Supplementary Information

Epigenetic mechanisms, T-cell activation, and *CCR5* genetics interact to regulate T-cell expression of CCR5, the major HIV-1 coreceptor

Materials and Methods

A. Study participants and sources of blood samples.

Samples were from the following sources as summarized in the following table.

Sources of Study Subjects/Samples	Description in SI	n	Data shown in Figures/Tables
Healthy donors, buffy coats from blood bank	(i)	39	Fig 2A-J, Fig 3A-D, Fig 4A-D, Fig. 5B-E and 5K-L, Fig 6A, Fig 7E-F, Fig 8A, Fig S3, Fig S6B
HIV infected patients on virally suppressive ART from SCOPE	(ii)	85	Fig 6B-C, Fig 7B-D, Fig 8B-C, Fig S5, Table S2.
Durban (South Africa) hospital employees	(iii)	84	Fig. S9A right panel
Acute HIV infection from NYU cohort	(iv)	29	Fig 8A
EC/VC from SCOPE cohort, NYU cohort, and SAMMC	(v)	42	Fig 8A
Children exposed perinatally to HIV from the Ukraine	(vi)	135	Fig S9A, left panel
HIV-infected women from Ukraine	(vi)	81	Fig S6A
CAPRISA from Durban, South Africa of black female sex workers	(vii)	189	Fig.7G-H, Fig.8D-E, Fig. S9B
HGDP-CEPH population	(viii)	354	Fig.7G-H
HIV infected individuals from Johannesburg, South Africa	(ix)	78	Fig.7G
Chimpanzees from Texas Biomedical Research Institute	(x)	92	Fig 7F,G

SCOPE, Study of the Consequences of the Protease Inhibitor Era; NYU, New York University; SAMMC, San Antonio Military Medical Center; CAPRISA, Centre for the AIDS Programme of Research in South Africa. HGDP-CEPH: Human Genome Diversity Project (HGDP)-CEPH

(i) Healthy donors were from the University of Texas Health Science Center at San Antonio, TX, USA. For some experiments, buffy coats were from anonymous HIV-uninfected blood donors obtained from South Texas Blood and Tissue Center, San Antonio, TX, USA. Genomic DNA isolated from buccal swabs from healthy donors was used to assess *CCR5* methylation status in epithelial cells.

(ii) 85 HIV-infected persons receiving viral load (VL) suppressive combination antiretroviral therapy (ART) with stably suppressed VL for greater than 6 months. Two evaluations were conducted on the peripheral blood mononuclear cells (PBMC) from these subjects: *CCR5* methylation status and extensive flow cytometric analyses. Methylation status and flow cytometric analyses were from PBMC derived on the same day. These study participants were part of the Study of the Consequences of the Protease Inhibitor Era (SCOPE), a clinic-based cohort of >1000 chronically HIV-infected individuals followed at the University of California, San Francisco (UCSF). Additional demographic and clinical characteristics of these participants are described below.

Characteristics	
Age — Median (IQR), years old	45.6 (39.8 – 51.9)
Gender — no. (%)	
Male	64 (75.3)
Female	21(24.7)
Race — no. (%)	
European American	50 (58.8)
African Americans	20 (23.5)
Other race	15 (17.6)
CD4+ T cell counts (cells/mm^3) — Median (IQR)	398 (273 – 561)
Pre-ART CD4+ (cells/mm^3) — Median (IQR)	175 (117 – 268)
Pre-ART VL — Median (IQR), copies	49,966 (19,000 – 104,102)

IQR, interquartile range

(iii) Study participants were employees from hospitals in Durban, South Africa who volunteered for this study and in whom HIV-1 serostatus was evaluated. 24 individuals were HIV-infected and 60 were uninfected.

(iv) 29 participants who were accrued during acute HIV-1 infection at New York University (NYU), approximately 58 days (median=58 days, IQR=43-80 days) after the estimated date of infection (EDI). Median age was 35 years (IQR, 30-40). Eleven

(37.9%), 12 (44.8%), and 5 (17.2%) were European-Americans, African-Americans and of other ethnicities, respectively. These HIV-infected persons received ART during acute HIV infection approximately 58 days (median) post-EDI, i.e., methylation status was analyzed on the same day that suppressive combination ART was initiated. The post-ART methylation status was evaluated approximately 239 days after initiation of ART (median=239 days, IQR=223-336 days), and all subjects had undetectable VL at time of evaluation of post-ART methylation status. The median (IQR) of the pre- and post-ART CD4+ T cell counts were 515 (431-639) and 677 (576-838) cells/mm^3, respectively (*P* = .001, difference between pre- and post-ART CD4+ T cell counts by paired t-test). The mean (SD) and median (IQR) plasma HIV RNA levels (VL) at time of ART initiation was 128,309 (258,296) and 32,912 (8,809-78,785) copies/ml, respectively.

(v) 42 HIV controllers (HIC) were categorized as elite controllers (EC) and viremic controllers (VC) using definitions as described previously (1, 2). These subjects came from 3 sources: SCOPE cohort (University of California, San Diego, CA), New York University, NY, and San Antonio Military Medical Center, San Antonio, TX.

(vi) HIV-uninfected or infected Ukrainian children exposed perinatally to HIV and HIV-infected mothers who were part of a previously described mother-child cohort of perinatal infection (3). The HIV-negative children were selected on the criteria that neither they nor their mothers had received antiretroviral prophylaxis (n=44). The HIV-infected children were also selected on the same basis (n=91). We used this criterion

for selection of the children to obviate any potential confounding effect of antiretroviral treatment on methylation status.

(vii) Subjects from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) cohort who were part of the CAPRISA 002 acute infection study (4). They comprised of a high-risk group of female sex workers (FSW) from Durban, KwaZulu-Natal, South Africa, an area with a very high prevalence of HIV infection (4). The characteristics of this cohort were described previously (4). These FSWs were enrolled into the cohort when they were seronegative. During the intensive prospective follow-up period, a subset of women seroconverted, whereas others remained seronegative for >2 yrs. (highly exposed uninfected) (4). Genotype frequency for *CCR5* −4223C/T polymorphism (rs553615728) and its linkage disequilibrium with *CCR5* HHA haplotype were studied in 26 FSWs who subsequently seroconverted and 120 FSWs who remained HIV negative with ≥2 years follow up. Subjects enrolled in other acute infection cohorts and followed at CAPRISA were also evaluated for genotype frequency.

We also evaluated methylation status in PBMC from FSWs who were therapy-naïve and for whom samples were available at two timepoints during early HIV disease (n=29). The median time interval between the first and second PBMC samples analyzed for methylation status was 295 days (IQR=113 days).

(viii) Human Genome Diversity Project (HGDP)-CEPH: We genotyped the *CCR5* −4223C/T polymorphism from a representative panel (n=354) from the HGDP-CEPH Cell Line panel. This panel comprised 1,065 individuals from 57 human populations with

minimal admixture. The characteristics of this panel are as described previously (5, 6), and also at the following web site: http://research.marshfieldclinic.org/genetics/Freq/DiversityPanel.xls.

(ix) HIV-1 infected individuals from Johannesburg, South Africa: This study group comprised HIV-1-infected black South African individuals, recruited from Soweto and Johannesburg, who were categorized into three groups according to the extent of their HIV-1 control or disease. (a) Long term non-progressors (LTNP) (n=11) were HIV-1-infected individuals with VL>10 000 RNA copies/ml for a period of ≥7 years, who were not on ART and had no demonstrable CD4+ T cell decline. (b) HIV-1 controllers (HICs) (n=41) were defined as HIV-1-infected individuals with viral loads (VL)<2000 HIV RNA copies/ml in the absence of ART. Among them 11 had VL<50 HIV RNA copies/ml in the absence of ART which were maintained for at least two clinic visits. (c) Progressors (n=66) comprised HIV-1-infected individuals requiring ART and demonstrating CD4+ T cell decline to <300 cells/mm^3 and with VL>10,000 RNA copies/ml. Further details of the HIV-1-infected individual subgroups are available in the accompanying table; CD4+ T cell counts and VLs represent the values obtained from individual participants at their first clinic visit after enrollment into this study.

Characteristics of HIV-1-infected black South African study participants classified according to different phenotypes of HIV-1 disease control

Group	n	Age (years)*	Gender (% females)	CD4+ T cell counts (cells/µl)*	Viral load (HIV RNA copies/ml)*
Long term nonprogressors (LTNPs)	11	41 (34 – 44)	81.8	663 (635 – 749)	54,375 (13,415 – 77,820)
HIV controllers	41	36 (31 – 40)	87.8	722 (580 – 883)	360 (124 – 860)
Progressors	66	38 (34 – 41)	81.8	177 (146 – 210)	38,444 (19,970 – 107,347)

*median (interquartile range) are shown

(x) Chimpanzees from Texas Biomedical Research Institute, San Antonio, TX: CCR5 expression in sorted CD4+ T cells was assessed by flow cytometry using PBMCs from 12 chimpanzees. These chimpanzees were 20.7 (IQR, 20.2 – 22.8) years old at time of the sampling and 5 of them were female. The HIV and Hepatitis virus infection status for these animals is unknown. *CCR5* −4223C/T genotype data was obtained from 80 chimpanzees that were previously described by us (6).

B. Cell isolation, flow cytometry, and FACS analyses.

PBMCs were isolated from fresh whole blood through density gradient sedimentation on Histopaque-1077 (Sigma). Cell viability was assessed by Trypan Blue staining. CD3+, CD4+ or CD8+ CD45RO−CD45RA+ and/or CD45RO+CD45RA− T cells were isolated using AutoMACS (Miltenyi Biotec) through specific labeling kits using negative and positive selection according to standard protocols (Pan T- Cell Isolation, CD45RA and CD45RO Microbeads Kits, Miltenyi) and depending on the purpose of the experiment. Sample acquisition was performed with a FACSCantoII (BD Bioscience) and analyzed

by FACS Diva v6.1 (BD Biosciences). Unless otherwise noted, all FACS analyses were conducted on freshly isolated PBMC and T cells. Gates that encompassed live lymphocytes were set based on forward- and side-scatter measurements. For quantification, ≥10,000 gated events were collected for each sample. The following mouse-derived anti-human monoclonal antibodies (mAb) and appropriate controls were obtained from BD: CD3-Pacific blue or PE, CD4-Amcyan, CD8-Pe-Cy7 or CD8-PE, CCR5 (Clone 2D7)-FITC, CD45RO-APC, CD45RA PE, HLA-DR-APC-Cy7 and CD38 PerCP- Cy5.5. In some experiments, pan T cells were isolated using the AutoMACS as described above and sorted based on the expression of CCR5 (positive and negative) in the following groups: i) CD3+CD4+CD45RO+ and CD3+CD4+CD45RO- and ii) central memory (CD45RA-CD45RO+CCR7+) and effector memory (CD45RA-CD45RO+CCR7-). T cell sorting was conducted by a FACSAria III (BD Biosciences) at the Institutional Flow Cytometry core at UTHSCSA. Purity of the isolated T cells after final sorting was >98%. We used similar protocols for determining CCR5 expression on CD4+ T cells in chimpanzees.

Additional flow cytometry analyses on subjects with ART-induced VL suppression were conducted in the following manner on cryopreserved PBMC. The frequencies of activated (CD38+ HLA-DR+) CD4+ and CD8+ T cells were measured by flow cytometry at the University of California San Francisco Core Immunology Laboratory, using methods optimized and validated for frozen PBMCs (7). Cryopreserved PBMCs were rapidly thawed in warm media, counted on a Guava PCA with the Viacount assay (Millipore), and washed and stained the same day. Thawed cells were stained with

Aqua Amine Reactive Dye (AARD, Life Technologies) to discriminate dead cells, washed, and then stained. To perform the immunophenotyping, two different protocols were followed: one to assess levels of T-cell activation/differentiation and CCR5 levels (Method #1) and one to determine the maturational status of T cells (i.e. naïve/memory differentiation status) (Method #2).

For Method #1, the following fluorescently-conjugated monoclonal antibodies were used: CD3-Pacific Blue, CD38-PE, HLA-DR-FITC, CD4-PE-Cy7, and CCR5-PE-Cy5 (all from BD Biosciences), CD8-Qdot 605 (Life Technologies), and CD57 Alexa 647 (BioLegend). Cells were labeled with AARD for 20 minutes at room temperature followed by addition of surface markers. Staining was done at 37^{o}C for 15 minutes, washed and fixed in 0.5% formaldehyde (Polyscience), and held at 4^{o}C until analysis. A fluorescent-minus one control (FMO) was used for CD38 and HLA-DR to determine the cut-off for positive staining.

