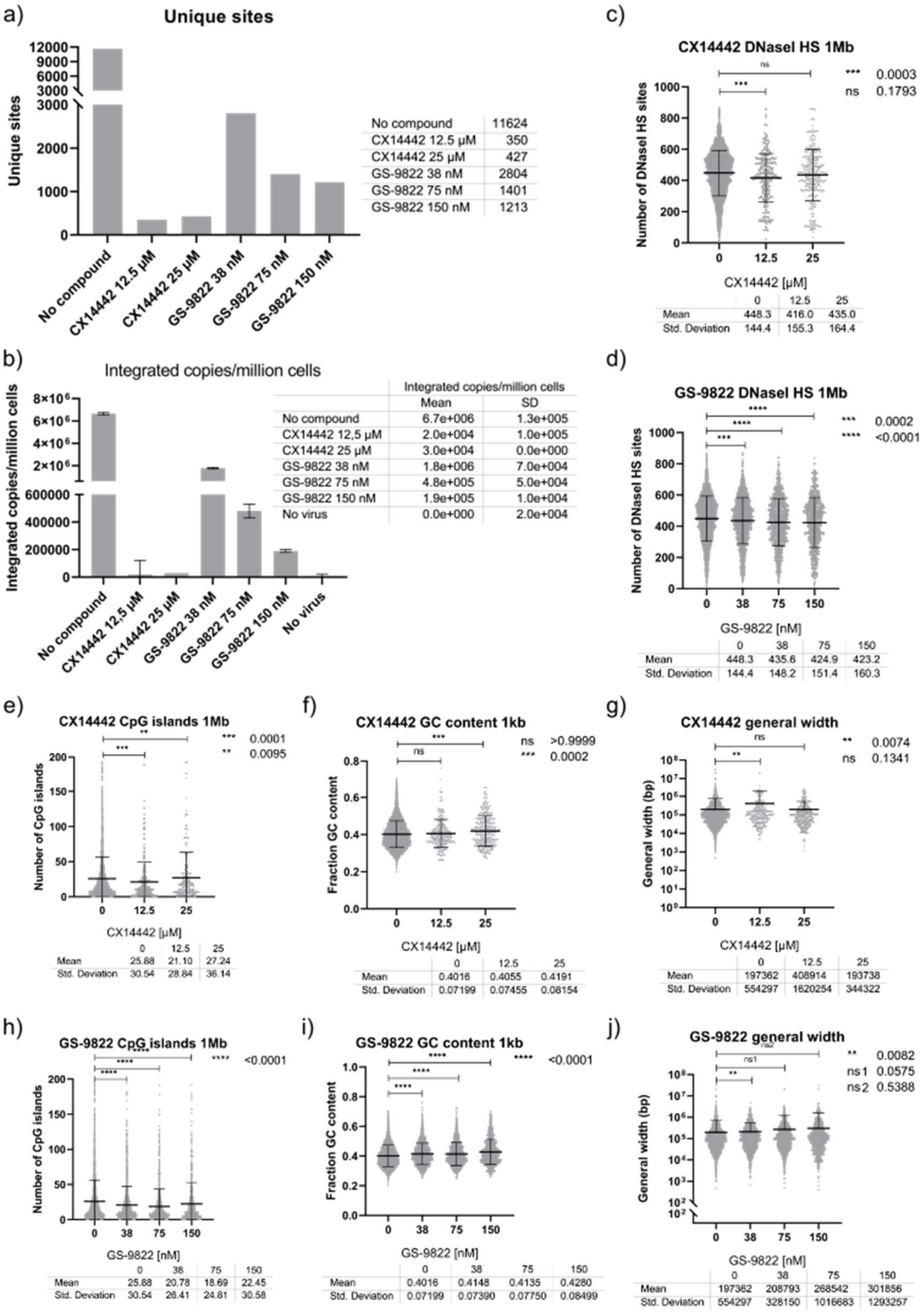
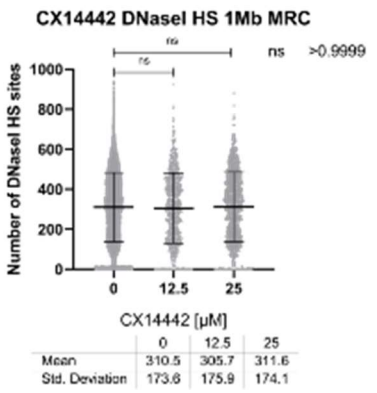
Supplementary Figures:

