidney, and other organs10,12,13. Rhabdoid tumors that cannot be cured through surgery are generally chemotherapy resistant and are almost always lethal14. Rhabdoid tumors exhibit inactivating mutations of SMARCB1, generally as a result of genomic rearrangements of the 22q11.2 chromosomal locus15. These mutations may be inherited as part of the rhabdoid tumor predisposition syndrome but are not thought to involve chromosomal instability13. Whereas SMARCB1 mutations are sufficient to cause rhabdoid tumors in mice16, human rhabdoid tumors have been observed to have multiple molecular subtypes and rearrangements of additional chromosomal loci that are poorly understood10,17,18. These findings...

1Molecular Pharmacology Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, USA. 2Cancer Biology & Genetics Program, Sloan Kettering Institute, Memorial Sloan Kettering Cancer Center, New York, New York, USA. 3Program in Bioinformatics and Integrative Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, Massachusetts, USA. 4Joint BSC-CRG-IRB Research Program in Computational Biology, Barcelona Supercomputing Center (BSC-CNS), Barcelona, Spain. 5The Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge, UK. 6Institute for Computational Biomedicine, Weill Cornell Medical College, New York, New York, USA. 7Antitumor Assessment Core Facility, Memorial Sloan Kettering Cancer Center, New York, New York, USA. 8Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. 9New York Genome Center, New York, New York, USA. 10Bioinformatics Core, Memorial Sloan Kettering Cancer Center, New York, New York, USA. 11Northwestern University Feinberg School of Medicine, Ann & Robert H. Lurie Children’s Hospital of Chicago, Chicago, Illinois, USA. 12Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, New York, USA. 13Department of Oncology, St. Jude Children’s Research Hospital, Memphis, Tennessee, USA. 14Department of Pathology, Boston Children’s Hospital, Boston, Massachusetts, USA. 15Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, Massachusetts, USA. 16Department of Biochemistry, University of Cambridge, Cambridge, UK. 17The Wellcome Trust Sanger Institute, Hinxton, Cambridge, UK. 18Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain. 19Weill Cornell Medical College, Cornell University, New York, New York, USA. 20Department of Pediatrics, Memorial Sloan Kettering Cancer Center, New York, New York, USA. 21Present address: Department of Pediatric Oncology/Hematology, Charité-Universitätsmedizin Berlin, Berlin, Germany.

© 2017 Nature America, Inc., part of Springer Nature. All rights reserved.
suggest that additional genetic elements and molecular mechanisms may contribute to the pathogenesis of rhabdoid tumors.

In humans, nearly half of the genome comprises sequences derived from transposons, including both autonomous and nonautonomous mobile genetic elements\(^1\). Most human genes encoding enzymes that might mobilize transposons appear to be catalytically inactive, with the exception of L1 long interspersed repeated sequences, which appear to induce structural genomic variation in human neurons and adenocarcinomas\(^2\)–\(^4\); Mariner transposase-derived SETMAR, which functions in DNA repair\(^5\); and Transib-like DNA transposase RAG1/2, which catalyzes somatic recombination of V(D)J receptor genes in lymphocytes\(^6\). In particular, aberrant activity of RAG1/2 in lymphoblastic leukemias and lymphomas can induce the formation of chromosomal translocations that generate transforming fusion genes\(^7\)–\(^9\). The identities of similar genomic rearrangements and the mechanisms by which they may be formed in childhood and adult solid tumors are unknown, but the existence of additional human recombinases that can induce somatic-DNA rearrangements has long been hypothesized\(^10\).

Recently, human PGBD5 and THAP9 have been found to catalyze transposition of synthetic DNA transposons in human cells\(^11\)–\(^13\). The physiologic functions of these activities are currently not known. PGBD5 is distinguished by its deep evolutionary conservation among vertebrates (~500 million years) and developmentally restricted expression in tissues from which several childhood solid tumors, including rhabdoid tumors, are thought to originate\(^14\)–\(^16\). PGBD5 is transcribed as a multi-exonic noncoding transcript from a gene encoding a full-length transposase that has become immobilized on human chromosome 1 (refs. 30,31). Genomic transposition activity of PGBD5 requires distinct aspartate residues in its transposase domain as well as specific DNA sequences containing inverted terminal repeats similar to those of piggyBac transposons from the lepidopteran Trichoplusia ni\(^30\). These findings, combined with the recent evidence that PGBD5 can induce genomic rearrangements that inactivate the HPRT1 gene\(^32\), prompted us to investigate whether PGBD5 might induce site-specific DNA rearrangements in human rhabdoid tumors that share developmental origin with cells that normally express PGBD5.

RESULTS

Human rhabdoid tumors exhibit genomic rearrangements associated with PGBD5-specific signal-sequence breakpoints

First, we analyzed the expression of PGBD5 in large well-characterized cohorts of primary childhood and adult tumors (Supplementary Fig. 1a). We observed that PGBD5 was highly expressed in a variety of childhood and adult solid tumors, including rhabdoid tumors, but not in acute lymphoblastic or myeloid leukemias (Supplementary Fig. 1a). The expression of PGBD5 in rhabdoid tumors was similar to that in the embryonal tissues from which these tumors are thought to originate, but it was not significantly associated with currently defined molecular subgroups or patient age at diagnosis (Supplementary Fig. 1a–f).

To investigate potential PGBD5-induced genomic rearrangements in primary human rhabdoid tumors, we performed de novo structural-variant analysis of whole-genome paired-end Illumina sequencing data for 31 individually matched tumors and normal paired blood specimens from children with extracranial rhabdoid tumors, which are generally characterized on the basis of inactivating SMARCB1 mutations\(^16\). Owing to their repetitive nature, sequences derived from transposons present challenges to genome analysis. Thus, we reasoned that genome analysis approaches that do not rely on short-read alignment algorithms, such as the local assembly-based algorithm laSv and the tree-based sequence-comparison algorithm SMuFin, might identify genomic rearrangements that otherwise might escape conventional algorithms\(^33,34\).

Using this assembly-based approach, we observed recurrent rearrangements of the SMARCB1 gene on chromosome 22q11 in nearly all cases examined, in agreement with the established pathogenic function of inactivating mutations of SMARCB1 in rhabdoid tumor originesis (Fig. 1a). In addition, we observed previously unrecognized somatic deletions, inversions, and translocations involving focal regions of chromosomes 1, 4, 5, 10, and 15 (median of three per tumor), which were recurrently altered in more than 20% of cases (Fig. 1a and Supplementary Data Set 1). These results indicated that, in addition to the pathogenomic mutations of SMARCB1, human rhabdoid tumors are characterized by additional distinct and recurrent genomic rearrangements.

To determine whether any of the observed genomic rearrangements might be related to PGBD5 DNA transposase or recombine activity, we first used a forward genetic screen to identify PSS sequences that were specifically found at the breakpoints of PGBD5-induced deletions, inversions, and translocations that caused inactivation of the HPRT1 gene in a thioguanine resistance assay\(^35\). Using these PSS sequences as templates for supervised analysis of the somatic genomic rearrangements in primary human rhabdoid tumors, we identified specific PSS sequences associated with the breakpoints of genomic rearrangements in rhabdoid tumors (P = 1.1 × 10\(^{-10}\)), hyper-geometric test; Fig. 1b and Supplementary Fig. 2). By contrast, we observed no enrichment of the RAG1/2-recombination signal (RSS) sequences at the breakpoints of somatic rhabdoid tumor genomic rearrangements, although the RSS and PSS sequences were equally sized, a result consistent with the lack of RAG1/2 expression in rhabdoid tumors. Likewise, we did not find significant enrichment of PSS motifs at the breakpoints of structural variants and genomic rearrangements in breast carcinomas lacking PGBD5 expression, even though these breast carcinoma genomes were characterized by high rates of genomic instability (Supplementary Data Set 1). The PSS sequences observed in human rhabdoid tumors exhibited both similarities and differences as compared with those found in the forward genetic screen (Supplementary Fig. 2), thus suggesting that context-dependent factors may control PGBD5 activity.

