



Published in final edited form as:

Nat Genet. 2010 February ; 42(2): 181–185. doi:10.1038/ng.518.

Somatic mutation of EZH2 (Y641) in Follicular and Diffuse Large B-cell Lymphomas of Germinal Center Origin

Ryan D. Morin¹, Nathalie A. Johnson², Tesa M. Severson¹, Andrew J. Mungall¹, Jianghong An¹, Rodrigo Goya¹, Jessica E. Paul¹, Merrill Boyle², Bruce W. Woolcock², Florian Kuchenbauer², Damian Yap², R. Keith Humphries², Obi L. Griffith¹, Sohrab Shah², Henry Zhu³, Michelle Kimbara³, Pavel Shashkin³, Jean F. Charlot³, Marianna Tcherpakov³, Richard Corbett¹, Angela Tam¹, Richard Varhol¹, Duane Smailus¹, Michelle Moksa¹, Yongjun Zhao¹, Allen Delaney¹, Hong Qian¹, Inanc Birol¹, Jacqueline Schein¹, Richard Moore¹, Robert Holt¹, Doug E. Horsman⁴, Joseph M. Connors², Steven Jones¹, Samuel Aparicio², Martin Hirst¹, Randy D. Gascoyne⁴, and Marco A. Marra^{*,1,6}

¹ Genome Sciences Centre, BC Cancer Agency, 675 W. 10th Avenue, Vancouver, Canada

² BC Cancer Agency, 600 W. 10th Avenue, Vancouver, Canada

³ BPS Biosciences, 11526 Sorrento Valley Rd., Ste. A2, San Diego, CA 92121

⁴ Department of Pathology, University of British Columbia and BC Cancer Agency, 600 W. 10th Avenue, Vancouver, Canada

⁵ Division of Medical Oncology, University of British Columbia BC Cancer Agency, 600 W. 10th Avenue, Vancouver, Canada

⁶ Department of Medical Genetics, University of British Columbia, 675 W. 10th Avenue, Vancouver, Canada

Keywords

RNA-seq; WTSS; PRC2; genome; transcriptome; epigenetic

Follicular lymphoma (FL) and the GCB subtype of diffuse large B-cell lymphoma (DLBCL) derive from germinal center B-cells¹. Targeted re-sequencing studies have revealed mutations in various genes in the NFκB pathway^{2,3} that contribute to the activated B-cell

Users may view, print, copy, download and text and data- mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use: http://www.nature.com/authors/editorial_policies/license.html#terms

*To whom correspondence should be addressed.

Author Contributions: M.A.M., R.D.G., D.E.H., and J.M.C. conceived of the study and led the design of the experiments. R.D.M. performed the WGSS and WTSS analysis, produced Figures 1 and 2 and, with M.A.M., wrote the manuscript. N.A.J. prepared the samples, performed sample sorting and COO analysis and contributed to the text. O.G. and R.D.M. analyzed gene expression data. T.M.S., A.J.M. and J.E.P. performed sequence validation experiments and visual inspection of capillary sequence data. D.S. and M.M. constructed multiplexed libraries for deep re-sequencing of EZH2. H.Z., M.K., P.S., J.F.C., D.Y. and M.T. conducted enzymatic assays. I.B. performed statistical analysis and contributed to the manuscript. J.A. and S.J. produced the model for EZH2 and contributed to the manuscript. M.B. and B.W. prepared samples and performed FACS. F.K. and R.K.H. validated expression findings in the RNA. A.D., H.Q., R.C. and S.S. performed copy number analysis. A.T., Y.Z., R.H., M.H. and R.M. produced the sequencing libraries and performed the sequencing. R.V. processed raw sequencing data. R.G. identified candidate mutations. J.S., M.H., S.A. conceived of experiments and contributed to the text.

(ABC) DLBCL subtype but, thus far, few GCB-specific mutations have been identified⁴. Here, we report recurrent somatic mutations affecting the polycomb group oncogene⁶ *EZH2*, which encodes a histone methyltransferase responsible for tri-methylating H3K27. After the recent discovery of mutations in the histone H3K27me3 demethylase UTX in several cancer types⁵, *EZH2* is the second histone methyltransferase affected by mutations in cancer. These mutations, which replace a single tyrosine in the SET domain of the *EZH2* protein (Y641), occur in 21.7% of GCB DLBCLs, 7.2% of FLs and are absent from ABC DLBCLs. Our data are consistent with the notion that *EZH2* proteins with mutated Y641 have reduced enzymatic activity *in vitro*.

Advances in DNA sequencing technology have recently enabled the characterization of genomes and transcriptomes at sufficient resolution for identification of somatic point mutations^{7–9}. To develop new insight into novel mutations potentially contributing to B-cell Non Hodgkin lymphomas (NHL), we used Illumina technology to sequence genomic DNA and RNA purified from a malignant lymph node biopsy obtained from a FL patient (“FL Sample A”)(Methods). FL patient A qualified as a grade 1 FL that co-expressed CD10, BCL2 and BCL6 by immuno-histochemistry. This sample was chosen for sequence analysis because it exhibited an unusually simple karyotype (Supplementary Figure 1), lacking t(14;18)(q32;q21), and other large-scale alterations (Supplementary Figures 2–5; Supplementary Tables 1 and 2). We analyzed the exon sequences of this tumor for mutations in both the genome (WGSS) and transcriptome (WTSS)(Table 1 and Methods). Matched constitutional DNA from the FL patient was sequenced to reveal “germline” sequence variants (Methods). We produced 25.6 aligned gigabases (Gb) from the tumor genomic library, yielding 9.47-fold redundant base coverage on average, and an additional 2.2 Gb of aligned sequence from the WTSS library, yielding 18.86-fold redundant base coverage on average within exons (Table 1; Methods). We focused our analysis on novel changes predicted to affect protein-coding sequence (Methods). Among these variants (Supplementary Table 3), we discovered a mutation affecting exon 15 of the *EZH2* gene, which encodes a portion of the *EZH2* SET domain. *EZH2* is the catalytic component of the PRC2 complex, which is responsible for adding methyl groups to lysine 27 of histone 3 (H3K27)¹⁰, thereby repressing transcription at loci associated with histones bearing tri-methylated H3K27. We established that this mutation, which is predicted to replace Y641 with a histidine, was somatic in nature by confirming its presence in tumor DNA, and absence in constitutional non-tumor DNA (Methods). We also confirmed that the mutation was heterozygous in this patient sample.

