The landscape of somatic mutations in epigenetic regulators across 1000 pediatric cancer genomes

Robert Huether[1,*], **Li Dong**[2,*], **Xiang Chen**[1], **Gang Wu**[1], **Matthew Parker**[1], **Lei Wei**[1], **Jing Ma**[2], **Michael N. Edmonson**[1], **Erin K. Hedlund**[1], **Michael C. Rusch**[1], **Sheila A. Shurtleff**[2], **Heather L. Mulder**[3], **Kristy Boggs**[3], **Bhavin Vadordaria**[3], **Jinjun Cheng**[2], **Donald Yergeau**[3], **Guangchun Song**[2], **Jared Becksfort**[1], **Gordon Lemmon**[1], **Catherine Weber**[2], **Zhongling Cai**[2], **Jinjun Dang**[2], **Michael Walsh**[10], **Amanda L. Gedman**[2], **Zachary Faber**[2], **John Easton**[3], **Tanja Gruber**[2,10], **Richard W. Kriwacki**[4], **Janet F. Partridge**[6], **Li Ding**[7,8,9], **Richard K. Wilson**[7,8,9], **Elaine R. Mardis**[7,8,9], **Charles G. Mullighan**[2], **Richard J. Gilbertson**[5], **Suzanne J. Baker**[5], **Gerard Zambetti**[6], **David W. Ellison**[2], **Jinghui Zhang**[1], and **James R. Downing**[2]

[1]Department of Computational Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA

[2]Department of Pathology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA

[3]The Pediatric Cancer Genome Project Laboratory, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN Tennessee 38105, USA

[4]Department of Structural Biology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA

[5]Department of Developmental Neurobiology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, Tennessee 38105, USA

[6]Department of Biochemistry, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

[7]The Genome Institute, Washington University School of Medicine, St. Louis, Missouri, USA

[8]Department of Genetics, Washington University School of Medicine, 4444 Forest Park Ave, St Louis, Missouri 63108, USA

[9]Siteman Cancer Center, Washington University, St Louis Missouri, USA

Correspondence should be addressed to J.R.D. (james.downing@stjude.org) or J.Z. (jinghui.zhang@stjude.org).
*These authors contributed equally to this work and are co-first authors

[10]Department of Oncology, St. Jude Children's Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA

Abstract

Here we sequence 633 genes, encoding the majority of known epigenetic regulatory proteins, in over 1000 pediatric tumors to define the landscape of somatic mutations in epigenetic regulators in pediatric cancer. Our results demonstrate a marked variation in the frequency of gene mutations across 21 different pediatric cancer subtypes, with the highest frequency of mutations detected in high-grade gliomas, T-lineage acute lymphoblastic leukemia, medulloblastoma, and a paucity of mutations in low-grade glioma, and retinoblastoma. The most frequently mutated genes are *H3F3A, PHF6, ATRX, KDM6A, SMARCA4, ASXL2, CREBBP, EZH2, MLL2, USP7, ASXL1, NSD2, SETD2, SMC1A,* and *ZMYM3*. Importantly, we identify novel loss-of-function mutations in the ubiquitin-specific-processing protease 7 (USP7) in pediatric leukemia, which result in a decrease in deubiquitination activity. Collectively, our results help to define the landscape of mutations in epigenetic regulatory genes in pediatric cancer and yield a valuable new database for investigating the role of epigenetic dysregulations in cancer.

INTRODUCTION

Genome wide mutation profiling of pediatric cancer has yielded important insights into the molecular pathology of the major subtypes of cancer seen in children[1]. Two general observations to emerge from these studies are that pediatric cancers on average contain fewer somatic mutations than comparable tumors arising in adults; and that genes that encode proteins involved in epigenetic regulation are mutated at a high frequency in a subset of pediatric cancers. A striking example of the latter are mutations in histone 3 (*H3F3A*, encoding H3.3 and *HIST1H3B*, encoding H3.1) that cause a p.Lys27Met amino acid substitution in up to 78 % of diffuse intrinsic pontine glioma – a highly aggressive subtype of pediatric brain tumor[2,3]. Additional epigenetic regulators recurrently mutated in pediatric cancer include *CREBBP*, *EED*, *EP300*, *EZH2*, *PHF6*, and *SETD2* in acute lymphoblastic leukemia[4–6], *CHD7, HDAC9, KDM4C, KDM6A, MLL2, SMARCA4* and *ZMYM3* in medulloblastoma[7], and *ATRX* in neuroblastoma and high-grade glioma[2,8].

To extend these observations, we determine the frequency of somatic mutations in genes directly implicated in epigenetic regulation across each of the major subtypes of pediatric cancer as part of the St. Jude Children's Research Hospital-Washington University Pediatric Cancer Genome Project [1]. A total of 633 epigenetic regulatory genes in 1,020 pediatric cancers representing 21 different cancer subtypes including brain tumors, solid tumors and leukemias are sequenced. Our comprehensive analysis helps to define the landscape of mutations in epigenetic regulatory genes in pediatric cancer and provides a database that should be of significant value in elucidating the role of epigenetic dysregulation in cancer.

RESULTS

Somatic mutations in epigenetic regulatory genes

The 633 epigenetic regulatory genes analyzed in this study include enzymes that covalently modify histones including histone writers (n=159) and histone erasers (n=55); the proteins that bind histone writers (n=65) or histone erasers (n=20); histones (n=88); histone readers (n=116); chromatin remodelers (n=72); and enzymes that covalently modify DNA (n=58) (Fig. 1a, Supplementary Data 1). These genes were sequenced in 1,020 pediatric cancers representing 21 different cancer subtypes including brain tumors (4 subtypes), solid tumors (6 subtypes) and leukemias (11 subtypes) (Table 1). DNA samples from both tumor and matched germ line were analyzed by either whole genome sequencing (WGS, n=434), whole exome sequencing (WES, n=244) or custom designed capture sequencing of all coding exons of the 633 genes (CC, n=426) (Table 1 and Supplementary Data 2). The average read depth for WGS, WES and CC is 30x, 100x and 342x, respectively. Across the entire cohort, 96.7% of the coding exons of the 633 genes had coverage greater than 20x. Because of the variation in sequencing methods used across the cohort, we limited our mutation analyses to the detection of single nucleotide variants (SNVs) and small insertion/deletions (indels). This analysis yielded a >90% power to detect mutations that occurred with a mutant allele fraction (MAF) of 0.3, and thus focuses on mutations in the dominant malignant clone (Supplementary Fig. 1 and Supplementary Data 3–4). All identified non-silent coding region mutations were experimentally validated by an independent sequencing platform resulting in a total of 668 validated somatic mutations, with 62% (414) occurring with a MAF >30%.

Sixty-two of the 633 genes were recurrently mutated across the patient cohort, with an additional 128 genes mutated in a single case (Supplementary Fig. 2 and Supplementary Data 5). The pediatric tumors that had the highest frequency of mutations in epigenetic genes were high-grade gliomas (HGG), T-lineage acute lymphoblastic leukemia (TALL), and medulloblastoma (MB) (43%-59% of cases in these tumor subtypes had a mutation in an epigenetic gene in the dominant tumor clone, Fig. 1b). Osteosarcoma (OS) also exhibited high rates of mutation in epigenetic regulatory genes; however, the high background mutation rate in these tumors suggest that the majority of the epigenetic regulatory gene mutations in this cancer subtype were passenger rather than driver mutations (Supplementary Data 5). Importantly, several pediatric cancers were notable for almost a complete absence of mutations in epigenetic regulators including low-grade gliomas (LGG), retinoblastoma (RB), and infant leukemia (INF) (Fig. 1b). However, it is important to remember that the majority of pediatric infant leukemias contain a translocation involving the *MLL* gene and thus have an alteration in a key epigenetic regulator as part of the leukemia's initiating lesions [9].

Most frequently mutated epigenetic regulatory genes

The most frequently mutated epigenetic regulatory gene in pediatric cancer (mutated in 5 or more cases) were *H3F3A, PHF6, ATRX, KDM6A, SMARCA4, ASXL2, CREBBP, EZH2, MLL2, USP7, SETD2, ASXL1, NSD2, SMC1A,* and *ZMYM3* (Fig. 1c and Supplementary Table 1 and Supplementary Data 5). Although each of these genes has been implicated in

cancer, *USP7, SMC1A* and *ASXL2* have only been reported to be mutated in a single pediatric case each, and are rarely mutated in adult cancers (http://cancer.sanger.ac.uk/cosmic). Importantly, a majority of the top 15 mutated genes were found to be mutated in multiple different pediatric cancer subtypes. The only exceptions were mutations in *ASXL2, NSD2, PHF6, SETD2*, and *USP7* which were identified in leukemias, but not in brain or solid tumors. Mutations in at least one of the top 15 genes were found in 23% of the pediatric brain tumors, 15% of pediatric leukemias, but only 7% of pediatric solid tumors. When we extend this analysis to all recurrently mutated epigenetic regulators (mutated in 2 or more cases), brain tumors (30%) and leukemia (30%) share the highest frequency of cases containing mutations in epigenetic regulators, followed by pediatric solid tumors (17%).

Consistent with previous reports, the identified mutations in *PHF6*, *KDM6A*, *ATRX*, *MLL2*, *CREBBP*, *SETD2*, *SMARCA4*, *ASXL2*, *ASXL1*, and *ZMYM3* are predicted to result in a loss-of-function (Supplementary Table 1). By contrast, the NSD2 p.E1099K mutation has recently been shown to lead to enhanced histone methytransferase activity[10], whereas the p.K27M mutation in H3F3A eliminates the ability of this residue to undergo normal regulatory posttranslational modifications and confers a gain-of-function activity that leads to a block in the trimethylation of all H3 in the cell including the wild type protein[11,12]. Although both activating and inactivating mutations of EZH2 have been previously reported[6,13], we primarily detected EZH2 inactivating mutations in pediatric cancer. Lastly, although the functional significance of the identified cohesion subunit *SMC1A* missense mutations remains to be determined, some of the identified somatic mutations have been observed as germ line mutations in patients with Cornelia de Lange syndrome[14].

