YMTHE, Volume *29*

Supplemental Information

Evaluating the Genomic Parameters Governing rAAV-Mediated Homologous Recombination

Laura P. Spector, Matthew Tiffany, Nicole M. Ferraro, Nathan S. Abell, Stephen B. Montgomery, and Mark A. Kay

Supplemental Methods and Materials

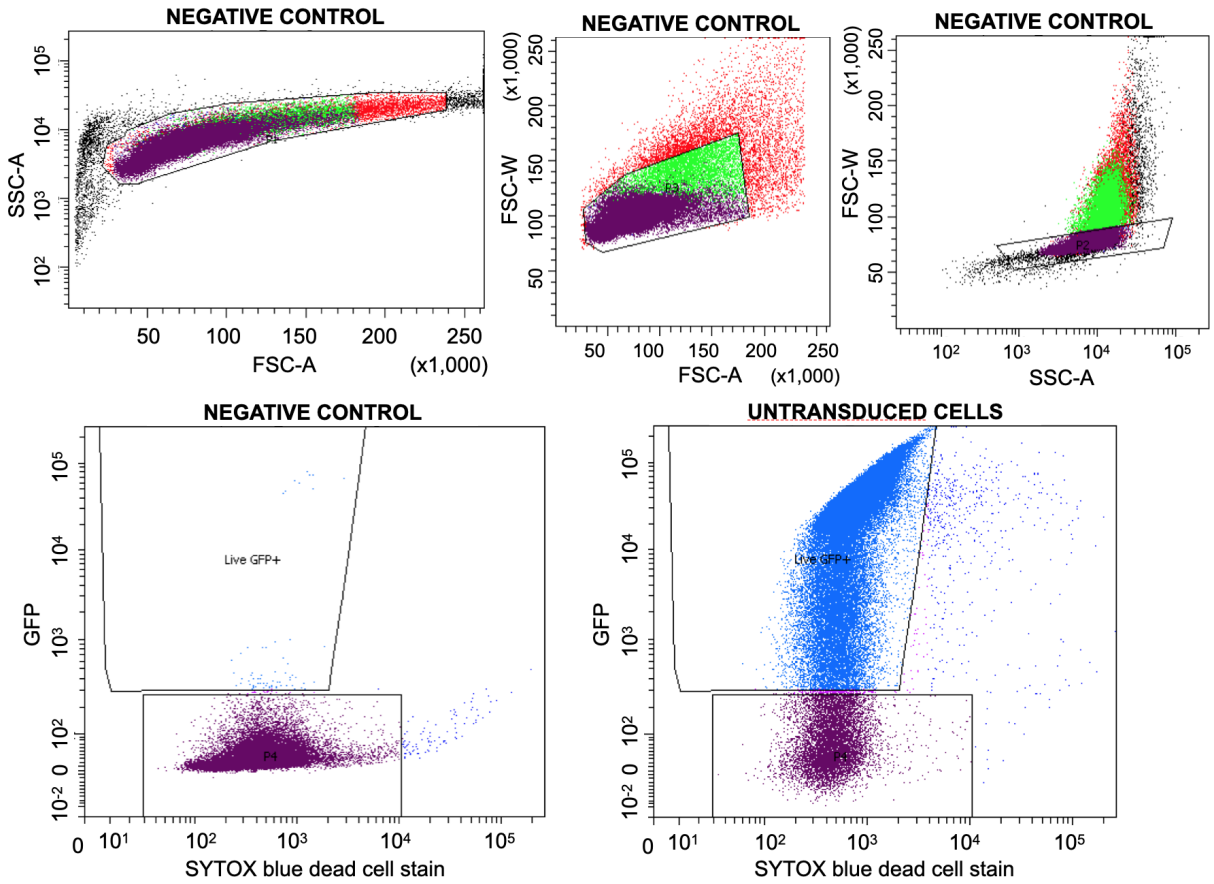
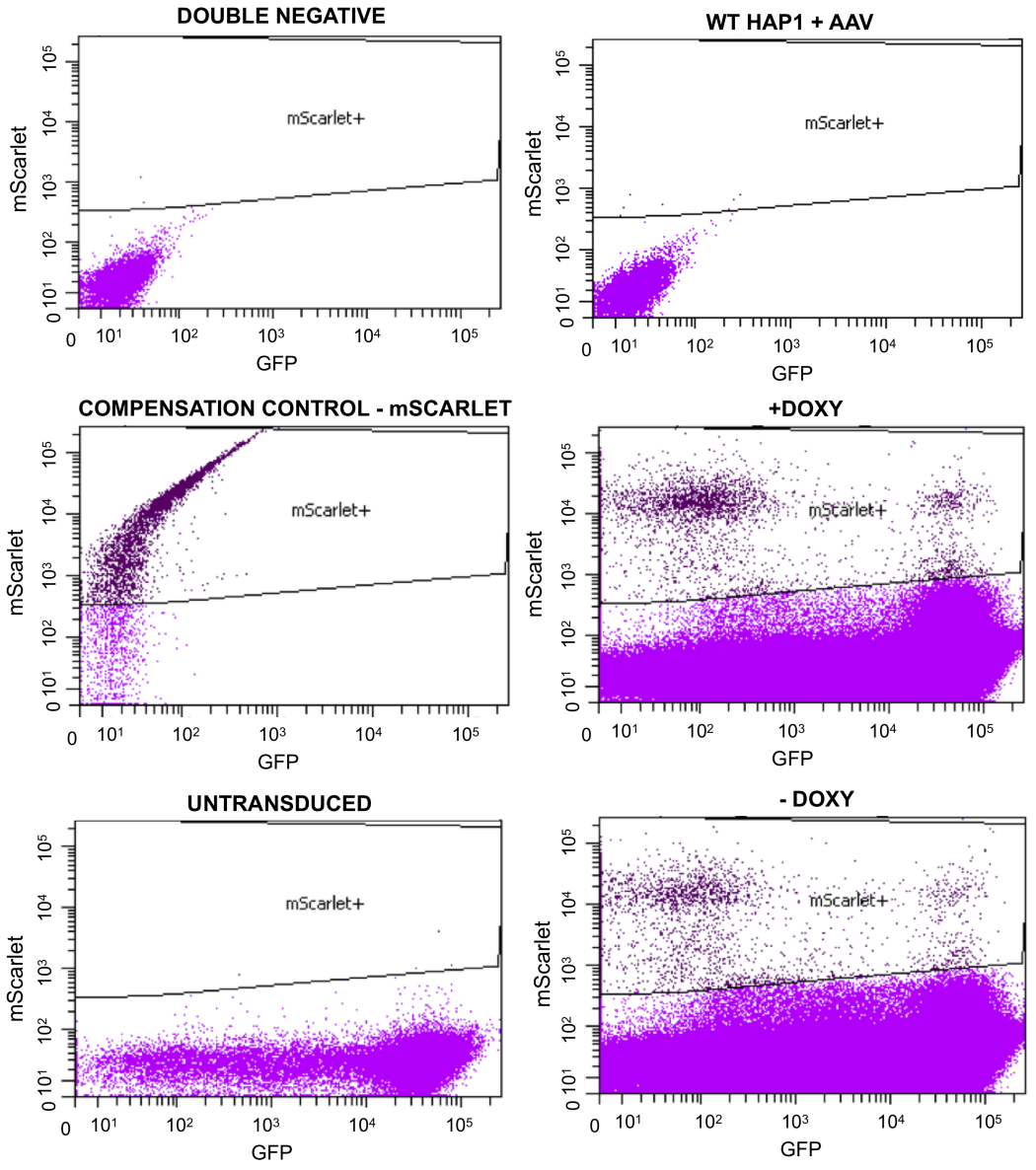