Overall, the majority of the observed rearrangements were deletions and translocations (Fig. 1a and Supplementary Fig. 3a). Notably, we found recurrent PSS-containing genomic rearrangements affecting the CNTNAP2, TENM2, TENM3, and TET2 genes (Fig. 1a–c, Supplementary Fig. 3c and Supplementary Data Set 1). Using allele-specific PCR followed by Sanger DNA sequencing, we confirmed three of the observed intragenic CNTNAP2 deletions and rearrangement breakpoints (Fig. 1c). Likewise, we confirmed the somatic nature of mutations of CNTNAP2 and TENM3 by allele-specific PCR in matched tumor and normal primary patient specimens (Supplementary Fig. 3d–h).

CNTNAP2, a member of the neurexin family of signaling and adhesion molecules, has previously been found to function as a tumor-suppressor gene in gliomas\(^36\). In agreement with the potential pathogenic functions of the apparent CNTNAP2 rearrangements in rhabdoid tumors identified in our analysis, CNTNAP2 has recently been reported to be recurrently deleted in an independent cohort of rhabdoid tumor patients\(^18\). By using comparative RNA-sequencing gene expression analysis in our cohort, we found that primary tumors
containing recurrent genomic rearrangements of CNTNAP2, as compared with those lacking CNTNAP2 rearrangements, were indeed associated with a significant decrease in CNTNAP2 mRNA expression (∆P = 0.017, t-test; Fig. 1d). Additional mechanisms, including as-yet-undetected mutations or silencing, may contribute to the loss of CNTNAP2 expression in apparently nonrearranged cases (Fig. 1d).

Interestingly, some of the observed genomic rearrangements with PSS-containing breakpoints in rhabdoid tumors involved SMARCBI deletions (Fig. 1a,b and Supplementary Data Set 1), thus suggesting that in a subset of rhabdoid tumors, PGBD5 activity itself may contribute to the somatic inactivation of SMARCBI in rhabdoid tumorigenesis. Similarly, we observed recurrent interchromosomal translocations and complex structural rearrangements containing breakpoints with the PSS motifs that involved SMARCBI (Fig. 1b and Supplementary Data Set 1), including chromosomal translocations, as previously observed through cytogenetic methods. For example, we verified the t(5;22) translocation by using allele-specific PCR followed by Sanger sequencing of the translocation breakpoint (Supplementary Fig. 3i,j). Together, these results indicated that human rhabdoid tumors exhibit recurrent genomic rearrangements that are defined by PSS breakpoint sequences specifically associated with PGBD5, at least some of which appear to be pathogenic and may be coupled with inactivating mutations of SMARCBI itself.

PGBD5 is physically associated with human genomic PSS sequences that are sufficient to mediate DNA rearrangements in rhabdoid tumor cells

In prior studies, human PGBD5 has been found to localize to cell nuclei. To test whether PGBD5 in rhabdoid tumor cells is physically associated with genomic PSS-containing sequences, as would be predicted for a DNA transposase that induces genomic rearrangements, we used chromatin immunoprecipitation followed by DNA sequencing (ChiP-seq) to determine the genomic localization of endogenous PGBD5 in human G401 rhabdoid tumor cells. We observed that human DNA regions bound by PGBD5 were significantly enriched in PSS motifs (P = 2.9 × 10−20, hypergeometric test), in contrast to scrambled PSS sequences of identical composition or functionally unrelated RSS sequences of equal size, neither of which showed significant enrichment (P = 0.28 and 1.0, respectively, hypergeometric test; Fig. 2a).

To test the hypothesis that PGBD5 can act directly on human PSS-containing DNA sequences and mediate their genomic rearrangement, we used the previously established DNA transposition reporter assay. Human embryonic kidney (HEK) 293 cells were transiently transfected with plasmids for expression of human GFP-PGBD5, hyperactive lepidopteran T. ni GFP-piggyBac DNA transposase or control GFP in the presence of reporter plasmids for expression of the

Figure 1 Human rhabdoid tumors exhibit genomic rearrangements associated with PGBD5-specific signal-sequence breakpoints. (a) Aggregate Circos plot of somatic structural variants identified in 31 human rhabdoid tumors by using laSV, as marked for PSS-containing breakpoints (outer ring, structural rearrangements; red outline, rearrangements occurring in ≥3 of 31 samples and highlighted in red for rearrangements with recurrence frequency >13%). (b) Recurrently rearranged genes are labeled. (c) Aggregates of all CNVs and structural variants in 31 human rhabdoid tumors are depicted. Recurrently rearranged CNVs (inner line, as color-coded), translocations and complex structural rearrangements containing recurrent genomic rearrangements of CNTNAP2, as compared with those lacking CNTNAP2 rearrangements, were indeed associated with a significant decrease in CNTNAP2 mRNA expression (∆P = 0.017, t-test; Fig. 1d). Additional mechanisms, including as-yet-undetected mutations or silencing, may contribute to the loss of CNTNAP2 expression in apparently nonrearranged cases (Fig. 1d).

Interestedly, some of the observed genomic rearrangements with PSS-containing breakpoints in rhabdoid tumors involved SMARCBI deletions (Fig. 1a,b and Supplementary Data Set 1), thus suggesting that in a subset of rhabdoid tumors, PGBD5 activity itself may contribute to the somatic inactivation of SMARCBI in rhabdoid tumorigenesis. Similarly, we observed recurrent interchromosomal translocations and complex structural rearrangements containing breakpoints with the PSS motifs that involved SMARCBI (Fig. 1b and Supplementary Data Set 1), including chromosomal translocations, as previously observed through cytogenetic methods. For example, we verified the t(5;22) translocation by using allele-specific PCR followed by Sanger sequencing of the translocation breakpoint (Supplementary Fig. 3i,j). Together, these results indicated that human rhabdoid tumors exhibit recurrent genomic rearrangements that are defined by PSS breakpoint sequences specifically associated with PGBD5, at least some of which appear to be pathogenic and may be coupled with inactivating mutations of SMARCBI itself.

PGBD5 is physically associated with human genomic PSS sequences that are sufficient to mediate DNA rearrangements in rhabdoid tumor cells

In prior studies, human PGBD5 has been found to localize to cell nuclei. To test whether PGBD5 in rhabdoid tumor cells is physically associated with genomic PSS-containing sequences, as would be predicted for a DNA transposase that induces genomic rearrangements, we used chromatin immunoprecipitation followed by DNA sequencing (ChiP-seq) to determine the genomic localization of endogenous PGBD5 in human G401 rhabdoid tumor cells. We observed that human DNA regions bound by PGBD5 were significantly enriched in PSS motifs (P = 2.9 × 10−20, hypergeometric test), in contrast to scrambled PSS sequences of identical composition or functionally unrelated RSS sequences of equal size, neither of which showed significant enrichment (P = 0.28 and 1.0, respectively, hypergeometric test; Fig. 2a).

To test the hypothesis that PGBD5 can act directly on human PSS-containing DNA sequences and mediate their genomic rearrangement, we used the previously established DNA transposition reporter assay. Human embryonic kidney (HEK) 293 cells were transiently transfected with plasmids for expression of human GFP-PGBD5, hyperactive lepidopteran T. ni GFP-piggyBac DNA transposase or control GFP in the presence of reporter plasmids for expression of the
neomycin-resistance gene (NeoR) flanked by a human PSS sequence, as identified from rhabdoid tumor rearrangement breakpoints (Supplementary Figs. 2 and 3 and Supplementary Data Set 1), lepidopteran piggyBac inverted terminal repeat (ITR) transposon sequence\(^3\), or control plasmids lacking flanking transposon elements (Fig. 2b). Clonogenic assays of transfected cells in the presence of G418 to select for neomycin-resistant cells with genomic reporter integration demonstrated that GFP–PGBD5, but not control GFP, exhibited efficient activity toward reporters containing terminal repeats with the human PSS sequences but not control reporters lacking flanking transposon elements (\(P = 5.0 \times 10^{-5}\), t-test; Fig. 2c,d). This activity was specific, because the lepidopteran GFP-piggyBac DNA transposase, which efficiently mobilizes its own piggyBac transposons, did not mobilize reporter plasmids containing human PSS sequences (Fig. 2c,d).