To determine *EZH2* mutational status in DLBCL samples, we next used WTSS to sequence the transcriptomes of 31 DLBCL patient samples and seven DLBCL cell lines. Based on cell of origin (COO) expression classification¹¹, the primary lymphoma samples were classified as belonging to either the ABC (n=12), GCB (n=15) or unclassifiable subtypes (n=2) (Methods; Supplementary Table 4). Coverage of *EZH2* in the WTSS libraries was consistently high, ranging in the patient samples from 5.3-fold to 187-fold redundant base-pair coverage (median 48.7-fold). Coverage of codon 641 ranged from 5-fold to 295-fold (median 46.5-fold; Supplementary Table 4). We identified Y641 mutations in four of these 31 patient samples and four of the cell lines (Table 2). No other mutations in *EZH2* were

detected and all mutations appeared to be heterozygous. The striking recurrence of these mutations suggested that mutation of Y641 in *EZH2* was a common feature of lymphoma. Notably, despite a median base coverage depth of 11.4-fold in the UTX locus, we found no evidence for UTX mutations in these libraries.

We determined the prevalence of Y641 mutations in both FL and DLBCL tumors by Sanger sequencing the exon containing this codon in 251 FL samples, of which 30 had matched DLBCL samples taken at histological transformation and 320 primary DLBCL samples (including the original 31 patient samples)(Supplementary Table 5). This revealed a total of 18 FL and 35 DLBCL samples with heterozygous Y641 mutations. Of note, all Y641 mutations detected by WTSS demonstrated clear evidence for expression of both alleles (Supplementary Table 4). To search for additional sites mutated in this gene, we also sequenced all exons of *EZH2* in 24 FL patients (in addition to FL patient A) and found only one example of an *EZH2* mutation not affecting Y641 (Figure 1; Supplementary Table 5). This mutation, affecting N635, was found in conjunction with a Y641 mutation and we confirmed that both mutations were in a cis orientation. We confirmed that these mutations were somatic in the 7 FL (including “Sample A”) and 2 DLBCL patients for which germline DNA was available.

To exclude the possibility that such mutations can also occur in non-malignant germinal center B-cells or other types of lymphoma, we sequenced this region of exon 15 in eight CD77⁺-enriched centroblast samples from reactive tonsils and 23 reactive lymph nodes (a source of normal B-cells) and 80 samples of other lymphoma types using both Sanger sequencing and targeted ultra-deep Illumina re-sequencing (Methods; Supplementary Figure 6; Supplementary Table 6). We also sequenced WTSS libraries generated from two additional normal centroblast samples (Supplementary Table 4). Consistent with our hypothesis that Y641 mutations are unique to malignant B-cells, none of these samples showed evidence for mutations at Y641 or elsewhere within the sequenced region (Table 3). Notably, all of the DLBCL samples with known COO and which were also positive for *EZH2* mutations were of the GCB subtype and not the ABC subtype. This revealed a significant enrichment of Y641 mutations among the GCB subtype of DLBCLs (Table 3; n=18/83 GCB vs 0/42 ABC, P = 0.00168, two-tailed Fisher’s Exact Test).

We next assessed the effect various Y641 mutations would have on the structure, and potentially the function, of the *EZH2* SET domain by generating a computational model (Supplementary Figure 7) using the crystal structure of the highly conserved MLL1 SET domain¹² as the structural template (Methods). Our model indicates that Y641 interacts with the lysine 27 side chain of the H3 histone tail, as has been suggested in other SET domain proteins¹³. Though no *EZH2* SET domain mutations have been reported in humans, detailed mutant phenotypes have been described in *Drosophila*. A mutation altering the tyrosine orthologous to *EZH2* Y641 has been characterized in the *Drosophila* ortholog *E(z)* in an allele known as “E(z)¹”. *Drosophila* E(z)¹ mutant protein was found to be incapable of trimethylating H3K27 *in vitro*¹⁴.

We sought to directly determine whether *EZH2* with mutated Y641 impacts the catalytic activity of PRC2 in a cell-free methylation assay. Individual clones, each containing one of

the four most frequently detected mutations (Figure 1), were first expressed along with the other components of PRC2. PRC2 complexes were purified and tested *in vitro* for H3K27 tri-methylation activity using ELISA and an antibody specific for H3K27me3 (Methods). The results (Figure 2) indicated that, compared to wild-type EZH2, all four Y641 mutants consistently demonstrated a marked reduction (~7-fold) in their ability to tri-methylate the H3K27 peptide. This biochemical result suggested that the four predominant Y641 variants observed in our sequencing study could confer reduced ability of PRC2 complexes to tri-methylate H3K27 *in vivo*.

Other reports have suggested that increased abundance of *EZH2* mRNA correlates with cancer progression in tissues in which *EZH2* expression is normally low or undetectable, such as breast and prostate^{6,15}. However, *EZH2* mRNA is known to be abundant in normal germinal center B-cells¹⁶, and a conditional knock-out of the mouse *EZH2* ortholog indicated that the SET domain is required for early B-cell development, including rearrangement of the immunoglobulin heavy chain (*IGH*) locus¹⁷. Given the apparent requirement for *EZH2* in germinal center B-cells, it is possible that the mechanism by which *EZH2* contributes to lymphomagenesis is distinct from the apparently straightforward increases in *EZH2* mRNA abundance observed in breast⁶ and prostate¹⁵ cancers. Expression of both *EZH2* and *BM11* (the catalytic component of PRC1) have been linked to the degree of malignancy of B-cell NHL and perturbations in the balance of the quantities of these two proteins has been suggested as an early event in lymphomagenesis¹⁸. However, mutation of *EZH2* has not, to date, been implicated in B-cell malignancies or any other cancer.

Though the biological mechanism is not known, our findings suggest that *EZH2* Y641 mutations, and possibly a reduction in H3K27 tri-methylation, have a role in the pathogenesis of GCB lymphomas. The well-studied Phe/Tyr switch¹⁹ site is known to regulate the number of methyl groups a SET domain-containing protein can add without compromising its overall catalytic activity. Although the Y641 residue is distinct from the Phe/Tyr switch site, the result of our *in vitro* experiment does not rule out the possibility that these mutations may alter the product (or target) specificity of EZH2. Our finding is particularly timely in light of recent studies demonstrating enhanced DNA methylation at PRC2 targets in lymphoma as compared to normal B-cells^{20,21}. H3K27 tri-methylation via PRC2 can be a precursor to DNA methylation and, in some cases, DNA methyltransferase may be physically coupled with PRC2²². Hence, Y641 mutations may contribute to the differential DNA methylation that has been observed at polycomb targets in FL²⁰ and DLBCL²¹.