Supplementary Figure 1.

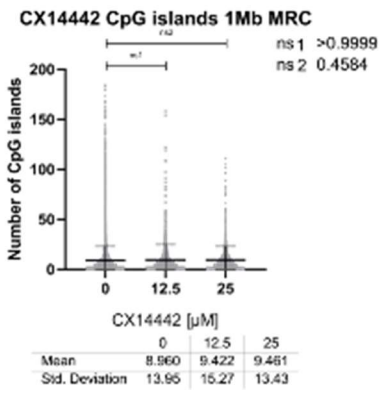


k)



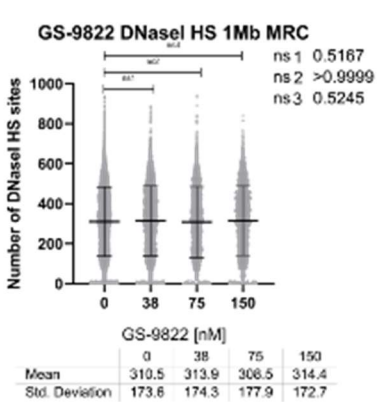
	0	12.5	25
Mean	310.5	305.7	311.6
Std. Deviation	173.6	175.9	174.1

l)



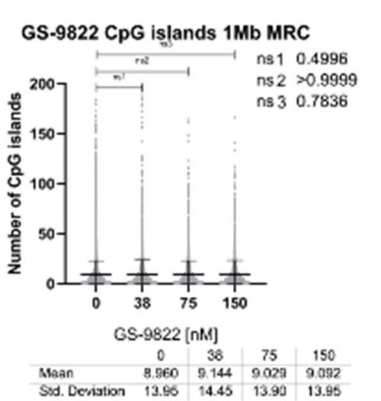
	0	12.5	25
Mean	8.960	9.422	9.461
Std. Deviation	13.95	15.27	13.43

m)



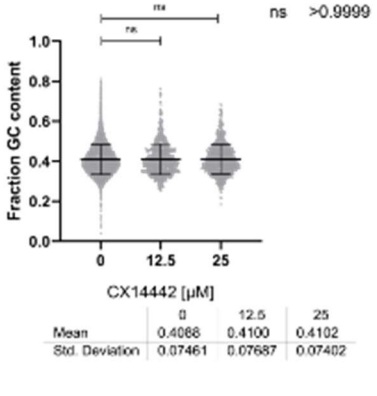
	0	38	75	150
Mean	310.5	313.9	306.5	314.4
Std. Deviation	173.6	174.3	177.9	172.7

n)



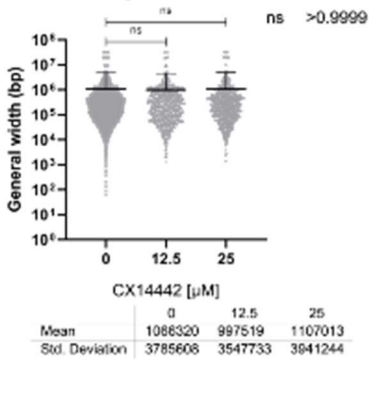
	0	38	75	150
Mean	8.960	9.144	9.029	9.092
Std. Deviation	13.95	14.45	13.90	13.95

o)