Figure S1: Selection of polyclonal, lentivirus-infected cell population by fluorescence-activated cell sorting. HAP1 cells were infected with lentivirus particles and sorted following the addition of 100ng/mL doxycycline to the cell culture medium. GFP+/SYTOX blue- cells were retained (blue population, bottom right panel). Negative control refers to cells not exposed to doxycycline prior to sorting. Untransduced cells refers to the polyclonal population of lentivirus-infected cells not transduced with barcoded rAAV library, exposed to doxycycline, for consistency with Figures S2 and S3. Top panels show gating of singlets. Channels are FITC-A (GFP, Y-axis in bottom panels) and BV421-A (SYTOX blue dead cell stain, X-axis in bottom panels).



Tube: DOX 1_001

Population	#Events	%Parent	%Total
All Events	2,500,000	####	100.0
scatter	2,399,046	96.0	96.0
singlet 1	1,842,224	76.8	73.7
singlet 2	1,792,281	97.3	71.7
P1	1,777,019	99.1	71.1
mScarlet+	4,160	0.2	0.2

Tube: NO DOX 1_002

Population	#Events	%Parent	%Total
All Events	2,909,800	####	100.0
scatter	2,791,494	95.9	95.9
singlet 1	2,290,714	82.1	78.7
singlet 2	2,229,568	97.3	76.6
P1	2,207,134	99.0	75.9
mScarlet+	4,515	0.2	0.2

Figure S2: Representative first sorting for mScarlet positive cells with controls. The polyclonal, lentivirus-infected HAP1 cell line was transduced with barcoded rAAV library and sorted following the addition of 100ng/mL doxycycline to the cell culture medium. mScarlet+ cells were retained, as shown in the rightmost two panels, upper quadrants (dark purple population). Double negative cells represent cells not exposed to doxycycline prior to sorting. Untransduced cells represent cells not transduced with barcoded rAAV library, exposed to doxycycline. WT+AAV represent wildtype HAP1 cells, which lack a target site, transduced with barcoded rAAV library and exposed to doxycycline. Channels are PE-CF594-A (mScarlet, Y-axis) and FITC-A (GFP, X-axis).

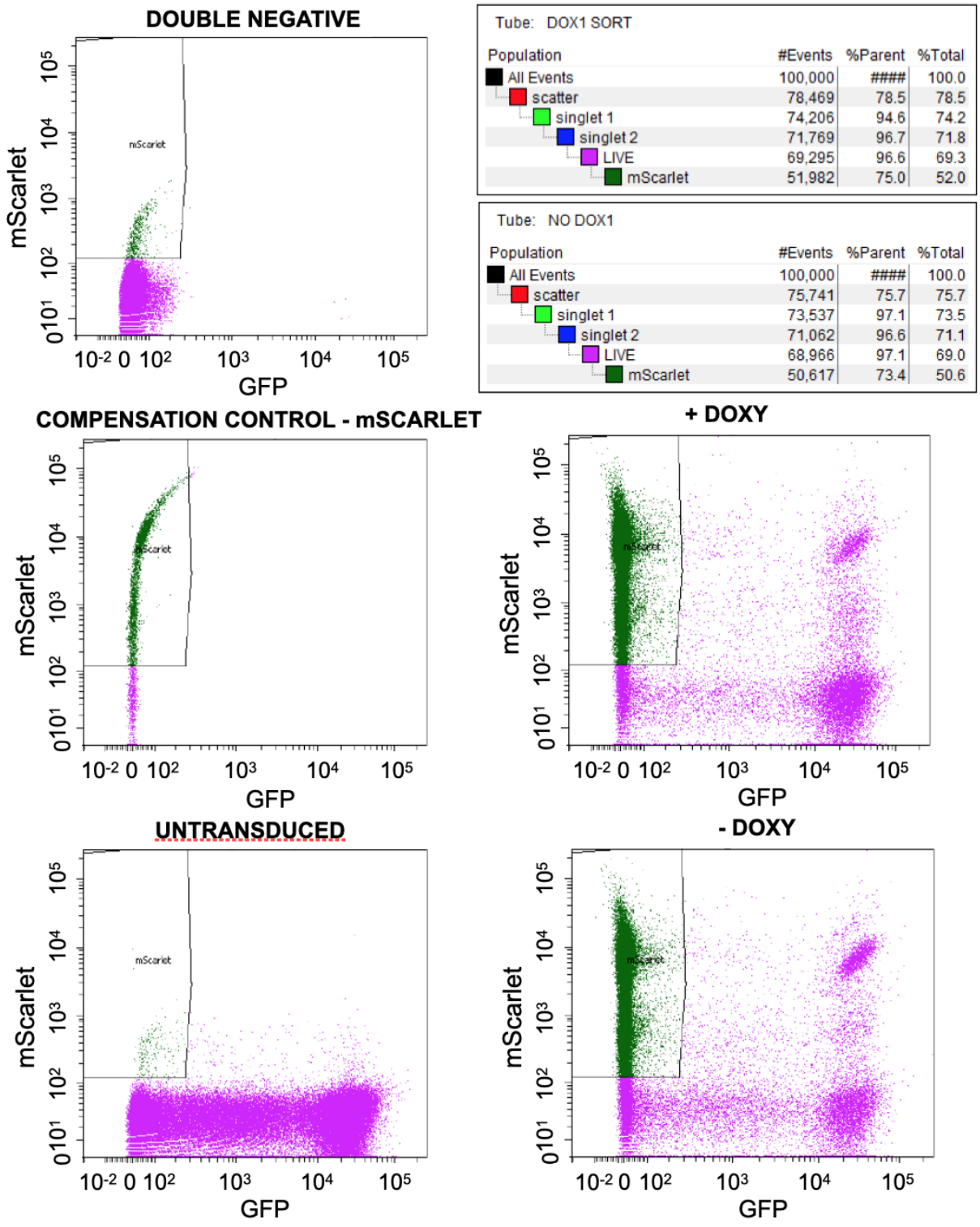


Figure S3: Representative second sorting for mScarlet positive/GFP negative cells with controls. Cells sorted as shown in Figure S2 were again sorted following the addition of 100ng/mL doxycycline to the cell culture medium. This time, mScarlet+/GFP- cells were retained, as shown in the rightmost two panels, upper left quadrant (green population). Double negative cells represent cells not exposed to doxycycline prior to sorting. Untransduced cells represent cells not transduced with barcoded rAAV library, exposed to doxycycline. Channels are PE-CF594-A (mScarlet, Y-axis) and FITC-A (GFP, X-axis).