In conclusion, we identified novel recurrent somatic mutations affecting the *EZH2* SET domain and have associated these with FL and DLBCL cases of only the GCB subtype. Our data indicate that mutation of the *EZH2* SET domain is among the most frequent genetic events observed in GCB malignancies after t(14;18)(q32;q21). The mutated tyrosine corresponds to a key residue in the active site of the EZH2 protein and, consistent with functional studies of a comparable mutation in the *Drosophila E(z)* ortholog, we show that PRC2 complexes containing mutated EZH2 protein exhibit reduced H3K27 tri-methylation activity *in vitro*. We have shown that a wild-type copy of EZH2 is present in all samples with Y641 mutations and have also detected expression of both alleles in the mutant samples

profiled by transcriptome sequencing. This, along with the fact that all lymphomas with mutations in *EZH2* appear to have a mutation affecting Y641, sets *EZH2* apart from the pattern of mutational inactivation seen in the case of *UTX*, which appears to behave as a tumor suppressor gene⁵. Aside from the recurrence of inactivating mutations in *UTX*, *EZH2* is the only protein affecting H3K27 methylation status to be identified as a target of somatic mutation in cancer, and the first in which recurrent mutations of the SET domain appear restricted to a specific lymphoma subtype.

Methods

Sample Acquisition

Two samples from the initial patient (FL patient A) were utilized. Both had ~70% tumor content based on the co-expression of CD19 and lambda by flow cytometry and were “fresh” frozen at source. The first was taken at the time of diagnosis and was used for WTSS and WGSS. The second was acquired at the time of progression and was flow sorted to >95% purity. It was analyzed by karyotype and fluorescence *in situ* hybridization (FISH) for the presence of a translocation t(14;18) using the dual color, dual fusion probe. It was also analyzed for copy number alterations by array comparative hybridization (aCGH) and fingerprint profiling²⁶ and for loss of heterozygosity (LOH) by Affymetrix 500K array. All DLBCL samples profiled by WTSS were fresh frozen biopsies having >50% tumor content by flow cytometry. All other specimens used in this study were obtained at the time of diagnosis and were derived from archived fresh frozen tissue or frozen tumor cell suspensions. Germline DNA was obtained from peripheral blood in live patients and from CD19-negative sorted tumor cell suspensions using Miltenyi magnetic beads (Miltenyi Biotec, Bergisch Gladbach, Germany) for deceased patients. All lymphoma samples were diagnosed according to the World Health Organization criteria of 2008 by an expert hematopathologist (R.D.G.). Benign specimens included reactive pediatric tonsils or purified CD77-positive centroblasts sorted from reactive tonsils using Miltenyi beads (Miltenyi Biotec, Bergisch Gladbach, Germany). The tumor specimens were collected as part of a research project approved by the University of British Columbia-British Columbia Cancer Agency Research Ethics Board (BCCA REB) and are in accordance with the Declaration of Helsinki. Informed consent was obtained from all patients whose samples were profiled using WTSS or WGSS. Our protocols stipulate that this data will not be released into the public domain but can be made available via a tiered-access mechanism to named investigators of institutions agreeing by a materials transfer agreement to honor the same ethical and privacy principles required by the BCCA REB.

Preparation and sequencing of Illumina libraries

RNA was extracted from a total lymph node section using AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA, USA) and DNaseI treated. For whole transcriptome shotgun sequencing (WTSS/RNA-seq) analysis, we used a modified method similar to the protocol we have previously described⁷. Briefly, PolyA⁺ RNA was purified using the MACS mRNA isolation kit (Miltenyi Biotec, Bergisch Gladbach, Germany), from 5–10ug of DNaseI-treated total RNA as per the manufacturer’s instructions. Double-stranded cDNA was synthesized from the purified polyA⁺ RNA using the Superscript Double-Stranded cDNA

Synthesis kit (Invitrogen, Carlsbad, CA, USA) and random hexamer primers (Invitrogen) at a concentration of 5 μ M. The cDNA was fragmented by sonication and a paired-end sequencing library prepared following the Illumina paired-end library preparation protocol (Illumina, Hayward, CA, USA).

Genomic DNA for construction of whole genome shotgun sequencing (WGSS) libraries was prepared from the same biopsy material using the Qiagen AllPrep DNA/RNA Mini Kit (Qiagen, Valencia, CA, USA). DNA quality was assessed by spectrophotometry (260/280 and 260/230) and gel electrophoresis before library construction. DNA was sheared for 10 min using a Sonic Dismembrator 550 with a power setting of “7” in pulses of 30 seconds interspersed with 30 seconds of cooling (Cup Horn, Fisher Scientific, Ottawa, Ontario, Canada), and analyzed on 8% PAGE gels. The 200–300bp DNA fraction was excised and eluted from the gel slice overnight at 4°C in 300 μ l of elution buffer (5:1, LoTE buffer (3 mM Tris-HCl, pH 7.5, 0.2 mM EDTA)-7.5 M ammonium acetate), and was purified using a Spin-X Filter Tube (Fisher Scientific), and by ethanol precipitation. WGSS libraries were prepared using a modified paired-end protocol supplied by Illumina Inc. (Illumina, Hayward, USA). This involved DNA end-repair and formation of 3' Adenosine overhangs using Klenow fragment (3' to 5' exo minus) and ligation to Illumina PE adapters (with 5' overhangs). Adapter-ligated products were purified on QIAquick spin columns (Qiagen, Valencia, CA, USA) and PCR-amplified using Phusion DNA polymerase (NEB, Ipswich, MA, USA) and 10 cycles with the PE primer 1.0 and 2.0 (Illumina). PCR products of the desired size range were purified from adapter ligation artifacts using 8% PAGE gels. DNA quality was assessed and quantified using an Agilent DNA 1000 series II assay (Agilent, Santa Clara CA, USA) and Nanodrop 7500 spectrophotometer (Nanodrop, Wilmington, DE, USA) and DNA was subsequently diluted to 10nM. The final concentration was confirmed using a Quant-iT dsDNA HS assay kit and Qubit fluorometer (Invitrogen, Carlsbad, CA, USA). For sequencing, clusters were generated on the Illumina cluster stations using v1 cluster reagents. Paired-end reads were generated using v3 sequencing reagents on the Illumina GA_{ii} platform following the manufacturer's instructions. Image analysis, base-calling and error calibration were performed using v1.0 of Illumina's Genome analysis pipeline. Paired-end WTSS and WGSS libraries were sequenced to 36, 50 or 76 cycles. The WGSS library comprised a mixture of 13 flow cell lanes of 36-nt reads, 16 lanes of 50-nt reads and 6 lanes of 76-nt reads.

