**Primary CD4+T cell purification and cell culture**

Primary CD4+ T cells from elite controllers received from the CODEX cohort (Table 1) were purified from peripheral blood mononuclear cells (PBMCs) using human CD4+ T cell Isolation Kit (Miltenyi Biotec). Briefly, 10 million PBMCs were resuspended in buffer (2 - 8 ℃) containing phosphate-buffered saline (PBS), pH 7.2, 0.5% bovine serum albumin (BSA), and 2 mM EDTA in the volume of 40 µL. 10 µL of CD4+ T cell Biotin-Antibody Cocktail were subjected to resuspended PBMCs and the mixture was incubated at 4 ℃ for 5 min. 30 µL of buffer was again added to the mixture followed by adding 20 µL of CD4+ T cell MicroBead Cocktail. The mixture was then incubated at 4 ℃ for 10 min. 400 µL of buffer was added to fill the volume up to 500 µL. Cell suspension was applied to the MACS column (LS), which has been rinsed with 3 mL of buffer. Flow-through containing the enriched CD4+ T cells were collected. The column was washed with 3 mL of buffer and the flow-through was again collected and combined with the previous collected flow-through. 1x PBS containing 2% fetal bovine serum was added to enriched CD4+ T cells up to a 50 mL volume. Cells were centrifuged at at 300 g (break low) at 20 ℃ for 10 min and discarded supernatant. Cells were resuspended in 3 mL RBC lysis buffer () at the room temperate for 10 min. 1x PBS containing 2% fetal bovine serum was added to RBC-lysed CD4+ T cells up to a 50 mL volume. Cells were centrifuged at at 300 g (break low) at 20 ℃ for 10 min and discarded supernatant. Cell pellets were resuspended in RPMI 1640 medium (Gibco) supplemented with 10% fetal bovine serum (Gibco), 1% Penicillin-Streptomycin (Gibco), 1% sodium pyruvate (Gibco) and 1% GlutaMAX (100x) (Gibco).

The human primary CD4+ T cells were purified from the healthy donor (Établissement Français Du Sang). Briefly, 1x phosphate-buffered saline (PBS) was added to buffy coat up to a 160 mL volume. 8 mL RosetteSepTM Cocktail (Stemcell Technologies) was slowly added into the mixture and incubated at room temperature for 20 min. The mixture was diluted by using 1x PBS containing 2% fetal bovine serum (Gibco) up to 240 mL and distributed into 8 50 mL falcon tubes containing 10 mL Lymphocytes separation medium (eurobio, CMSMSL01-01). The mixture was centrifuged at 1,200 g (slow acceleration and brake off) at 20 ℃ for 20 min. Enriched cell layer was harvested with a pipette and transferred to new 50 mL falcon tubes. Collected cells were diluted by 1x PBS containing 2% fetal bovine serum up to a 50 mL volume. Cells were centrifuged at 300 g (break low) at 20 ℃ for 10 min and discarded supernatant. Cells were resuspended in 3 mL RBC lysis buffer () at the room temperate for 10 min. 1x PBS containing 2% fetal bovine serum was added to RBC-lysed CD4+ T cells up to a 50 mL volume. Cells were centrifuged at at 300 g (break low) at 20 ℃ for 10 min and discarded supernatant. Cell pellets were resuspended in RPMI 1640 medium (Gibco) supplemented with 10% fetal bovine serum (Gibco), 1% Penicillin-Streptomycin (Gibco), 1% sodium pyruvate (Gibco) and 1% GlutaMAX (100x) (Gibco).

**Transfection and viral Infection**

The barcoded HIV (pHCC1) library was produced as described before (Chen et al. 2017; Chen et al. 2018). To prepare viral stocks, two million HEK 293T cells in 10 cm dishes were transfected with 3 µg p8.2, 2 µg pMD2G and 6 µg barcoded pHCC1 (Chen et al. 2017; Chen et al. 2018). After 16 hours, the medium was replaced. The supernatant was collected 72 hours post transfection and filtered through a 0.45 µm pore membrane (Merck). The vector was concentrated by using ultracentrifugation (Beckman Coulter LE-80K). Supernatant was discarded and pellet was resuspended with 250 µL PBS containing Ca2+ and Mg2+ at 4 ℃ for 3 h. Viral stocks were stored at -80 °C. Transfection efficiency was validated by measuring the percentage of GFP(+) cells from barcoded HIV-infected HEK 293T cells by FACS analysis.