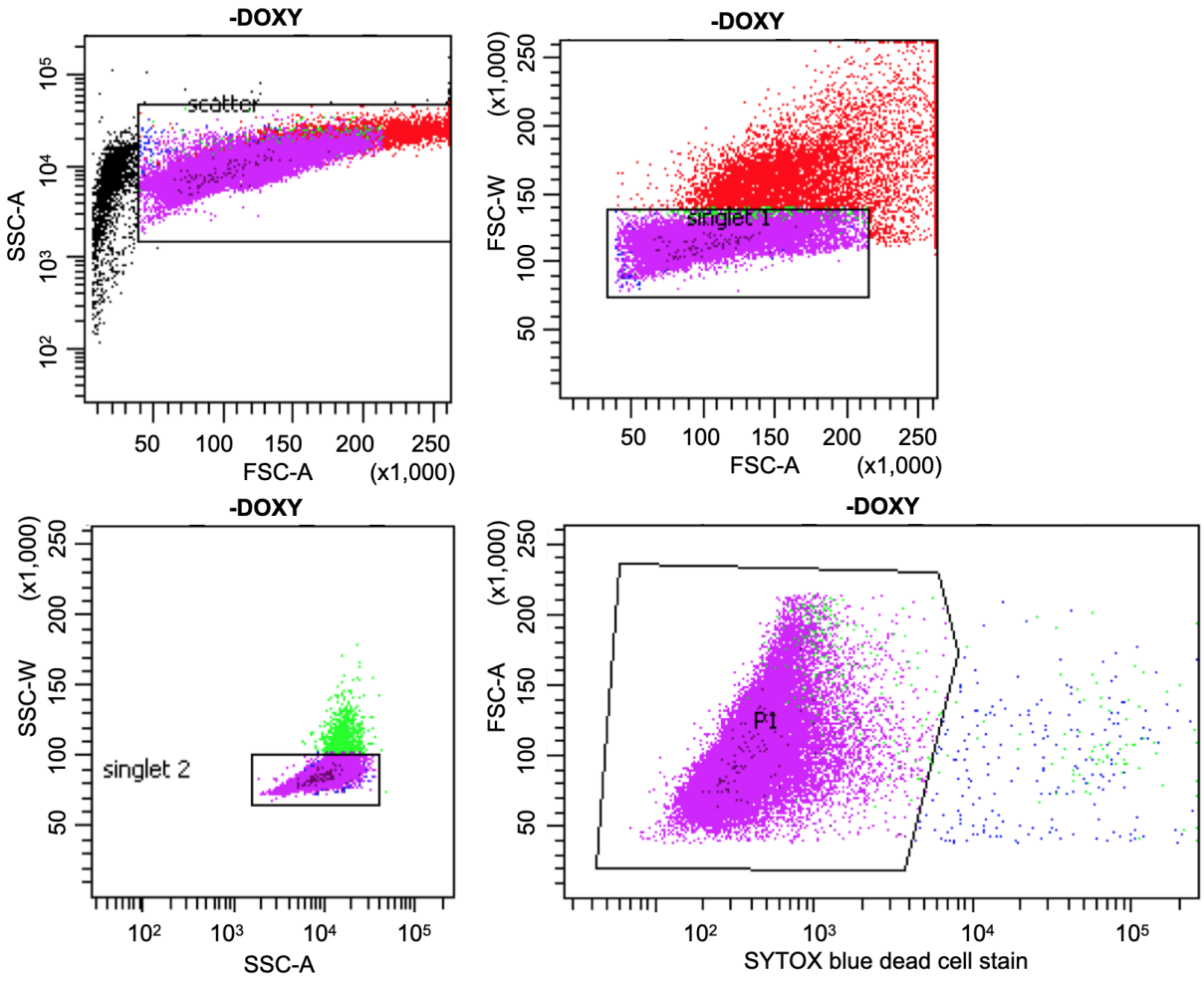


Figure S4: Representative gating in fluorescence-activated cell sorting experiments for singlets and live cells using forward scatter/side scatter (FSC and SSC channels) and SYTOX blue dead cell stain (BV421-A channel), respectively.