Targeted ultra-deep re-sequencing using read indexing

This procedure describes the individual PCR amplification of EZH2 exon 15, indexing of individual amplicons, and subsequent pooling and sequencing. Individual indexes allow the deconvolution of reads deriving from individual samples in multiplexed libraries such that many samples can be concurrently sequenced in the same library. Genomic DNA from individual samples was normalized to 5ng/ μ L and 5ng of each sample was PCR amplified using Phusion DNA polymerase (New England Biolabs, Ipswich, MA, USA) in 96-well format using gene specific primers (Primer EZH2_015R3 and Primer EZH2_015F, Supplementary Table 7) to produce ~300bp amplicons. Hot Start PCR conditions: 98°C, 60s, then 36 cycles (98°C-10s, 60°C-15s, 72°C-30s), final extension 72°C, 5min. Amplicons were cleaned using AMPure beads (Beckman Coulter, CA, USA) on a Biomek F/X

(Beckman Coulter, Fullerton, CA, USA) and eluted with 40uL elution buffer EB (QIAGEN, USA). Cleaned amplicons were QC tested on a 1.2% SeaKem LE Agarose gel (Cambrex, East Rutherford, NJ, USA) using 1X TAE buffer. Bands were quantified by the QBit™ Fluorometer (Invitrogen, Carlsbad, CA, USA) high sensitivity assay. Approximately 500ng of each amplicon DNA sample was then phosphorylated and end-repaired in 50uL reactions at room temp, 30min (T4 DNA Pol 5U, DNA Pol I (Klenow) 1U, T4 PNK 100U, dNTP mix 0.4mM (Invitrogen). End-repair reactions were cleaned using AMPure beads and dATP was added to the 3'-ends using Klenow (exo-) 5U and 0.2mM dATP in 1X Klenow Buffer (Invitrogen) with 30-min incubation at 37°C in a Tetrad thermal cycler (MJ Research, USA). DNA was again cleaned on AMPure beads using a Biomek FX. Adapter ligation (10:1 ratio) was completed with 0.03uM Adapter (Multiplexing Adapters 1 and 2, Supplementary Table 7), 100ng DNA, T4 DNA Ligase 5U, 0.2mM ATP, 1X T4 DNA Ligase Buffer (Invitrogen) for 30min @ room temp. Adapter-ligated DNA was cleaned using AMPure beads on a Biomek FX. A selection of DNA samples were quantified on a QBit™ (Invitrogen). Phusion DNA polymerase, 15-cycle indexing enrichment PCR was performed using Primers 1.0 and 2.0 (IDT, USA) and 96 custom indexing primers (indexes shown in Supplementary Table 6). The PCR program was as follows: 98°C for 60s followed by 15 cycles of 98°C, 10s, 65°C, 15s, 72°C, 30s. The PCR products were cleaned using AMPure beads and eluted in 40uL elution buffer EB (QIAGEN, USA). Quality of product was assessed by QC gels: 1.75% SeaKem LE agarose 1X TAE, (0.2uL of every amplicon) and on Bioanalyzer-1000 (Agilent Technologies, Santa Clara, CA, USA). All 96 ~400bp amplicons from each plate were then pooled (15uL of each well) into a separate 1.5mL microfuge tube. Hence, one tube represents a plate of 96 pooled and indexed PCR products from 96 distinct DNA templates. The 400bp DNA size fraction was purified using 8% PAGE gels (1X TAE) and eluted from the gel slice overnight at 4°C in 400 µl of elution buffer (5:1, LoTE buffer (3 mM Tris-HCl, pH 7.5, 0.2 mM EDTA)-7.5 M ammonium acetate). Gel pieces were filtered using a Spin-X Filter Tube (Fisher Scientific, Pittsburgh, PA, USA). DNA was precipitated using ethanol and was quantified using an Agilent DNA 1000 series II assay (Agilent Technologies, Santa Clara CA, USA) then diluted to 10nM. The final concentration was confirmed using a Quant-iT dsDNA HS assay kit and QBit™ fluorometer (Invitrogen). An individual library was constructed from each indexed sample (comprising amplicons from up to 96 distinct template DNAs). Each of these libraries was sequenced on a single flowcell lane.

SNV analysis of tumor DNA and RNA sequence

All reads were aligned to the human reference genome (hg18) or (for WTSS) to a genome file that has been augmented with a set of all exon-exon junction sequences using the MAQ aligner²⁷ v0.7.1. Candidate single nucleotide variants (SNV) were identified in the aligned genomic sequence reads and the transcriptome (WTSS) reads using an approach similar to one we have previously described⁷. One key difference in our variant calling in this study is the application of a Bayesian SNV identification algorithm (“SNVmix”) currently under development by our group (version 0.11.7; http://compbio.bccrc.ca/?page_id=204)²⁸. This approach is able to identify SNVs with a minimum coverage of two high quality (Q20) bases. All sites assessed as being polymorphisms (SNPs) were disregarded, including variants matching a position in dbSNP or the personal genomes of Venter²⁹, Watson, the anonymous Asian³⁰ and Yoruban³¹ individuals. Additionally, all candidate mutations also

found in the genomic sequence from this patient's germline DNA were ignored. For the targeted re-sequencing experiment, coverage was generally greater than 1000x read depth at codon 641. Hence, we used all unambiguously mapped reads spanning this site to determine the percentage of reads with a high quality mismatch (Illumina base quality ≥ 20). These percentages are reported for each sample in the supplementary information (Supplementary Table 6).

Amplicon sequencing for SNV identification and Sanger sequence validation

Exon 15 of *EZH2* was PCR amplified from genomic DNA using *EZH2*_ASP_1 and *EZH2*_ASP_2. Priming sites for M13 Forward -21 and M13 reverse were added to their 5' ends to allow direct Sanger sequencing of amplicons. Unless otherwise stated, amplicons were produced from genomic DNA from both tumor and matched normal patient DNA. All capillary traces were analyzed using Mutation Surveyor and all variants were visually inspected to confirm their presence in tumor and absence from germline traces.

Computational modeling of EZH2 wild-type and mutant SET domain

The *EZH2* SET domain sequence was used to search for the structural template for homology modeling in the Protein Data Bank. The available crystal structure of the MLL1 SET domain (PDB ID 2w5z)¹² was identified as the best template (with sequence identities of 39% for the SET domain and no alignment gaps). A three dimensional model of the *EZH2* SET domain was constructed via the protein modeling server SWISS-MODEL³². Because MLL1 is a H3K4 binding protein, there was some concern that the target lysine residue of *EZH2* (K27) may not reside in the same conformation. Another concern is that the MLL1 crystal structure is in an open conformation and this conformation has reduced methyltransferase activity compared to the closed ones. The conformation change may shift the position of Y641. To confirm these we built alternative models using other structures as templates. We used the H3K9 binding proteins EHMT1 (PDB ID 2RFI), DIM-5 (1PEG), SUV39H2 (2R3A) and G9a (2O8J), as well as the H3K36 binding protein SETD2 (3H6L). The striking overlap of the conserved tyrosine residue corresponding to Y641 confirms that the position of Y641 remains unchanged in all proteins regardless of an open or closed conformation. The co-crystallized H3 peptides in 1PEG and 2RFI helped us confirm that the conformations of K4 and K9 are quite similar in those models. Therefore, we assume that the K27 in *EZH2* will pose a conformation close to that shown in the model.