The most frequently mutated epigenetic proteins in pediatric cancer function within a network of 8 epigenetic regulatory complexes that include the Set1 (Compass/Compass-like)[15], mixed lineage leukemia (MLL)[16], activating signal cointegrator-2 containing (ASCOM)[17], nucleosome remodeling and deacetylation (NuRD)[18], polycomb repressor 2 (PRC2)[19], the SWI/SNF containing (BAF/PBAF)[20], CREBBP/EP300 (CREB) complex[21], and the DNMT1/USP7/UHRF1 (DUU)[22] (Fig. 2 and Supplementary Data 5–6). Nearly half of all proteins contained within these complexes are mutated at least once in pediatric cancer. No significant differences were detected in the frequency of mutations within the BAF/PBAF and inter-related MLL/ASCOM/compass complexes across the pediatric cancer subtypes analyzed. By contrast, over half of the mutations within the CREB, PRC2 and NuRD complexes occurred in pediatric leukemias, and all of the mutations in the DNMT1/USP7/UHRF1 (DUU) complex, which regulates DNA methylation and histone deubiquitination, were identified in leukemias. Of particular note, novel mutations were observed in the ubiquitin-specific-processing protease 7 (*USP7*).

Loss-of-function mutations in USP7

The de-ubiqutinase USP7 has been suggested to lead to the stabilization of several nuclear proteins including the tumor suppressor p53[23], PTEN[24], the DNA methyl transferase DNMT1[22], and histone H2B[25]. Nine *USP7* mutations were detected in 8 patients in our study (Fig. 3a). There were five frame shift mutations (T177fs, V203fs, R340fs, D380fs, D483fs) that would encode truncated proteins that lack the full catalytic domain, and four

missense mutations (C300R, D305G, A381T, Q821R). Three of the missense mutations occurred within the catalytic domain and based on the crystal structure of USP7, reside at the binding interface between the catalytic domain of USP7 and ubiquitin (Fig. 3b), a region that when mutated has been shown to impair ubiquitin binding[26]. C300R is predicted to structurally perturb one side of the USP7 ubiquitin binding pocket and A381T and D305G alter interactions with key ubiquitin binding residues (Supplementary Fig. 3–5). All except one of the *USP7* mutations (A381T) were found in TALL resulting in an overall mutation frequency of 8% in TALL. Of the 7 TALL cases with a *USP7* mutation, none had somatic mutations in *TP53*.

To directly assess the functional consequences of the *USP7* mutations identified in pediatric ALL, we transfected wild type and mutant USP7 (C300R and D305G) into 293T cells and assessed their effect on the level of mono-ubiquitinated H2B-K120, a known target of USP7[25]. Transfection resulted in similar levels of expression of the wild type and mutant USP7 proteins (Fig. 3c, 3d and Supplementary Fig. 6). As expected, enforced over-expression of wild type USP7 led to marked reduction in the amount of mono-ubiquitinated H2B-K120 (Fig. 3e, 3f and Supplementary Fig. 6). By contrast, expression of the USP7 mutants failed to alter the level of mono-ubiquitinated H2B-K120 (Fig. 3e, 3f and Supplementary Fig. 6).

Discussion

By performing sequence analysis on the entire genomic complement of genes that encode epigenetic regulatory proteins in over 1000 pediatric cancer samples, we have generated an initial view of the somatic mutational landscape of these genes across 21 different pediatric cancer subtypes, including the predominant forms of leukemia, brain tumors and solid malignancies seen in the pediatric population. Although our analysis is limited to SNVs and indels, these results demonstrated a marked variation in the frequency of mutations seen in the three major pediatric tumor types, with 30% of pediatric brain tumors and leukemias containing mutations and only 17% of pediatric solid tumors. Moreover, specific subtypes of brain tumors and leukemias exhibited an exceptionally high frequency of mutations in epigenetic regulator genes including 46% of high-grade gliomas with mutations in Histone H3 (this frequency increasing to 78% for pontine gliomas); 43% of the medulloblastomas and 56% of T-lineage ALLs with mutations in histone writers, erasers, and readers. At the other end of the spectrum were low-grade glioma and retinoblastoma, two tumor types that had almost no mutations within epigenetic regulatory genes.

Not only did the frequency of somatic mutation of these genes vary across the tumor types, but also the specific genes mutated showed some variation between tumor types. Focusing on the most commonly mutated genes, which function as part of eight key epigenetic protein complexes including PRC2, NuRD, MLL, ASCOM, Compass, BAF/PBAF, CREB, and DUU, we observed that over half of the mutations within the CREB, PRC2 and NuRD complexes occurred in pediatric leukemias, and all of the mutations in the DUU complex were identified in leukemias. By contrast, no significant differences were detected in the frequency of mutations within the BAF/PBAF and inter-related MLL/ASCOM/compass complexes across the pediatric cancer subtypes analyzed.

Within the DUU complex, we identified the recurrent somatic mutation of the *USP7* gene which encodes a deubiquitinase that interacts with p53, MDM2, DNMT1/UHRF1 and histones. Although rare somatic mutations of *USP7* have been found in adult cancers (http://cancer.sanger.ac.uk/cosmic), our structural modeling predicts that they would be well tolerated and thus likely represents passenger mutations (Supplementary Fig. 7 and Supplementary Table 2). By contrast, in pediatric cancer the majority of the *USP7* mutations identified are loss-of-function mutations including frame shift mutations within the catalytic domain that would encode truncated USP7 proteins and missense mutations that have reduced deubiquitinase activity. Importantly, the pediatric *USP7* mutations were exclusively detected in leukemias, with 6 of the 7 leukemias containing a mutation classified as non-ETP (or standard) T-ALL (6/46 [13%] of non-ETP T-ALLs contain a mutation in this gene). Defining the key intracellular proteins affected by the altered USP7 function and how these changes specifically contribute to the establishment of the non-ETP TALL malignant clone remains to be investigated.

