Supplemental Figure 1. Validation of experimental positive controls. (A) Representative western blot of positive control protein knockdown in human CD34+ HSPCs after transfection with gRNAs. β-actin loading control. (B) Editing efficiency of control gRNAs, n = 8. (C) Proliferation time course normalized to gAAVS1. Significance determined by Dunnett’s multiple comparisons test, n = 6. (D) Progression of erythroid differentiation through expression of CD71 and Glycophorin A (GlyA). (E) Representative flow plots of day 14 erythroid differentiation cultures with positive control gRNAs. (F) Erythroid differentiation after 14 days in culture, expressed as fraction of mature erythroid/total erythroid normalized to gAAVS1. Significance determined by paired t-test, n = 6.

Supplemental Figure 2. Genes in high proliferation group show expansion of edited cells. Genes were evenly split into low, mid, and high groups based on AAVS1 normalized proliferation. (A) Initial editing of each group, showing no editing bias before proliferation assay in two replicate arrays. (B) Change in the percentage of edited alleles before and after expansion in 7-day proliferation culture. Significance determined by One-way ANOVA and Tukey’s multiple comparisons test, p < 0.05.

Supplemental Figure 3. Proliferation assay time course. Proliferation time course normalized to gAAVS1 on day 3 (first measurement). n=167 biological samples.

Supplemental Figure 4. Correlation of proliferation assay biological replicates. Heat map of Pearson R correlation coefficients for each biological replicate compared against each other. Each replicate represents an array with different configurations of gRNAs, so that each gene was tested 3-4 times.

Supplemental Figure 5. Correlation of proliferation assay technical replicates. Pearson R correlation coefficients for two technical replicates within each biological replicate. Each technical replicate contains two assay replicates.

Supplemental Figure 6. Correlation of erythroid differentiation assay biological replicates. Heat map of Pearson R correlation coefficients for each biological replicate compared against each other. Each replicate represents an array with different configurations of gRNAs, so that each gene was tested 3-4 times. Replicates 1.1 and 1.2 are repeated runs of the same configuration.

Supplemental Figure 7. Genes included in arrayed screen are overall enriched for tumor suppressor activity. Frequency distribution of (A) proliferation normalized to gAAVS1 or (B) combined score for all edited genes. Significance of the median increase over gAAVS1 control determined by Wilcoxon signed-rank test, p < 0.05 n = 108.

Supplemental Figure 8. Design of machine learning tumor suppressor classification. Flow chart depicting design of random forest machine learning model for prediction of chromosome 7 tumor suppressor genes.

Supplemental Figure 9. Quality control measures for machine learning model. (A) AUC (Area Under The Curve) ROC (Receiver Operating Characteristics) curve. Error bar represents 90% CI. (B) Out-of-bag (OOB) error rate frequency distribution, representing the distribution of OOB error rate of 100 bootstraps of the random forest model. mean = 0.290, n=100. Represents the average error for each iteration using predictions from the trees that do not contain it within their respective bootstrap sample. (C,D) Importance of classification variables across 100 bootstrap iterations. Each classification variable is defined by the technology, cell line/mutation signature, and reference(Author.year).

Supplemental Figure 10. Heat map of variables used to consider myeloid TSG status in all genes included in experimental analysis. Genes are ranked by experimental combined score, machine learning score, and CDR status equally weighted. Columns 1 and 2 are binary variables where yellow = yes and purple = no; the remaining columns are z-scores of the experimental and machine learning classification variables.