***In vitro* EZH2 H3K27 tri-methylation assay**

Mutant constructs were generated using site-directed mutagenesis of the Refseq *EZH2* (NM_004456) with an N-terminal His tag. Wild-type *EZH2* and each of the four Y641 mutant constructs were co-expressed along with wild-type AEPB2, EED, SUZ12 and RbAp48 in SF9 cells using a baculovirus expression system (pVL1392, cloned using *BamHI* and *EcoRI*). Together, these five proteins associate to form an enzymatically active PRC2 complex *in vitro*. Expression of *EZH2* protein from each of the four mutant constructs was confirmed by Western blot and detected using anti-*EZH2*. Assay plates are coated with biotinylated histone H3 (21-44) peptide. Purified PRC2 was added to the plate along with S-adenosylmethionine (in the assay buffer) to detect enzyme activity. Methylated histone H3

was measured using a highly specific mouse-derived monoclonal antibody, which recognizes only the tri-methylated K27 residue of histone H3³³ (Active Motif, Catalog Number 39535.). The secondary antibody, which is labeled with Europium, was detected using time-resolved fluorescence (620nm). PRC2 methylase activity of each mutant (and wild-type EZH2) was tested at varying purified PRC2 amounts (between 0 and 200ng).

Cell lines

DB³⁴, KARPAS 422³⁵, SU-DHL-6³⁶ and WSU-DLCL2³⁷ are cell lines obtained from DSMZ and all “OCI-LY”³⁸ lines were obtained from Dr. L Staudt, NIH.

Cell-of-origin (COO) determination

Total RNA was reversed transcribed (one cycle) and hybridized to U133-2 Plus arrays according to the manufacturer’s protocol (Affymetrix). CEL files were normalized using robust multi-chip analysis (RMA). Cell of origin (COO) was calculated using model scores for ABC and GCB derived from the 185-gene model described by Lenz et al.¹¹ and the Bayesian formula described by Wright et al³⁹.

Copy number analysis of tumor DNA

BAC Array comparative genomic hybridization was performed as previously described⁴⁰ and did not identify any significant alterations. The tumor DNA was analyzed for large copy number alterations using an Affymetrix 500K SNP array as previously described⁴¹ using peripheral blood as a matched normal comparator. The sequencing data was also used to directly infer the presence of large-scale deletions and amplifications. This was accomplished by probabilistic identification of deviations in the proportion of unambiguously mapped reads between the normal and tumor genomic libraries as previously described²⁸. Due to the lower amount of sequence reads from the matched normal tissue, aligned reference reads were first used to define genomic bins of equal coverage containing 200 mapped reads. The sequencing depth of the normal genome provided 684,029 bins with a median size of 3953 bp, representing 2.942 gigabases of the hg18 assembly. A hidden Markov model (HMM) was used to classify and segment continuous regions into five discrete states: copy number loss (HMM 1), neutral (HMM 2), gain (HMM 3), amplification (HMM 4), and high-level amplification (HMM 5) using methodology outlined previously⁴². All segments and their HMM states are included in Supplementary Table 2 and displayed in Supplementary Figure 5.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This study was funded in part by grants from the National Cancer Institute Office of Cancer Genomics (see below), the National Cancer Institute of Canada (NCIC) Terry Fox Foundation New Frontiers Program Project Grant (grant no. 016003/grant type 230/project title: Biology of Cancer: Follicular Lymphoma as a Model of Cancer Progression) and Genome Canada/Genome BC Grant Competition III (project title: High Resolution Analysis of Follicular Lymphoma Genomes) to J.M.C., M.A.M., R.D.G. and D.E.H. and was supported by The Terry Fox Foundation (grant no. 019001). In addition, N.A.J. is a research fellow of the Terry Fox Foundation through an

award from the NCIC (019005) and the Michael Smith Foundation for Health Research (MSFHR) (ST-PDF-01793). M.A.M. is a Terry Fox Young Investigator and a Michael Smith Senior Research Scholar. A.J.M. is supported by a Fellowship Award from The Leukemia & Lymphoma Society. R.D.M. is a Vanier Scholar (Canadian Institutes for Health Research) and is also supported by a MSFHR senior graduate fellowship. The laboratory work for this study was undertaken at the Genome Sciences Centre, British Columbia Cancer Research Centre and the Centre for Translational and Applied Genomics, a program of the Provincial Health Services Authority Laboratories. The authors thank the BC Cancer Foundation and the Lion's Club International for their support. The authors gratefully acknowledge D. Gerhard for helpful discussions. Special thanks to C. Suragh and A. Drobniec for expert project management assistance. This project has been funded in whole or in part with federal funds from the National Cancer Institute, National Institutes of Health, under contract no. NO1-CO-12400. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products or organizations imply endorsement by the US Government.