Similarly, understanding how the other identified mutations alter the epigenetic landscape of a cell and contribute to transformation remains to be determined. This will require not only elucidating the effect of each mutation on the function of the encoded protein, but also determining how the mutant protein affects the epigenetic regulatory complexes in which it functions. This would require future investigation of how the altered function is influenced by the baseline epigenetic state of the target cell of transformation, and how this altered function complements other somatic mutations that are required for the development of overt cancer. The database developed by our work will help to focus these studies on the cell lineages that correspond to the tumor types in which specific mutations are detected.

Methods

Patients and samples

The use of human tissues for sequencing was approved by the institutional review boards of St Jude Children's Research Hospital, Memorial Sloan-Kettering Cancer Center, and Washington University in St Louis (St Jude IRB# FWA00004775, Protocol# XPD09-018). Written informed consent and/or assent were obtained from patients and/or legal guardians at the time of the surgical resection or bone marrow biopsy. Matched normal samples were obtained either from peripheral blood, bone marrow or adjacent normal tissue. All leukemia samples have 70% blasts. The tumor content for the four subtypes of brain tumors, HGG, LGG, MB and EPD, exceeds 50%, 67%, 90% and 95%, respectively. The tumor purity for solid tumors ranges from 48% to 96%.

Identification of genes involved in chromatin modification

We searched multiple data sources in order to identify the proteins that 1) bind a histone peptide, 2) modify nascent histone amino acids, 3) are part of established complexes involved in histone modification, 4) reorganize nucleosomes, or 5) modify or bind modified genomic DNA. A core set of proteins were identified that are known to directly modify histones or DNA[13,27,28], bind directly to modified or nascent histones[29,30], or alter chromatin state[31]. To expand our list to include additional homologs, we searched UniProt

database[32] for the known histone reader domains (Bromo, Tandem Bromo, Chromo, PHD, Tandem PHD, Tudor, Tandem, PWWP, MBT, WD40, ADD, Ankyrin Repeats, ZF-CW, and 14-3-3)[29] and catalytic modification domains (such as SET and Jumonji)[33]. The list was further expanded to include proteins within known complexes and potential complexes[15–22,31]. Potential epigenetic protein complexes were identified by using the core set of genes to search the STRING database (species=9606 and required score >900) for interaction partners[34]. The large numbers of proteins identified were culled down by manually verifying that the interaction to a search protein was functionally relevant to histone or, DNA modification or chromatin remodeling. All proteins were assigned a functional class (Writers, Erasers, Reader, Remodel chromatin, Modify DNA, Histone family, Binding histone eraser, Binding histone writer). In the case where proteins can be grouped in multiple classes, each protein was only assigned to the highest functional class available.

Sequencing and experimental verification

Whole-genome (n=434) and whole-exome (n=244) sequencing and analysis were described in detail elsewhere[6,7]. For 426 cases analyzed by custom capture, libraries used for the enrichment were constructed from repli-G WGA DNA (Qiagen) with TruSeq DNA sample prep kits (Illumina), following manufactures recommendations. Probe set for capturing all coding exons of the 633 chromatin modifying genes was designed using Design Studio (Illumina). The resulting probe set was then synthesized and provided as part of a TruSeq Custom Enrichment kit (Illumina). Library hybridization and enrichment of the targeted regions was conducted using the manufacturer's instructions. The enriched libraries were then sequenced on a HiSeq 2000 (Illumina) using V3 Chemistry (PE100 protocol), with 24 samples pooled per lane. Sequence data were analyzed using the same methods as those for WGS and WES. For cases that were not subjected to whole-genome or whole-exome sequencing, their TP53 mutation status was analyzed by Sanger sequencing of coding exons using an ABI3730.

The majority of the putative variants were validated by NGS amplicon sequencing using the Nextera XT library prep kit (Illumina) and sequenced on the MiSeq (Illumina). Following an effective validation protocol[35], the MiSeq paired-end 150-cycle protocol was performed with variants called by MiSeq reporter. - A subset of the putative variants was validated by amplicon Sanger Sequencing using an ABI3730 and the BigDye 3.1 cycle sequencing kit (Applied Biosystems). Amplicons used for validation were generated from WGA DNA prepared independently of the material used for the custom enrichment, with oligos designed by software based on Primer3[36]. The PCR was performed with 20ng of WGA DNA input using the AmpliTaq Gold 360 Master Mix (Life Technologies) per manufacturer's instructions. Samples with existing WGS or WES data corroborating the SNVs or indels observed in the targeted enrichment data were considered to be validated.

Functional and statistical significance of top 15 genes

Loss-of-function (LOF) mutations include indels or SNVs that result in frame shift, nonsense or affect splice sites. Functional significance of a missense mutation is determined for USP7 by majority rule (>=50% predict deleterious) using Polyphen[37]

(probably_deleterious and greater), Sift[38] (deleterious), and Mutation Assessor[39] (medium assignment or greater). Known activating mutations were annotated based on literature search.

Mutational significance was calculated for recurrently mutated genes. The background mutation rate (BMR) for WGS samples were estimated based on mutations in non-coding, non-repetitive regions (i.e. Tier3 data) and the disease specific median BMR estimate was used for the BMR of WES and Capture samples from the same disease type. The probability of a gene mutated in a specific sample under the null hypothesis of random background mutation was estimated from the amino acid length and the BMR of this sample. The probability of observing a gene mutated in at least *n* samples under the null hypothesis of random background mutation was estimated using a one tail Poisson binomial distribution.

Analysis of novel mutations in leukemia

To determine whether the mutations identified in leukemia are novel, we first searched PubMed for all genes with recurrent SNVs with and without the term "cancer". These genes were also used to mine the COSMIC dataset v63 (downloaded Feb 18, 2013). To further classify novel genes within leukemia, a similar PubMed search was performed with the search term "leukemia". This was accompanied by data mining of COSMIC to identify genes with mutations associated with the term "haematopoietic_and_lymphoid_tissue". The associated publications were reviewed to determine whether the mutations in published literature were identified in pediatric or adult patients.

Structural modeling and epigenetic regulator network

The structure of the catalytic domain of USP7 (Fig. 3b) bound to ubiquitin aldehyde (PDB: 1NBF) was obtained from the PDB[26,40]. Mutations and graphics were generated using Pymol[41]. The network graph was generated in Cytoscape[42].

USP7 mutagenesis and transfection

To demonstrate loss-of-function, two missense mutations identified in TALL (C300R and D305G) were introduced by site-direct mutagenesis (Agilent, Santa Clara, CA) on a wild type *USP7* cDNA construct (plasmid pCl-neo-Flag-HAUSP was deposited by Dr. Bert Vogelstein at addgene). The following primers were used 5′-CATGATGTTCAGGAGCTTCGTCGAGTGTTGCTCGA-3′ for C300R-F and 5′-TCGAGCAACACTCGACGAAGCTCCTGAACATCATG-3′ for C300R-R and 5′-GCTTTGTCGAGTGTTGCTCGGTAATGTGGAAAATAAGATGA-3′ for D305G-F and 5′-TCATCTTATTTTCCACATTACCGAGCAACACTCGACAAAGC-3′ for D305G-R. All constructs were sequenced for verification (Supplementary Fig. 8). 3×10^5 293T cells (ATCC, Cat# CRL-11268) per well of a 6-well plate were cultured in DMEM (Lonza, Walkersville, MD) with 10% of FBS (Sigma, Atlanta, GA). 2ug of plasmid DNA was transfected into cells with X-tremeGENE HP DNA transfection reagent (Roche, Indianapolis, IN).

Western blot

Total protein of 293T cells was extracted at 72 hours post transfection. Protein levels of USP7, total H2B and H2B ubiquityl Lys120 were detected by western blot with indicated antibodies. Human HAUSP (USP7) (Catalog # PA5-17179) and GAPDH (Catalog # MA5-15738) antibodies were purchased from Thermo Scientific (Rockford, IL), human H2B (catalog # 39126) and H2BK120ub1 (Catalog # 39624) antibodies were purchased from Active Motif (Carlsbad, CA). Secondary goat anti-rabbit (Catalog # ab97051) or anti-mouse (Catalog # ab97265) antibodies were purchased from abcam (Cambridge, MA). Briefly, the blots were incubated in the 1:1000 diluted primary antibodies overnight at 4 degree C and followed by incubating in 1:5000 diluted secondary antibodies. The protein bands were detected by SuperSignal West Femto Maximum Sensitivity Substrate (Catalog # 34096) purchased from Thermo Scientific (Rockford, IL).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

References

1. Downing JR, et al. The Pediatric Cancer Genome Project. Nat Genet. 2012; 44:619–22. [PubMed: 22641210]

2. Schwartzentruber J, et al. Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. Nature. 2012; 482:226–31. [PubMed: 22286061]

3. Wu G, et al. Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. Nat Genet. 2012; 44:251–3. [PubMed: 22286216]

4. Mullighan CG, et al. CREBBP mutations in relapsed acute lymphoblastic leukaemia. Nature. 2011; 471:235–9. [PubMed: 21390130]

5. Holmfeldt L, et al. The genomic landscape of hypodiploid acute lymphoblastic leukemia. Nat Genet. 2013

6. Zhang J, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. Nature. 2012; 481:157–63. [PubMed: 22237106]

7. Robinson G, et al. Novel mutations target distinct subgroups of medulloblastoma. Nature. 2012; 488:43–8. [PubMed: 22722829]

8. Cheung NK, et al. Association of age at diagnosis and genetic mutations in patients with neuroblastoma. JAMA. 2012; 307:1062–71. [PubMed: 22416102]

9. Neff T, Armstrong SA. Recent progress toward epigenetic therapies: the example of mixed lineage leukemia. Blood. 2013; 121:4847–53. [PubMed: 23649466]

10. Jaffe JD, et al. Global chromatin profiling reveals NSD2 mutations in pediatric acute lymphoblastic leukemia. Nat Genet. 2013; 45:1386–91. [PubMed: 24076604]

11. Lewis PW, et al. Inhibition of PRC2 activity by a gain-of-function H3 mutation found in pediatric glioblastoma. Science. 2013; 340:857–61. [PubMed: 23539183]

12. Chan KM, et al. The histone H3.3K27M mutation in pediatric glioma reprograms H3K27 methylation and gene expression. Genes Dev. 2013; 27:985–90. [PubMed: 23603901]

13. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012; 150:12–27. [PubMed: 22770212]