To determine whether endogenous PGBD5 can mediate genomic rearrangements in rhabdoid cells, we transiently transfected human G401 rhabdoid cells with the NeoR transposon reporter plasmids and determined their chromosomal integration by using flanking-sequence exponential anchored (FLEA) PCR to amplify and sequence-specific segments of the human genome flanking transposon-integration sites\(^30\) (Fig. 2e and Supplementary Fig. 4). Similar assays in HEK293 cells lacking PGBD5 expression did not induce measurable genomic integration of reporter transposons (Fig. 2c,d). In contrast, we observed that endogenous PGBD5 in G401 rhabdoid tumor cells was sufficient to mediate integration of transposon-containing DNA into human genomic PSS-containing sites (Fig. 2f and Supplementary Tables 1 and 2). This activity was specifically observed for transposon reporters with intact transposons but not those in which the essential 5’-GGCTAAACCC-3’ hairpin structure was mutated to 5’-ATATTAACCC-3’ (location of mutation underlined; Supplementary Table 1). Thus, PGBD5 physically associates with human genomic PSS sequences that are sufficient to mediate DNA rearrangements of synthetic reporters in rhabdoid tumor cells.

**PGBD5 expression in genomically stable primary human cells is sufficient to induce malignant transformation in vitro and in vivo**

Recurrent somatic genomic rearrangements in primary rhabdoid tumors associated with PGBD5-specific signal-sequence breakpoints, their targeting of tumor-suppressor genes, and their specific activity as genomic-rearrangement substrates suggest that PGBD5 DNA transposase activity might be sufficient to induce tumorigenic mutations...
that contribute to malignant cell transformation. To determine whether PGBD5 can act as a human-cell-transforming factor, we used established transformation assays of primary human foreskin B) and retinal pigment epithelial (RPE) cells immortalized with telomerase. Primary RPE and BJ cells at passage 3–5 are immortalized by the expression of human TERT telomerase in vitro, undergo growth arrest after contact inhibition, and do not form tumors after transplantation into immunodeficient mice in vivo. Prior studies have established the essential requirements for their malignant transformation through the concomitant dysregulation of p53, Rb, and Ras pathways. Thus, transformation of primary human RPE and BJ cells enables detailed studies of human PGBD5 genetic mechanisms that cannot be performed in mouse or other heterologous model systems.

To test whether PGBD5 has transforming activity in human cells, we used lentiviral transduction to express GFP-PGBD5 and control GFP transgenes in telomerase-immortalized RPE and BJ cells, at levels 1.1- to 5-fold and 1.5- to 8-fold higher than those in primary rhabdoid tumor specimens and cell lines, respectively (Fig. 3a,b). We observed that GFP-PGBD5-expressing, but not nontransduced or GFP-expressing, RPE and BJ cells formed refractile colonies in monolayer cultures and exhibited anchorage-independent growth in semisolid cultures, a hallmark of cell transformation (Fig. 3c,d). When transplanted into immunodeficient mice, GFP-PGBD5-expressing RPE and BJ cells formed subcutaneous tumors with latency and penetrance similar to those observed in cells expressing both mutant HRAS and the SV40 large T antigen, which dysregulates both p53 and Rb pathways (Fig. 3f and Supplementary Fig. 5). Importantly, both RPE and BJ cells transformed with GFP-PGBD5 had stable diploid karyotypes when they were passaged in vitro (Supplementary Fig. 6). By contrast, expression of the distantly related lepidopteran GFP-piggyBac DNA transposase, which exerts specific and efficient transposition activity on lepidopteran piggyBac transposon sequences (Fig. 2d), did not transform RPE cells (Fig. 3e), in spite of being equally expressed (Supplementary Fig. 7a). These results indicated that the PGBD5 transposase can specifically transform human cells in the absence of chromosomal instability both in vitro and in vivo.

PGBD5-induced cell transformation requires DNA transposase activity

To test whether the cell-transforming activity of PGBD5 requires the enzymatic activity of its transposase, we used PGBD5 point mutants that are either proficient or deficient in DNA transposition reporter assays. Thus, we compared p.Glu373Ala and p.Glu365Ala PGBD5 mutants, which retain wild-type transposition activity, with p.Asp168Ala, p.Asp194Ala, and p.Asp386Ala, or their double mutant (DM) p.[A Asp194Ala]+[Asp386Ala] and triple mutant (TM) p.[A Asp168Ala]+[A Asp194Ala]+[A Asp386Ala], which occur at residues required for efficient DNA transposition in vitro, in agreement with their evolutionary conservation and putative function as the DDD/E catalytic triad for phosphodiester-bond hydrolysis. After confirming stable and equal expression of these PGBD5 mutants in RPE cells by protein blotting (Fig. 4a), we assessed their transforming activity with contact inhibition assays in monolayer cultures and transplantation into immunodeficient mice. Whereas ectopic expression of wild-type GFP-PGBD5 induced efficient and fully penetrant cell transformation, the p.Asp168Ala, p.Asp194Ala, DM, and TM deficient in transposition...
function in reporter assays did not induce contact inhibition in vitro or tumor formation in vivo (Fig. 4b,d). By contrast, transposition-proficient p.Glu373Ala and p.Glu365Ala mutants exhibited transforming activity equivalent to that of wild-type GFP-PGBD5 (Fig. 4b,d). Importantly, we confirmed that the catalytic mutants of GFP-PGBD5 on average retained their chromatin-locating proficiency compared with that of wild-type PGBD5, as assessed with ChiP−seq (Fig. 4c). Although the p.Asp368Ala mutant exhibited decreased transposition activity in reporter assays in vitro, its expression induced wild-type transposition activity in vivo (Fig. 4d). This result suggested that the transposition activity of PGBD5 might involve noncanonical DNA transposition or recombination reactions, in agreement with the dispensability of some catalytic residues for certain types of DNA transposase–induced DNA rearrangements. Thus, cell transformation induced by PGBD5 requires its nuclease activity.

**Transient expression of PGBD5 is sufficient for PGBD5-induced cell transformation**

If PGBD5 can induce transforming genomic rearrangements, then transient exposure to PGBD5 should be sufficient to heritably transform human cells. To test this prediction, we generated doxycycline-inducible PGBD5-expressing RPE cells and performed protein blotting, -2° graf

**Figure 5** Transient PGBD5 transposase expression is sufficient to transform human cells. (a) Tumor volume of RPE cells as a function of time in primary (1°, light-gray box) and secondary (2°, dark-gray box) transplants, with PGBD5 expression induced by doxycycline (black), as indicated. RPE cells were treated with doxycycline in vitro for 10 d before transplantation. Red arrowhead denotes withdrawal of doxycycline from the diet. Inset, protein blot of PGBD5 protein and actin control, in cells derived from tumors after primary transplant. (b) Representative photomicrographs of hematoxylin- and eosin-stained tumor sections from doxycycline-inducible PGBD5-expressing RPE tumors after continuous (+Dox) and discontinuous (−Dox) doxycycline treatment, with 10 mice per experimental group. (c) Protein blot of PGBD5 in G401 and A204 rhabdoid tumor cells after depletion of PGBD5 with two independent shRNAs, as compared with nontransduced cells and control cells expressing shGFP. The prefix ‘sh’ denotes shRNAs. Actin, loading control. (d) Relative number of viable G401 and A204 cells 72 h after PGBD5 shRNA depletion. Errors bars, s.d. of three independent experiments.
PGBD5-induced transformation requires DNA end-joining repair

If PGBD5-induced cell transformation involves transposase-mediated genomic rearrangements, then this process should depend on the repair of DNA double-strand breaks (DSBs) generated by the DNA-recombination reactions. Genomic rearrangements induced by transposases of the DDE/E superfamily involve transsterification reactions, which generate DSBs that are predominantly repaired by DNA nonhomologous end-joining (NHEJ) in somatic cells, as is the case for human V(DJ) rearrangements induced by the RAG1/2 recombinase. To test whether PGBD5-induced cell transformation requires NHEJ, we used isogenic RPE cells that were wild type or deficient in the NHEJ cofactor PAXX (encoded by C9orf142), which stabilizes the NHEJ repair complex and is required for efficient DNA repair.

In contrast to defects in other NHEJ components, such as LIG4, PAXX deficiency does not appreciably alter cell growth or viability but significantly decreases NHEJ efficiency without requiring TP53 inactivation to survive. Thus, we generated RPE cells that expressed doxycycline-inducible PGBD5 and were C9orf142+/+ or C9orf142−/−, and confirmed the induction of PGBD5 and lack of PAXX expression by protein blotting (Fig. 6a). Doxycycline-induced expression of PGBD5 in C9orf142−/− but not isogenic C9orf142+/+ RPE cells caused the accumulation of DNA-damage-associated phosphorylated histone H2AX (γH2AX) (Fig. 6b and Supplementary Fig. 8b), apoptosis-associated cleavage of caspase 3 (Fig. 6c and Supplementary Fig. 8a), and cell death (Supplementary Fig. 8c). We confirmed the requirement of NHEJ for the repair of PGBD5-induced rearrangements by using Xrcc5-deficient mouse embryonic fibroblasts (data not shown).

Importantly, PGBD5-mediated induction of DNA damage and cell death in NHEJ-deficient C9orf142−/− cells, as compared with isogenic NHEJ-proficient C9orf142+/+ cells, was nearly completely rescued by the p.[Asp168Ala]+[Asp194Ala]+[Asp386Ala] alteration of residues required for the transposase activity of PGBD5 (Fig. 6d). Thus, NHEJ DNA repair is required for the survival of cells expressing active PGBD5.

PGBD5-induced cell transformation involves site-specific genomic rearrangements associated with PGBD5-specific signal-sequence breakpoints

The requirements for PGBD5 enzymatic transposase activity, cellular NHEJ DNA repair, and the ability of transient PGBD5 expression to promote cell transformation are all consistent with the generation of heritable genomic rearrangements that mediate PGBD5-induced tumorigenesis. To determine the genetic basis of PGBD5-induced cell transformation, we sequenced whole genomes of PGBD5-induced tumors as well as control GFP-expressing and nontransduced RPE cells, by using massively parallel paired-end Illumina sequencing (Supplementary Data Set 1). For the rabbitoid tumor genome analysis, we used the assembly-based algorithm laSV as well as conventional techniques (Supplementary Table 3, Supplementary Figs. 9–11 and Supplementary Data Set 1). This analysis led to the identification of distinct genomic rearrangements, specifically in PGBD5-induced tumor cell genomes, as compared with those of control GFP-transduced and nontransduced RPE cells (Fig. 7a). The identified rearrangements were characterized by intrachromosomal deletions with a median length of 183 bp, in agreement with their apparent limited detectability through conventional genome analysis methods, as well as inversions, duplications, and translocations (Supplementary Fig. 12a–c and Supplementary Data Set 1).

As with genomic rearrangements found in primary human tumors (Fig. 1), the genomic rearrangements found in PGBD5-transformed RPE cells revealed significant enrichment of PSS motifs at the breakpoints of PGBD5-induced tumor structural variants (P = 7.2 × 10−3), hypergeometric test; Fig. 7b and Supplementary Data Set 1). By contrast, the breakpoints of structural variants in the GFP-control RPE-cell genomes, presumably at least in part because of normal genetic variation, exhibited no enrichment in PSS motifs (P = 0.37). We independently verified these findings by using the direct tree-graph–based comparative SmuFin analysis method (Supplementary Fig. 12a and Supplementary Data Set 1). In addition, we validated five of these rearrangements by using variant and wild-type allele-specific PCR followed by Sanger DNA sequencing of rearrangement breakpoints, to confirm that they were specifically present in PGBD5-transformed but not control GFP-transduced RPE cells (Supplementary Fig. 12d–h). Additionally, we did not find genomic-rearrangement breakpoints containing RSS sequences that were targeted by the RAG1/2 recombinase, which is not expressed in RPE cells. We also did not find evidence of structural alterations of the annotated human MER75 and MER85 piggyBac-like transposable elements, in agreement with the distinct evolutionary history of human PGBD5 (ref. 30).

To identify genomic rearrangements that might be functionally responsible for PGBD5-induced cell transformation, we analyzed the recurrence of PGBD5-induced genomic rearrangements in ten...
different RPE tumors from independent transduction experiments in individual mouse xenografts. We detected 59 PGBD5-induced structural variants per tumor, 42 (71%) of which were deletions, 36 (61%) of which affected regulatory intergenic elements, and 13 (22%) of which contained PSS motifs at their breakpoints (Supplementary Data Set 1). In particular, we identified recurrent and clonal PSS-associated rearrangements of WWOX, including duplication of exons 6–8 (Fig. 7d). WWOX is a tumor-suppressor gene that controls p53 signaling. We confirmed the duplication of exons 6–8 of WWOX by PCR and Sanger DNA sequencing (Fig. 7d), and tested its functional consequence on WWOX protein expression by protein blotting (Fig. 7e). Remarkably, this mutation resulted in low-level expression of the dominant-negative or gain-of-function activity of mutant WWOX in RPE-cell transformation. We observed this mutation in two out of ten independent RPE tumors, a result consistent with its probable pathogenic function in PGBD5-induced cell transformation.

To determine its function in PGBD5-induced RPE-cell transformation, we depleted endogenous WWOX and ectopically expressed wild-type WWOX in nontransformed wild-type and WWOX-mutant PGBD5-induced RPE-cell tumors (Supplementary Fig. 13a,d). In agreement with the tumorigenic function of PGBD5-induced mutations of WWOX, we found that WWOX inactivation was necessary but not sufficient to maintain clonogenicity of PGBD5-transformed RPE tumor cells in vitro (Supplementary Fig. 13b,c,e,f). Thus, PGBD5-induced cell transformation involves site-specific genomic rearrangements that are associated with PGBD5-specific signal-sequence breakpoints that recurrently target regulatory elements and tumor-suppressor genes (Fig. 7f).

**DISCUSSION**

Here, we found that primary human rhabdoid tumor genomes exhibit signs of PGBD5-mediated DNA recombination involving recurrent mutations of previously elusive rhabdoid tumor-suppressor genes (Fig. 1). These genomic rearrangements involve breakpoints associated with the PSS sequences, which are sufficient to mediate DNA rearrangements in rhabdoid tumor cell lines and physical recruitment of endogenous PGBD5 transposase (Fig. 2). The enzymatic activity of PGBD5 is both necessary and sufficient to promote similar genomic rearrangements in primary human cells, thus causing their malignant transformation (Figs. 3–7).

PGBD5-induced genomic rearrangements exhibit a defined architecture, including characteristic deletions, inversions, and complex rearrangements distinct from those generated by other known mutational processes. We observed an imprecise relationship between PSS sequences and genomic-rearrangement breakpoints, with evidence of incomplete ‘cut and paste’ DNA transposition, in agreement with potentially aberrant targeting of PGBD5 nuclease activity. Although our structure–function studies suggested that PGBD5 induces...
genomic rearrangements in conjunction with the canonical NHEJ apparatus, it is possible that PGBD5 activity may also promote other DSB-repair pathways, such as alternative microhomology-mediated end-joining (Supplementary Fig. 14). We confirmed that the putative catalytic aspartate mutants of PGBD5 on average maintained the chromatin localization of wild-type PGBD5. It is also possible that these residues contribute to cell transformation, owing to their interaction with cellular cofactors or assembly of DNA-regulatory complexes, or yet-unknown nuclease-independent functions that contribute to cell transformation.

PSS-associated genomic rearrangements induced by PGBD5 in rhabdoid tumors are reminiscent of McClintock’s ‘mutable loci’ induced by DNA transposase–mediated mutation of the Ds locus, which controls position-effect variegation in maize. Insofar as nuclease substrate accessibility is controlled by chromatin structure and conformation, PGBD5-induced genomic rearrangements indeed may be coupled to developmental regulatory programs that control gene expression and specification of cell fate, as suggested by their strong association with developmental regulatory DNA elements in our analysis. The association of PGBD5-induced rearrangements may involve sequence-specific recognition of human genomic PSS sequences or alternatively may be determined by their accessibility or the presence of cellular cofactors, as determined by the cellular developmental states.

Importantly, the spectrum of PGBD5-induced genomic rearrangements and their PSS sequences identified in this study should provide a useful approach for the functional characterization of tumor genomes and identification of cancer-causing genomic alterations. In the case of rhabdoid tumors, the association of SMARCB1 mutations with additional recurrent genomic lesions, such as structural alterations in CNTP2, TEM2, and TET2 genes, which regulate developmental and epigenetic cell-fate speciation, may lead to the identification of additional mechanisms of childhood cancer pathogenesis, including those that cooperate with the dysregulation of SWI–SNF–BAF-mediated chromatin and nucleosome remodeling induced by SMARCB1 loss. While the current study was under review, an additional genome analysis of rhabdoid tumors was described, and the results independently identified recurrent mutations at CNTP2 and other loci in human rhabdoid tumors. Notably, the recurrence patterns of PGBD5-induced genomic rearrangements in rhabdoid tumors indicate that, even for rare cancers, comprehensive tumor genome analyses will be necessary to define the spectrum of causal genomic lesions and potential therapeutic targets. Our results also indicated that improved genome-analysis methods, such as SMuFin and IaSv used in our work, and confirmation of their sensitivity and specificity will be needed to elucidate tumorigenic genome rearrangements. Similarly, given the existence of distinct molecular subtypes of rhabdoid tumors, it will be important to determine the extent to which PGBD5-induced genome remodeling contributes to this phenotypic diversity.

In summary, PGBD5 defines a distinct class of oncogenic mutators that contribute to cell transformation due not to mutational activation but instead to their aberrant induction and chromatin targeting, thereby inducing site-specific transforming genomic rearrangements. Our data identified PGBD5 as an endogenous human DNA transposase that is sufficient to fully transform primary immortalized human cells in the absence of chromosomal instability. Given the expression of PGBD5 in various childhood and adult solid tumors, owing to its aberrant or co-opted tissue expression, we anticipate that PGBD5 may also contribute to the pathogenesis of these cancers. Similarly, it will be important to investigate the functions of PGBD5 in normal vertebrate and mammalian development, given its ability to induce site-specific somatic genomic rearrangements in human cells. Finally, the functional requirement for cellular NHEJ DNA repair in PGBD5-induced cell transformation might facilitate the development of rational therapeutic strategies for rhabdoid and other tumors involving endogenous DNA transposases.

**METHODS**

Methods, including statements of data availability and any associated accession codes and references, are available in the online version of the paper.

**ACKNOWLEDGMENTS**

We are grateful to A. Gutierrez, M. Mansour, D. Bauer, T. Look, H. Zhu, C. Feschotte, M. Kharaas, J. Petretto, and M. Gil Mir for critical discussions, and J. Gilbert for editorial advice. We thank T. Westbroek (Baylor College of Medicine) and M. Alazd (MD Anderson Cancer Center) for materials. This work was supported by funding from NIH K08 CA166660, P30 CA008748, U54 OD020355, U11 TR000457, P50 CA140146, Spanish Ministerio de Economia y Competitividad SAF2014-60293-R, Cancer Research UK, the Wellcome Trust, the Starr Cancer Consortium, the Burroughs Wellcome Fund, the Sarcoma Foundation of America, the Matthew Larson Foundation, the Josie Robertson Investigator Program, and the Rita Allen Foundation. A.G.H. is supported by the Berliner Krebsgesellschaft e.V. and the Berlin Institute of Health. A.K. is supported as a Damon Runyon–Richard Lumsden Foundation Clinical Investigator.

**AUTHOR CONTRIBUTIONS**

A.G.H. contributed to the study design and collection and interpretation of the data. B.K. performed ChIP–seq, whole-genome sequencing, and FLEA PCR data analysis. I.Z. analyzed tumor genome–sequencing data with IaSv, E.I., C. Reed, A.E., and E.S. performed in vitro transformation assays and vector design and cloning. I.C.M. performed experiments and analyzed data. E.R.-F., S.G., M.P., C.E.M., A.-K.E., M.S., K.A., C. Reeves, N.D.S., D.T., and Z.W. analyzed genome-sequencing data. A.N.B. and S.P.J. contributed to creation of PAXX-deficient cells and study design. E.D.S. contributed to mouse-xenograft study design. M.G. performed statistical analysis of data sets. C.R.A. performed histological analysis of tumor samples. E.P., C.W.M.R., H.S., E.M., and S.A.A. contributed to study design. A.K. contributed to study design, data analysis, and interpretation. A.K. and A.G.H. wrote the manuscript, to which all authors contributed.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

Reprints and permissions information is available online at [http://www.nature.com/reprints/index.html](http://www.nature.com/reprints/index.html).
Fine mapping of V(D)J recombinase mediated Oncogene regulation: an oncogenic super-enhancer formed Malignant rhabdoid tumours of the kidney

Integrated (epi)-genomic analyses identify subgroup-specific Site- and allele-specific polycomb dysregulation in T-cell

Sequencing of neuroblastoma identifies chromothripsis and
ONLINE METHODS

Reagents. All reagents were obtained from Sigma-Aldrich if not otherwise specified. Synthetic oligonucleotides were obtained from Eurofins (Eurofins MWG Operon) and were purified by HPLC, as listed in Supplementary Table 5. Antibodies are listed in Supplementary Table 6.

Plasmid constructs. Human PGBD5 cDNA (NM_024554.3) was cloned into the lentiviral vector in frame with N-terminal GFP to generate pRevLuc103-GFP-PGBD5 (GeneCopoeia). pRevLuc-Lv103 encoding GFP was used as a negative control in all experiments. Plasmid encoding the hyperactive T. ni piggyBac transposase, as originally cloned by N. Craig and colleagues45, was obtained from System Biosciences and was cloned into pRevLuc-Lv103. The plasmids pBABE-neo-largeT, pBABE-puro-H-Ras, psPAK2, and pMD2.G were obtained from Addgene. Missense GFP-PGBD5 mutants were generated through site-directed mutagenesis, according to the manufacturer's instructions (QuikChange Lightning), as previously described36. Doxycycline-inducible pINDUCER21 vector was a kind gift from T. Westbrook46 and was used to generate pINDUCER21-PGBD5 through Gateway cloning, according to the manufacturer's instructions (Fisher Scientific). Lentiviral shRNA and doxycycline-inducible WWOX expression vectors were a kind gift from M. Aldaz47. pLKO.1 shRNA vectors targeting PGBD5 (TRCN0000138412, TRCN0000135121) and control shGFP were obtained from the RNAi Consortium (Broad Institute). The PB-EF1-IRES-NEO transposon reporter plasmid was used as previously described16. pBS-EF1-IRES-NEO was created by cloning the EF1-IRES-NEO cassette from PB-EF1-IRES-NEO into the plBueScript plasmid and was modified by PCR mutagenesis to replace the T. ni piggyBac inverted terminal repeat with the PGBD5 signal sequence (CTGGGAAATCGG). All newly generated plasmids are available from Addgene (URLs).

Production and purification of anti-PGBD5 antibody. Synthetic peptide from human PGBD5 (NM_024554.3) ELQLLSIVPGDRDLQP5DSFTGPTRC was used to immunize mice (Lampire Biological Products). Hybridoma clones were screened through enzyme-linked immunosorbent assays, and hybridoma supernatants were purified with Protein A affinity chromatography to generate the 10A8-11-7-P-5 antibody used in protein blotting (Supplementary Table 6).

Lentivirus production and cell transduction. Lentivirus production was carried out as previously described48. Briefly, HEK293T cells were transfected with TransIT-TL1 with a 2:1:1 ratio of the lentiviral vector and psPAK2 and pMD2.G packaging plasmids, according to the manufacturer's instructions (Mirus). Virus supernatant was collected 48 and 72 h after transfection, pooled, filtered, and stored at −80 °C. RPE and BJ cells were transduced with virus particles at a multiplicity of infection (MOI) of 5 in the presence of 8 µg/ml hexadimethrine bromide. Transduced cells were selected for 2 d with puromycin hydrochloride (RPE cells at 10 µg/ml and BJ cells at 2 µg/ml) or G418 sulfate (2 mg/ml), depending on the vector-mediated resistance. For pINDUCER21 viruses, cells were transduced at an MOI of 1 and were isolated through fluorescence-activated cell sorting (FACS) (BD Bioscience). For inducible expression of WWOX, RPE cells were transduced with lentiviruses encoding tetOn-advanced-WWOX and selected with G418 sulfate (2 mg/ml) for 10 d. For shRNA depletion of WWOX, cells were transduced with lentiviruses encoding PGIPZ-shWWOX or pGIPZ-shScramble control and were selected with puromycin hydrochloride (10 µg/ml) for 2 d.

Cell culture. Low-passage RPE and BJ cells, and human tumor cell lines were obtained from the American Type Culture Collection (ATCC). Covyrl1427 RPE cells have been described previously41. The identity of all cell lines was verified by STR analysis (Genetics DNA Laboratories), and absence of Mycoplasma sp. contamination was determined with a Lonza MycoAlert system. Cell lines were cultured in 5% CO2 in a humidified atmosphere at 37 °C in Dulbecco’s Modified Eagle’s medium with high glucose (DMEM-HG) supplemented with 10% FBS and antibiotics (100 U/ml penicillin and 100 µg/ml streptomycin). Clonogenic assays of RPE cells were carried out in DMEM/F-12 medium. To assess the number of viable cells, cells were trypsinized, resuspended in medium and sedimented at 800g for 5 min. Cells were then resuspended in PBS, and 10 µl was mixed in a 1:1 ratio with 0.4% trypan blue (Thermo Fisher) and counted with a hemocytometer (Hauser Scientific).

Transposon reporter assay. The transposon reporter assay was performed with the pBS-EF1-IRES-NEO vector in HEK293 cells, as described previously39.

Quantitative RT–PCR. RNA was isolated with an RNeasy Mini kit according to the manufacturer’s instructions (Qiagen). cDNA was synthesized with a SuperScript III First-Strand Synthesis System according to the manufacturer’s instructions (Invitrogen). qRT–PCR was performed with KAPA SYBR FAST PCR polymerase with 20 ng template and 200 nM primers, according to the manufacturer’s instructions (KAPA Biosystems). PCR primers are listed in Supplementary Table 5. Ct values were calculated with ROX normalization in ViIA 7 software (Applied Biosystems).

Protein blotting. To analyze protein expression by protein blotting, 1 million cells were suspended in 80 µl of lysis buffer (4% SDS, 7% glyceral, 1.25% β-mercaptoethanol, 0.2 mg/ml bromophenol blue, and 30 mM Tris–HCl, pH 6.8) and incubated at 95 °C for 10 min. Cell suspensions were lysed with a Covaris S220 adaptive focused sonicator, according to the manufacturer’s instructions. Lysates were cleared by centrifugation at 16,000g for 10 min at 4 °C. Clarified lysates (30 µl) were resolved with SDS–PAGE and electroblotted on Immobilon FL, PVDF membranes (Millipore). Membranes were blocked with Odyssey Blocking buffer (Li-Cor) and blotted with the antibodies listed in Supplementary Table 6. Blotted membranes were visualized on an Odyssey CLx fluorescence scanner, according to the manufacturer’s instructions (Li-Cor), with goat secondary antibodies conjugated to IRDye 800CW or IRDye 680RD (Supplementary Table 6).

Flow cytometry of cleaved caspase-3. Cells were fixed with neutral-buffered formalin for 10 min on ice, washed with PBS, resuspended in 0.1% Triton X-100 in PBS, and incubated for 15 min at room temperature. Permeabilized cells were washed twice with PBS and resuspended in 100 µl of Hank’s balanced salt solution (HBSS) with 0.1% bovine serum albumin and 2 µl of Alexa Fluor 647–conjugated antibody against cleaved caspase-3 (Supplementary Table 6). Cells were incubated for 30 min at room temperature in the dark, washed twice with PBS and stained with 1 µg/ml DAPI. Cells were analyzed on a Fortessa LSR, as previously described (BD Bioscience)49,50.

Histological staining. Histologic processing and staining was done as previously described51,52. Briefly, cell lines were plated on eight-well glass Millicell EZ chamber slides at 5,000 cells/well, grown for 24 h, and fixed with 4% paraformaldehyde for 10 min at room temperature (Millipore). Tumor xenograft tissue was fixed with 4% paraformaldehyde for 24 h at room temperature. Tissues were embedded in paraffin with an ASP6025 tissue processor (Leica), sectioned at 5 µm with a RM2265 microtome (Leica), and collected on SuperfrostPlus slides (Fisher Scientific). Tissue sections were deparaffinized with EZPrep buffer (Ventana Medical Systems). Antigen retrieval was performed with Cell Conditioning 1 buffer (Ventana Medical Systems), and sections were blocked for 30 min with Background Buster solution (Innovex). Primary antibodies were applied for 5 h at 1 µg/ml (Supplementary Table 6). Secondary antibodies were applied for 60 min.

For immunohistochemistry staining, diaminobenzidine (DAB) detection was performed with a DAB detection kit according to the manufacturer’s instructions (Ventana Medical Systems). Slides were counterstained with hematoxylin, and a cover slip was mounted with Permount (Fisher Scientific).

For immunofluorescence staining, the detection was performed with streptavidin–HRP D (Ventana Medical Systems) and subsequent incubation with tyramide Alexa Fluor 647, as prepared according to the manufacturer’s instructions (Invitrogen). Slides were then counterstained with 5 µg/ml DAPI for 10 min, and a cover slip was mounted with Mowiol (Sigma–Aldrich).

Image acquisition. Bright-field images were acquired on an Axio Observer microscope (Carl Zeiss Microimaging). Epifluorescence images were acquired with an EVOS FL microscope (Thermo Fisher). Slides were scanned with a Panoramic 250 slide scanner, and images were analyzed with the Panoramic Viewer (3DHistech).

Karyotype analysis. Five million cells were grown for 24 h before harvesting. Cultures were treated with 0.005 µg/ml colcemid for 1 h at 37 °C, resuspended

doi:10.1038/ng.1866
in 75 mM KCl for 10 min at 37 °C, and fixed in methanol/acetic acid (3:1). Cells were transferred onto slides, stained in 0.08 μg/ml DAPI in citric acid buffer for 3 min, and mounted in Vectashield solution (Vector Labs). For each cell line, a minimum of 15 metaphases were counted.

**Anchorage independence assay.** One million RPE and BJ cells stably transduced with lentiviral vectors were expanded in 10-cm tissue culture plates until fully confluent. At confluence, cells were microscopically inspected for the occurrence of refractile colonies within the cell monolayer. For growth in semisolid medium, one million cells were resuspended in 2 ml of medium mixed with 2 ml of Matrigel (BD Bioscience). Cell suspensions were plated in 12-well tissue culture plates (200 μl per well). Semisolid suspensions were cultured for 10 d before scoring.

**Xenografts.** All mouse experiments were carried out in accordance with institutional animal protocols, as approved by the Memorial Sloan Kettering Cancer Center Institutional Animal Care and Use Committee. Ten million RPE and BJ cells were suspended in 200 μl Matrigel (BD Bioscience) and injected subcutaneously into the left flanks of 6-week-old female NOD.Cg-Prkd(ciscid)/Il2rg(tm1Wjl)/SzJ mice (Jackson Laboratory). Tumor growth was monitored with caliper measurements, and tumor volume was calculated with the formula 3.14159 × length × width × depth/6.000. Mice were sacrificed by CO2 asphyxiation 3 d after transplantation or when tumor size exceeded 2,000 mm3. For secondary xenografts, primary xenografts were manually dissected and dissociated with 2 ml of collagenase in PBS for 30 min at 37 °C. Dissociated cell suspensions were filtered with 40-μm nylon-mesh filters and cryopreserved with 10% DMSO, 40% FBS, and 50% DMEM-HG. For doxycycline treatment of mice, animals were fed 250 mg doxycycline chow, which was replaced weekly (Harlan). Photographs of mice and tumors were taken with a Nikon D3100 camera (Minato). Mouse experimental sample sizes were determined to achieve 80% power to detect a five-fold difference, by using the K-sample rank test. In mouse experiments with doxycycline treatment, we used randomization to assign animals to treatment groups. Mouse tumor size measurements were performed with blinding.

**Analysis of published gene expression arrays.** The R2 visualization and analysis platform (URLs) was used to reanalyze published HG-U133 Plus 2.0 microarray gene expression data from normal and tumor human tissues. The analyzed gene expression data sets are listed in Supplementary Table 7.

**Flanking sequence exponential anchored (FLEA) PCR.** Transposon mapping with FLEA PCR was done as previously described31. **Chromatin immunoprecipitation and sequencing (ChIP-seq).** ChIP was performed as previously described34. Briefly, cells were fixed in 1% formalin in PBS for 10 min at room temperature. Glycine (125 mM final concentration) was added to the cells, and cells were washed twice in ice-cold PBS and resuspended in SDS lysis buffer (1% SDS, 10 mM EDTA, and 50 mM Tris-HCl, pH 8.1). Lysates were sonicated with a Covaris S220 adaptive focused sonicator to obtain 100- to 500-bp chromatin fragments (Covaris). Lysates containing sheared chromatin fragments were resuspended in 0.01% SDS, 1% Triton X-100, 1.2 mM EDTA, 16.7 mM Tris-HCl, pH 8.1, and 167 mM NaCl. Rabbit anti-PGBD5 antibody was coupled to Protein A and Protein G Dynabeads according to the manufacturer’s protocol (Thermo Fisher Scientific). Lysates and antibody-coupled beads were incubated overnight at 4 °C. Precipitates were washed sequentially with ice-cold low-salt washing solution (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, and 150 mM NaCl), high-salt washing solution (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 20 mM Tris-HCl, pH 8.1, and 500 mM NaCl), LiCl washing solution (0.25 M LiCl, 1% IGEPA, 0.1% deoxycholic acid, 1 mM EDTA, and 10 mM Tris-HCl, pH 8.1), and Tris-buffered EDTA washing solution (1 mM EDTA and 10 mM Tris-HCl, pH 8.1), then eluted in elution buffer (1% SDS and 0.1 M NaHCO3). ChIP–seq libraries were generated with the NEBNext ChIP–seq Library Prep kit according to the manufacturer’s protocol (New England BioLabs). Libraries were sequenced on Illumina HiSeq 2500 instruments, with 30 million 2 × 50-bp paired-end reads.

**ChIP-seq analysis.** Reads were trimmed for both quality and adaptor sequences, and paired reads were removed if either read length became <20 nt. Bowtie2 (v2.2.2) with default parameters was used to align the reads to the human reference assembly hg19, and PCR and optical duplicates were removed with Picard (URLs). Genomic segments enriched for ChIP over input signal were classified with MACS (v1.4) with the default parameters, and genomic ‘blacklisted’ regions were subsequently filtered (URLs). Signals in enriched regions were then normalized to segment length and sequencing depth.

**Whole-genome DNA sequencing.** Genomic DNA was extracted with a PureLink Genomic DNA Mini Kit according to the manufacturer’s instructions (Thermo Fisher Scientific). Genome sequencing libraries were constructed with a TrueSeq Nano library kit, according to the manufacturer’s protocol (Illumina). Genomes were sequenced on Illumina HiSeq X instruments, with 2 × 150-bp paired-end reads. For analysis of primary patient rhabdoid tumor genomes, sequencing files were downloaded from the TARGET Data Matrix, as previously described35. Reads were aligned to the GRCh37 human reference with the Burrows–Wheeler Aligner (BWA an and BWA MEM for GATK and laSV analyses, respectively) and processed with the best-practices pipeline, which included marking of duplicate reads with Picard tools, realignment around indels, and base recalibration via Genome Analysis Toolkit (GATK) (ver. 3.2.2)15–26.

**Alignment-based mutational and structural variant analysis.** MuTect (v1.1.4)37, LoFreq (v2.0.0)38 (single-nucleotide variants (SNVs) only), Strelka (v1.0.13)39 (both SNVs and indels), Pindel (v0.2.5), and Scalpel (v0.4)39 (indels only) were used with the default filtering criteria, as implemented in each of the programs. Triallelic SNVs and common germline variants (≥1% MAF in 1000 Genomes Project release 3 or the Exome Aggregation Consortium server (URLs)), and a blacklist of recurrent artifactcalls seen in HapMap samples sequenced and analyzed with the same methodology, were filtered out. The union of all SNV and indel calls was annotated with snpEff, snpSift40 and GATK VariantAnnotator according to the annotations from ENSEMBL, COSMIC, 1000 Genomes Project, and ExAC41,62. Copy-number variants (CNVs) were detected with BIC-seq2 CNV changepoints. SVs with split-read support (tumor only), with at least one coinciding (within 500 bp) CNV changepoint called by two or more tools or called by CREST, are marked as higher confidence. The annotation with gene overlap (ReSeq, Cancer Gene Census), including prediction of potential effects on genes (for example, disruptive/exonic, intronic, and intergenic) and with annotated transposons, was done with bedtools47.

**laSV.** De novo assembly-based laSV33 was used with the following parameters: -s 15 -k 63 -p 3. Structural variants supported by fewer than four reads or with allele frequencies below 10% were filtered. Variant recurrence was measured in 100-kb bins with bedtools47. Circos plots were generated with Circos (version 0.67-4)39.

**SMuFin.** SMuFin was used with default parameters, as previously described42. SMuFin results included, SNVs as well as small (indels) and SVs. Large SVs were defined as SVs identified with a single breakpoint, for which the SV length exceeded the length of the underlying variant block called by SMuFin. Breakpoints supported by fewer than four reads were filtered. SV size was estimated on the basis of the assumption that SVs were caused by single genomic events.