References

1. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000; 403 (6769):503–511. [PubMed: 10676951]
2. Compagno M, et al. Mutations of multiple genes cause deregulation of NF-kappaB in diffuse large B-cell lymphoma. *Nature*. 2009; 459 (7247):717–721. [PubMed: 19412164]
3. Kato M, et al. Frequent inactivation of A20 in B-cell lymphomas. *Nature*. 2009; 459 (7247):712–716. [PubMed: 19412163]
4. Bea S, et al. Diffuse large B-cell lymphoma subgroups have distinct genetic profiles that influence tumor biology and improve gene-expression-based survival prediction. *Blood*. 2005; 106 (9):3183–3190. [PubMed: 16046532]
5. van Haafden G, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet*. 2009
6. Kleer CG, et al. EZH2 is a marker of aggressive breast cancer and promotes neoplastic transformation of breast epithelial cells. *Proc Natl Acad Sci U S A*. 2003; 100 (20):11606–11611. [PubMed: 14500907]
7. Morin R, et al. Profiling the HeLa S3 transcriptome using randomly primed cDNA and massively parallel short-read sequencing. *Biotechniques*. 2008; 45 (1):81–94. [PubMed: 18611170]
8. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods*. 2008; 5 (7):621–628. [PubMed: 18516045]
9. Ley TJ, et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature*. 2008; 456 (7218):66–72. [PubMed: 18987736]
10. Kirmizis A, et al. Silencing of human polycomb target genes is associated with methylation of histone H3 Lys 27. *Genes Dev*. 2004; 18 (13):1592–1605. [PubMed: 15231737]
11. Lenz G, et al. Stromal gene signatures in large-B-cell lymphomas. *N Engl J Med*. 2008; 359 (22):2313–2323. [PubMed: 19038878]
12. Southall SM, Wong PS, Odho Z, Roe SM, Wilson JR. Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol Cell*. 2009; 33 (2):181–191. [PubMed: 19187761]
13. Dillon SC, Zhang X, Trievel RC, Cheng X. The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biol*. 2005; 6 (8):227. [PubMed: 16086857]
14. Joshi P, et al. Dominant alleles identify SET domain residues required for histone methyltransferase of Polycomb repressive complex 2. *J Biol Chem*. 2008; 283 (41):27757–27766. [PubMed: 18693240]
15. Varambally S, et al. The polycomb group protein EZH2 is involved in progression of prostate cancer. *Nature*. 2002; 419 (6907):624–629. [PubMed: 12374981]
16. Raaphorst FM, et al. Cutting edge: polycomb gene expression patterns reflect distinct B cell differentiation stages in human germinal centers. *J Immunol*. 2000; 164 (1):1–4. [PubMed: 10604983]
17. Su IH, et al. Ezh2 controls B cell development through histone H3 methylation and Igh rearrangement. *Nat Immunol*. 2003; 4 (2):124–131. [PubMed: 12496962]

18. van Kemenade FJ, et al. Coexpression of BMI-1 and EZH2 polycomb-group proteins is associated with cycling cells and degree of malignancy in B-cell non-Hodgkin lymphoma. *Blood*. 2001; 97 (12):3896–3901. [PubMed: 11389032]
19. Couture JF, Dirk LM, Brunzelle JS, Houtz RL, Trievel RC. Structural origins for the product specificity of SET domain protein methyltransferases. *Proc Natl Acad Sci U S A*. 2008; 105 (52): 20659–20664. [PubMed: 19088188]
20. O’Riain C, et al. Array-based DNA methylation profiling in follicular lymphoma. *Leukemia*. 2009
21. Martin-Subero JI, et al. New insights into the biology and origin of mature aggressive B-cell lymphomas by combined epigenomic, genomic, and transcriptional profiling. *Blood*. 2009; 113 (11):2488–2497. [PubMed: 19075189]
22. Vire E, et al. The Polycomb group protein EZH2 directly controls DNA methylation. *Nature*. 2006; 439 (7078):871–874. [PubMed: 16357870]
23. Hans CP, et al. A significant diffuse component predicts for inferior survival in grade 3 follicular lymphoma, but cytologic subtypes do not predict survival. *Blood*. 2003; 101 (6):2363–2367. [PubMed: 12424193]
24. Thompson JD, Gibson TJ, Higgins DG. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics*. 2002; Chapter 2(Unit 2 3)
25. Pasqualucci L, et al. Hypermutation of multiple proto-oncogenes in B-cell diffuse large-cell lymphomas. *Nature*. 2001; 412 (6844):341–346. [PubMed: 11460166]
26. Krzywinski M, et al. A BAC clone fingerprinting approach to the detection of human genome rearrangements. *Genome Biol*. 2007; 8 (10):R224. [PubMed: 17953769]
27. Li H, Ruan J, Durbin R. Mapping short DNA sequencing reads and calling variants using mapping quality scores. *Genome Res*. 2008; 18 (11):1851–1858. [PubMed: 18714091]
28. Shah SP, et al. Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution. *Nature*. 2009; 461 (7265):809–813. [PubMed: 19812674]
29. Levy S, et al. The diploid genome sequence of an individual human. *PLoS Biol*. 2007; 5 (10):e254. [PubMed: 17803354]
30. Wang J, et al. The diploid genome sequence of an Asian individual. *Nature*. 2008; 456 (7218):60–65. [PubMed: 18987735]
31. Bentley DR, et al. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature*. 2008; 456 (7218):53–59. [PubMed: 18987734]
32. Kopp J, Schwede T. The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res*. 2004; 32 (Database issue):D230–234. [PubMed: 14681401]
33. Baskind HA, et al. Functional conservation of asx12, a murine homolog for the *Drosophila* enhancer of trithorax and polycomb group gene *asx*. *PLoS One*. 2009; 4 (3):e4750. [PubMed: 19270745]
34. Beckwith M, Longo DL, O’Connell CD, Moratz CM, Urba WJ. Phorbol ester-induced, cell-cycle-specific, growth inhibition of human B-lymphoma cell lines. *J Natl Cancer Inst*. 1990; 82 (6):501–509. [PubMed: 2313723]
35. Dyer MJ, Fischer P, Nacheva E, Labastide W, Karpas A. A new human B-cell non-Hodgkin’s lymphoma cell line (Karpas 422) exhibiting both t(14;18) and t(4;11) chromosomal translocations. *Blood*. 1990; 75 (3):709–714. [PubMed: 2297573]
36. Epstein AL, et al. Biology of the human malignant lymphomas. IV. Functional characterization of ten diffuse histiocytic lymphoma cell lines. *Cancer*. 1978; 42 (5):2379–2391. [PubMed: 214220]
37. Al-Katib AM, et al. Bryostatins 1 down-regulates *mdr1* and potentiates vincristine cytotoxicity in diffuse large cell lymphoma xenografts. *Clin Cancer Res*. 1998; 4 (5):1305–1314. [PubMed: 9607591]
38. Tweeddale ME, et al. The presence of clonogenic cells in high-grade malignant lymphoma: a prognostic factor. *Blood*. 1987; 69 (5):1307–1314. [PubMed: 3567358]
39. Wright G, et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A*. 2003; 100 (17):9991–9996. [PubMed: 12900505]

40. Cheung KJ, et al. Genome-wide profiling of follicular lymphoma by array comparative genomic hybridization reveals prognostically significant DNA copy number imbalances. *Blood*. 2009; 113 (1):137–148. [PubMed: 18703704]
41. Delaney AD, Qian H, Friedman JM, Marra MA. Use of Affymetrix mapping arrays in the diagnosis of gene copy number variation. *Curr Protoc Hum Genet*. 2008; Chapter 8(Unit 8):13. [PubMed: 18972373]
42. Shah SP, et al. Integrating copy number polymorphisms into array CGH analysis using a robust HMM. *Bioinformatics*. 2006; 22 (14):e431–439. [PubMed: 16873504]