14. Mannini L, Liu J, Krantz ID, Musio A. Spectrum and consequences of SMC1A mutations: the unexpected involvement of a core component of cohesin in human disease. Hum Mutat. 2010; 31:5–10. [PubMed: 19842212]

15. Smith E, Lin C, Shilatifard A. The super elongation complex (SEC) and MLL in development and disease. Genes Dev. 2011; 25:661–72. [PubMed: 21460034]

16. Schuettengruber B, Martinez AM, Iovino N, Cavalli G. Trithorax group proteins: switching genes on and keeping them active. Nat Rev Mol Cell Biol. 2011; 12:799–814. [PubMed: 22108599]

17. Goo YH, et al. Activating signal cointegrator 2 belongs to a novel steady-state complex that contains a subset of trithorax group proteins. Mol Cell Biol. 2003; 23:140–9. [PubMed: 12482968]

18. Ramirez J, Hagman J. The Mi-2/NuRD complex: a critical epigenetic regulator of hematopoietic development, differentiation and cancer. Epigenetics. 2009; 4:532–6. [PubMed: 19923891]

19. Richly H, Aloia L, Di Croce L. Roles of the Polycomb group proteins in stem cells and cancer. Cell Death Dis. 2011; 2:e204. [PubMed: 21881606]

20. Reisman D, Glaros S, Thompson EA. The SWI/SNF complex and cancer. Oncogene. 2009; 28:1653–68. [PubMed: 19234488]

21. Wu RC, et al. Regulation of SRC-3 (pCIP/ACTR/AIB-1/RAC-3/TRAM-1) Coactivator activity by I kappa B kinase. Mol Cell Biol. 2002; 22:3549–61. [PubMed: 11971985]

22. Felle M, et al. The USP7/Dnmt1 complex stimulates the DNA methylation activity of Dnmt1 and regulates the stability of UHRF1. Nucleic Acids Res. 2011; 39:8355–65. [PubMed: 21745816]

23. Sarkari F, Sheng Y, Frappier L. USP7/HAUSP promotes the sequence-specific DNA binding activity of p53. PLoS One. 2010; 5:e13040. [PubMed: 20885946]

24. Song MS, et al. The deubiquitinylation and localization of PTEN are regulated by a HAUSP-PML network. Nature. 2008; 455:813–7. [PubMed: 18716620]

25. van der Knaap JA, et al. GMP synthetase stimulates histone H2B deubiquitylation by the epigenetic silencer USP7. Mol Cell. 2005; 17:695–707. [PubMed: 15749019]

26. Hu M, et al. Crystal structure of a UBP-family deubiquitinating enzyme in isolation and in complex with ubiquitin aldehyde. Cell. 2002; 111:1041–54. [PubMed: 12507430]

27. Kouzarides T. Chromatin modifications and their function. Cell. 2007; 128:693–705. [PubMed: 17320507]

28. Bountra C, Oppermann U, Heightman TD. Animal models of epigenetic regulation in neuropsychiatric disorders. Curr Top Behav Neurosci. 2011; 7:281–322. [PubMed: 21225415]

29. Yun M, Wu J, Workman JL, Li B. Readers of histone modifications. Cell Res. 2011; 21:564–78. [PubMed: 21423274]

30. Khare SP, et al. HIstome--a relational knowledgebase of human histone proteins and histone modifying enzymes. Nucleic Acids Res. 2012; 40:D337–42. [PubMed: 22140112]

31. Lans H, Marteijn JA, Vermeulen W. ATP-dependent chromatin remodeling in the DNA-damage response. Epigenetics Chromatin. 2012; 5:4. [PubMed: 22289628]

32. Magrane M, Consortium U. UniProt Knowledgebase: a hub of integrated protein data. Database (Oxford). 2011; 201110.1093/database/bar009

33. Arrowsmith CH, Bountra C, Fish PV, Lee K, Schapira M. Epigenetic protein families: a new frontier for drug discovery. Nat Rev Drug Discov. 2012

34. Jensen LJ, et al. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009; 37:D412–6. [PubMed: 18940858]

35. Zhang J, et al. Whole-genome sequencing identifies genetic alterations in pediatric low-grade gliomas. Nat Genet. 2013; 45:602–12. [PubMed: 23583981]

36. Rozen S, Skaletsky H. Primer3 on the WWW for general users and for biologist programmers. Methods Mol Biol. 2000; 132:365–86. [PubMed: 10547847]

37. Adzhubei IA, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–9. [PubMed: 20354512]

38. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009; 4:1073–81. [PubMed: 19561590]

39. Reva B, Antipin Y, Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Res. 2011; 3910.1093/nar/gkr407

40. Berman HM, et al. The Protein Data Bank. Nucleic Acids Res. 2000; 28:235–42. [PubMed: 10592235]

41. Schrödinger, L. The PyMOL Molecular Graphics System. 2011. Version 1.3

42. Cline MS, et al. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc. 2007; 2:2366–82. [PubMed: 17947979]
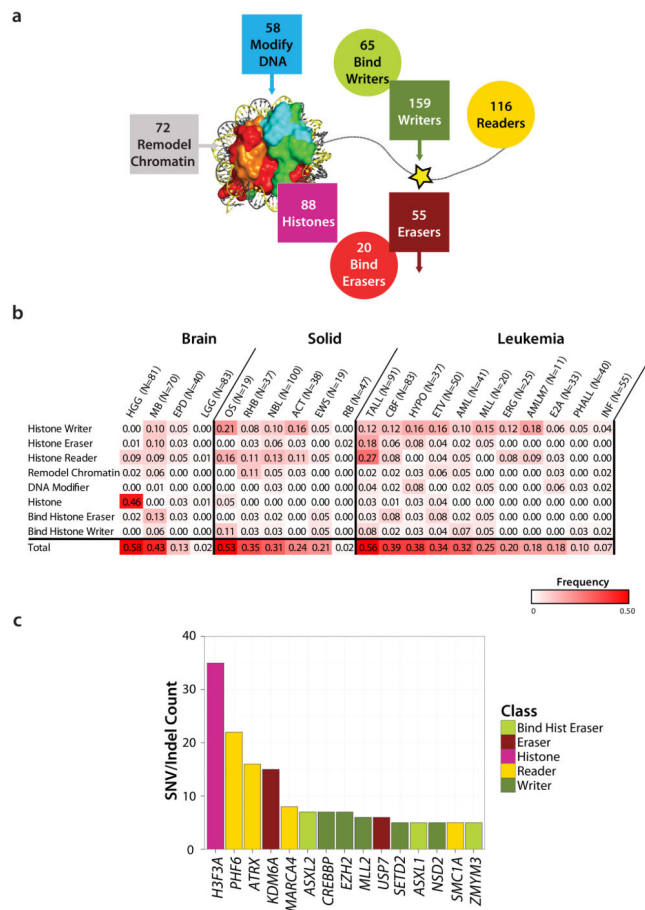
Figure 1. The landscape of somatic mutations in epigenetic regulators in 21 pediatric cancer subtypes

(**a**) Eight classes of epigenetic genes were interrogated across the cohort (Histone Writer, Bind Histone Writer, Histone Eraser, Bind Histone Eraser, Histone, Histone Reader, Chromatin Modifier, and DNA modifier), with the numbers of genes within each class indicated. (**b**) Fraction of tumors in each cancer subtype with at least one mutation in each class of epigenetic genes. Only sequence mutations (i.e. SNVs and indels) with a mutant allele fraction >0.3 (i.e. present in the dominant clone) were included in the analysis (**c**) Top 15 most frequently mutated genes colored coded by class.
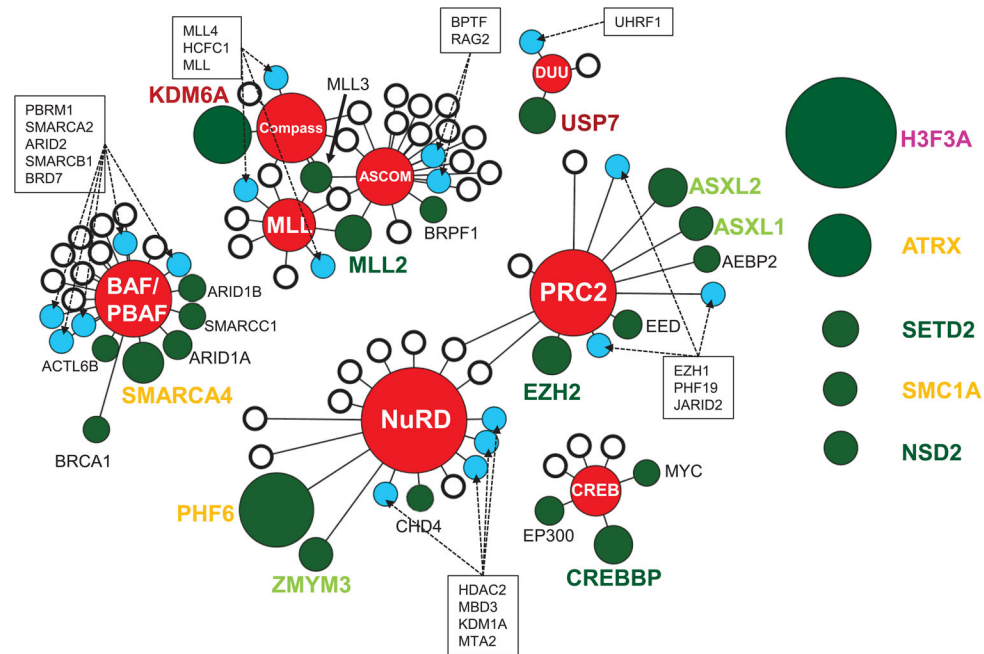
Figure 2. Epigenetic complexes affected by recurrently mutated proteins in pediatric cancer
A subset (35%) of the recurrently mutated epigenetic regulatory proteins (green circles) function within one or more of eight key epigenetic protein complexes (red nodes). Individual somatic mutation were also detected in additional components of these complexes (blue circles), whereas other components were never found to be mutated within our patient cohort (white circles). The size of each green circle is proportional to the number of mutated samples. The distance between the circles and the central complex node indicates whether the protein is a core (short) or transient (long) component of the complex. Recurrently mutated proteins that do not belong to one of these core complexes are presented on the right as unattached circles. The color of each protein name conforms to the color scheme for epigenetic regulatory classes presented in Figure 1.
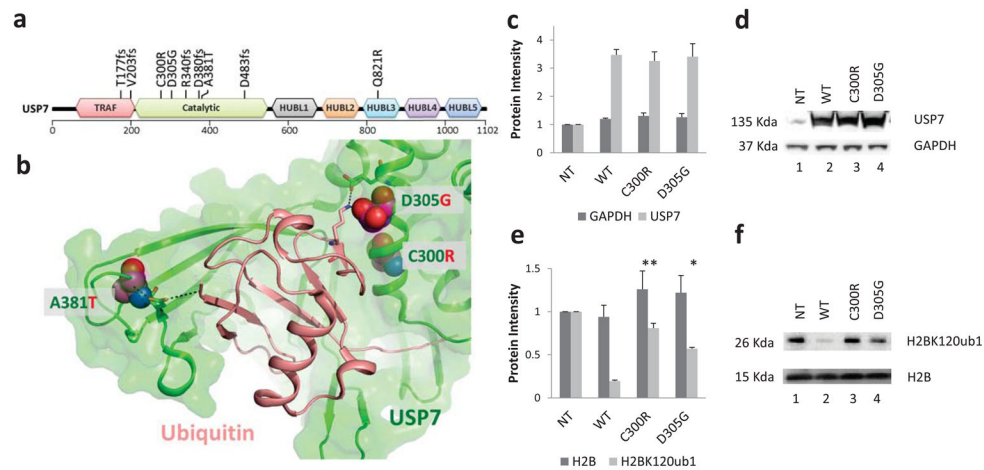
Figure 3. Novel ALL-specific mutations of USP7