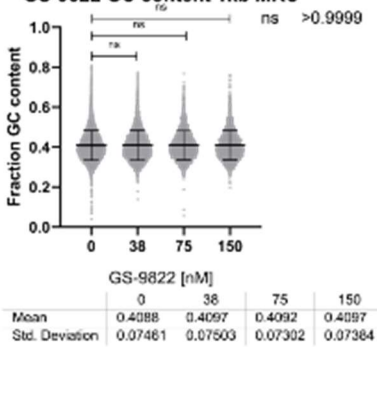
	0	12.5	25
Mean	0.4068	0.4100	0.4102
Std. Deviation	0.07461	0.07687	0.07402

p)



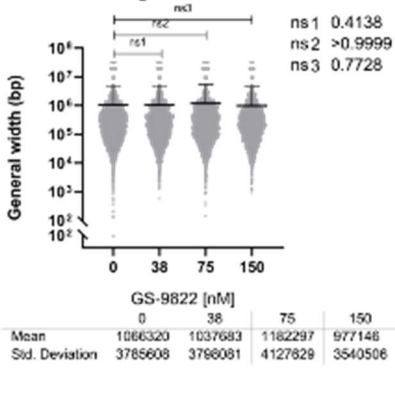
	0	12.5	25
Mean	1086320	997519	1107013
Std. Deviation	3785608	3547733	3941244

q)



	0	38	75	150
Mean	0.4088	0.4097	0.4092	0.4097
Std. Deviation	0.07461	0.07503	0.07302	0.07384

r)

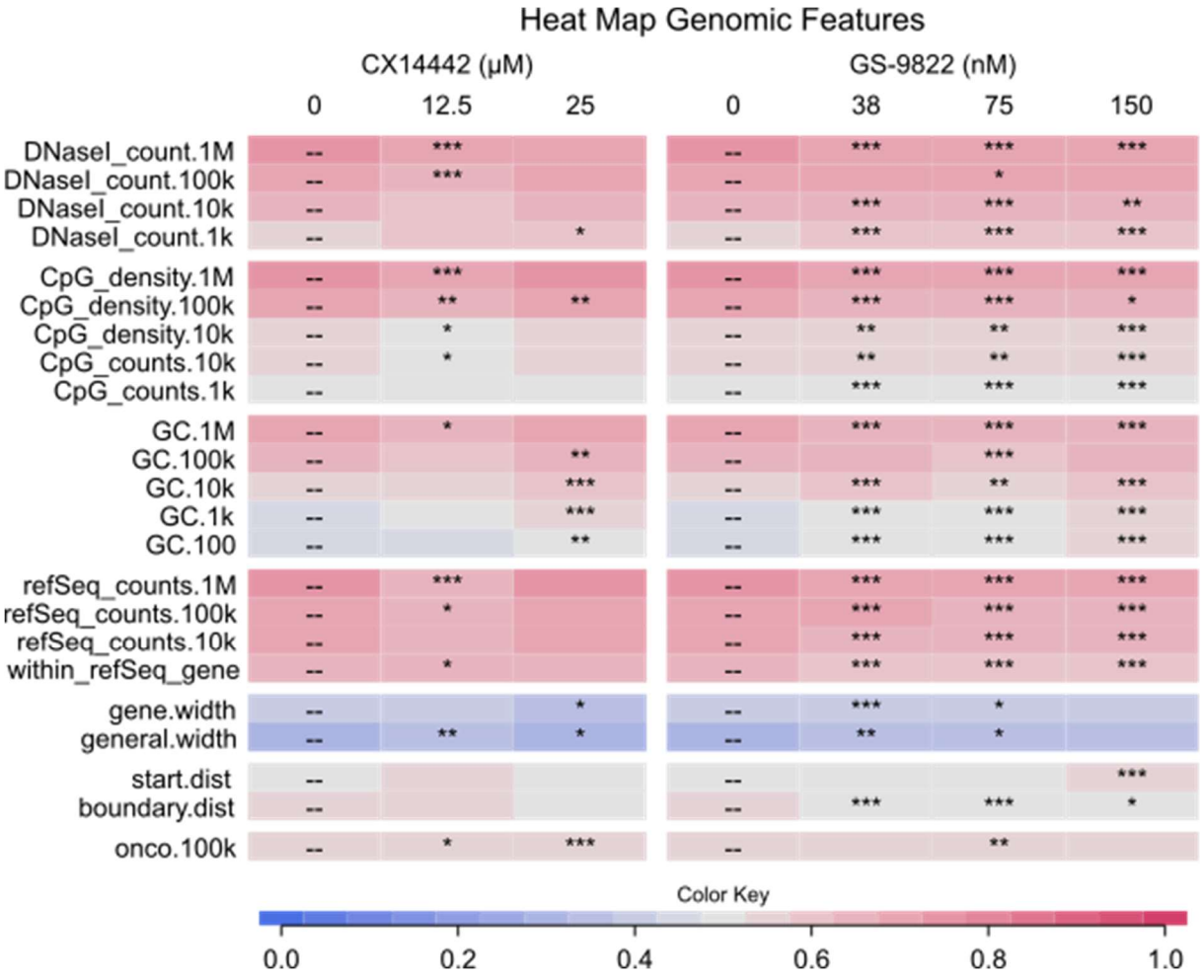


	0	38	75	150
Mean	1086320	1037683	1182297	977146
Std. Deviation	3785608	3796081	4127829	3540506

Suppl. Fig. 1. GS-9822 retargets integration away from gene dense regions.

a) Number of unique sites as determined by NGS analysis. b) Number of integrated copies/million cells determined on genomic DNA of NGS samples using Alu LTR and CCR5 qPCR. c-r) Impact of treatment with CX14442 or GS-9822 on integration sites with respect to genomic markers of gene density (at long range) or active chromatin. In each graph individual integration sites are shown per condition with mean values and standard deviations plotted on top and values tabulated below each graph. Samples were compared to the no compound condition using a Kruskall-Wallis test and p-values are listed beside the graphs. c, d, k, m) Number of DNase I hypersensitive sites within a 1 Mb range of integration sites (c, d) and matched random controls (k, m). e, h, l, n) Number of CpG islands within a 1Mb range for each integration site (e, h) and matched random controls (l, n). f, i, o, q) Relative GC content of the 1 kb region surrounding each integration site (f, i) and matched random controls (o, q). g, j, p, r) Width of genes and intergenic regions surrounding each integration site (g, j) and matched random controls (p, r). All samples were obtained from a single transduction experiment.

Supplementary Figure 2.

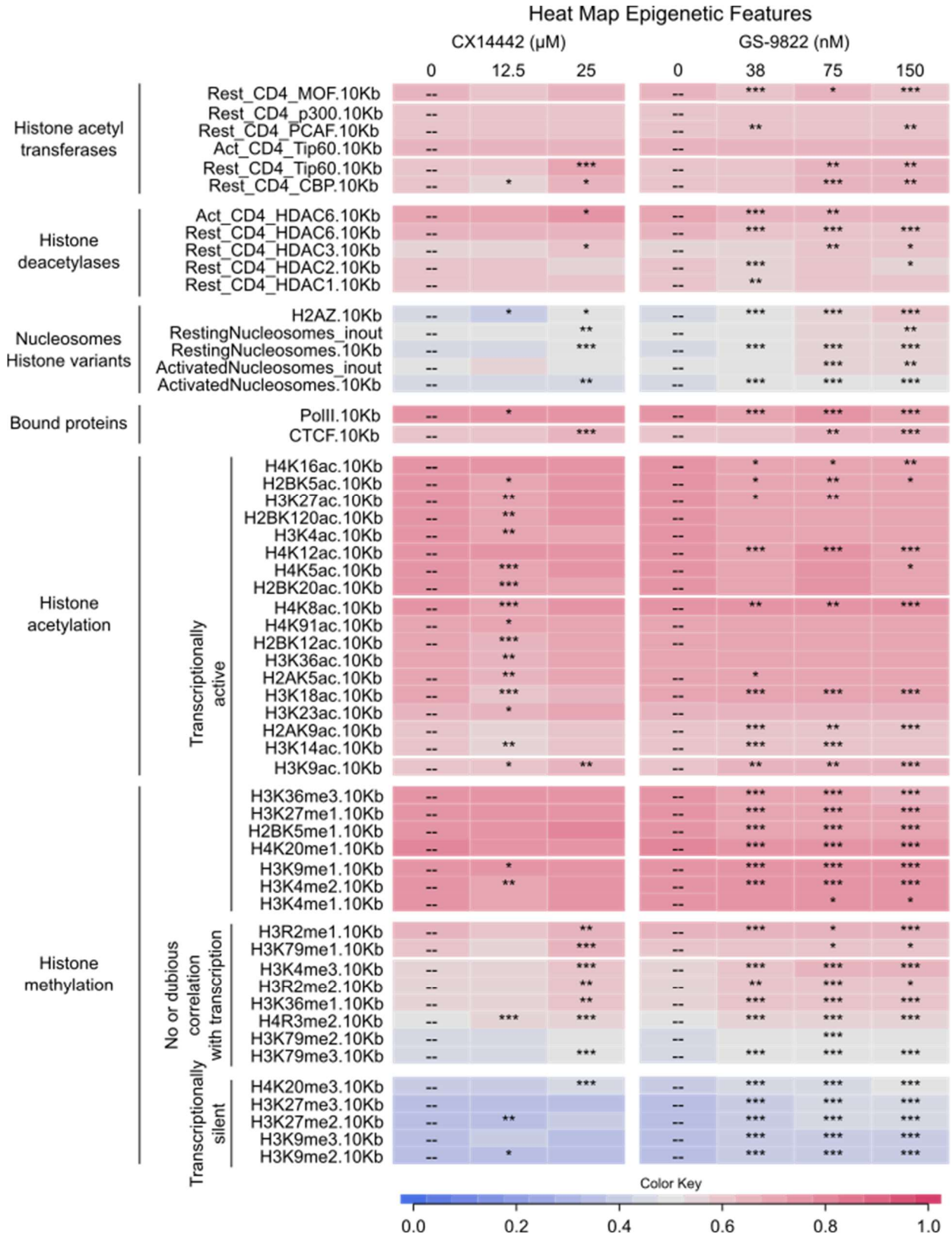


Suppl. Fig. 2. Heat map representation of genomic features of lentiviral integration sites after treatment with CX14442 and GS-9822.

Illumina Miseq sequencing data were analyzed using the INSPIIRED software and represented as a heat map (1, 2). Colors indicate whether a genomic feature is favored (red) or disfavored (blue) for integration as compared to computer generated matched random controls (MRCs) using a receiver operating characteristic (ROC) curve area. Each integration site is compared to its MRC and according to the rank of the integration site, a number is assigned (1 if the feature is favored at the integration site over the MRC, 0 if it is disfavored and 0.5 if the feature is equal for the two

sites). The analysis was done for all integration sites after which an average was calculated. These ROC curve areas are then statistically analyzed using Wald type test statistics which are referred to the Chi Square distribution (*=$p<0.05$, **$p=0.01$, ***=$p<0.0001$).

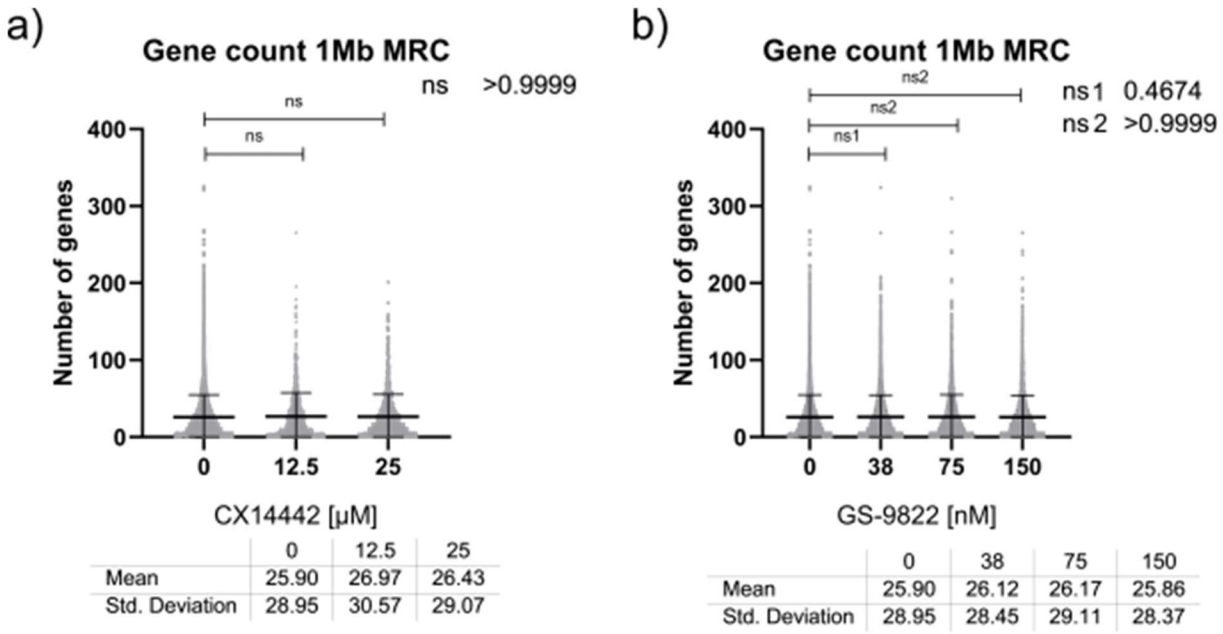
Supplementary Figure 3.



Suppl. Fig. 3. Heat map representation of epigenetic features of lentiviral integration sites after treatment with CX14442 and GS-9822.

After Illumina Miseq sequencing, data were analyzed using the INSPIIRED software, yielding the heat map shown (1, 2). Color codes were as described in Suppl. Fig. 2. The ROC curve areas were statistically analyzed using Wald type test statistics which are referred to the Chi Square distribution (*=p<0.05, **p=0.01, ***=p<0.0001).

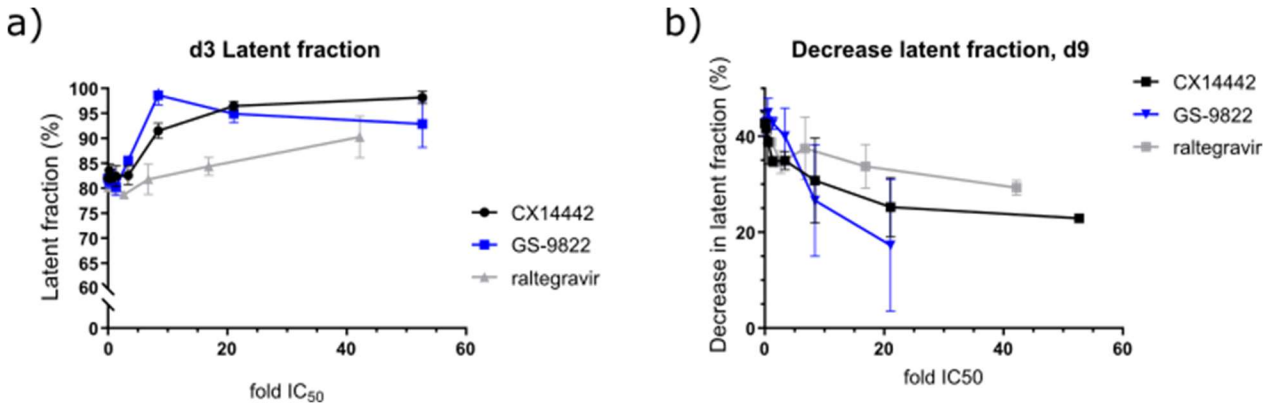
Supplementary Figure 4.



Sup. Fig. 4 GS-9822 retargets integration away from gene dense regions; matched random controls.

SupT1 cells were transduced for 3 days with CH-SFFV-eGFP-P2A-fLuc in the presence or absence of varying drug concentrations and kept in culture for at least 10 days. Next genomic DNA was extracted for Illumina Miseq integration site sequencing and data were analyzed via the INSPIIRED platform. a, b) Graphs plotting the number of genes counted within a 1 Mb range of matched random controls generated for each integration site for samples treated with CX14442 and GS-9822, respectively. Annotated data was obtained using the University of California Santa Cruz (UCSC) Genome Browser website (http://genome.ucsc.edu, UCSC Known Genes) (3, 4). Mean values and standard deviations are plotted on top. Samples were compared to the no compound condition using a Kruskall-Wallis test.

Supplementary Figure 5.



Supplementary Figure 5. Treatment with CX14442 or GS9822 but not raltegravir increases immediate latency and decreases reactivation from latency in cells transduced with the HIV-1 OGH vector.

Data of one representative experiment out of 4 plotting the average of duplicate measurements with standard deviation. SupT1 cells were transduced with a 1/20000 dilution of HIV-1 OGH. Cells were treated with increasing concentrations of CX14442, GS9822 or raltegravir. Compound concentrations are plotted as a fold of the IC_{50}. The data represented here is the same as in Fig. 4 d and Fig. 5 c, but here the data is represented on a linear axis, rather than a logarithmic one. a) The latent fraction (percentage of single mKO2 positive cells/(percentage of single mKO2 positive cells + percentage of double positive cells)*100) or (quadrant C/ (quadrant B + quadrant C)) as shown in Fig. 3.c, was calculated three days post transduction. b) Upon reactivation, the latent fraction decreases and the plotted decrease in latent fraction is calculated by subtracting the latent fraction in the TNFα treated condition from the latent fraction in the non-treated condition 9 days after transduction or 24 hours after reactivation with TNFα.

References:

1. Sherman E, Nobles C, Berry CC, Six E, Wu Y, Dryga A, Malani N, Male F, Reddy S, Bailey A, Bittinger K, Everett JK, Caccavelli L, Drake MJ, Bates P, Hacein-Bey-Abina S, Cavazzana M, Bushman FD. 2017. INSPIIRED: A Pipeline for Quantitative Analysis of Sites of New DNA Integration in Cellular Genomes. Mol Ther - Methods Clin Dev 4:39–49.

2. Berry CC, Nobles C, Six E, Wu Y, Malani N, Sherman E, Dryga A, Everett JK, Male F, Bailey A, Bittinger K, Drake MJ, Caccavelli L, Bates P, Hacein-Bey-Abina S, Cavazzana M, Bushman FD. 2017. INSPIIRED: Quantification and Visualization Tools for Analyzing Integration Site Distributions. Mol Ther - Methods Clin Dev 4:17–26.

3. Hsu F, Kent JW, Clawson H, Kuhn RM, Diekhans M, Haussler D. 2006. The UCSC known genes. Bioinformatics 22:1036–1046.

4. Haeussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, Gibson D, Diekhans M, Clawson H, Casper J, Barber GP, Haussler D, Kuhn RM, Kent WJ. 2019. The UCSC Genome Browser database: 2019 update. Nucleic Acids Res 47:D853–D858.