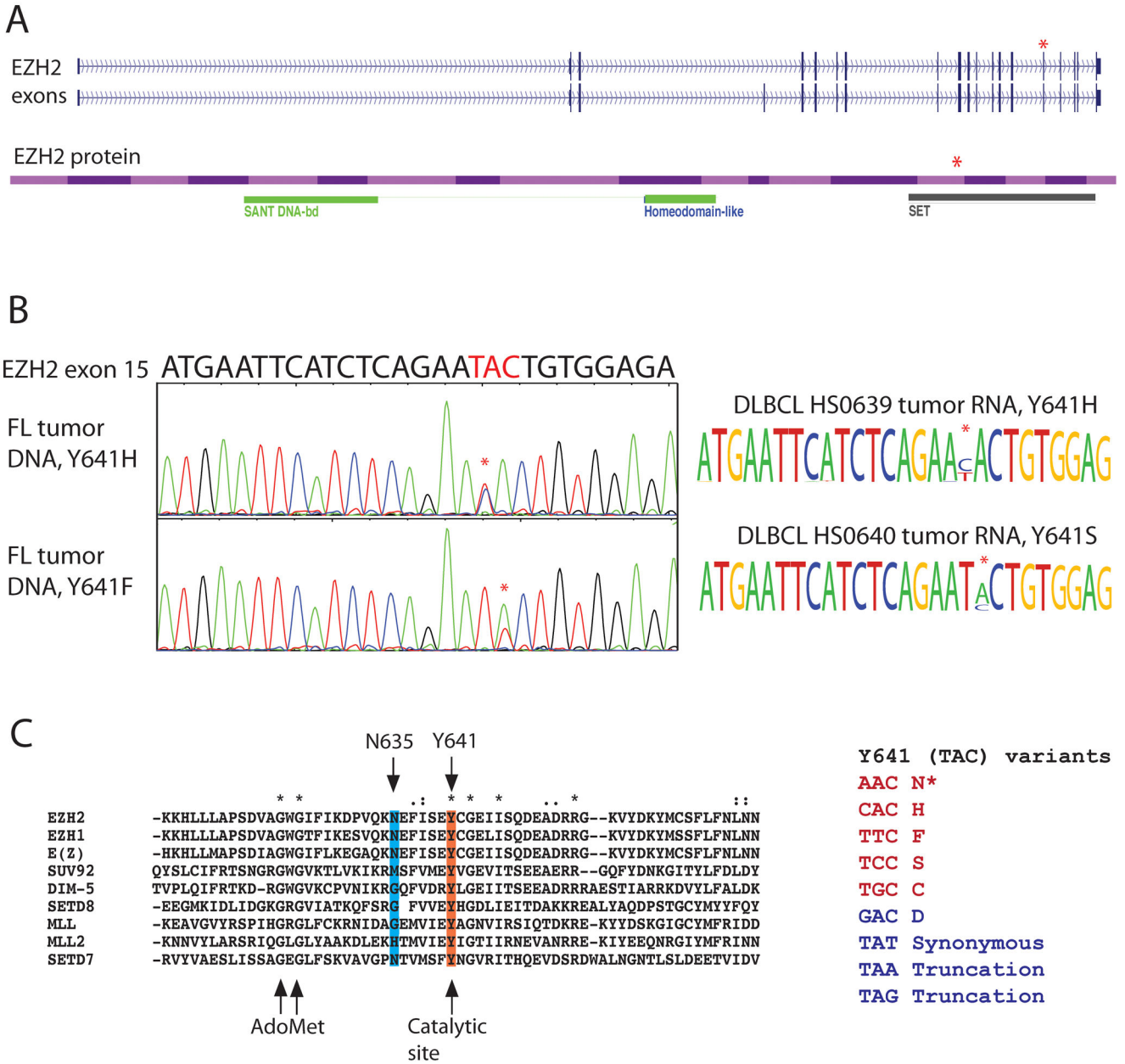


Figure 1. Recurrent mutations of Y641 in EZH2

(A) Genomic organization of the EZH2 locus, alternative exons and protein domain structure. The location of the mutation affecting Y641 in exon 15 of the *EZH2* gene and protein is indicated with a red asterisk. (B) Illustration of sequencing results. Three of the five distinct mutations and amino acid replacements in codon 641 from different lymphoma samples as detected by capillary sequencing (left) or Illumina WTSS (right). (C) A multiple alignment of *EZH2*, *EZH1* (its paralog), the *Drosophila* ortholog *E(Z)* and six other human SET domain proteins demonstrates the intra and inter-species sequence conservation of SET domains. Conservation codes reported by ClustalX are shown above²⁴. The predominant mutation in *EZH2* affects a key tyrosine in the catalytic site of the SET domain (orange)

conserved in the *Drosophila* ortholog *E(Z)*. With one exception, all *EZH2* mutations in FL and DLBCL alter this amino acid. The exception was a double mutant (FL), with a second somatic mutation affecting N635 (blue). All mutants comprise 5 of the 8 possible non-synonymous variants of this codon (lower right, in red). Notably, the five observed amino acid changes were not found at equal frequencies. We detected a slight enrichment for Y641F (49%) followed by Y641S (21%), Y641N (15%) and Y641H (13%) and only a single example of Y641C (2%)(Supplementary Table 5). Of the unobserved variants (D, blue), two would result in a truncated protein and the third would introduce an aspartate residue. The pattern and nature of these changes (A->G, A->T, T->G, T->A), indicated to us that these mutations do not likely arise from AID-induced somatic hypermutation at this locus²⁵.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