For Method #2, these mAbs were employed: CD3-Pacific Blue, CD27-APC-Alexa 750, CD28-APC, CD45RA-PE-Cy7 (BD Biosciences), CD8-Qdot 605 and CD4-PE-Texas Red (Life Technologies). The staining procedure was similar to the one described for Method #1, except that labeling was done at 4˚C for 30 min.

Cells that were prepared from either method were analyzed on a BD LSR II flow cytometer. For some regression analyses shown, the percentage of naïve CD8+ T cells

was calculated as the addition of the percentage of $CD8^+CD27^+CD28^+CD45RA^+CD57^+$ plus the percentage of $CD8^+$ $CD27^+CD28^+CD45RA^+CD57^-$ T cells.

C. Bisulfite conversion of genomic DNA.

DNA methylation status was assessed by bisulfite genomic sequencing (BGS) and also by pyrosequencing; both methods utilize bisulfite-modified DNA as template. For BGS, genomic DNA (isolated by QIAamp DNA blood minikit, QIAGEN) was bisulfite-modified and PCR amplified; individual clones were sequenced according to the protocol described by Clark *et al.* (8) with modifications. Sodium bisulfite concentration was increased to 5 M (pH of reaction: 5.0), and the antioxidant hydroquinone was used at a final concentration of 125 mM and time reduced to 5 h to increase kinetics of reaction and decrease template degradation; mutagenesis was carried out at 55˚C for 5 h with intermittent denaturation cycles at 95˚C for 1 min every hour. In some studies, bisulfite modification was performed using a commercially available kit (EZ DNA methylation Direct, Zymo Research), following the manufacturer's instructions. The amount of input DNA ranged between 200 ng and 2 µg. Methylation status of CpG sites in *CCR5* was assessed by Sanger sequencing and pyrosequencing, as described in **sections F-H** below.

D. *CCR5* numbering system and CpG nomenclature.

Our *CCR5* nomenclature, numbering system, and haplotype classification scheme were described previously (9). The numbering system designates the first nucleotide of the *CCR5* translation start site (i.e. the "adenine" base of the ATG start codon) as +1 and

the preceding nucleotide that is immediately upstream as –1. Upstream and downstream *CCR5* promoters are designated as Pr2 and Pr1, respectively (9). The first *CCR5* CpG dinucleotide that is upstream of the *CCR5* open reading frame is designated as CpG# –1. The position of the cytosine residue is at 46413998 in hg19 and is based on the *CCR5*-HHC haplotype. The CpG that is immediately upstream of CpG# –1 is designated as CpG# –2, and so on. The location of the CpG sites in the *CCR5* cis-regions along with their flanking sequences is shown in **Table S1.** CpG sites created by private polymorphisms were not numbered.

E. *CCR5* genotyping and haplotype assignment.

Genotyping of *CCR5* single nucleotide polymorphisms (SNP) and assignment of *CCR5* haplotypes was as described previously using Taqman® allele discrimination assays and PCR-RFLP technique (10, 11). The *CCR5* −4223C/T polymorphism (rs553615728) was originally identified by Picton et al. (12) with a minor allele frequency of 5.1% (genotype frequency of 11.4%) in South Africa Africans (SAA) and 0.00% in South Africa Caucasians (SAC) and overlaps with the cytosine residue of the *CCR5* −41 CpG site (**Figure 1C**). This SNP was genotyped in the following populations.

Cohort	Population	# studied	# *CCR5 −4223* data available	# *CCR5 HHA* data available	Data shown in Figure
HGDP-CEPH*	AF	127	127	127	Fig 7G,H
	M-E	57	57	n/a	Fig. 7G
	C-S-A	23	23	n/a	Fig. 7G
	E-A	80	80	n/a	Fig. 7G
	EU	48	48	n/a	Fig. 7G
	OC	19	19	n/a	Fig. 7G
Chimp	Chimp	80	80	n/a	Fig. 7G
HIV+	LTNP	11	11	n/a	Fig. 7G

	HIC	41	41	n/a	Fig. 7G
	Progressors	66	66	n/a	Fig. 7G
CAPRISA	HIV- (FSW)	120	120	117	Fig, 7G,H
	HIV+ (FSW)	26	26	26	Fig. 7G,H
	Acute HIV infection cohorts (not FSWs)	43	43	43	Fig. 7G,H

* Abbreviations used- HGDP-CEPH, Human Genome Diversity Project-CEPH populations; AF: Africa; M-E: Middle East; C-S-A: Central South Asia; E-A, East Asia; EU, Europeans; OC, Pacific Ocean; Chimp, Chimpanzees; HIC, HIV controllers; LTNP, long-term nonprogressors; n/a, not available. FSW, Female sex workers

F. Bisulfite genomic sequencing (BGS) of *CCR5 cis*-regions.

BGS of different *cis*-regulatory regions was performed on bisulfite-modified genomic DNA **(Section C)**. Primers and PCR conditions are listed in **Tables S3 and S4,** respectively. Of note, for certain PCR amplicons, a nested PCR approach was undertaken, whereas in other instances a semi-nested approach was used. PCR cycling conditions were as follows: initial stage of 80˚C for 4 min (addition of *Taq* polymerase under hot start conditions) followed by 94˚C for 2 min and 40 cycles of (94˚C for 10 sec, annealing temp. for 30 sec, 72˚C for extension time indicated) and a final step of 72˚C for 10 min. PCR mix (25 µl) consisted of PCR buffer, 3 mM $MgCl_2$, 280 µM of each dNTP, 0.45 µM for each primer, 2.5 U Taq polymerase (Invitrogen) and 2-4 µl of bisulfite modified DNA. Second and third rounds of PCR were performed using dilutions (1:100 to 1:10000) of 1st PCR product to reduce PCR bias. PCR products were gel-purified following the manufacturer's protocol (QIAquick Gel Extraction Kit, QIAGEN) and cloned into a pCR2.1 TOPO cloning vector (Invitrogen); recombinant plasmids were used to transform chemically competent *E. coli* (TOPO® One Shot® competent cells, Invitrogen). Sequencing of constructs was done using M13 Forward and M13 Reverse primers using a commercial service (Macrogen Inc., South Korea).

G. Amplification of *CCR5 cis*-regions for pyrosequencing.

To quantify the methylation status of CpGs that reside in the *CCR5*-upstream *cis*-region, an amplicon of 164 bp was designed to map the methylation status of the CpGs# –41 to –37, respectively. For *CCR5*-Pr2 methylation analysis, a 173 bp fragment encompassing four CpG sites (CpGs# –31 to –28) located in *CCR5*-Pr2 was amplified. For quantification of *CCR5*-intron2 methylation, a 296 bp amplicon containing 6 CpGs (CpGs# –6 to –1) in *CCR5*-intron 2 was designed. For simplicity in the figures shown in main text as well as here, DNA methylation content for CpGs# –6 and –1 in intron 2 was omitted although qualitatively the data were concordant with the contiguous CpGs for each fragment. Pyrosequencing assays in which a group of individuals were compared were conducted at the same time to mitigate batch effects.

The primer sequences for PCR amplification of the indicated regions are shown in **Table S3**. The PCR reaction mix (30 µl) consisted of PCR buffer, 3 mM $MgCl_2$, 200 µM of each dNTP, 6 pmol of forward primer, 6 pmol of reverse primer, 0.75 U *Platinum* Hot-start polymerase (Invitrogen), and 2-4 µL of bisulfite-treated DNA. PCRs were performed under “hot-start” conditions to reduce non-specific amplification. For amplification of *CCR5*-Pr2 (CpGs# –31 to –28), the PCR cycling conditions were: initial stage of 95°C for 15 minutes; followed by 45 cycles of 95°C for 30 sec, 55°C for 30 sec, and 72°C for 30 sec; and a final stage of 72°C for 5 min. For amplification of other regions, similar conditions were used except the annealing temperature was 60°C. Successful PCR amplification was confirmed by gel electrophoresis. Five µl of PCR

reaction product was mixed with 1 µl of xylene cyanol sample loading buffer and loaded onto a 2.5% w/v agarose gel.

H. Pyrosequencing of *CCR5 and IFNG cis*-regions.

Pyrosequencing was employed as a strategy to quantify methylation levels in the genomic DNA and was performed by a commercial firm (Epigendx, Worcester, MA). PCR products were pyrosequenced with the PSQ 96HS system (Pyrosequencing, Qiagen), according to established methods, using the sequencing primers shown in **Table S3** ("FS" primers). Pyrosequencing is a sequencing method based on real-time monitoring of DNA synthesis and has been successfully used to assess DNA methylation levels (13, 14). A very well-studied artifact of bisulfite sequencing method is the PCR bias that results in a disproportional amplification of one allele compared to the other allele (15). To test for absence of this bias, low methylated DNA control and *in vitro* methylated DNA (80-90%) were mixed at different ratios followed by bisulfite modification, PCR, and pyrosequencing analysis; highly corroborative results were obtained (**Fig. S11**). To cross-validate data obtained by pyrosequencing and BGS, percent methylation was plotted for each CpG by both methods; these results showed a high degree of correlation (data shown for *CCR5*-upstream in **Table S5**).

We assessed the percentage of *INFG* DNA methylation at six CpG sites at positions −425, −315, −183, −8, −2 and +42 (relative to the first nucleotide of *IFNG* ATG translational start site) by pyrosequencing using a commercial service (EpigenDx, MA).

The average methylation content of 6 CpG sites from three independent donors was included in the analyses.

I. Statistical analyses.

Two-tailed paired *t*-Student tests were applied to evaluate differences in methylation content of pre- vs. post-ART samples (New York University cohort). Where appropriate, two-tailed parametric (non-paired *t*-test), non-parametric (Mann-Whitney) tests and Chi-square tests were also used to assess differences in methylation or CCR5 surface expression. Pearson's or Spearman's correlation coefficients were computed. Multiple linear regression analyses were conducted and adjusted for covariates. A Cochran-Armitage test for trend was used to evaluate whether the likelihood of possessing a specific *CCR5* haplotype (HHE) or genotype (HHC/HHC) decreased or increased with increasing methylation status of *CCR5 cis*-regions. We defined higher vs. lower methylation content as a methylation content above vs. below the median methylation content of all *Wt/Wt CCR5* haplotypes. Explained variability was derived from R^2 values calculated from bivariate correlation analyses. All statistical analyses were conducted using Stata, version 10 (StataCorp).

J. Analysis of publicly available whole genome sequence data.

Co-associations between different transcription factors and co-factors in *CCR5* loci were determined using publicly available data, ChIP-Seq datasets from GM12878 lymphoblastoid cells (abbreviated as GM) were obtained from Gene Expression Omnibus (GEO) series GSE30263 (16), GSE32465 (17), and GSE31477 and were

visualized in the Human Epigenome Browser (18). The Y-axis indicates the score that measures quantitative data obtained for the corresponding track. DNase I HS data are from GEO series GSE18927 and GSE29692 (19, 20). The MeDIP-Seq data used for generating the wiggle plots depicted in **Fig. S2B** were retrieved from GEO series GSE16368 (21). The nucleotide coordinates used for generating the plots shown in **Figure 1B** are chr3:46408899-46415450 (hg19).

K. Multi-species alignment of *CCR5* loci.

CCR5 cis-regulatory regions of various primate and non-primate species, including human, chimpanzee, orangutan, sooty mangabey, sabaeus, rhesus macaque, marmoset, dog, cow, and mouse were downloaded from NCBI were aligned using the ClustalW algorithm and the Malign program of the Vista suite (22).

L. Bioinformatic analyses of transcription factor binding sites and electrophoretic mobility shift assays.

To predict the transcription factors (TFs) that differentially bind haplotype-specific polymorphisms that alter the CpG sites, we used JASPAR webserver with default parameters (23). Representative TFs with significant JASPAR score differences that have been filtered according to their cell- or tissue-specific expression (biogps.org) and putative function are shown in **Fig. S7**. We used the schema proposed by Medvedeva et al., to assign the activator/repressor function to the TFs (24).

Antibody supershift assays were performed using oligonucleotides spanning -41 CpG site and *in vitro* transcribed and translated (IVTT) products generated from a CREB1 expression vector using previously described protocols (25). Anti-CREB1 and anti-MyoD antibodies were obtained from Santa Cruz Biotechnology, Dallas, TX. Sequences of the oligomers used for the EMSAs are shown in **Table S3.**

M. qRT-PCRs and allelic expression imbalance (AEI) studies.

qRT-PCRs for *CCR5* exon 1 (full-length transcripts) and total transcripts and *CD4* mRNA were performed using Taqman® gene expression assays according to previously published methods (25). We used a pyrosequencing-based assay to assess AEI in CCR5 transcripts. For this experiment, we used purified T-cell subsets from healthy donors who were heterozygous for HHA and HHE and determined the HHA- and HHE-specific expression. T-cell receptor activation protocol was as described previously (25). The validity of the pyrosequencing assay for the determination of −2135C/T allelic expression levels was confirmed by sequencing PCR products obtained from plasmid DNAs containing −2135T and −2135C combined in specific ratios; there was a linear relationship between the expected and measured signals **(Fig. S6C)**. The sequences of the PCR amplification primers and the sequencing primer are shown in **Table S3**. The conditions for the amplification were initial denaturation at 94°C for 2 min, 35 cycles at 94°C for 10 sec, 58°C for 30 sec, and 72°C for 1 min, followed by an extension time of 72°C for 7 min.

N. Th1 polarization experiments.

Th1 polarization was carried out using a previously described protocol (26). In brief, media was supplemented with rIL-12 (2.5ng/mL; Peprotech), anti-Human IL-4 Ab (5ug/mL; BD Biosciences), and anti-Human IL-10 (5ug/mL; BD Biosciences). For Th1 polarization, purified CD4+ CD45RA+ T cells were resuspended at $1x10^6$ cells/mL and were activated for 3 days with immobilized anti-CD3 (BD Biosciences) and anti-CD28 (BD Biosciences) antibodies in the presence or absence of Th1-specific polarizing conditions. The cells were removed from stimulation and cultured for an additional 4 days with or without polarizing conditions. Cells were cultured in RPMI 1640 (Invitrogen) supplemented with glutamax, penicillin-streptomycin, 10% FBS (Life Sciences) and recombinant IL-2 (50U/mL; BD Biosciences). Cells were collected at the time points indicated for isolation of DNA and RNA.

O. 5-azadC treatment of Jurkat T-cell line and primary T cells.

Jurkat T cell line (obtained from the American Type Culture Collection) was treated with 0.02, 0.2, 1 and 2 μM of 5-azadC for 72 h; replenishment of this chemical was done daily. Quantification of (i) CCR5 exon-1-containing isoforms, and (ii) total CCR5 mRNA isoforms and (iii) CD4 transcripts was assessed by qPCR using predesigned or custom Taqman® gene expression assays as previously described (25). CCR5 protein expression was visualized by confocal microscopy (LSM510 confocal microscope equipped with a 488 laser, using mAb mouse antihuman CCR5 (1:20 dilution, mouse anti-CCR5 IgG2a,κ, BD Pharmingen), followed by AlexaFluor488 staining with a conjugated rabbit antimouse IgG secondary Ab (1:100 dilution, Invitrogen). Images were generated in the Optical Imaging Core facility, UTHSCSA. For reproducibility and

comparison purposes, all microscope settings were kept identical between different groups. Actin staining was used as control to normalize fluorescence signals. Densitometric analysis (of relative fluorescent intensities) is shown in **Fig. 5J** was performed with the NIH Image J software.

To assess a synergistic effect of activation-induced demethylation and chemically-induced demethylation on CCR5 surface expression, PBMCs were activated as described above but stimulation was performed in the presence of 1μM 5-azadC. This chemical was added at the beginning of the anti-CD3/CD28 activation procedure and after 72h of T-cell receptor (TCR) stimulation, when cells were washed and expanded in the absence of antibodies. T cell cultures with no 5-azadC added were included as controls.

P. Quantification of CCR5 and IFNγ mRNA levels by RNA-Seq

Manufacturer's recommended protocols were used except where indicated. Total RNA from samples was extracted using TRIzol® Reagent with the PureLink® RNA Mini Kit cells (Life Technologies, Grand Island, NY). All the samples were assessed using Bioanalyzer 2100 (Agilent) and had RNA integrity number (RIN) > 9 and the average quantity of RNA was >500 ng. RNA-seq library preparation was performed using TruSeq RNA sample preparation kit v2 (Illumina) per manufacturer's protocol. The libraries were quantified using qPCR and 8 pM from each sample were pooled (12 samples per lane of a flow cell) for cluster generation using TrueSeq PE Cluster kit v3-cBot-HS (Illumina, San Diego, CA). The cluster generation was followed by 2 x 101 bp

paired-end sequencing on an Illumina HiSeq™ 2000 machine using TruSeq SBS Kit v3-HS. Sequencing reads were obtained and quality filtered using the standard Illumina pipeline. Decoding of the indexes was performed to assign sequences from 12 multiplexed samples per lane to each of the sample files. The average number of paired-end (PE) reads per sample was ~12M.

Sequence reads were aligned to the UCSC hg19 build of the *Homo sapiens sapiens* reference genome using TopHat v1.4.1 (27) with default parameters except with UCSC hg19 known gene reference annotation file. HTSeq framework, v0.5.3p3 [http://www-huber.embl.de/users/anders/HTSeq/] was used to get the gene counts of the aligned reads mapped to each gene in the reference genome annotation file provided by Illumina iGenomes project for UCSC hg19 build of the *Homo sapiens* genome. HTSeq-count was run with no for stranded option and intersection-nonempty as overlap resolution mode. The gene count data was generated for each sample for 23239 unique, well curated gene features in the reference annotation file. This gene count matrix was normalized using DESeq package in R (28). Two-step quality control was performed on the matrix before normalization. The only data used from these analyses were those for CCR5 and IFNγ mRNA gene counts.

Supplementary Figures

Fig. S1. Epigenetic models that may underpin the four characteristic features of CCR5 surface expression on T cells. (A) *Feature 1: inter-T cell type differences in CCR5 expression.* High vs. low methylation status of *CCR5 cis*-regions serve as a determinant of low vs. high CCR5 expression, respectively on CD45RO- (CD45RA+) vs. CD45RO+ (CD45RA-) T cells or effector vs. central memory T cells (TEM vs. TCM). The lollipops with diamond, circle and square heads depict methylation status of three *cis*-regulatory regions, namely *CCR5*-Pr1, *CCR5*-Pr2 and *CCR5*-intron 2, respectively. Black, methylated; white, unmethylated. **(B)** *Feature 2: inter-individual differences in CCR5 expression.* $CCR5^{high}$ vs. $CCR5^{low}$ expressers have low vs. high methylation, respectively, in the same *CCR5 cis*-regions that track expression of CCR5 on CD45RA+ (CD45RO−) vs. CD45RA− (CD45RO+) T cell subsets shown in model 1. **(C)** *Feature 3: activation-associated upregulation of CCR5 expression.* T-cell activation (TCA) induces demethylation of *CCR5 cis*-regions, resulting in upregulation of CCR5 on T cells. However, *CCR5 cis*-regions may exhibit enhanced susceptibility vs. resistance to undergoing demethylation upon TCA, epigenetic traits that may result in high ($CCR5^{high}$) vs. low ($CCR5^{low}$) CCR5 expression in the face of low (TCA^{low}) and high (TCA^{high}) TCA levels, respectively. **(D)** *Feature 4: CCR5 promoter haplotype-dependent differences in surface CCR5 levels. CCR5* haplotypes contain polymorphisms that by creating or destroying CpG sites may result in *cis*-regions that are more susceptible (e.g. *CCR5*-HHE) vs. resistant (e.g. *CCR5*-HHA) to undergo activation-induced demethylation, despite comparable levels of T-cell activation. Thus, *CCR5* haplotypes that associate

with increased vs. decreased HIV-AIDS susceptibility may manifest increased vs. decreased susceptibility, respectively, to undergo TCA-associated demethylation.

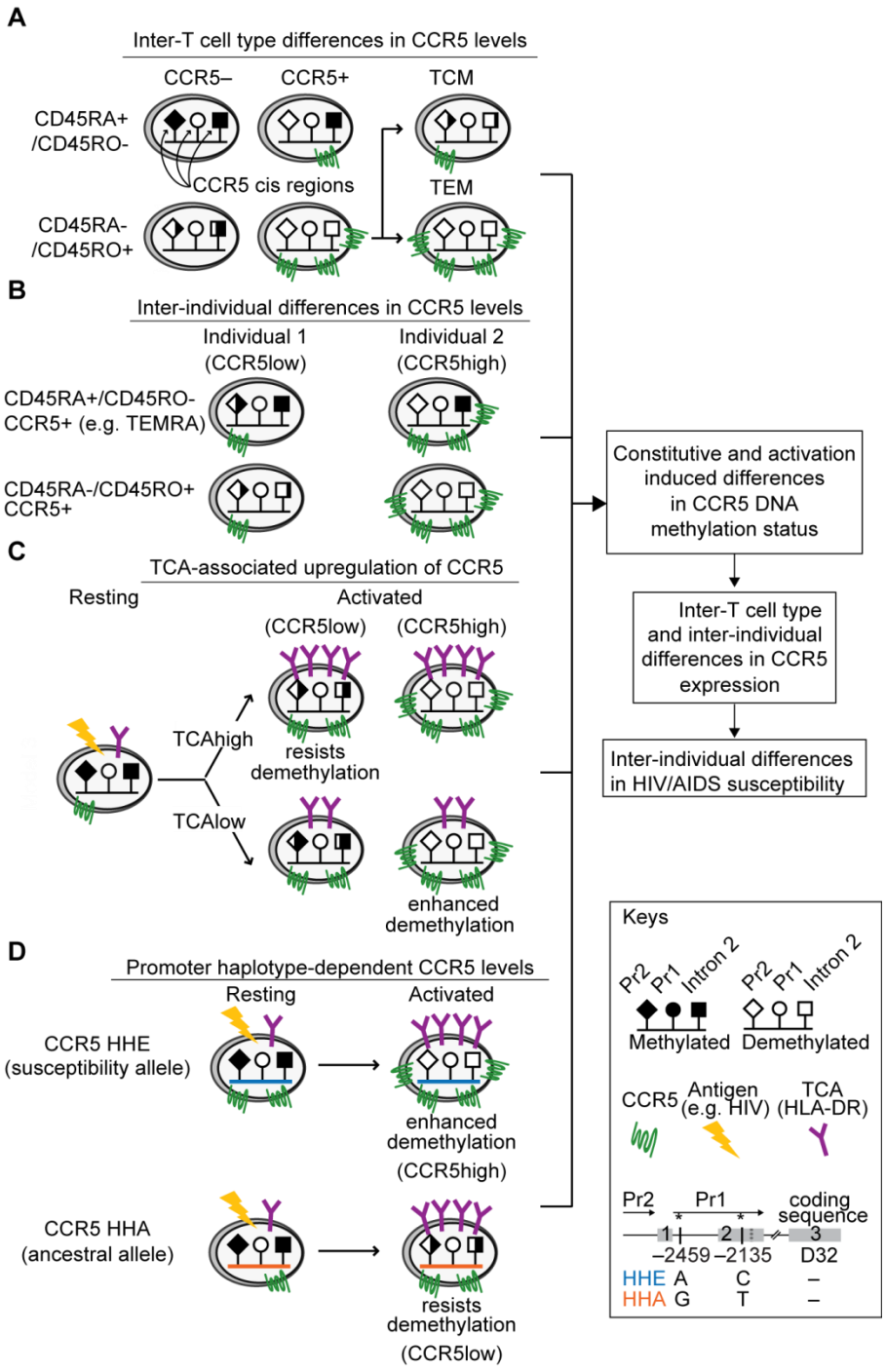


Fig. S2. DNA methylation patterns in *CCR5 cis*-regions in Jurkat T cells obtained by bisulfite genomic sequencing (BGS) and in primary T cells obtained by MeDIP-Seq. (A) Top left, Schema of *CCR5* gene showing the three exon structure, dual promoters, and location of the CpG dinucleotides (short and long ticks) upstream of the ORF. CpG numbering is shown in detail in **Table S1**. **Bottom left**, Jurkat T cells are phenotypically comparable to naïve T cells (29) and have a CCR5 mRNA/protein profile similar to that of naïve T cells (25, 30). Methylation content of the indicated regions was determined by BGS. These data are in agreement with those from Weirda et al (31). Other details are as in main **Figure 2**. *Asterisks* indicate the haplotype-specific CpG sites **(Table S1)**. *Lack of circles*, No data available from the indicated CpG site in the sequenced clone. **Right panel:** *CCR5*-Pr2 is demethylated following 5-azadC treatment of Jurkat cells. Jurkat T cells were treated with the indicated concentration of 5-azadC and *CCR5*-Pr2 methylation levels were determined by BGS. Untreated Jurkat T cells contained ~92% methylated CpGs in *CCR5*-Pr2 whereas treated Jurkat T cells contained ~44% methylated CpGs in *CCR5*-Pr2 suggesting that 5-azadC has strong demethylating effect in this region. **(B) Top**, gene and methylation schema. **Bottom,** wiggle plots represent Me-DIP sequence data derived from PBMC (orange), CD4+ CD45RA+ naïve T cells (green), and CD4+ CD45RO+ memory (mem) T cells (red), obtained from three independent donors (TC007, TC003, and TC009). Note that different cell types from a single donor show a similar trend in methylation levels. For example, PBMC and T cell subsets from donor TC009 show the highest methylation levels relative to other donors. CTCF peaks from lymphoblastoid cell line GM12878 (GM) are shown as reference on the top of the panel (blue) and have blue-colored

vertical dashed lines drawn through them. The Y-axis shows the number of reads mapped. **(C)** Pairwise comparison of CD4+CD45RA+ T cells Me-DIP track to CD4+CD45RO+ T cells Me-DIP track from donor TC009. The values shown are the log value of CD45RA+ divided by CD45RO+ (log_2(track1/track2)). This comparison indicated that methylation content of the *CCR5*-upstream, the *CCR5*-Pr2 and *CCR5*-intron 2 is greater in CD45RA+ T cells than in the CD45RO+ T cells. Notably, there were no differences in methylation levels in the open reading frame (*P* =.718) between the two cell types. The data shown were retrieved from GEO series GSE16368 (21) and visualized on the Washington University Epigenome Browser.

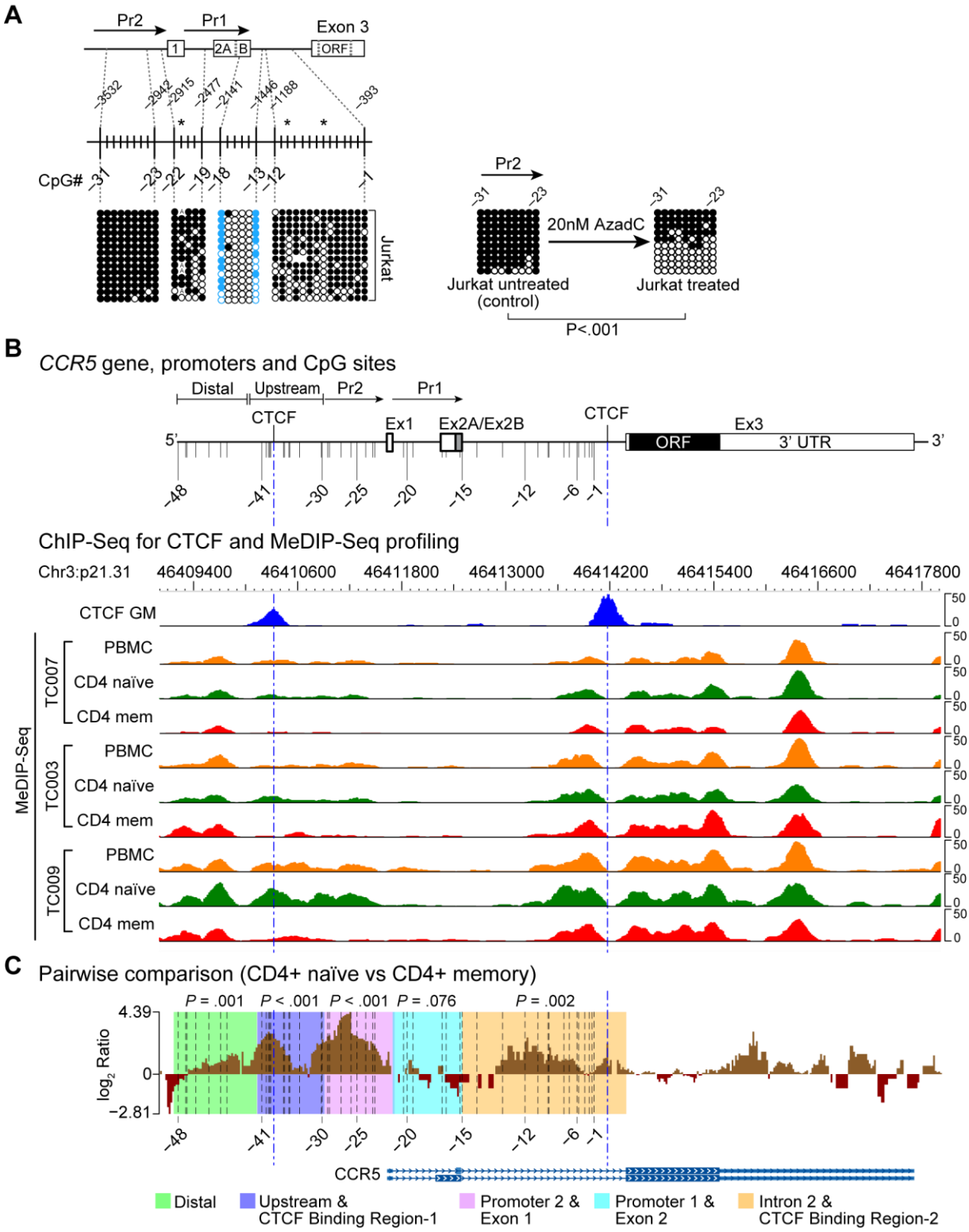

A
Pr2
Pr1
Exon 3
1
2A
B
ORF
-3532
-2942
-2915
-2477
-2141
-1446
-1188
-393
CpG#
-31
-23
-22
-19
-18
-13
-12
-7
Jurkat
Pr2
-31
-23
20nM AzadC
-31
-23
Jurkat untreated (control)
Jurkat treated
P<.001
B
CCR5 gene, promoters and CpG sites
Distal
Upstream
Pr2
Pr1
CTCF
Ex1
Ex2A/Ex2B
CTCF
Ex3
5'
ORF
3' UTR
3'
-48
-41
-30
-25
-20
-15
-12
-6
-1
ChIP-Seq for CTCF and MeDIP-Seq profiling
Chr3:p21.31
46409400
46410600
46411800
46413000
46414200
46415400
46416600
46417800
CTCF GM
MeDIP-Seq
TC007
PBMC
CD4 naïve
CD4 mem
TC003
PBMC
CD4 naïve
CD4 mem
TC009
PBMC
CD4 naïve
CD4 mem
50
0
C
Pairwise comparison (CD4+ naïve vs CD4+ memory)
P = .001
P < .001
P < .001
P = .076
P = .002
4.39
log2 Ratio
0
-2.81
-48
-41
-30
-25
-20
-15
-12
-6
-1
CCR5
Distal
Upstream & CTCF Binding Region-1
Promoter 2 & Exon 1
Promoter 1 & Exon 2
Intron 2 & CTCF Binding Region-2

Fig. S3. Relationship between CCR5 DNA methylation status, transcripts and surface expression levels. (A) Differences in methylation levels of C*CR5 cis*-regions in sorted HLA-DR+ CCR5+CD45RO+ CD4+T cells from donors who constitutively expressed higher vs. lower CCR5 levels. *P* values derived by Student's *t*-test compare the average methylation levels of representative CpGs, as assessed by pyrosequencing, in each of the indicated *cis*-regions. **(B-C)** Representative methylation profiles (assessed by pyrosequencing) in the three indicated *CCR5 cis*-regions in CCR5+ versus CCR5− CD45RA+ (panel A) and CD45RO+ (panel B) CD4+ T cells, derived from donors who constitutively expressed higher versus lower surface CCR5. CCR5+CD45RA+ T cells are indicative of TEMRA T cells and are known to express higher levels of CCR5 (32). Lines connecting the datapoints are provided for better visualization of data. **(D)** Differences in methylation content (assessed by pyrosequencing) in individuals who constitutively expressed higher (n=4) *versus* lower (n=4) CCR5 levels on sorted CD4+ T cells (higher vs. lower, CCR5 levels were %CCR5+ in CD4+ T cells: 15.5±6.66 *versu*s 5.8±1.84, respectively; *P* = .031). Data are shown for CCR5− (left) and CCR5+ (right) CD4+ CD45RO+ T cells. **(E)** CCR5 surface expression in sorted CD45RA+CD45RO− CD3+ T cells from two representative donors with differing CCR5 surface expression. **(F)** Pyrosequencing data depicting percent CpG methylation of *CCR5*-Pr2 (*left*, average of CpG sites −31 to −28), and of intron 2 (*right*, average of CpG sites −5 to −2). Consistent with the CCR5 expression profile, methylation levels were lower in the donor g relative to donor f. **(G)** qRT-PCR was used to quantify levels of CCR5 exon 1-containing transcripts as well as CCR5 total transcripts as reported previously (25). Data represents fold difference of the full-length

exon-1-containing CCR5 transcripts or total CCR5 transcripts with donor f used as the reference. Donor g expressed high surface CCR5 and accordingly also expressed higher amounts of CCR5 mRNA isoforms. Data is representative of one of two experiments. *P* values are indicated for the percent methylation levels between the donors with high and low CCR5 expression. The error bars shown in panels A and E indicate SEM.

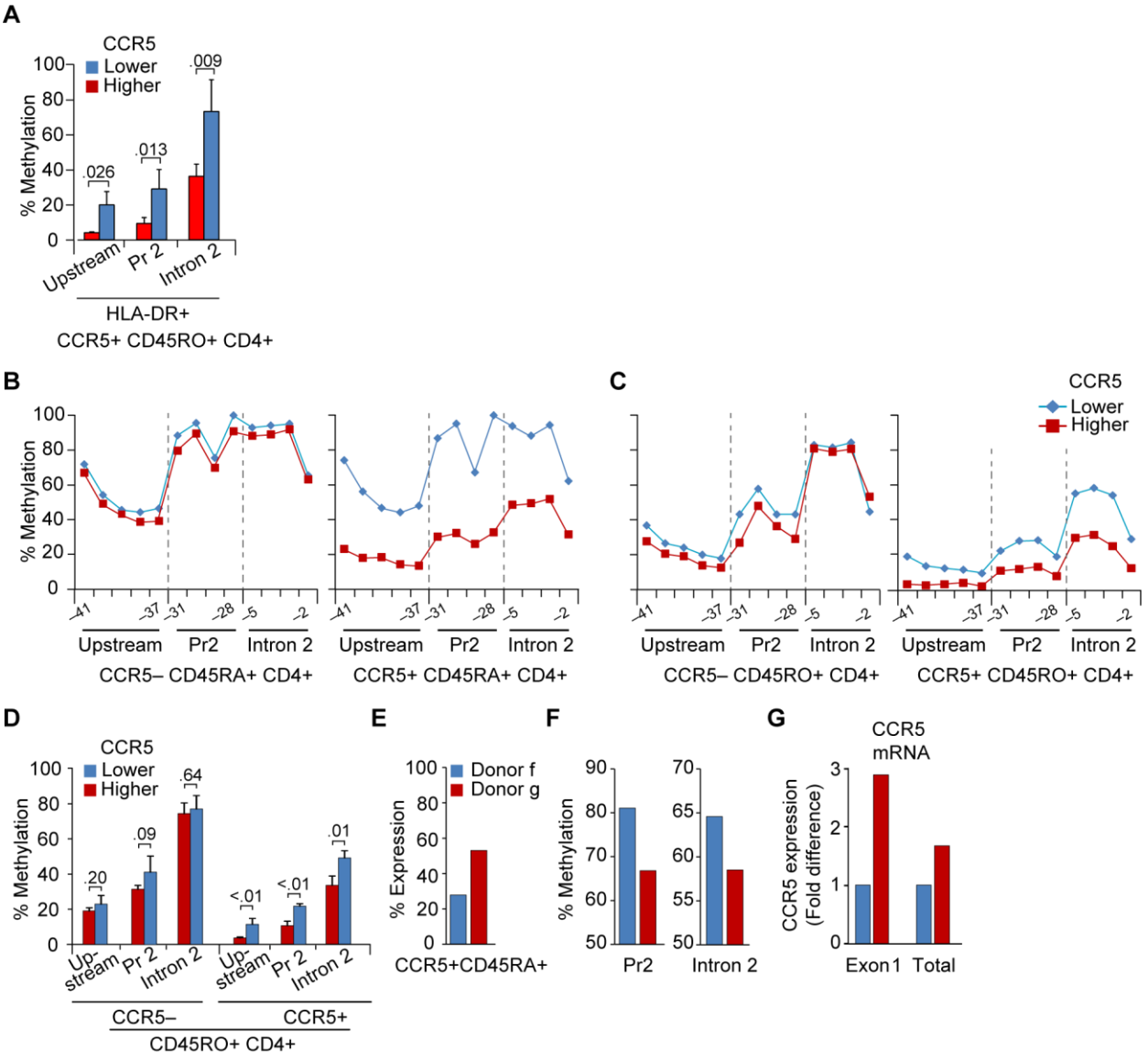


Fig. S4. 5-azadC induces CCR5 protein expression in Jurkat T cells. Cells were treated with the indicated concentration of 5-azadC and CCR5 expression was assessed using confocal microscopy. A representative experiment of Jurkat T cells treated with 0.02 µM 5-azadC is shown.

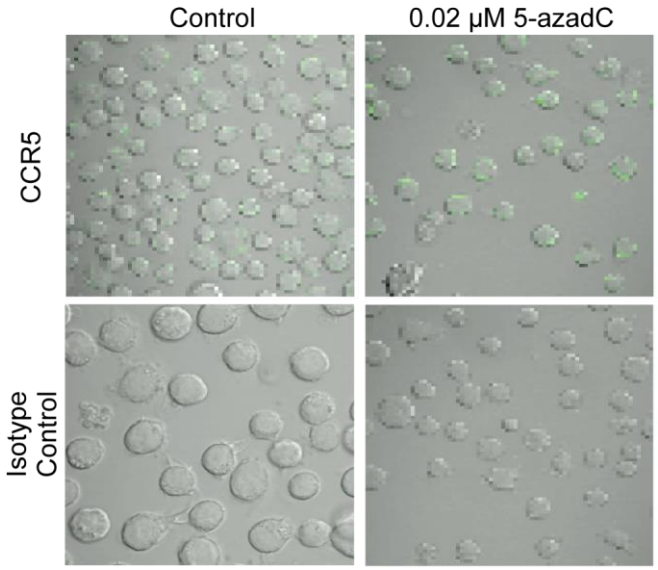


Fig. S5. Relationship between T-cell activation and expression of CCR5 on CD4+ and CD8+ T cells. (A) Activation levels (assessed as %CD38+HLA-DR+ T cells) and **(B)** CCR5 expression (measured as MFI and %CCR5+ expressing T cells) in CD4+ and CD8+ T cells were determined by flow cytometry in 85 HIV-infected persons treated with virally suppressive ART. Patients had stably suppressed viral loads for greater than 6 months and were from the SCOPE cohort. Significance values by Student's *t*-test. Error bars indicate SEM. **(C-D)** The plots depict the positive correlations (Spearman's correlation coefficients) between T-cell activation (%CD38+HLA-DR+) in CD4+ (C) or CD8+ (D) T cell subsets with CCR5 expression. *P* and rho correlation values are shown. Data shown in this figure were derived from 85 HIV infected patients on virally suppressive ART from SCOPE cohort.

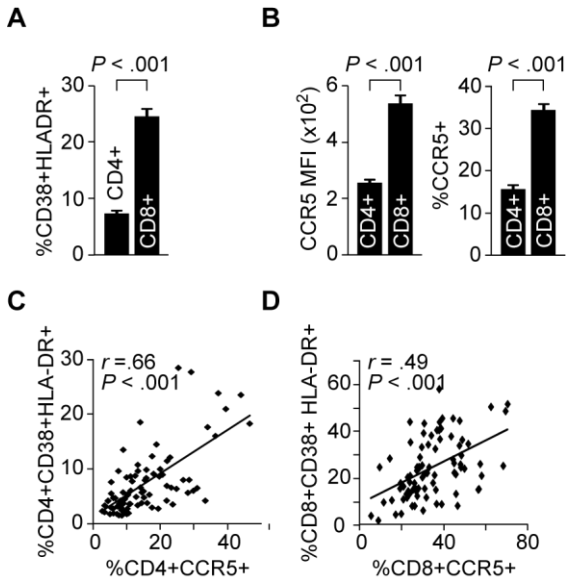


Fig. S6. Allele specific methylation and allelic expression imbalance (AEI) patterns of *CCR5* HHA and HHE alleles. (A) *CCR5* haplotype-specific methylation patterns in therapy-naive HIV-infected women from the Ukraine**.** Average methylation content of *CCR5*-Pr2 (CpGs# −31 to −28) in 81 HIV+ women of European descent categorized into those bearing HHA-containing or lacking genotypes (left) or those bearing HHE-containing or lacking genotypes (right). Error bars indicate SEM. **(B)** Allelic expression imbalance (AEI) of CCR5 mRNA**.** Basal (constitutive) and activation-induced allelic imbalance in CCR5 mRNA expression in two healthy donors bearing the HHA/HHE genotype (denoted as donors *h* and *i)*. Pyrosequencing was used to determine allele-specific differences in expression to assess the contribution of HHA- and HHE-specific mRNA to the total pool of CCR5 transcripts, at baseline (0 h) in naïve T cells and after 120 h of the TCR stimulation. The −2135T/C (rs1799988) *CCR5* SNP in exon 2 (9) was used to distinguish HHA- vs. HHE-specific mRNA. Prior to activation, donor *h* had comparable levels of HHA- and HHE-specific mRNA levels, whereas donor *i* had higher levels of HHA-specific compared with HHE-specific mRNA levels. However, after activation, a greater amount of HHE-specific mRNA was expressed compared with HHA-specific mRNA. These data indicated that the *CCR5*-HHE haplotype is more sensitive to activation-induced demethylation than the *CCR5*-HHA haplotype, and that this differential sensitivity to demethylation was associated with increased vs. decreased haplotype-specific gene expression. **(C)** Validation of the pyrosequencing assay used for AEI determination. Plasmid DNAs bearing the −2135C or −2135T SNP were mixed at specific ratios and amplified by PCR and the products were subjected to pyrosequencing. The chart shows a linear regression line, which indicates a linear

relationship between measured and expected levels of the −2135T allele in a plasmid mixture, suggesting no PCR bias in amplification of the alleles and validating pyrosequencing as a reliable method for quantifying −2135 T/C alleles.

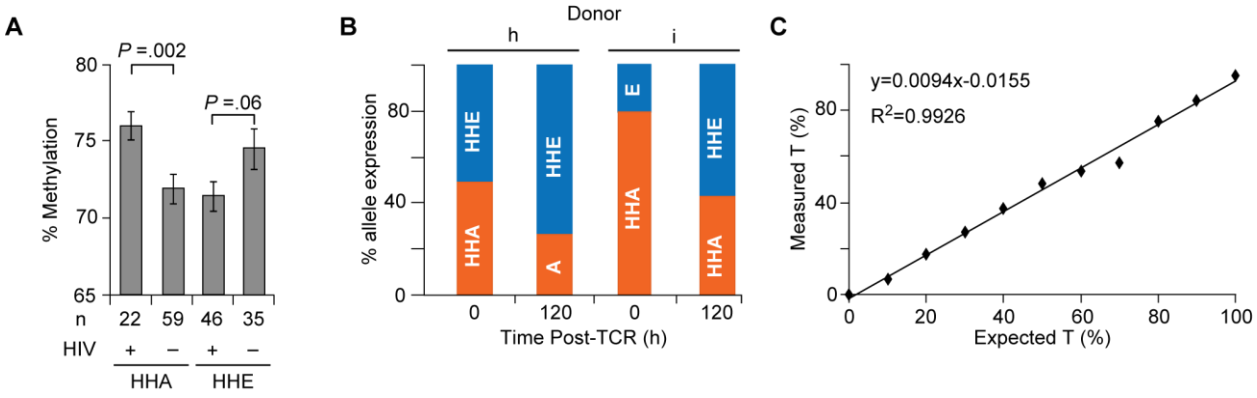


Fig. S7. Haplotype-specific SNPs at CpG dinucleotide sites and their predicted functional effects. Top, Gene map showing the 3-exon structure of *CCR5*, *CCR5* promoters (horizontal arrows) and open reading frame (ORF) residing in exon 3. Numbering of *CCR5* gene is according to Mummidi et al. (9). CpG numbering is according to **Table S1**. CpGs that were most 5' and 3' for each of the PCR fragments that were subjeted to BGS are indicated in *green*. **Middle**, CpG dinucleotides specific to or shared among the *CCR5* haplotypes (HHA to HHG) that generate distinct *CCR5* epihaplotypes. *Red* color denotes CpG sites; *,CpG site polymorphisms relative to the HHC haplotype, shaded in gray. Note that only CpG# –33 was HHC-specific; a number followed by a letter "b" or "c" was assigned to the rest of the CpGs present in other *CCR5* haplotypes but not in HHC. HHA*1, shaded in *blue*, is predominantly present in persons of African origin whereas HHA*2, shaded in *yellow*, that bears the −4223T SNP that disrupts the −41 CpG site is restricted to persons in Southern Africa. Note that only a subset of individuals with the indicated haplotypes carried the corresponding CpG sites **(Table S1).** For example, not all individuals bearing the HHA haplotype will contain the −13b CpG site, whereas the −13b CpG site is present exclusively in HHA-bearing individuals. ND, not determined. **Bottom**, Predicted functional effects of the haplotype-specific SNPs at the indicated CpG sites. We used position weight matrix method as implemented in JASPAR algorithm (23) to assess the role of the haplotype-specific SNPs in altering TF binding specificity at the CpG sites. Candidate TFs that are predicted to differentially bind the indicated CpG sites are shown. *Upward arrows*, gain of TF binding or increased affinity of binding as indicated by JASPAR score, relative to the corresponding CpG site. *Downward arrows*, loss of TF binding or decreased affinity

of binding as indicated by JASPAR score, relative to the corresponding CpG site. Both CREB1 and ELK-1 (indicated in bold) binding sites contain a CpG site in their core consensus motifs that are TGACGTCA and CCGGAARY, respectively (23). The plus (+) or the minus (—) symbols indicate whether the TF is an activator or a repressor of transcription according to Medvedeva et al. (24). TFs that exhibit both activator and repressor function are indicated as (+,—). Although no function is assigned to POU2F2 and NFATC2/NFAT1, they are known to be involved in *CCR5* transcriptional regulation (25) and T-cell activation pathways (33), respectively.

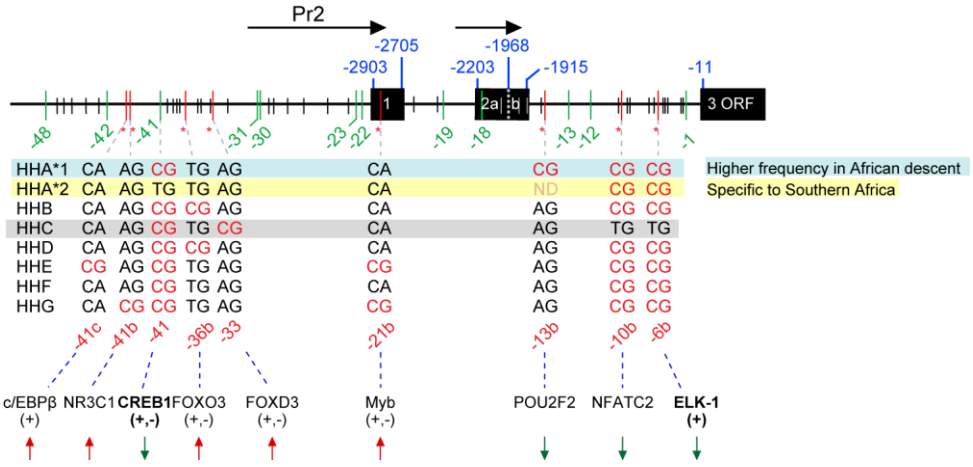


Fig. S8. **Polymorphism at the CpG -41 disrupts a CREB1 binding site in the *CCR5*-upstream region**. **(A)** Logo indicating the CREB1 nucleotide binding profile from JASPAR database (23). **(B)** Alignment of the region bearing CpG −41 polymorphism (rs553615728) with the CREB1 consensus site (Cons). The CpG site is shown in red. The *CCR5* −4223T that is present in HHA*2 is indicated in *green*. **(C)** EMSA demonstrating differential binding of CREB1 to −41 CpG and −41 TpG bearing oligomers. Radiolabeled oligomers bearing either −41 CpG or −41 TpG polymorphisms were incubated without (lanes 1, 5) or with *in vitro* transcribed and translated (IVTT) CREB1 (lanes 2-4 & lanes 6-7). The identity and the specificity of the bound complexes were established by addition of anti-CREB1 antibody (lane 3) that showed slower migration ("super-shift"). Anti-Myo-D antibody is used as a negative control (lane 4). #, free probe; *triangles*, IVTT CREB1-probe complexes; *, supershift; *closed circles*, non-specific bands). **(D)** Cell type specific *in vivo* enrichment of CREB1 in the region overlapping with the −41CpG site. Wiggle plots represent CREB1 ChIP-Seq data derived from K562 (pink) and GM12878 (orange) cell lines. Also shown are the CTCF ChIP-Seq peaks from GM12878. Additional details are as in **Fig. 1**.

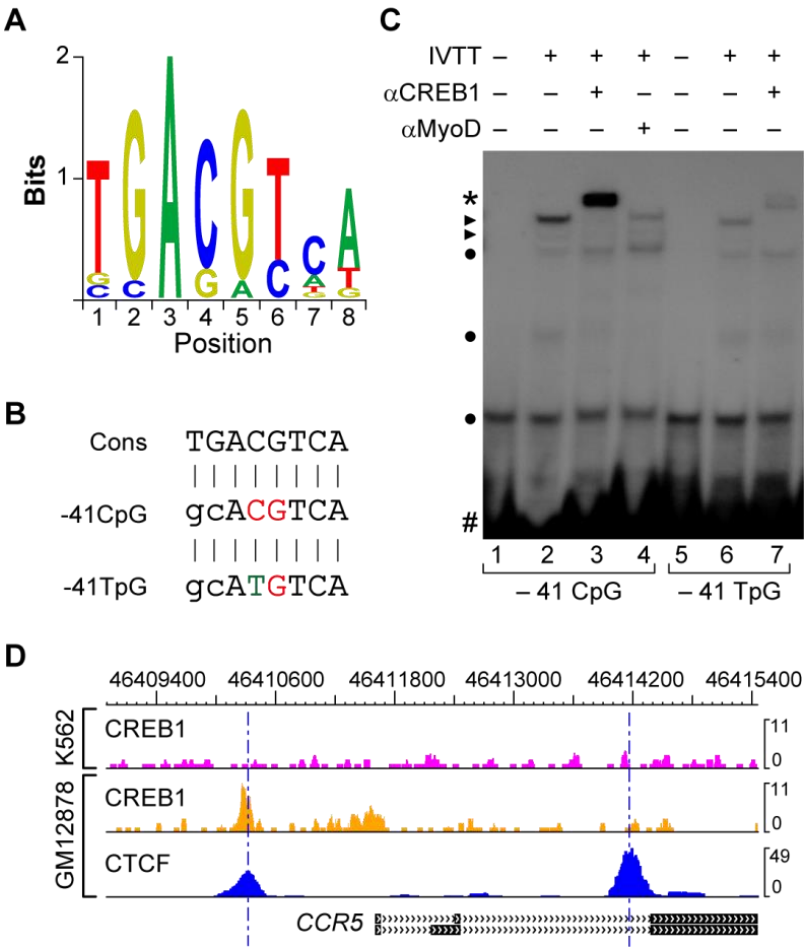
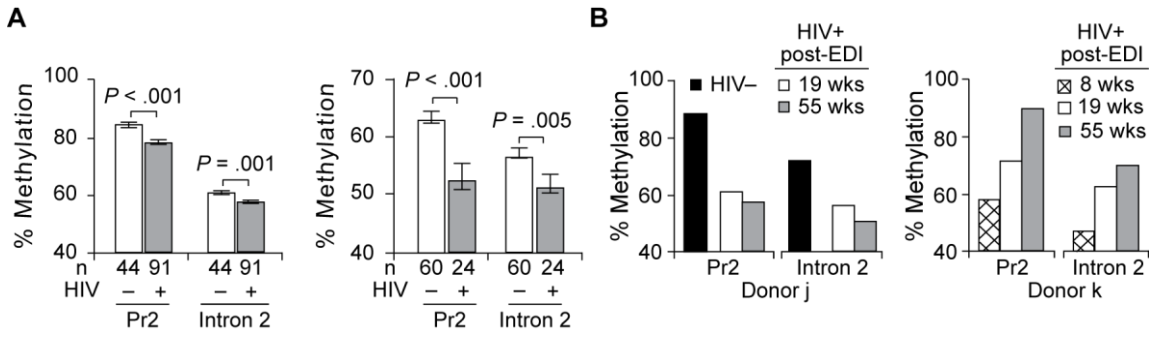

A
Bits
Position
B
Cons TGACGTCA
-41CpG gcACGTCA
-41TpG gcATGTCA
C
IVTT
αCREB1
αMyoD
– 41 CpG
– 41 TpG
D
46409400
46410600
46411800
46413000
46414200
46415400
K562
CREB1
GM12878
CREB1
CTCF
CCR5

Fig. S9. Association of DNA methylation status in *CCR5 cis*-regions according to HIV-1 serostatus, and pre- vs. post-infection. (A) Association of methylation status in *CCR5*-Pr2 (average of methylation levels at CpG# −31 to −28), *CCR5*-intron 2 (average of CpG# −5 to −2) with HIV-1 serostatus in Ukrainian children exposed perinatally to HIV-infected mothers who did not receive antiretroviral prophylaxis or antiretroviral therapy (*left panel*) and health care workers from Durban, South Africa (*right panel*). Numbers of HIV-uninfected and –infected subjects are shown below the bars. Error bars indicate SEM. **(B)** Change in *CCR5* methylation status during early HIV infection**.** *Left panel***,** percent methylation of *CCR5*-Pr2 and *CCR5*-intron2 from a study participant ~16 weeks before estimated date of infection (EDI) (in black bar labeled as HIV−) and 9 and 55 weeks post-EDI (donor j). *Right panel*, same as the previous panel, but in another HIV-infected patient (donor k) at three different time points post-seroconversion: 8, 19 and 55 wks post-EDI; pre-seroconversion sample was unavailable in this donor. Methylation data were obtained at 2 similar time points in donors j and k (i.e., 19 and 55 weeks post-EDI). Data shown in the B panel were derived from CAPRISA cohort.



S10. Hypothetical model by which the extent and distribution of *CCR5* methylation content may underpin inter-T cell type differences in CCR5 mRNA isoforms and surface expression. (A) CCR5− CD45RA+ T cells. The presence of a CTCF/cohesin binding site and a stalled RNA polymerase II imparts an insulator function to a *cis*-regulatory region downstream of the evolutionarily conserved CpG site at −41 and blocks interactions (red **X**) between the distal region ("enhancer") of *CCR5* from the downstream *cis*-regions by formation of a cohesin/CTCF-mediated "DNA loop" between the *CCR5*-upstream and *CCR5*-intron 2. This "DNA loop" and methylation of *CCR5*-Pr2 and -intron 2 represses the transcriptional activity required for production of exon 1-containing transcripts (e.g., preventing binding of Pol II to Pr2; red X) whereas the core of *CCR5*-Pr1 is constitutively demethylated (flanked by blue circles) and may underpin the constitutive expression of exon 1-lacking CCR5 mRNA isoform in this cell type. **(B)** CCR5-negative CD45RO+ T cells. Activation of CD45RA+ T cells leads to induction of the transcription factors (TFs) such as CREB1 (34) and Oct-2 (25) that may promote demethylation of *CCR5*-Pr2 and recruitment of additional transcription cofactors and allow the Pol II to escape from the stalled state. However, the methylated state of intron 2 prevents expression of full-length CCR5 transcripts. **(C)** CCR5-positive CD45RA+ T cells (e.g., TEMRA cells) and CD45RO+ T cells. Further activation may lead to recruitment of additional TFs (e.g., c/EBPβ (35)) that lead to demethylation of intron 2, promoting interactions between the *CCR5*-Distal (enhancer) and intron 2, and juxtaposing the enhancer, promoters, and intron 2 leading to formation of full-length CCR5 transcripts and surface CCR5 expression. The occupancy of *CCR5* locus by CTCF/cohesin complexes and their role in enhancer-promoter interactions in CD45RO+

T cells awaits further studies. Only relevant CREB1, c/EBPβ and Oct-2 binding sites are shown in the schematic.

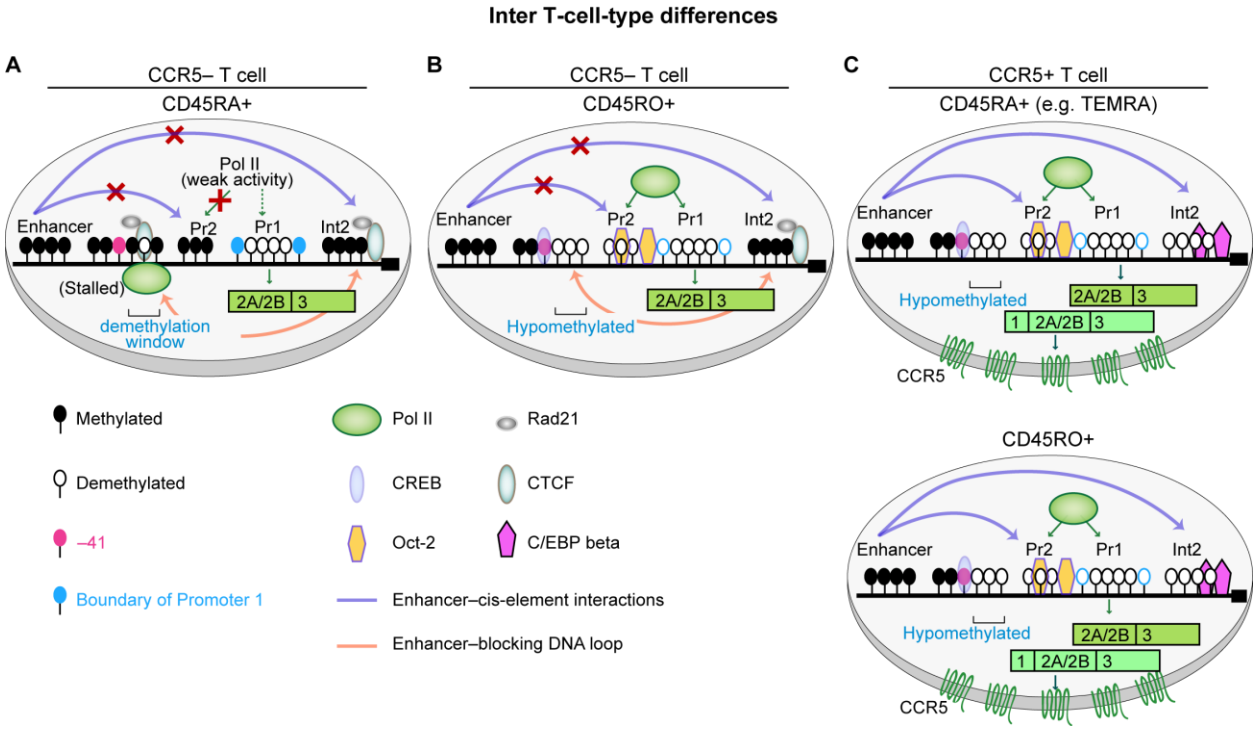
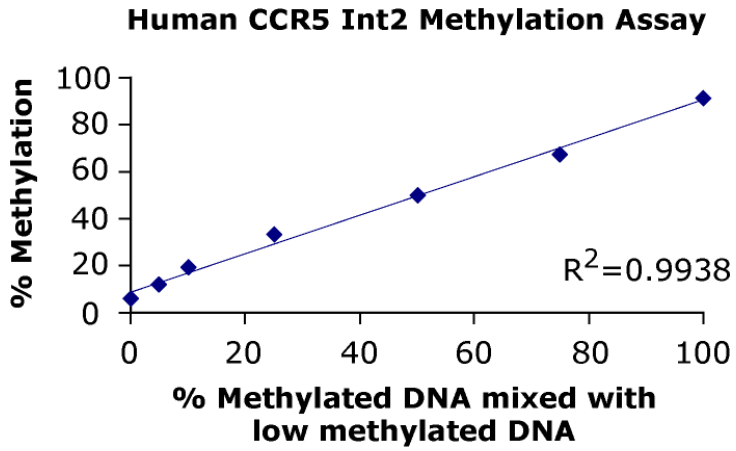


Fig. S11. Pyrosequencing is a reliable method to assess methylation levels in *CCR5 cis*-regions. Correlation between % methylation of the average of 4 CpGs in *CCR5* intron 2 (Int2; CpGs# −5 to −2) measured through pyrosequencing and the stoichiometric ratio of methylated DNA mixed with low methylated DNA. The strong linear correlation (R^2 > .99) suggests that the pyrosequencing assay had negligible bias in PCR amplification of methylated and unmethylated DNA and that this methodology can be used for quantitative detection of DNA methylation levels. A similar strategy was used to validate other pyrosequencing assays used in this study.



Supplementary Tables:

Table S1. CpG numbering system in the *CCR5 cis*-regulatory regions

CCR5 cis-region	CpG #	Position (hg19)	5' and 3'-Flanking Sequence	Pyro Assay
Distal	–48	46409208	ggactatagggg**cg**caccaccacatc	
	–47	46409309	aagttatccact**cg**ccttggcttccc	
	–46	46409294	ggcatgagccac**cg**cacttggcctag	
	–45	46409412	ctttttacagcc**cg**tctcactgttga	
	–44	46409564	ccatcctgttga**cg**atgctctgaaaa	
	–43	46409697	ggagggatggca**cg**aaacaccctcca	
	–42	46409770	ctatggggtgtc**cg**aatgtacttaat	
	–41c	46410036	aaagtggagtaa**cg**cacactgcaaag	
	–41b	46410137	atagaatcatgt**cg**tatttagggtgg	
Upstream	–41	46410171	ccaggtctagca**cg**tcatttaacaga	Upstream
	–40	46410222	gaagtatcttgc**cg**aggtcacacagc	
	–39	46410251	tcagcagcacag**cg**tgtgtgactccg	
	–38	46410263	cgtgtgtgactc**cg**agcctgctccgc	
	–37	46410274	ccgagcctgctc**cg**ctagcccacatt	
	–36b	46410306	tgggggtgagta**cg**tcttcacatcct	
	–36	46410422	acagtgttggtc**cg**gcagcctccggg	
	–35	46410432	tccggcagcctc**cg**gggggttctgcac	
	–34	46410483	ctatctatctttcgaaaaaccaacgt	
	–33	46410495	tcgaaaaaccaa**cg**ttgtatttatgc	
	–32	46410605	taaacagtcccc**cg**agggtgggtgcc	
Promoter 2	–31	46410857	agaacattttct**cg**atgattcgctgt	Pr2
	–30	46410865	ttctcgatgatt**cg**ctgtccttgtta	
	–29	46410931	ccctatatgggg**cg**gggttgggggtg	
	–28	46410952	gggtgtcttgat**cg**ctgggctatttc	
	–27	46411083	agacagagactc**cg**gtgaaccaattt	
	–26	46411199	ctagatgctggc**cg**tggatgcctcat	

	–25	46411268	ccaatacccagacgagaaagctgagg	
	–24	46411369	ggattgggggcacgtaattttgctgt	
	–23	46411447	gagccaaggtcacggaagcccagagg	
	–22	46411474	catcttgtggctcgggagtagctctc	
Exon 1 and Intron 1	–21b	46411542	gatgtcaccaaccgccaagagagctt	
	–21	46411803	gtttagctcacccgtgagcccatagt	
	–20	46411845	acaggttttttcgtttacagagaac	
	–19	46411912	ggataggggatacgggagagtggag	
Promoter 1	–18	46412248	gtgagaaaagcccgtaaataaacttt	
	–17	46412295	tctgattcttttcgccttcaatacac	
	–16	46412454	agctgagacatccgttcccctacaag	
	–15	46412478	agaaactctccccggtaagtaacctc	
Intron 2	–14	46412648	atgagtttcagacggttataacatca	
	–13b	46412708	aaagggagcaatcgtattttaataac	
	–13	46412943	ttaaattacaaacgccaaataaaatt	
	–12	46413201	tgtggtggcagacgaaacatttttta	
	–11	46413351	acttgggcacccggccatttcactc	
	–10b	46413418	taattctctttcgaggactgagagg	
	–10	46413466	gtgcaggcttcccgcattcaaaatcg	
	–9	46413478	cgcattcaaaatcggttgcttactag	
	–8	46413650	agtatgtgccctcgaggcctcttaat	
	–7	46413711	ggctaactctagcgtcaataaaaatg	
	–6b	46413743	ctgagttgcagccgggcatggtggct	
	–6	46413801	aggcaggaggatcgcttgagcccagg	Intron2
	–5	46413819	agcccaggagttcgagaccagcctgg	
	–4	46413890	ttggtgtggtggcgcctgtagtcccc	
	–3	46413943	attgcttgagccgggatgatccagg	
	–2	46413974	ttggtgtggtggcgcctgtagtcccc	
	–1	46413998	ctccagcctgggcgacagagtgagac	

The *CCR5* gene and CpG numbering system is shown. Left-most column indicates the *CCR5 cis*-regions (color-coded). Next column is the CpG numbering system with CpG –1 representing the first CpG upstream of the coding sequence. The next column shows the nucleotide coordinates of the CpG sites

and the numbering is based on hg19; the nucleotide number corresponds to the cytosine residue of the indicated CpG dinucleotide. CpG sites not present in HHC because of a polymorphism are highlighted in *yellow*. The HHC–specific CpG is highlighted in *orange*. The –6b CpG (rs2856764A/C) and –10b CpG (rs2254089C/T) are absent in HHC and present in all other haplotypes; –13b CpG (rs9282632A/C) is present in some HHA; –21b CpG is present in some HHE and HHG (rs2227010A/G); –33 CpG is present only in HHC (rs2856757A/C); –36b CpG (rs41499550C/T) is present in some HHB and HHD; –41b CpG (rs41490645A/C) is present in some HHG; –41c CpG (rs7637813A/G) is present in some HHE. Only some HHA (designated as HHA-2) bear rs553615728T that disrupts the –41 CpG site. Data were obtained by sequencing an ~9 kb region of the *CCR5* locus from multiple individuals with known haplotypes and by genotyping ~200 individuals who were homozygous for *CCR5* haplotypes. The SNP data corroborated with dBSNP build 137 (http://www.ncbi.nlm.nih.gov/SNP/). The column labeled Pyro assay refers to the region analyzed with pyrosequencing assays used in this study.

Table S2. Association of methylation content of *CCR5 cis*-regions with CCR5 surface levels on CD8+ T cells after controlling for *CCR5* genotype and levels of T-cell activation.

	%CCR5			CCR5 MFI		
***Cis*-region**	**Coeff.**	**95% CI**	***P* value**	**Coeff.**	**95% CI**	***P* value**
Model 1: Covariates − none						
Pr2	−0.70	−1.10 - −0.31	.001	−11.40	−19.40 - −3.40	.006
Model 2: Covariates – none						
Intron 2	−1.21	−1.65 - −0.78	< .001	−23.12	−31.81 - −14.43	< .001
Model 3: Covariates – none						
Pr2	0.06	−0.47 - 0.59	.82	5.42	−5.06 - 15.90	.31
Intron 2	−1.27	−1.91 - −0.62	< .001	−27.98	−40.77 - −15.18	< .001
Model 4: Covariates − %CD8+ CD38+ HLADR+ T cells and %CD8+ naïve T cells						
Pr2	0.25	−0.23 - 0.74	.30	9.54	−0.06 - 19.14	.05
Intron 2	−0.87	−1.47 - −0.26	.005	−20.18	−32.11 - −8.26	.001
Model 5: Covariates – as in model 4 plus *CCR5* haplotypes						
Pr2	0.23	−0.28 - 0.74	.36	10.40	0.10 - 20.71	.05
Intron 2	−0.78	−1.45 - −0.11	.02	−20.74	−34.25 - −7.22	.003
Model 6: Covariates – as in model 5 plus demographic and other clinical variables						
Pr2	0.47	−0.15 - 1.10	.13	11.44	−1.26 - 24.13	.08
Intron 2	−0.85	−1.57 - −0.13	.02	−20.97	−35.61 - −6.33	.006

Data are from 85 VL-suppressed HIV-infected subjects. Pr2 refers to the average methylation levels of 5 CpG sites −31 to −28 in *CCR5*-Pr2. Coeff. Coefficient. Intron 2 refers to the average methylation levels of 5 CpG sites −6 to −2 in *CCR5*-intron 2. *CCR5* haplotypes: HHA, HHC, HHE, HHG*1, HHF*2 and HHG*2, (haplotypes with >5% allele frequency in the cohort). Demographic and other clinical variables refer to age, gender, ethnicity, HCV infection status, duration of ART in months, and pre-treatment CD4+ T cell counts. MFI, mean fluorescent intensity.

Table S3. Nucleotide sequences of oligomers used in this study and bisulfite converted targeted sequences for pyrosequencing assays.

Name	Sequence (5' to 3')
Primers used for bisulfite genomic sequencing (BGS)*	
distalS3	GTTTGGTTTGGTTTTGAGATATAGT
distalA3	TCTCAATTTACATAACTCTTATTCTC
upstrS4	GGTTATAGAATTATGTAGTATTTAGGG
upstrA4	TACTACAACAAAACTCAATAACATAATC
nestdistS3	AGTTTGTTGTTATTTAGGTTGGAGT
nestupstS4	TATTTAGGGTGGAAGTTGTTTTA
nestdistA3	TAACTCTTATTCTCTTTTTATTATTAAATACA
nestupstA4	AAAACTCAATAACATAATCATAACAAAAA
S10	GTGGGTGTTTAAAAGGTTTTATATTTG
A9	CTATTTTATTACCCTTCAAACCAATAAAC
A13	ACAAAATACCCTCTAAACTTCC
A12	TCTCTACTCATCCCACTACACAAAATC
S3	GTAATTTTGTTGTTTGGGGTTTTATTTG
A5	AAAATCCTAAACTTCACATTAACCCTAT
S8	GGAAGTTTAGAGGGTATTTTGTGGTT
A6	CTCCCTACACCTTAAACTAAACAAC
S6	ATAGGGTTAATGTGAAGTTTAGGATTTT
2AII	CTTAAACTAAACAACTAAAAAAAACCCC
2SII	ATGTGGGTTTTTGATTAGATGAATG
A3	ATTTAATTACCCTCCATAAATACAAACTAT
S4	GTTGTTTAGTTTAAGGTGTAGGGAG
2AIII	CCCTCCATAAATACAAACTATTTTATACATC
2SIII	TGTAGGGAGTTTGAGATTTATAGGG
*Primer combinations are noted in **Table S4**	
Primers for PCR amplification and pyrosequencing	
CCR5-Upstream (CpGs# –41 to –37): [Amplicon Size=164 bp]	
UPST-FPB	biotin-TTTAGGGTGGAAGTTGTTTTAGG
UPST-RP	ACTCACCCCCAAAAAACAATATA
UPST-FS2	CCAAACTTCCATTTCCT
UPST-FS1	CCCAAAAAACAATATGAACTA
CCR5-Pr2 (CpGs# –31 to –28): [Amplicon Size=173 bp]	
Pr2FP	TTTGAATTGTATATATGGGATGAA
Pr2RPB	biotin-TACTTAAAAAAAACCAAAACAATATAA
Pr2FS1	ATATATGGGATGAATTAGAA
Pr2FS2	TGAGTTTTGTTGTAGTATAG
CCR5-Intron2 (CpGs# –6 to –1) [Amplicon Size=296 bp]	
Intron2FP	GGTTAGTATTTTAGGAGGTTGAGGTAG
Intron2RPB	biotin-CAAACATAATACAACTCAACCTTT
Intron2FS1	TTATTAGGAGGTTCAGGTAGG
Intron2FS2	TAGTTTGGTGTGGTGG
Intron2FS3	TAGGTTGTAGTGAGTTATGA
Abbreviations: F=Forward, R=Reverse, P=PCR primer, S=Sequencing Primer, B=Biotin. Note: Biotin group was linked to 5'-end of primer and were HPLC purified.	

Table S3 (continued)	
Bisulfite-converted target sequence for each pyrosequencing primer*	
UPSTW1FS1	TTATTTTTATTTATTYGTTTTGGTTTTTTAAAGTGTTGGGAATTYGTTTTGGTTTTTTAAAGTGTTGGGA
UPSTW1FS2	TTATYGTATTTGGTTTAGATATTTGTTTTT
UPSTW1FS3	TTATTTGTTTTTTTTATAGTTYGTTTTATTGTTGATTGATATTGATT
UPSTW1FS4	TAATATTTTTTTTATTTTGTTGAYGATGTTTTGAAAATATGGTTTAGAA
UPSTW2FS1	YGAAATATTTTTTAATATGGGTATGG
UPSTW2FS2	YGAATGTATTTAATAATAAAAAGAGA
UPSTDEMETHYLRS2	CATCTATTAAATAACRTACTAAACCTAAAACAACTTCCACC
UPSTDEMEHYLRS1	ACRAAACAAACTCRAAATCACACACRCTATACTACTAACTTACTATATAACCTCRACAAAATACTTCCACTACCCAAACT
PR2 FS1	TATTTTTTYGATGATTYGTTGTTTTTGTTATGATTATGTTAT
PR2 FS2	ATATATGTTTTTATATGGGGYGGGGGTGGGGGTGTTTTGATYGTTGGGT
Intron2 FS1	GGATYGTTTGAGTTTAGGAGTTYGAGATTAGTTTGGGTAATATAGTGT
Intron2 FS2	YGTTTGTAGTTTTTAGTTATTTGGAGGGGTGAGGTGAGAGGATTGTTTGAGTTYGGGATGGTTTAGGTTGTAGTGAGTT
Intron2 FS3	TYGTGTTATTGTATTTTAGTTTGGGYGATAGAGTGAGATTTTGTTTTATAA
*Bisulfite-converted target sequence is indicated. R=Purine Y=Pyrimidine	
Primers used for pyrosequencing to determine AEI	
−2135F	NNNATCCAGTGAGAAAAGCCCGTAAAT
−2135R	biotin- TTGAAGGCGAAAAGAATCAGA
−2135 Seq	AAAAGCCCGTAAATAAAC
EMSA Oligomers	
CCR5-CREB1* (wt)	CCAGGTCTAGCA**CG**TCATTTAACA
CCR5-CREB1 (mu)	CCAGGTCTAGCA**TG**TCATTTAACA
*−41 CpG site and the corresponding mutated site indicated in bold	

Table S4. Primer combinations and PCR cycling conditions used for BGS.

	PCR round	Primer combination	Annealing Temp. (˚C)	Extension Time (sec)
***CCR5 cis*-region**				
Distal	1	distalS3 & distalA3	65	120
	2	nestdistS3 & nestdisttA3	60	90
Upstream	1	upstrS4 & upstrA4	65	120
	2	nestupstS4& nestupstA4	60	90
Promoter 2	1	S10 & A9	60	90
	2	S10 & A13	60	80
Exon 1 and Intron1	1	S10 & A12	58	120
	2	S3 & A5	58	60
	3	S8 & A5	58	45
Exon 2 and Intron 2 (5' end)	1	S6 & A6	58	120
	2	2SII & 2AII	54	90
Intron 2 (3' end)	1	S4 & A3	58	120
	2	2SIII & 2AIII	58	120

Table S5. Validation of pyrosequencing and BGS assays for *CCR5*-upstream region

CpG Site	Coefficient of Determination (r^2)	*P* value (2-Tailed)
−41	.907	<.001
−40	.794	.001
−39	.930	<.001
−38	.931	<.001
−37	.962	<.001

Correlation coefficient (r^2) between the data obtained for BGS and pyrosequencing at the indicated CpG sites. *P* value was calculated taking the non-directional probability (two-tailed *t*-test).

Acknowledgements

We thank members of the Veterans Affairs (VA) Center on AIDS and HIV infection and the Center for Personalized Medicine for technical support (M. V. Soledad, J. Camargo, J. Sigala, Z. Lou, A. Carrilllo, E. Koppelaar, A. O'Sullivan, Una Aluyen, C. Peña, Y. Li, S. McCurdy, K. Begum, R. Maldonado, D. Narendra, A. Herr, N. Chopra, C. Ratliff, W. Pate, P. Ingale, S. Luo, J. Castiblanco); statistical input (N. Harper, T. Le and R. Dhanda); and advice (G. Catano, B. Taylor, and M. Gerardi). We thank K. M. Gorena for help in the Flow Cytometry Core that is supported by UTHSCSA, NIH-NCI P30 CA54174 (CTRC at UTHSCSA), and UL1RR025767 (CTSA). The work was supported by the Veterans Affairs (VA) Research Center for AIDS and HIV Infection and VA Center for Personalized Medicine grant (IP1 CX000875-01A1) and awards from the NIH (R37-AI046326 and R01-AI043279), Veterans Affairs (VA MERIT award) and San Antonio Life Sciences Institute to S.K.A. S.K.A is also a recipient of the Elizabeth Glaser Scientist Award, the Burroughs Wellcome Clinical Scientist Award in Translational Research, and the Doris Duke Distinguished Clinical Scientist Award and the Senior Scholar Award from the Max and Minnie Tomerlin Voelcker Fund. S.M. is supported by a Department of Veterans Affairs MERIT award (I01BX000975). The cohort and investigators from the Ukraine were supported by the U.S. Civilian Research and Development Foundation (CDRF) UKB1-2931-DN-08 to L.S.K. The CAPRISA 002 study was supported by the NIAID, NIH (grant U19 AI 51794). V.R. was a recipient of a CAPRISA training fellowship and a Fogarty AITRP fellowship (TWO-0023) and a KwaZulu-Natal Research Institute for TB and HIV (K-RITH) travel award. Thumbi Ndung'u is supported by the Hasso Plattner Foundation and holds the South African

DST/NRF Research Chair in Systems Biology of HIV/AIDS. Special acknowledgements to the following members of the CAPRISA Acute Infection Study team: C. Williamson, L. Morris, C. Gray and W. Hide. Research utilizing samples from the Johannesburg and Soweto cohorts was supported by the South African Research Chairs Initiative of the Department of Science and Technology and National Research Foundation of South Africa (C.T.T holds the South African DST/NRF Research Chair in HIV Vaccine Translational Research) and the Claude Leon Foundation (A.P. was a recipient of a Postdoctoral Fellowship). Patient recruitment from Soweto was also in part supported by an award from the National Institutes of Health grant RO1HL090312 (PI R.E. Chaisson). The cohort from NYU was supported by awards from the NIH P01AI057127 and P30AI027742. The work from UCSF was supported in part by UCSF/Gladstone CFAR (P30 AI027763), the Clinical and Translational Science Institute (UL1 RR024131), the Center for AIDS Prevention Studies (P30 MH62246), and the CFAR Network of Integrated Systems (R24 AI067039). P.W.H is a recipient of a Doris Duke Charitable Foundation Clinical Scientist Development Award (2008047). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We are indebted to the subjects who participated in these studies and who made this work possible. Because of space constraints, we regret our inability to cite additional excellent work. German Gornalusse dedicates this paper to Jorge P. Gornalusse (10/14/38-10/30/09).

References for SOM

1. Deeks SG & Walker BD (2007) Human immunodeficiency virus controllers: mechanisms of durable virus control in the absence of antiretroviral therapy. *Immunity* 27(3):406-416.
2. Dolan MJ*, et al.* (2007) CCL3L1 and CCR5 influence cell-mediated immunity and affect HIV-AIDS pathogenesis via viral entry-independent mechanisms. *Nature immunology* 8(12):1324-1336.
3. Shostakovich-Koretskaya L, *et al.* (2009) Combinatorial content of CCL3L and CCL4L gene copy numbers influence HIV-AIDS susceptibility in Ukrainian children. *Aids* 23(6):679-688.
4. van Loggerenberg F*, et al.* (2008) Establishing a cohort at high risk of HIV infection in South Africa: challenges and experiences of the CAPRISA 002 acute infection study. *PloS one* 3(4):e1954.
5. Rosenberg NA*, et al.* (2002) Genetic structure of human populations. *Science* 298(5602):2381-2385.
6. Gonzalez E, *et al.* (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. *Science* 307(5714):1434-1440.
7. Sinclair E*, et al.* (2006) Protective immunity to cytomegalovirus (CMV) retinitis in AIDS is associated with CMV-specific T cells that express interferon- gamma and interleukin-2 and have a CD8+ cell early maturational phenotype. *The Journal of infectious diseases* 194(11):1537-1546.
8. Clark SJ, Statham A, Stirzaker C, Molloy PL, & Frommer M (2006) DNA methylation: bisulphite modification and analysis. *Nature protocols* 1(5):2353-2364.
9. Mummidi S*, et al.* (2000) Evolution of human and non-human primate CC chemokine receptor 5 gene and mRNA. Potential roles for haplotype and mRNA diversity, differential haplotype-specific transcriptional activity, and altered transcription factor binding to polymorphic nucleotides in the pathogenesis of HIV-1 and simian immunodeficiency virus. *J Biol Chem* 275(25):18946-18961.
10. Gonzalez E*, et al.* (1999) Race-specific HIV-1 disease-modifying effects associated with CCR5 haplotypes. *Proc Natl Acad Sci U S A* 96(21):12004-12009.
11. Catano G*, et al.* (2011) Concordance of CCR5 genotypes that influence cell-mediated immunity and HIV-1 disease progression rates. *The Journal of infectious diseases* 203(2):263-272.
12. Picton AC, Paximadis M, & Tiemessen CT (2010) Genetic variation within the gene encoding the HIV-1 CCR5 coreceptor in two South African populations. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* 10(4):487-494.
13. Tost J & Gut IG (2007) DNA methylation analysis by pyrosequencing. *Nature protocols* 2(9):2265-2275.
14. Tost J & Gut IG (2007) Analysis of gene-specific DNA methylation patterns by pyrosequencing technology. *Methods in molecular biology* 373:89-102.
15. Warnecke PM*, et al.* (1997) Detection and measurement of PCR bias in quantitative methylation analysis of bisulphite-treated DNA. *Nucleic acids research* 25(21):4422-4426.
16. Wang H*, et al.* (2012) Widespread plasticity in CTCF occupancy linked to DNA methylation. *Genome research* 22(9):1680-1688.

17. Gertz J*, et al.* (2013) Distinct properties of cell-type-specific and shared transcription factor binding sites. *Molecular cell* 52(1):25-36.
18. Zhou X*, et al.* (2011) The Human Epigenome Browser at Washington University. *Nat Methods* 8(12):989-990.
19. Neph S*, et al.* (2012) An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* 489(7414):83-90.
20. Thurman RE*, et al.* (2012) The accessible chromatin landscape of the human genome. *Nature* 489(7414):75-82.
21. Bernstein BE*, et al.* (2010) The NIH Roadmap Epigenomics Mapping Consortium. *Nat Biotechnol* 28(10):1045-1048.
22. Frazer KA, Pachter L, Poliakov A, Rubin EM, & Dubchak I (2004) VISTA: computational tools for comparative genomics. *Nucleic acids research* 32(Web Server issue):W273-279.
23. Mathelier A*, et al.* (2014) JASPAR 2014: an extensively expanded and updated open-access database of transcription factor binding profiles. *Nucleic acids research* 42(Database issue):D142-147.
24. Medvedeva YA*, et al.* (2014) Effects of cytosine methylation on transcription factor binding sites. *BMC genomics* 15:119.
25. Mummidi S*, et al.* (2007) Production of specific mRNA transcripts, usage of an alternate promoter, and octamer-binding transcription factors influence the surface expression levels of the HIV coreceptor CCR5 on primary T cells. *Journal of immunology* 178(9):5668-5681.
26. Sandig H*, et al.* (2009) Fibronectin is a TH1-specific molecule in human subjects. *The Journal of allergy and clinical immunology* 124(3):528-535, 535 e521-525.
27. Trapnell C, Pachter L, & Salzberg SL (2009) TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* 25(9):1105-1111.
28. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome biology* 11(10):R106.
29. Manger B, Weiss A, Weyand C, Goronzy J, & Stobo JD (1985) T cell activation: differences in the signals required for IL 2 production by nonactivated and activated T cells. *Journal of immunology* 135(6):3669-3673.
30. Wu L*, et al.* (1997) CCR5 levels and expression pattern correlate with infectability by macrophage-tropic HIV-1, in vitro. *J Exp Med* 185(9):1681-1691.
31. Wierda RJ*, et al.* (2012) Epigenetic control of CCR5 transcript levels in immune cells and modulation by small molecules inhibitors. *Journal of cellular and molecular medicine* 16(8):1866-1877.
32. Oswald-Richter K*, et al.* (2007) Identification of a CCR5-expressing T cell subset that is resistant to R5-tropic HIV infection. *PLoS pathogens* 3(4):e58.
33. Macian F (2005) NFAT proteins: key regulators of T-cell development and function. *Nature reviews. Immunology* 5(6):472-484.
34. Kuipers HF*, et al.* (2008) CC chemokine receptor 5 gene promoter activation by the cyclic AMP response element binding transcription factor. *Blood* 112(5):1610-1619.
35. Rosati M, Valentin A, Patenaude DJ, & Pavlakis GN (2001) CCAAT-enhancer-binding protein beta (C/EBP beta) activates CCR5 promoter: increased C/EBP beta and CCR5 in T lymphocytes from HIV-1-infected individuals. *Journal of immunology* 167(3):1654-1662.