2 million primary CD4+ T cells in 24-well plates were activated by 1 µg/mL PHA (Remel Europe Ltd.) / 50 U/mL IL2 (Roch) in a 1 mL volume 3 d before infection. Cells were infected by barcoded viral inocula with MOI around 0.1. The medium was replaced by 0.5 mL fresh RPMI containing IL2 d 3 post infection. The efficiency of infection was monitored d 4 post infection by FACS analysis (see **FACS analysis** for detail). Cells were FACS-sorted d 5 post infection and immediately resuspended in 350 µL Buffer RLT (supplied in AllPrep DNA/RNA Mini Kit, Qiagen) containing 1% β-mercaptoethanol (β-ME) for DNA and RNA extraction (see **DNA extraction and inverse PCR** for detail).

**FACS analysis**

Flow cytometry for validation of purity of CD4+ T cells and infection efficiency was performed on a MACSQuant Analyzer 16 (Miltenyi Biotec). Before FACS analysis, cells were washed and resuspended in 500 µL phosphate‐buffered saline (PBS) containing 2 % fetal bovine serum and 2 mM EDTA and centrifuged at 400 g at room temperature for 5 min. Discard supernatant. Pellets were resuspended in 100 µL PBS containing 2 % fetal bovine serum, 2 mM EDTA, 5 µL anti-human CD3 (BioLegend) and 5 µL anti-human CD4 (SONY) at room temperature for 20 min. Cells were washed again using 500 µL phosphate‐buffered saline (PBS) containing 2 % fetal bovine serum and 2 mM EDTA and centrifuged at 400 g at room temperature for 5 min. Pellets were resuspended in 1% perfluoroalkoxy alkanes (PFA) for FACS analysis.

**DNA extraction and inverse PCR**

Genomic DNA from barcoded HIV-infected primary CD4+ T cells was with the AllPrep DNA/RNA Mimi Kit (Qiagen) in a final 50 µL volume. 50 µL genome DNA was split into two proportion to prepare the libraries followed by high-throughput sequencing to identify HIV integration sites. The same B-HIVE protocol (Chen et al., 2017; Chen et al., 2018) was used. Of note, 50 µL of the products from first round of nested PCR were split into five proportions, with which five independent second round of nested PCR were performed. Eventually, the PCR products with same indices were pooled for high-throughput sequencing. The sequence of each primer is described in **Table X**. Integration sites retrieved from the B-HIVE technology are referred to as “De novo EC IS” and “De novo HD IS” in this study.

**mRNA purification and whole transcriptome amplification**

Total RNA from infected cell pools was extracted with the AllPrep DNA/RNA Micro Kit (Qiagen) and mRNA was purified from total RNA with an Oligotex mRNA Mini Kit (Qiagen). Amplification of RNA and DNA barcodes followed by high-throughput sequencing was performed based on the previous protocol (Chen et al., 2017; Chen et al., 2018). The sequence of each primer is described in **Table X**.

**HIV-1 Integration Site position (IS) from HIV-1-infected elite controllers and HIV-1-infected individuals receiving cART treatment**

Integration sites from HIV-1-infected elite controllers collected from Jiang et al. (2020) (Jiang et al. 2020) were referred to as “Long term EC IS” in this study. Of note, integration sites present in the same position from the same study patient were considered as an identical provirus at the analytical step.

Integration sites from HIV-1-infected individuals receiving cART treatment collected from Einkauf et al. (2022) (Einkauf et al. 2022) were referred to as “cART IS untreat”, “cART IS short treatment” and “cART IS long treatment” depending on the period of cART treatment: short treatment is referred to as the treatment no more than 1 yr; long treatment is referred to as the treatment is more than 1 yr. “cART all IS” is referred to all IS from cART-treated individuals disregarding of the period of the treatment.