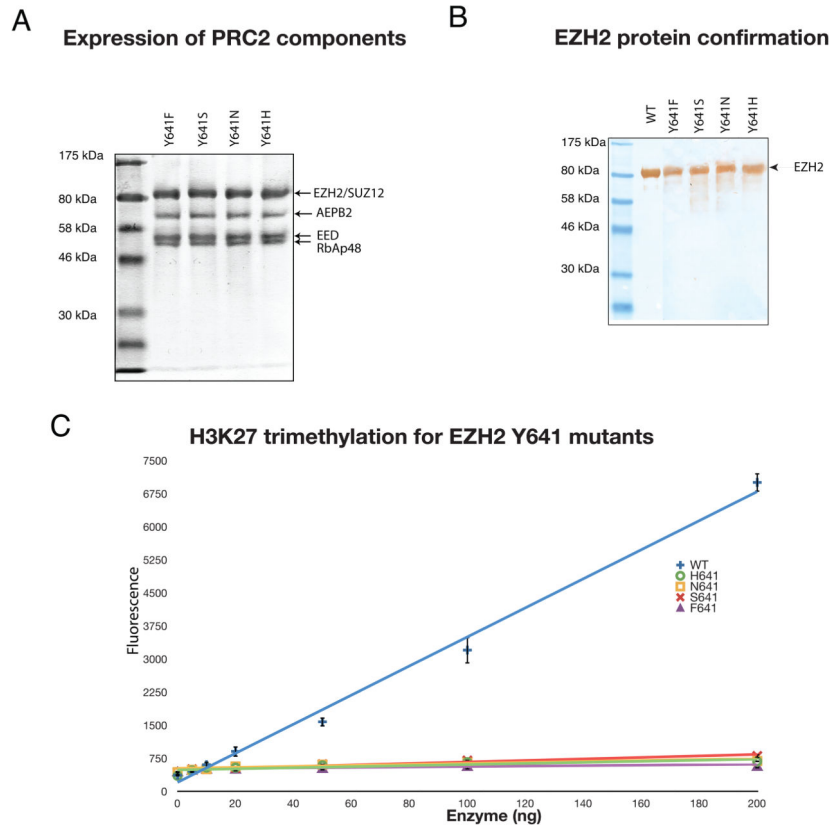


Figure 2. *In-vitro* assembly and functional analysis of PRC2 with mutant and wild-type *EZH2* (A) Wild-type *EZH2* and each of the four Y641 mutants were co-expressed along with wild-type AEPB2, EED, SUZ12 and RbAp48 in SF9 cells using a baculovirus expression system (Methods). Together, these five proteins associate to form an enzymatically active PRC2 complex *in vitro*. The purified complex from the SF9 cells showed strong expression of each of these proteins and confirmed their association and assembly into PRC2. (B) Expression of *EZH2* protein from each of the four mutant constructs was confirmed by Western blot. (C) The purified complex was then assayed using biotinylated histone H3 (21-44) peptide along with S-adenosylmethionine (in the assay buffer) to detect enzyme activity. Methylated histone H3 was measured using a highly specific antibody, which recognizes only the trimethylated K27 residue of histone H3 (Methods). The secondary antibody, which is labeled with Europium, was detected using time-resolved fluorescence (620nm). PRC2 methylase activity of each mutant (and wild-type *EZH2*) was tested at varying purified PRC2 amounts (between 0 and 200ng). The specific activity for the four mutants was calculated to be 0.001, 0.0012, 0.0011 and 0.0009 pmol/min/ug for the H, N, S and F mutants, respectively (mean = 0.00105). The wild-type enzyme (blue) showed a specific activity of 0.0071 (~6.8-fold greater). Error bars reflect the standard deviation of triplicate measurements.

Table 1

Summary of exonic sequence coverage in the genome and transcriptome sequence of FL patient A

Library Description	Raw reads (pairs)	Mapped reads	Total sequence coverage (bp)	Mean coverage depth of exons
FL sample A, matched germ line genomic DNA	93,473,829	163,216,278	7,986,844,356	2.80
FL sample A, tumor WTSS	51,729,560	63,262,348	2,277,444,528	18.9
FL sample A, tumor genomic DNA	351,666,782	563,762,488	27,024,661,976	9.47

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 2

Location and effect of all mutations in *EZH2* in FL and DLBCL determined by WTSS

Sample ID	Sample type or cell line name	AGE	SEX	t(14;18)	Genomic Position	Mutation*	Effect
HS0804	FL (Sample A)	44	F	No	chr7:148139661	T->C	Y->H
HS0639	DLBCL	60	M	Yes	chr7:148139661	T->C	Y->H
HS0648	DLBCL	92	F	No	chr7:148139661	T->A	Y->N
HS0640	DLBCL	68	F	Yes	chr7:148139660	A->C	Y->S
HS0942	DLBCL	73	M	Yes	chr7:148139661	T->A	Y->N
HS0798	DB	45	M	Yes	chr7:148139661	T->A	Y->N
HS0841	KARPAS 422	73	F	Yes	chr7:148139661	T->A	Y->N
HS0900	SU-DHL-6	43	M	Yes	chr7:148139661	T->A	Y->N
HS0901	WSU-DLCL2	41	M	Yes	chr7:148139660	A->T	Y->F
HS1163	OCL-LY1	NA	NA	Yes	chr7:148139661	T->A	Y->N

Abbreviations: WTSS, whole transcriptome shotgun sequencing; FL, follicular lymphoma; DLBCL, diffuse large B cell lymphoma

* All observed mutations are heterozygous and mutation is reported on the negative strand

Table 3Frequency of *EZH2* Y641 mutations in lymphoma and benign samples

Sample type	# of samples analyzed	# of samples with <i>EZH2</i> Y641 mutation	Prevalence of Y641 mutation
FL			
Grade 1	133	10	7.5%
Grade 2	60	4	6.7%
Grade 3	28	2	7.1%
Total FL	221	16	7.2%
FL and DLBCL pairs*			
FL	30	2	6.7%
DLBCL	30	4	13.3%
Total FL-derived	60	6	10.0%
DLBCL			
GCB	83	18	21.7%
PMBCL	24	1	4.2%
ABC	42	0	0%
U	25	0	0%
Non-GCB	22	0	0%
Not available [§]	124	12	9.7%
Total primary DLBCL	320	31	9.7%
MCL	25	0	0%
SLL	30	0	0%
PTCL	25	0	0%
Cell lines			
GCB cell lines [¥]	7	5	71.4%
ABC cell lines [¥]	2	0	0%
Benign			
Reactive lymph node	23	0	0%
Purified CD77 ⁺ centroblasts [¶]	8	0	0%
Total primary NHL samples	681	53	7.8%
Total cell lines	9	5	55.5%
Total benign	31	0	0%

Abbreviations: FL, follicular lymphoma; DLBCL, diffuse large B cell lymphoma; GCB, germinal center B cell sub-type DLBCL defined by gene expression profiling (GEP); PMBCL, primary mediastinal B cell lymphoma; ABC, activated B cell sub-type of DLBCL defined by GEP; U, unclassifiable sub-type DLBCL defined by GEP; Non-GCB, non-germinal center type of DLBCL defined by immunohistochemistry using the Hans criteria²³; MCL, mantle cell lymphoma; SLL, small lymphocytic lymphoma; PTCL, peripheral T cell lymphoma not otherwise specified

[§] Affymetrix array analysis was not performed; hence, COO information is unavailable

* FL and DLBCL pairs were samples derived from the same patient pre (FL) and post transformation (DLBCL).

‡ GCB cell lines: mutated *EZH2*: DB, KARPAS 422, SU-DHL-6 and WSU-DLCL2 and OCI-LY1; wild type *EZH2*: OCI-LY7 and OCI-LY19.

‡‡ ABC cell lines: wild type *EZH2*: OCI-LY3 and OCI-LY10.

¶ CD77+ centroblasts were purified based on CD77 selection from reactive tonsils.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript