SUPPLEMENTAL METHODS

Gene expression data analysis

Gene expression profiling was previously performed on 181 FL samples from the LLMPP cohort using the Affymetrix (Santa Clara, CA, USA) U133A and U133B arrays (1). The R statistical environment (v. 2.12.1, www.R-project.org) was used for data pre-processing, statistical and clustering analysis. The samples were normalised using GC-RMA (2) and filtered to select the most reliable and variable probes. Differential expression analysis was performed using LIMMA (3). The Benjamini and Hochberg (4) multiple hypothesis testing correction procedure was applied to control the false discovery rate (FDR). Differentially expressed genes were identified using an adjusted p-value < 0.01 in combination with a fold-change of at least 1.

REFERENCES

1. Dave SS, Wright G, Tan B, Rosenwald A, Gascoyne RD, Chan WC, et al. Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. N Engl J Med. 2004;351:2159-69.
2. Wu Z, Irizarry RA, Gentleman R, Martinez-Murillo F, Spencer F. A Model-Based Background Adjustment for Oligonucleotide Expression Arrays. Journal of the American Statistical Association. 2004;99:909-17.
3. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004;3:Article3.
4. Benjamini Y, Hochberg, Y. Controlling the false discovery rate: A practical and powerful approach to ultiple testing. Journal of the Royal Statistical Society. 1995;57:289-300.

SUPPLEMENTAL RESULTS

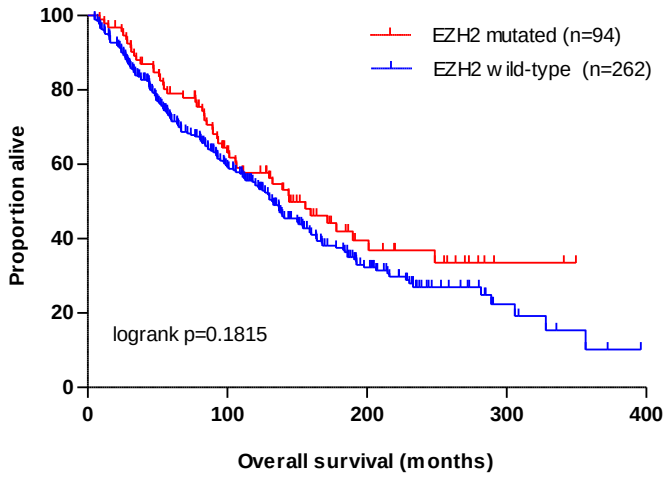
The majority of *EZH2* mutations represent clonal events

In order to determine whether the *EZH2* mutations represented clonal or subclonal mutations, we used our previous array-based methylation profiling data of 164 FL biopsies (49 *EZH2* mutated, 115 wt; VAF range: 2.5-48.5) where the difference in tumor methylation to benign lymph node controls ($\Delta\beta$ value) provided an estimate of the tumor content. [(1)] We observed a significant correlation (Pearson test, r=0.7620, p<0.0001) between $\Delta\beta$ values and *EZH2* VAF (supplemental *Figure 2*), suggesting that the majority of cases with low *EZH2* VAF corresponded to cases with low tumor content in the biopsy sample and are likely to represent genuine clonal events. However, we detected 6 cases with low VAF (≤7%) and relatively high $\Delta\beta$ values suggesting that *EZH2* mutations were subclonal in these cases. Considering the possibility that *EZH2* mutations have an effect on DNA methylation profiles, it is important to note that next generations sequencing offers a more robust/reliable means of clonality assessment compared to the methylation based analysis.

REFERENCES

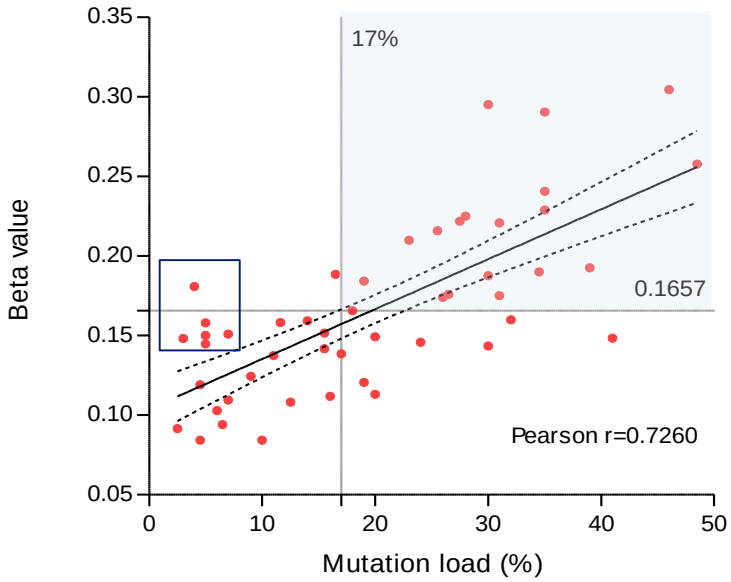
1. O'Riain C, O'Shea DM, Yang Y, Le Dieu R, Gribben JG, Summers K, et al. Array-based DNA methylation profiling in follicular lymphoma. Leukemia 2009;23(10):1858-66.

