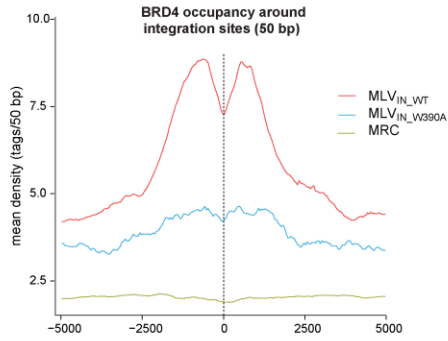


Figure S5

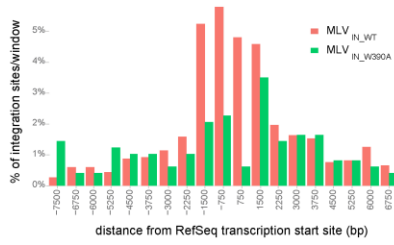
a



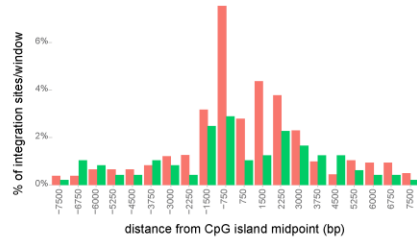
b

Vector	Type	Total Sites	% TSS within 2kb	% TSS within 4kb	% CpG within 2kb	% CpG within 4kb	% DHS within 2kb	% DHS within 4kb
MLV _{IN_WT}	insertion	1831	22.9	29.7	22.9	28.6	54.5	69.2
MLV _{IN_W390A}	insertion	485	10.7***	16.7***	10.1***	16.0***	38.3***	53.6***

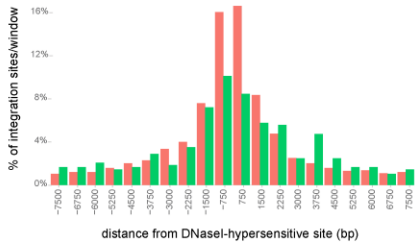
c



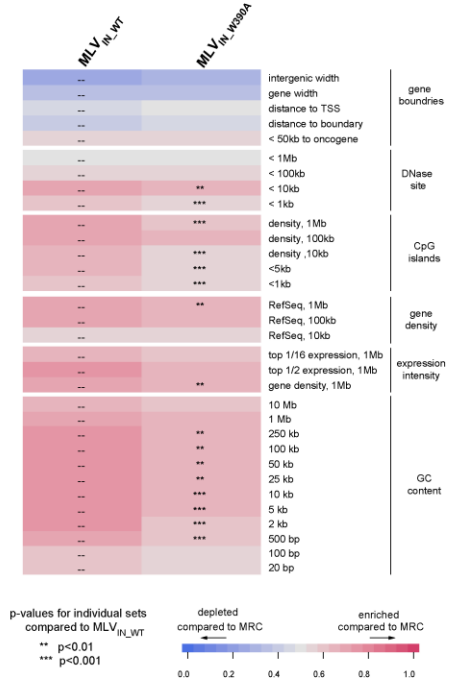
d



e



f



g

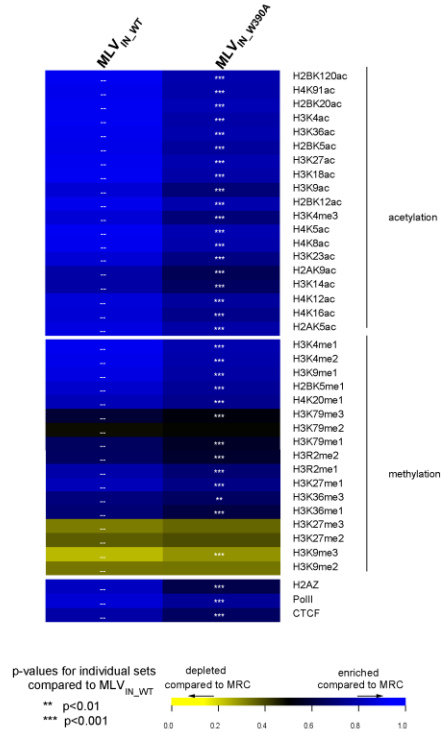


Figure S5: Integration site distribution analysis for MLV_{IN_WT} and MLV_{IN_W390A} in human primary CD4+ T cells

(a) Mean background-subtracted ChIP-seq read density for BRD4 generated in CD4+ T-cells in 50bp bins for a 10kb window around MLV_{IN_WT} and MLV_{IN_W390A} integration sites obtained in CD4+ T-cells.

(b) MLV -based vector integration sites obtained from CD4+ T-cells and their genomic distribution. Integration percentages for 2-4kb windows around TSS, CpG island midpoints and DHS. TSS; transcription start sites, CpG; CpG-rich islands, DHS; DNase I-hypersensitive sites. All data sets reached significance $p < 0.001$ (pairwise Fishers test), compared to MRC and MLV_{IN_WT} .

(c-e) Integration frequencies in 750bp bins around TSS, CpG islands and DHS in CD4+ T-cells for MLV_{IN_WT} and MLV_{IN_W390A} .

(f-g) Heat maps summarizing the relation between vector integration site frequency and genomic (f) or epigenetic (g) features in CD4+ T-cells for MLV_{IN_WT} and MLV_{IN_W390A} . Features analyzed are shown to the right of the corresponding row of the heat map. Tile colors indicate whether a particular feature is favored or disfavored for integration of the respective data sets relative to their MRCs, as detailed in the colored ROC area scale at the bottom of the panel. p values (*) show significance of departures from MLV_{IN_WT} integration sites (** $p < 0.01$; *** $p < 0.001$, Wald statistics referred to χ^2 distribution).