**Regulatory element analysis.** Annotated regulatory elements were compiled from both ENCODE and NIH Roadmap Epigenomics Consortium (URLs). The analysis focused on distal DNase I–hypersensitivity sites, because distal sites have been shown to vary in a cell-type-specific manner, and DNase I sensitivity covers both active and poised regulatory elements. Cancer cell line
PGBD5 signal sequence (PSS) analysis. The position weight matrix (PWM) for the PSS and RSS were generated as previously described\(^\text{32}\). These PWMs were used to scan sequences around variant breakpoints (±50 bp) for both PSS and RSS with the sequence-motif-matching algorithm FIMO\(^\text{30}\). Additionally, PGBD5 signal-sequence motifs associated with structural variants were detected by analysis of 20-bp windows around variant breakpoints with MEME with default parameters\(^\text{21}\). Matches with a false discovery rate <0.1 and within 15 bp from the variant breakpoints were retained and counted. All variants associated with PSS motifs were manually verified. To construct the position-scrambled PSS, the perl rand function was used to generate ten independent position-scrambled PWMs.

Statistical analysis. All experiments were performed a minimum of three times with a minimum of three independent measurements. For comparisons between two sample sets, statistical analysis of means was performed with two-tailed unpaired Student’s t-tests. Survival analysis was done with the Kaplan–Meier method, as assessed with a log-rank test. For gene expression analysis, statistical significance was assessed with paired t-tests. False discovery was assessed at the 0.05 level with the step-down Dunnett method, as extended to general parametric models\(^\text{22,23}\). The significance of sequence-motif enrichment was assessed with hypergeometric tests. For significance analysis of association of structural variants with regulatory elements, Welch’s t-test was used. Calculations were performed with R statistical computing software\(^\text{24}\).

Code availability. Scripts used in this analysis are openly available at github (URLs).

Data availability. Genome and chromatin immunoprecipitation sequencing data have been deposited in the NCBI Sequence Read Archive and Gene Expression Omnibus databases (Bioproject 320056 and Data Set GSE81166, respectively). Analyzed data are openly available at the Zenodo digital repository (http://dx.doi.org/10.5281/zenodo.50633), as summarized in Supplementary Table 8.

44. Li, X. et al. piggyBac transposase tools for genome engineering. *Proc. Natl. Acad. Sci. USA* 110, E2279–E2287 (2013).
45. Westbrook, T.F., Stegmeier, F. & Elledge, S.J. Dissecting cancer pathways and vulnerabilities with RNAi. *Cold Spring Harb. Symp. Quant. Biol.* 70, 435–444 (2005).
46. Ferguson, B.W. et al. The cancer gene WWOX behaves as an inhibitor of SMAD3 transcriptional activity via direct binding. *BMC Cancer* 13, 593 (2013).
47. Kentsis, A. et al. Autocrine activation of the MET receptor tyrosine kinase in acute myeloid leukemia. *Nat. Med.* 18, 1118–1122 (2012).
48. Fox, R. & Aubert, M. Flow cytometric detection of activated caspase-3. *Methods Mol. Biol.* 141, 47–56 (2001).
49. Sordet, O. et al. Specific involvement of caspases in the differentiation of monocytes into macrophages. *Blood* 100, 4464–4453 (2002).
50. Yarilin, D. et al. Machine-based method for multiplex in situ molecular characterization of tissues by immunofluorescence detection. *Sci. Rep.* 5, 9534 (2015).
51. Fujisawa, S., Turkekul, M., Barlas, A., Fan, N. & Manova, K. Double in situ detection of sonic hedgehog mRNA and p MAPK protein in examining the cell proliferation signaling pathway in mouse embryo. *Methods Mol. Biol.* 717, 257–276 (2011).
52. Henssen, A., Carson, J.R. & Kentis, A. Transposon mapping using flanking sequence exponential anchored (FLEA) PCR. *Protocol Exchange* http://dx.doi.org/10.1038/protoc.2015.071 (2015).
53. Krivtsov, A.V. et al. DKK1 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell* 14, 355–368 (2008).
54. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754–1760 (2009).
55. McKenna, A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 20, 1297–1303 (2010).
56. Cilibskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 211–219 (2013).
57. Wilk, A. et al. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic Acids Res.* 40, 11189–11201 (2012).
58. Saunders, C.T. et al. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics* 28, 1811–1817 (2012).
59. Ye, K., Schulz, M.H., Long, Q., Apweiler, R. & Nong, Z. Pindel: a pattern growth approach to detect breaks of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 25, 2865–2871 (2009).
60. Narsisi, G. et al. Accurate de novo and transmitted indel detection in exome-capture data using microassembly. *Nat. Methods* 11, 1033–1036 (2014).
61. Cingolani, P. et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEFN: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)* 6, 80–92 (2012).
62. Xi, R. et al. Copy number variation detection in whole-genome sequencing data using the Bayesian information criterion. *Proc. Natl. Acad. Sci. USA* 108, E1128–E1136 (2011).
63. Rausch, T. et al. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics* 28, i333–i339 (2012).
64. Wang, J. et al. CREST maps somatic structural variation in cancer genomes with base-pair resolution. *Nat. Methods* 8, 652–654 (2011).
65. Chen, K. et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat. Methods* 6, 677–681 (2009).
66. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* 26, 841–842 (2010).
67. Narzisi, G. et al. Accurate de novo and transmitted indel detection in exome-capture data using Splazer2. *Bioinformatics* 28, 619–627 (2012).
68. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645 (2009).
69. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018 (2011).
70. Bailey, T.L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208 (2009).
71. Bork, P. & Wall, J. The protein families database PFam. Bioinformatics 18(suppl 1), iii1–iii3 (2002).
72. Eddy, S.R. A probabilistic model of local sequence alignment. *J. Mol. Biol.* 268, 721–741 (1997).
73. Wang, L., Fickett, J.W. & Bailey, T.L. Alignment-free sequence comparison using spaced seed alignments. *Bioinformatics* 27, 2515–2521 (2011).
74. Fickett, J.W. & Tung, I. A database of protein families. *Nucleic Acids Res.* 29, 269–278 (2001).
75. Narzisi, G. et al. Accurate de novo and transmitted indel detection in exome-capture data using Splazer2. *Bioinformatics* 28, 619–627 (2012).
76. Krzywinski, M. et al. Circos: an information aesthetic for comparative genomics. *Genome Res.* 19, 1639–1645 (2009).
77. Grant, C.E., Bailey, T.L. & Noble, W.S. FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018 (2011).
78. Bailey, T.L. et al. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* 37, W202–W208 (2009).
79. Bork, P. & Wall, J. The protein families database PFam. Bioinformatics 18(suppl 1), iii1–iii3 (2002).
80. Eddy, S.R. A probabilistic model of local sequence alignment. *J. Mol. Biol.* 268, 721–741 (1997).
81. Wang, L., Fickett, J.W. & Bailey, T.L. Alignment-free sequence comparison using spaced seed alignments. *Bioinformatics* 27, 2515–2521 (2011).
Erratum: PGBD5 promotes site-specific oncogenic mutations in human tumors

Anton G Henssen, Richard Koche, Jiali Zhuang, Eileen Jiang, Casie Reed, Amy Eisenberg, Eric Still, Ian C MacArthur, Elias Rodriguez-Fos, Santiago Gonzalez, Montserrat Puiggròs, Andrew N Blackford, Christopher E Mason, Elisa de Stanchina, Mithat Gönen, Anne-Katrin Emde, Minita Shah, Kanika Arora, Catherine Reeves, Nicholas D Socci, Elizabeth Perlman, Cristina R Antonescu, Charles W M Roberts, Hanno Steen, Elizabeth Mullen, Stephen P Jackson, David Torrents, Zhiping Weng, Scott A Armstrong & Alex Kentsis

Nat. Genet.; doi:10.1038/ng.3866; corrected online 24 May 2017

In the version of this article initially published online, the affiliations for Jiali Zhuang listed an incorrect present address instead of an equal contribution. The error has been corrected in the print, PDF and HTML versions of this article.