SUPPLEMENTAL FIGURE S1



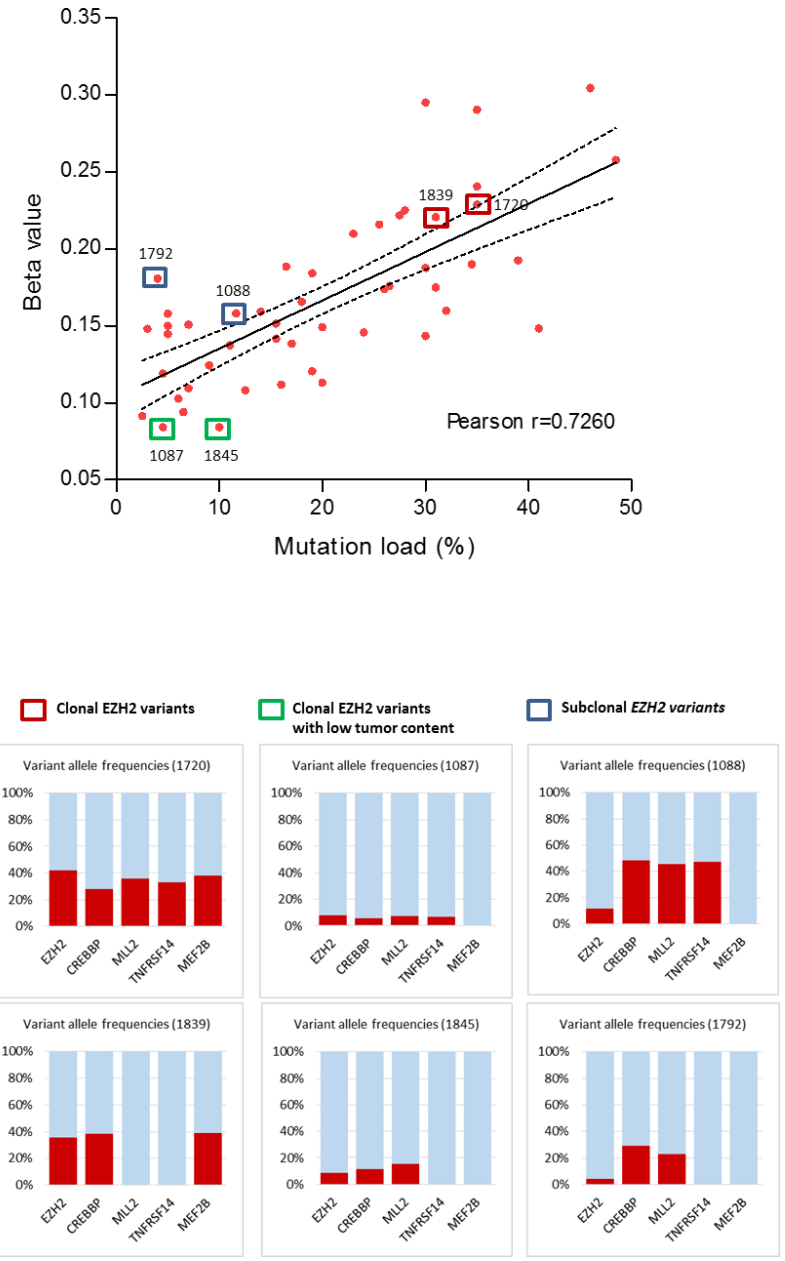
Supplemental Figure S1. Kaplan-Meier plots for overall survival by *EZH2* mutation status for the 356 patients from the Barts and the LLMPP cohorts with available clinical information. No significant differences were seen between the mutated (n=94) and wild-type (n=262) subgroups.

SUPPLEMENTAL FIGURE S2



Supplemental Figure S2. Significant correlation (*Pearson correlation, r=0.7620, p<0.0001*) between the *EZH2* variant allele frequencies (VAF) determined by deep sequencing and beta methylation values reflecting the tumor content of biopsies suggests that the majority of cases with low *EZH2* variant allele frequency corresponded to cases with low tumor content in the biopsy sample, indicating that *EZH2* mutations represent genuine clonal events. The 6 cases with low variant allele frequency (≤7%) highlighted in the blue square are characterized by high $\Delta\beta$ values suggesting that they are likely to represent subclonal mutations. A delta beta value of 0.1657 and a VAF of 17% represent the cut off levels selected for further analyses to define a gene expression signature (Figure 1C) using the cases with estimated high tumor content. (The mean ± 95% confidence interval of the linear regression is shown).

SUPPLEMENTAL FIGURE S3



Supplemental Figure S3. Illustrated is the concordance between the differential methylation and targeted resequencing approaches used for the assessment of clonality of *EZH2* mutations. The highlighted examples were predicted to represent clonal (1839, 1720), clonal with low tumor content (1087, 1845) and subclonal *EZH2* mutations (1088, 1792) based on correlation between the *EZH2* variant allele frequencies (VAF) and beta methylation values. Indeed, the distribution of VAFs for *CREBBP, MLL2, TNFRSF14* and *MEF2B* genes revealed by targeted resequencing supports these observations. *(Case 465 presented in Figure 1B as example for a subclonal EZH2 variant was not part of the cohort with methylation profiling data, therefore case 1792 was included as a representative subclonal EZH2 variant in this figure.)*