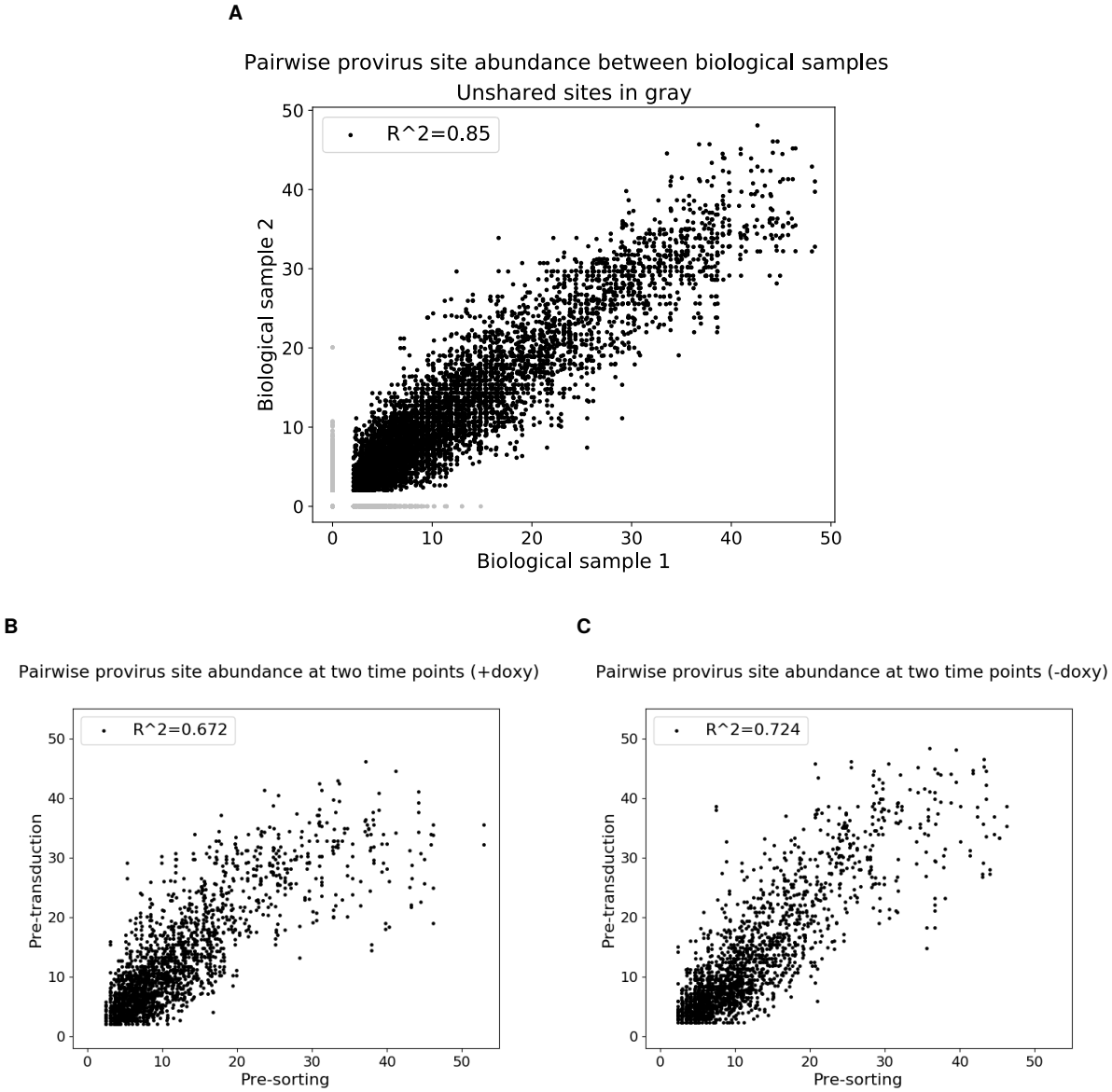


Figure S5: (A) Pairwise comparison of normalized provirus site abundance between all combinations of biological replicates, both within and between treatment groups (n=5,919). Sites present in only one of a pair of samples are plotted in gray (not shared n=1,995). (b and c) Pairwise comparison of normalized provirus site abundance between pre-transduction samples and passaging (pre-sorting) phase biological replicates in the same treatment group. +doxycycline sites n=1,941, -doxycycline sites n=1,906. R^2 calculated after fitting a linear model using ordinary least squares. Samples were mean centered and unit scaled before model fitting.

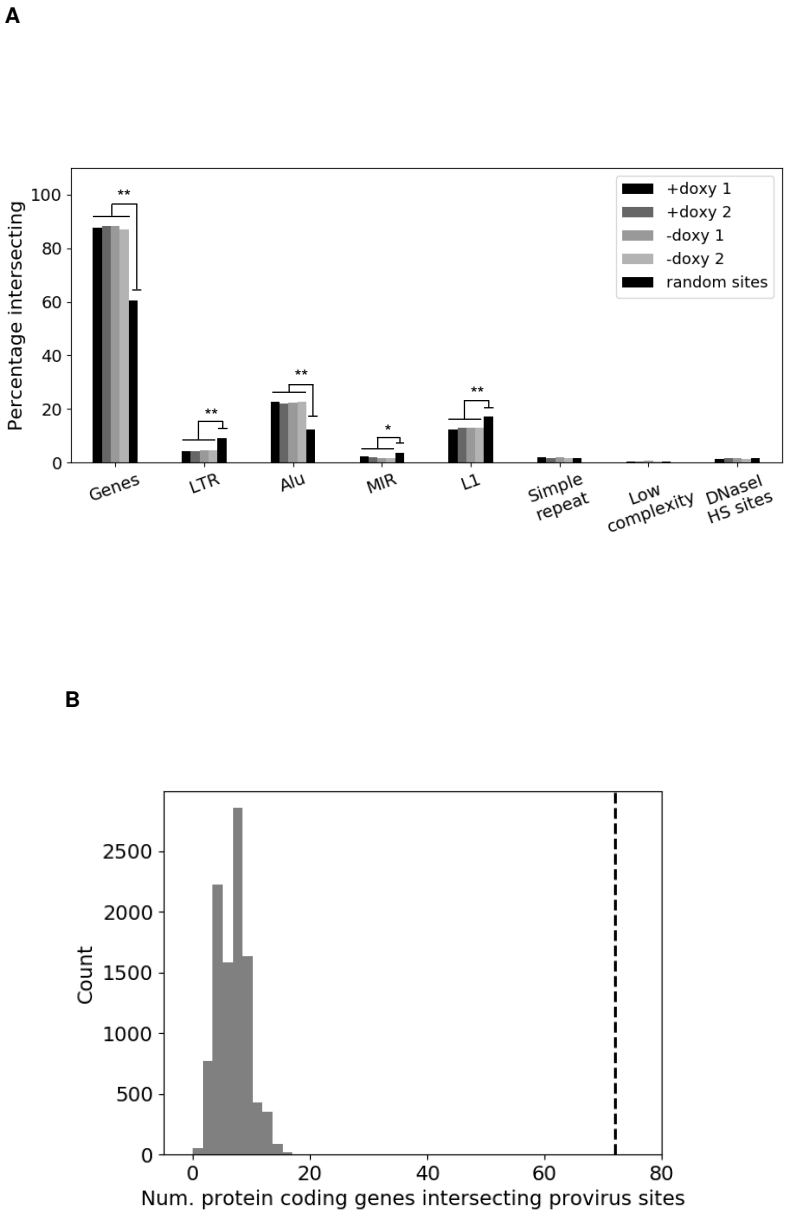


Figure S6: Confirming expected lentivirus integration preferences. (A) Percentage of 1,474 target sites recovered in each biological sample that intersect genes (GENCODEv31) and repeat elements (RepeatMasker). Random sites are provirus sites whose positions are randomly permuted along the same chromosome using bedtools[90] shuffle with the option -chrom. The sum of the percentage of sites intersecting all features for a single sample is greater than 100% due to the fact that some of the features overlap one another. P values are determined by one-way chi-square test for the distribution of intersecting and non-intersecting provirus sites compared to random positions in each category, requiring at least five counts in every category (observed and expected). *, $P < 0.05$. **, $P < 0.001$. Simple repeats: microsatellites. Low complexity repeats: poly-purine/poly-pyrimidine runs, simple tandem repeats, regions of high AT/GC content. (B) Distribution showing the number of times genes in a randomly selected protein coding gene set of the same size as the RIG gene set[54] (155 genes) intersect at least one provirus site in 10,000 permutations. Intersection of actual RIG gene set with provirus sites given by dotted black line, $P = 0.0$. P values determined as frequency at which the intersection with the random gene set is at least as large as the intersection with the RIG gene set, divided by the number of permutations.

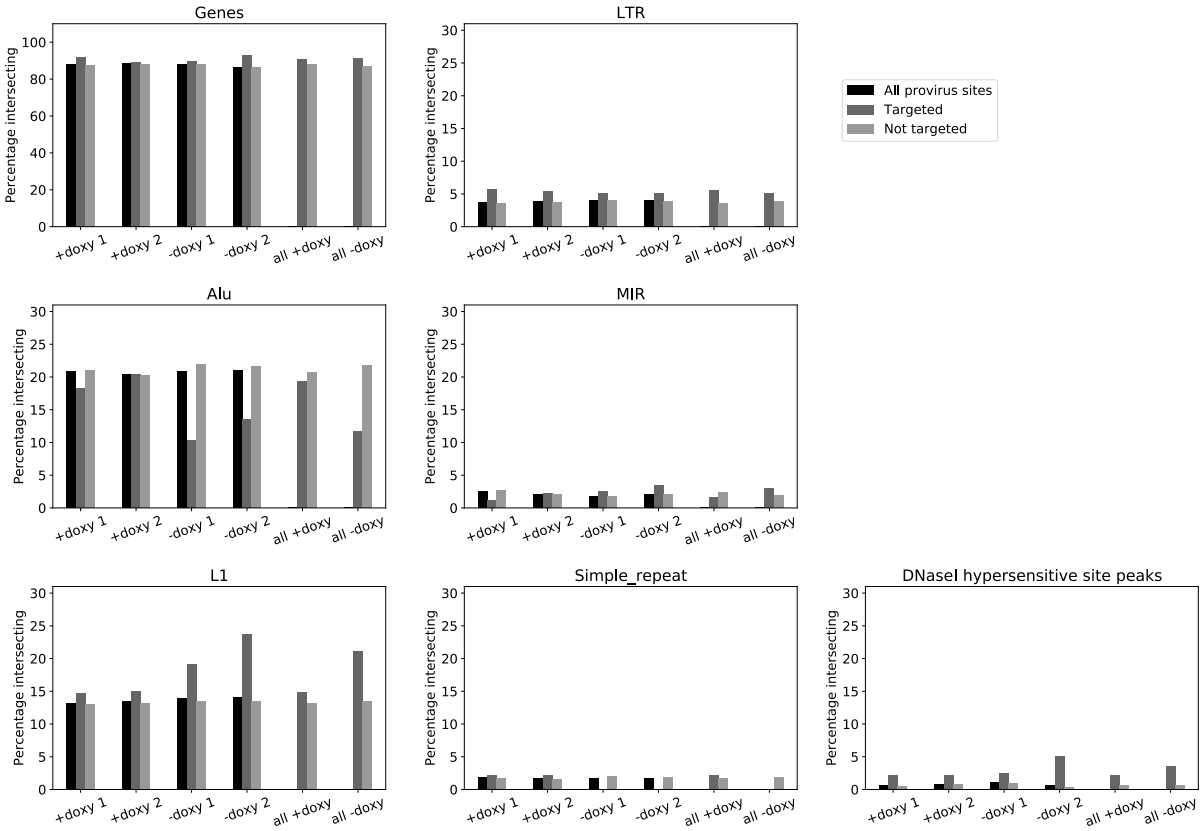


Figure S7: Association of targeted sites with chromosomal features. Percentage target sites recovered in each biological sample that intersect genes (GENCODEv31), repeat elements (RepeatMasker), and DNaseI-seq called peaks.[61] Shown are all provirus sites considered as potential targets, sites with at least one integrated barcode ('Targeted'), and sites with no integrated barcodes ('Not targeted'). The sum of the percentage of sites intersecting all features for a single sample is greater than 100% due to the fact that some of the features overlap one another. Repeat types with too small of sample size after intersection for statistical analysis were excluded (in this case, low-complexity repeats). Total counts are given in Table S2.

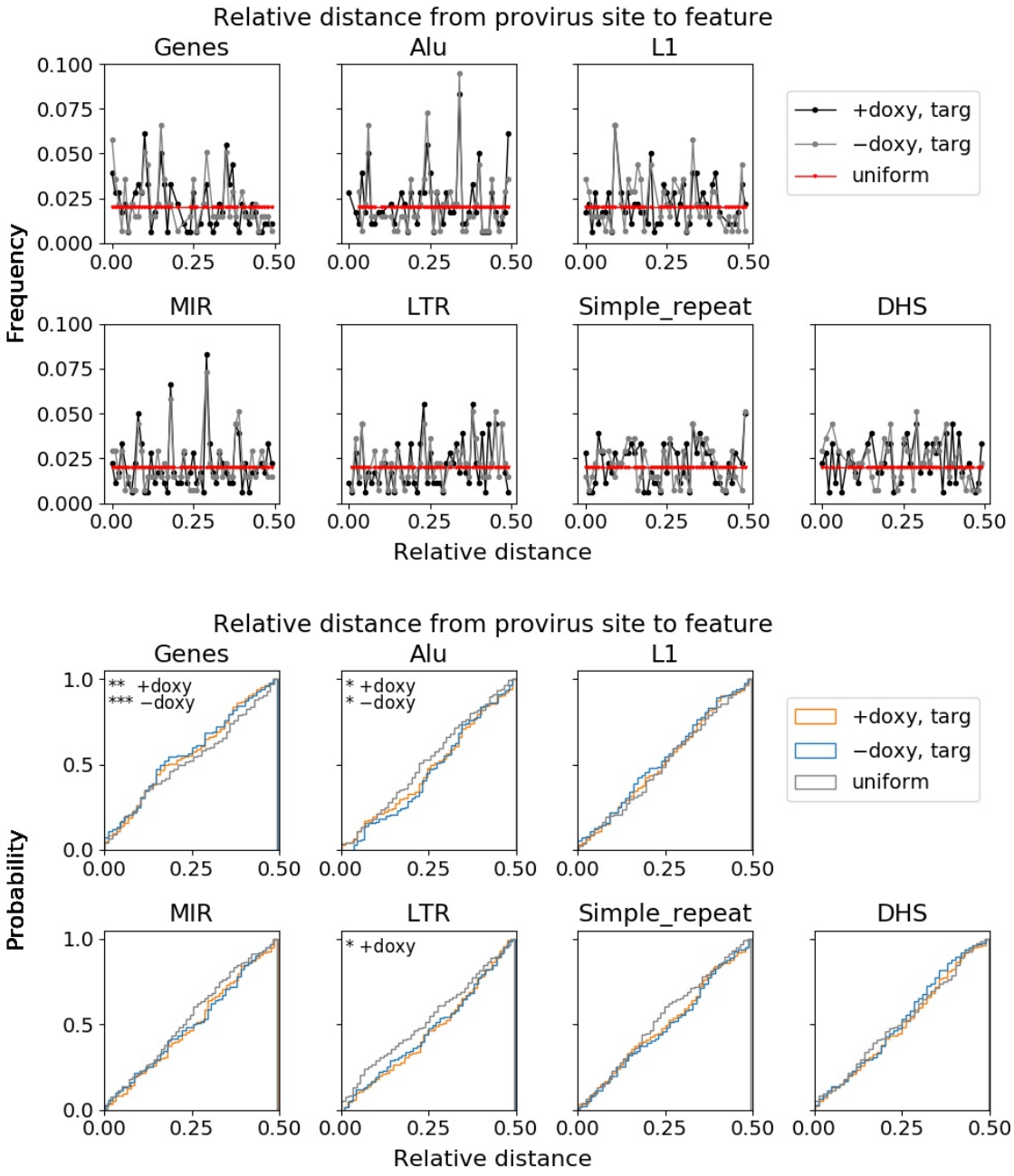


Figure S8: Relative distance of targeted sites to chromosomal features. Plots of frequency and probability (cumulative distribution functions) of relative distance from each provirus site to the nearest chromosomal feature, computed using bedtools reldist. P values shown were determined by a two-sided Kolmogorov-Smirnov test against a uniform distribution in [0,0.5]. *, $P < 0.05$. **, $P < 0.01$. ***, $P < 0.001$. No significant differences were observed in a two-sided Kolmogorov-Smirnov test comparing the two populations in each plot. +doxycycline n=180, -doxycycline n=136. **DHS**, DNaseI hypersensitive site peaks.

Table S1: (TSV table) GEO2R analysis of log fold change in expression in MCF12A breast epithelial cell line treated with 1ug/mL doxycycline or vehicle control for four days as measured by Affymetrix GeneChip (GSE45029).[98] There are three DNA-PK (*PRKDC*) probes and none showed significant upregulation or downregulation of DNA-PK with doxcycyline addition. **adj.P.Val**, *P*-value after adjustment for multiple testing using the Benjamini & Hochberg false-discovery rate method. **P.Value**, raw *P*-value. **t**, moderated t-statistic. **B**, B-statistic or log-odds that gene is differentially expressed. **logFC**, log2-fold change between treatment and control.

Biological rep.	+doxy 1	+doxy 2	-doxy 1	-doxy 2
Provirus sites	1024	989	933	955
Targeted provirus sites	88	93	78	59
Shared between replicates	61		43	
Shared between treatment groups	194			
DNA barcodes from barcode PCR	7690	6297	6765	5287
Mapped, accepted barcodes	933	969	1046	765

Table S2: Summary of site and barcode counts for each biological replicate. For sites shared between treatment groups, we consider concatenated replicates. See Materials and Methods for detailed description of intersection and filtering steps.

Expression level bin	Num. transcripts in bin	Median FPKM	FPKM IQR
all transcripts, low	50,972	0.0	0.05
all transcripts, med	3,795	13.11	7.81
all transcripts, high	3,400	48.60	57.11
provirus sites, low	396	1.32	3.73
provirus sites, med	395	15.08	7.44
provirus sites, high	396	44.52	39.97

Table S3: Number of all transcripts or transcripts intersecting provirus sites from Rodriguez-Castaneda et al.[59] that would be placed in each of the bins used in Figure 5, which were determined by splitting the genes that intersect provirus sites into equal sized bins. **IQR**, interquartile range.

State	RR	LCB	UCB	Num. targ.	Num. untarg.
PromU	1.848	0.147	23.289	0	2
PromD1	0.79	0.054	11.463	0	6
PromD2	1.173	0.314	4.382	2	17
Tx5'	1.397	0.979	1.994	33	244
Tx	0.792	0.4	1.568	8	103
Tx3'	1.416	1.02	1.966	41	304
TxWk	0.823	0.578	1.173	35	419
TxReg	1.705	0.751	3.869	5	28
TxEnh5'	1.071	0.46	2.494	5	47
TxEnh3'	0.424	0.028	6.45	0	12
TxEnhW	0.482	0.158	1.47	3	65
EnhA1	0.79	0.054	11.463	0	6
EnhA2	3.227	1.393	7.475	4	10
EnhAF	0.367	0.024	5.625	0	14
EnhW2	1.395	0.379	5.137	2	14
EnhAc	0.926	0.141	6.081	1	11
DNase	1.848	0.147	23.289	0	2
Het	1.848	0.147	23.289	0	2
PromP	1.858	0.309	11.182	1	5
PromBiv	1.848	0.147	23.289	0	2
ReprPC	0.552	0.037	8.261	0	9
Quies	0.743	0.532	1.038	41	528

Table S4: Relative risk ratio, lower (LCB) and upper (UCB) confidence interval bounds, and number intersecting segments for all K562 ChromHMM states intersecting targeted and untargeted provirus sites in +doxycycline treatment group, related to Figure 6. Targeted n=181 and untargeted n=1832.

State	RR	LCB	UCB	Num. targ.	Num. untarg.
PromU	2.287	0.181	28.869	0	2
PromD1	0.978	0.067	14.207	0	6
PromD2	0.324	0.021	5.029	0	20
Tx5'	1.735	1.177	2.558	29	224
Tx	0.973	0.467	2.025	7	92
Tx3'	1.025	0.671	1.567	24	300
TxWk	0.657	0.422	1.023	22	404
TxReg	1.341	0.452	3.98	3	28
TxEnh5'	1.756	0.816	3.779	6	42
TxEnh3'	0.525	0.034	7.992	0	12
TxEnhW	1.083	0.417	2.815	4	47
EnhA1	1.142	0.08	16.302	0	5
EnhA2	1.85	0.504	6.79	2	13
EnhAF	0.487	0.032	7.446	0	13
EnhW1	3.433	0.31	38.061	0	1
EnhW2	2.796	1.003	7.793	3	12
EnhAc	1.15	0.175	7.56	1	11
DNase	1.371	0.098	19.105	0	4
ZNF/Rpts	3.433	0.31	38.061	0	1
Het	2.287	0.181	28.869	0	2
PromP	0.978	0.067	14.207	0	6
PromBiv	3.433	0.31	38.061	0	1
ReprPC	0.487	0.032	7.446	0	13
Quies	0.905	0.629	1.302	37	511

Table S5: Relative risk ratio, lower (LCB) and upper (UCB) confidence interval bounds, and number intersecting segments for all K562 ChromHMM states intersecting targeted and untargeted provirus sites in -doxycycline treatment group, related to Figure 6. Targeted n=137 and untargeted n=1751.

State or marker	Estimate	Stderr	Pval	Source	Condition
TxEnh5'	-7.e-01	7.6e-01	3.6e-01	ChromHMM	+doxy
TxEnh3'	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
TxEnhW	-6.5e-01	9.4e-01	4.9e-01	ChromHMM	+doxy
EnhA1	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
EnhA2	-8.7e-01	9.3e-01	3.5e-01	ChromHMM	+doxy
EnhAF	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
EnhW1	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
EnhW2	4.e-01	5.3e-01	4.5e-01	ChromHMM	+doxy
EnhAc	-1.7e+01	1.5e+01	2.5e-01	ChromHMM	+doxy
DNase	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
PromU	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
ZNF/Rpts	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
Het	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
PromP	-1.2e+01	1.0e+01	2.5e-01	ChromHMM	+doxy
PromBiv	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
ReprPC	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
Quies	1.3e-01	1.8e-01	4.6e-01	ChromHMM	+doxy
PromD1	1.1e-14	2.7e+04	1.e+00	ChromHMM	+doxy
PromD2	-3.7e+00	3.1e+00	2.4e-01	ChromHMM	+doxy
Tx5'	-3.1e-01	2.4e-01	2.0e-01	ChromHMM	+doxy
Tx	7.2e-01	2.7e-01	8.5e-03	ChromHMM	+doxy
Tx3'	-4.8e-03	1.8e-01	9.8e-01	ChromHMM	+doxy
TxWk	1.3e-01	1.9e-01	4.9e-01	ChromHMM	+doxy
TxReg	-4.0e-01	8.0e-01	6.1e-01	ChromHMM	+doxy
H3K27ac	-6.5e-01	6.2e-01	3.e-01	Encode	+doxy
H3K27me3	-6.7e+00	6.6e+00	3.1e-01	Encode	+doxy
H3K36me3	-3.e-02	5.2e-01	9.5e-01	Encode	+doxy
H3K4me1	-8.4e-02	5.4e-01	8.8e-01	Encode	+doxy
H3K4me2	-7.5e-01	6.6e-01	2.6e-01	Encode	+doxy
H3K4me3	-3.1e+00	2.3e+00	1.8e-01	Encode	+doxy
H3K79me2	-3.2e-01	2.7e-01	2.4e-01	Encode	+doxy
H3K9me1	1.1e-14	2.7e+04	1.e+00	Encode	+doxy
TxEnh5'	-3.3e-02	4.3e-01	9.4e-01	ChromHMM	-doxy
TxEnh3'	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
TxEnhW	-9.9e-01	9.6e-01	3.0e-01	ChromHMM	-doxy
EnhA1	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
EnhA2	9.8e-01	4.9e-01	4.6e-02	ChromHMM	-doxy
EnhAF	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
EnhW1	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
EnhW2	4.4e-01	4.4e-01	3.2e-01	ChromHMM	-doxy
EnhAc	-3.3e+00	4.1e+00	4.2e-01	ChromHMM	-doxy
DNase	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
PromU	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
ZNF/Rpts	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
Het	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
PromP	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
PromBiv	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
ReprPC	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
Quies	-2.2e-01	2.2e-01	3.1e-01	ChromHMM	-doxy
PromD1	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy
PromD2	-3.7e-15	3.1e+04	1.e+00	ChromHMM	-doxy

State or marker	Estimate	Stderr	Pval	Source	Condition
Tx5'	-2.9e-01	2.5e-01	2.5e-01	ChromHMM	-doxy
Tx	2.6e-01	3.3e-01	4.4e-01	ChromHMM	-doxy
Tx3'	2.3e-01	2.1e-01	2.7e-01	ChromHMM	-doxy
TxWk	7.0e-04	2.4e-01	1.e+00	ChromHMM	-doxy
TxReg	6.e-01	5.0e-01	2.3e-01	ChromHMM	-doxy
H3K27ac	8.3e-01	3.6e-01	2.3e-02	Encode	-doxy
H3K27me3	-3.7e-15	3.1e+04	1.e+00	Encode	-doxy
H3K36me3	-1.9e-01	5.2e-01	7.1e-01	Encode	-doxy
H3K4me1	3.0e-01	4.2e-01	4.7e-01	Encode	-doxy
H3K4me2	6.e-01	5.0e-01	2.3e-01	Encode	-doxy
H3K4me3	-3.7e-15	3.1e+04	1.e+00	Encode	-doxy
H3K79me2	-3.3e-01	2.9e-01	2.4e-01	Encode	-doxy
H3K9me1	-3.7e-15	3.1e+04	1.e+00	Encode	-doxy

Table S6: Assessing the relationship between barcode heterogeneity and chromatin states and epigenetic measures. The presence of an overlapping ChromHMM segment or epigenetic peak was predicted by a logistic regression model, using as predictor variables the barcode heterogeneity at targeted sites and using as response variables the binary presence of an overlapping K562 chromatin state segment or epigenetic peak. Related to Figure 6.

.

State or marker	Estimate	Stderr	Pval	Source	Condition
TssA	nan	nan	nan	ChromHMM	+doxy
TxEnh5'	-1.5e-01	1.4e-01	3.0e-01	ChromHMM	+doxy
TxEnh3'	-7.2e-02	2.3e-01	7.5e-01	ChromHMM	+doxy
TxEnhW	-2.0e-01	1.1e-01	6.9e-02	ChromHMM	+doxy
EnhA1	-2.7e-01	2.9e-01	3.5e-01	ChromHMM	+doxy
EnhA2	-5.2e-01	4.2e-01	2.2e-01	ChromHMM	+doxy
EnhAF	-2.5e-01	3.8e-01	5.1e-01	ChromHMM	+doxy
EnhW1	-2.5e-01	2.1e-01	2.4e-01	ChromHMM	+doxy
EnhW2	-2.4e-02	2.3e-01	9.2e-01	ChromHMM	+doxy
EnhAc	-4.8e-01	4.7e-01	3.1e-01	ChromHMM	+doxy
DNase	-2.4e-01	6.6e-01	7.2e-01	ChromHMM	+doxy
PromU	nan	nan	nan	ChromHMM	+doxy
ZNF/Rpts	nan	nan	nan	ChromHMM	+doxy
Het	nan	nan	nan	ChromHMM	+doxy
PromP	-5.6e-01	3.8e-01	1.4e-01	ChromHMM	+doxy
PromBiv	nan	nan	nan	ChromHMM	+doxy
ReprPC	nan	nan	nan	ChromHMM	+doxy
Quies	6.7e-03	2.7e-02	8.0e-01	ChromHMM	+doxy
PromD1	nan	nan	nan	ChromHMM	+doxy
PromD2	-1.5e-01	1.3e-01	2.5e-01	ChromHMM	+doxy
Tx5'	-6.e-02	3.2e-02	6.8e-02	ChromHMM	+doxy
Tx	2.1e-01	6.2e-02	9.8e-04	ChromHMM	+doxy
Tx3'	4.9e-04	2.8e-02	9.9e-01	ChromHMM	+doxy
TxWk	5.5e-02	3.1e-02	7.9e-02	ChromHMM	+doxy
TxReg	-7.2e-02	9.2e-02	4.4e-01	ChromHMM	+doxy
H3k27ac	-6.8e-02	3.9e-02	8.2e-02	Encode	+doxy
H3k27me3	-9.0e-02	6.2e-02	1.5e-01	Encode	+doxy
H3k36me3	-2.5e-02	2.9e-02	3.8e-01	Encode	+doxy
H3k4me1	-7.5e-02	3.3e-02	2.4e-02	Encode	+doxy
H3k4me2	-8.0e-02	4.4e-02	6.8e-02	Encode	+doxy
H3k4me3	-7.0e-02	5.3e-02	1.9e-01	Encode	+doxy
H3k9ac	-6.2e-02	4.3e-02	1.5e-01	Encode	+doxy
H4k20me1	-1.2e-02	3.e-02	6.9e-01	Encode	+doxy
TssA	nan	nan	nan	ChromHMM	-doxy
TxEnh5'	1.2e-01	1.3e-01	3.6e-01	ChromHMM	-doxy
TxEnh3'	-2.e-01	2.5e-01	4.2e-01	ChromHMM	-doxy
TxEnhW	-1.3e-01	1.2e-01	2.8e-01	ChromHMM	-doxy
EnhA1	-3.8e-01	3.e-01	2.1e-01	ChromHMM	-doxy
EnhA2	8.4e-01	5.7e-01	1.5e-01	ChromHMM	-doxy
EnhAF	-3.5e-01	2.5e-01	1.6e-01	ChromHMM	-doxy
EnhW1	5.6e-01	2.3e-01	1.6e-02	ChromHMM	-doxy
EnhW2	5.3e-01	2.3e-01	2.5e-02	ChromHMM	-doxy
EnhAc	-4.1e-01	5.e-01	4.1e-01	ChromHMM	-doxy
DNase	nan	nan	nan	ChromHMM	-doxy
PromU	nan	nan	nan	ChromHMM	-doxy
ZNF/Rpts	nan	nan	nan	ChromHMM	-doxy
Het	nan	nan	nan	ChromHMM	-doxy
PromP	nan	nan	nan	ChromHMM	-doxy
PromBiv	nan	nan	nan	ChromHMM	-doxy
ReprPC	nan	nan	nan	ChromHMM	-doxy
Quies	-3.4e-02	3.0e-02	2.6e-01	ChromHMM	-doxy
PromD1	nan	nan	nan	ChromHMM	-doxy
PromD2	-1.1e-01	3.3e-01	7.5e-01	ChromHMM	-doxy

State or marker	Estimate	Stderr	Pval	Source	Condition
Tx5'	-5.4e-02	3.8e-02	1.6e-01	ChromHMM	-doxy
Tx	1.1e-01	7.5e-02	1.5e-01	ChromHMM	-doxy
Tx3'	3.3e-02	3.9e-02	4.e-01	ChromHMM	-doxy
TxWk	3.9e-02	4.3e-02	3.7e-01	ChromHMM	-doxy
TxReg	7.5e-02	1.1e-01	5.e-01	ChromHMM	-doxy
H3k27ac	3.5e-02	4.3e-02	4.1e-01	Encode	-doxy
H3k27me3	-8.0e-02	7.8e-02	3.0e-01	Encode	-doxy
H3k36me3	-8.4e-03	3.4e-02	8.0e-01	Encode	-doxy
H3k4me1	1.6e-02	3.6e-02	6.5e-01	Encode	-doxy
H3k4me2	3.1e-02	4.9e-02	5.2e-01	Encode	-doxy
H3k4me3	-2.9e-03	6.3e-02	9.6e-01	Encode	-doxy
H3k9ac	-5.e-02	5.4e-02	3.6e-01	Encode	-doxy
H4k20me1	2.7e-02	3.5e-02	4.5e-01	Encode	-doxy

Table S7: Estimates and standard errors resulting from fitting independent linear regression models using ChromHMM and Encode features, related to Figure 6. **nan** indicates where there were not enough intersections to fit the model.

State or marker	Estimate	Stderr	Pval	Source	Condition
TssA	nan	nan	nan	ChromHMM	+doxy
PromU	-5.7e+00	5.6e+00	3.1e-01	ChromHMM	+doxy
PromD1	-1.2e+01	1.7e+01	4.8e-01	ChromHMM	+doxy
PromD2	-4.e-01	3.5e-01	2.6e-01	ChromHMM	+doxy
Tx5'	-4.0e-02	3.3e-02	2.3e-01	ChromHMM	+doxy
Tx	1.5e-01	6.2e-02	1.8e-02	ChromHMM	+doxy
Tx3'	2.3e-03	3.2e-02	9.4e-01	ChromHMM	+doxy
TxWk	3.6e-02	3.4e-02	2.8e-01	ChromHMM	+doxy
TxReg	-1.7e-01	2.e-01	3.8e-01	ChromHMM	+doxy
TxEnh5'	-1.4e-01	1.0e-01	1.8e-01	ChromHMM	+doxy
TxEnh3'	-3.3e-01	9.4e-01	7.3e-01	ChromHMM	+doxy
TxEnhW	-2.e-01	1.1e-01	7.7e-02	ChromHMM	+doxy
EnhA1	-8.6e-01	7.2e-01	2.3e-01	ChromHMM	+doxy
EnhA2	-4.e-01	4.2e-01	3.5e-01	ChromHMM	+doxy
EnhAF	-7.0e-01	6.6e-01	2.9e-01	ChromHMM	+doxy
EnhW1	-7.7e-01	1.0e+00	4.5e-01	ChromHMM	+doxy
EnhW2	-2.5e-01	2.8e-01	3.8e-01	ChromHMM	+doxy
EnhAc	-1.6e+00	8.5e-01	6.8e-02	ChromHMM	+doxy
DNase	-5.2e-01	4.5e-01	2.5e-01	ChromHMM	+doxy
ZNF/Rpts	-3.9e+00	3.e+00	1.9e-01	ChromHMM	+doxy
Het	nan	nan	nan	ChromHMM	+doxy
PromP	-1.3e+00	1.0e+00	2.1e-01	ChromHMM	+doxy
PromBiv	nan	nan	nan	ChromHMM	+doxy
ReprPC	1.0e+00	2.4e+00	6.7e-01	ChromHMM	+doxy
Quies	1.7e-02	2.9e-02	5.5e-01	ChromHMM	+doxy
TssA	nan	nan	nan	ChromHMM	-doxy
PromU	-4.9e+00	6.0e+00	4.1e-01	ChromHMM	-doxy
PromD1	-1.5e+00	4.5e+00	7.5e-01	ChromHMM	-doxy
PromD2	-1.0e+00	1.5e+00	5.1e-01	ChromHMM	-doxy
Tx5'	-4.3e-02	3.9e-02	2.8e-01	ChromHMM	-doxy
Tx	8.7e-02	7.7e-02	2.6e-01	ChromHMM	-doxy
Tx3'	6.2e-02	4.7e-02	2.e-01	ChromHMM	-doxy
TxWk	7.e-03	4.9e-02	8.9e-01	ChromHMM	-doxy
TxReg	-4.7e-03	2.4e-01	9.8e-01	ChromHMM	-doxy
TxEnh5'	5.4e-02	1.1e-01	6.1e-01	ChromHMM	-doxy
TxEnh3'	-2.9e-01	9.3e-01	7.5e-01	ChromHMM	-doxy
TxEnhW	-3.7e-02	1.3e-01	7.8e-01	ChromHMM	-doxy
EnhA1	-2.3e-01	7.5e-01	7.6e-01	ChromHMM	-doxy
EnhA2	6.1e-01	4.3e-01	1.6e-01	ChromHMM	-doxy
EnhAF	-1.1e+00	6.2e-01	8.1e-02	ChromHMM	-doxy
EnhW1	2.7e+00	1.1e+00	1.1e-02	ChromHMM	-doxy
EnhW2	5.1e-01	2.9e-01	8.5e-02	ChromHMM	-doxy
EnhAc	2.6e-01	9.9e-01	7.9e-01	ChromHMM	-doxy
DNase	6.3e-01	4.6e-01	1.7e-01	ChromHMM	-doxy
ZNF/Rpts	-2.5e+00	3.2e+00	4.3e-01	ChromHMM	-doxy
Het	nan	nan	nan	ChromHMM	-doxy
PromP	2.4e+00	2.3e+00	3.e-01	ChromHMM	-doxy
PromBiv	nan	nan	nan	ChromHMM	-doxy
ReprPC	-2.0e+00	3.0e+00	5.e-01	ChromHMM	-doxy
Quies	-5.0e-02	3.3e-02	1.3e-01	ChromHMM	-doxy

Table S8: Estimates and standard errors resulting from fitting independent linear regression models using as predictor variables the barcode heterogeneity at targeted sites and using as response variables the proportion of 127 cell types assigned to a given state at each site. **nan** indicates where there were not enough intersections to fit the model.