(**a**) Location of the identified USP7 somatic mutations relative to the TRAF (tumor necrosis factor [TNF] receptor-associated factor), catalytic, and HUBL1-5 (USP7/HAUSP ubiquitin-like domain) domains (colored red, green, black, orange, teal, purple and blue respectively). Mutations C300R, D483fs, and Q821R occurred at mutant allele frequencies (MAF) below 30%, whereas all other mutations occurred with MAF >30% and thus represent the dominant malignant clone. (**b**) Location of the missense somatic mutations (C300R, D305G, and A381T: magenta space filled) within the USP7 catalytic domain (green cartoon) – Ubiquitin (peach cartoon) interface. Specific residues and interactions between USP7 and Ubiquitin are shown as sticks and black dots and further described in Supplementary Fig. 3–5. (**c** and **d**) 293T cells were transfected with USP7 WT or mutant constructs as indicated. Protein extracts were prepared at 72 hours post transfection and subjected to western blot analysis using antibodies to the indicated proteins. Bars represent mean of protein band intensities of 3 replicates ± S.E.M. (**e** and **f**). The level of Histone H2B ubiquityl Lys120 (H2BK120ub1) and total H2B were detected at 72 hours by immunoblot using an antibody specific for mono-ubiquitinated and total H2B. Bars represent mean of protein band intensities of 3 replicates ± S.E.M. NT, untransfected control. The statistical significance of the changes observed between wild type and USP7 mutants were assessed by t-test with ***** equal to a p<0.05 and ****** equal to p<0.01.

Table 1

Pediatric Tumor Dataset

Disease Type	WGS Pairs	Exome Pairs	Histone Capture Pairs	WGS/Exome Overlap	CC/WGS Overlap	CC/Exome Overlap	Total Sample Pairs
Brain Tumor							
High Grade Glioma (HGG)	36	45	3	0	3	0	81
Low Grade Glioma (LGG)	35	0	48	0	0	0	83
Medulloblastoma (MB)	36	0	49	0	15	0	70
Ependymoma (EPD)	40	0	0	0	0	0	40
Solid Tumor							
Neuroblastoma (NBL)	38	0	79	0	17	0	100
Retinoblastoma (RB)	4	0	46	0	3	0	47
Adrenocortical Carcinoma (ACT)	20	18	0	0	0	0	38
Rhabdomyosarcoma (RHB)	13	3	26	0	3	2	37
Osteosarcoma (OS)	19	0	2	0	2	0	19
Ewing's Sarcoma (EWS)	19	0	0	0	0	0	19
Leukemia							
Infant Acute Lymphoblastic Leukemia (INF ALL)	23	6	33	0	7	0	55
Mixed Lineage leukemia (MLL)	0	20	0	0	0	0	20
T-lineage ALL (TALL)	12	3	89	0	10	3	91
ETV/RUNX1 translocation ALL (ETV)	50	1	0	1	0	0	50
Philadelphia Chromosome-Positive ALL (PHALL)	24	18	0	2	0	0	40
TLS-ERG translocation ALL (ERG)	14	12	0	1	0	0	25
Hypodiploid ALL (HYPO)	20	17	4	0	4	0	37
E2A/PBX1 translocation ALL (E2A)	10	24	0	1	0	0	33
Core Binding Factor Acute Myeloid Leukemia (CBF AML)	17	66	4	0	4	0	83
Acute Megakaryoblastic Leukemia (AMLM7)	4	11	2	4	2	2	11
Other AML (AML)	0	0	41	0	0	0	41
Total Pairs	*434*	*244*	*426*	*9*	*70*	*7*	*1020*

Each sample represents a tumor/germline pair and is categorized by the type of cancer (Brain, Solid and Leukemia). The numbers of samples sequenced by each method (Whole-genome sequencing [WGS], whole exome [exome], or custom designed Histone capture sequencing [CC]) are listed by cancer subtype.