We use also the IS from cART treated individual from Maldarelli et al., 2014 and Wargner et al., 2014 paper, PMID: 24968937 and 25011556 respectively

Maldarelli

**Transient Transcriptome sequencing in Resting and activated CD4 T cells from healthy donors**

TT-seq library construction was done using protocol from (Schwalb et al., 2016). Briefly, purified CD4+ T cells were labeled in culture media for 5 min with 500µM 4-thiouridine (4sU, Sigma-Aldrich). Cells were harvested, spike-ins were added, RNA was purified using QIAzol lysis reagent (Qiagen) and fragmented on Bio-Ruptor with a High setting for one cycle 30s ON - 30s OFF. Fragmentation efficiency was analyzed on the Agilent Bioanalyzer 2100 with an RNA 6000 Pico kit (Agilent Technologies). Fragmented RNA was biotinylated with Biotin-HPDP EZ link (Thermo Fisher Scientific) and purified with chloroform and precipitated with isopropanol. Biotinylated RNA labeled with Streptavidin beads (µMacs Streptavidin Kit – Miltenyi Biotec.) with a ratio of 1:1 then purified on µMACs columns (Miltenyi Biotec). RNA cleaning was done with Agencourt RNAclean XP beads (Beckman Coulter) with a ratio 1.8:1. Sequencing libraries were prepared with the Ovation Universal RNA-seq library Kit (NuGEN) following the manufacturer's instructions. Libraries were paired-end sequenced (50bp) on an Illumina HiSeq 2500.

Reads were mapped to human genome GRCh37 (ftp://ftp.ensembl.org/pub/release-75/fasta/homo\_sapiens/dna/) using STAR (version STAR\_2.5.1b\_modified) (Dobin et al., 2013) with default parameters. Annotations from ENSEMBL (GRCh37 release 75) were used for genome indexing with STAR before mapping. We used RSEM (v1.2.29) (B. Li & Dewey, 2011) to compute Transcripts Per Million (TPM) values of all annotated Transcripts:

rsem-calculate-expression --paired-end --star --estimate-rspd --append-names --star-gzipped-read-file --strand-specific --forward-prob 1 --time --calc-pme -calc-ci

To fix thresholds of transcription we use control genes known to be expressed or not expressed in the same cell types used in this study (Supplementary). Based on the value of TPM of control genes we empirically set this value to 6 TPM. Differential expression was assessed using featureCount (v1.5.1) (Liao et al., 2014) and DESeq2 (version DESeq2\_1.12.4) (Rajkumar et al., 2015) from Bioconductor:

featureCounts -a hg19.gtf -p -t exon -g gene\_id -s 1 sorted.bam

**ChIP-seq of Histone modification of Resting and Activated CD4 T cell**

We use raw read from publicly available data from (Burren et al., 2017). Reads were mapped to the human genome (GRCh37) using BWA (bwa\_aln 0.7.10-r789) (H. Li & Durbin, 2009) with default parameters.

Fragment length were inferred using cross strand correlation method from MaSC.pl (Version 1.2.1) (Ramachandran, Palidwor, Porter, & Perkins, 2013), the mappability track were generated using GEM-mappability (build 1.315 (beta)) (Derrien et al., 2012).

Meta-gene average profiles were generated using deeptools 2 software (version 2.5.3) (Ramírez et al., 2016). First, normalized genome wide coverage (bigwig files) were generated using bamCompare with signal extraction scaling (SES) normalization method to subtract background:

bamCompare --bamfile1 {IP} --bamfile2 {Input} --scaleFactorsMethod SES --sampleLength 1000 --numberOfSamples 100000 --outFileName {IP.norm.bw} --pseudocount 0 --ratio subtract --binSize 10 --smoothLength 30 --minMappingQuality 20 --extendReads {fragment.length}

Then we constructed meta-gene matrices using deeptools computeMatrix and annotation file from ENSEMBL (release 75):

computeMatrix scale-regions --regionsFileName Homo\_sapiens.GRCh37.75.bed --scoreFileName {normalised.bw} --outFileName {output} --upstream 2000 --unscaled5prime 500 --regionBodyLength 2000 --unscaled3prime 500 --downstream 2000 --binSize 50 --sortRegions “descend” --sortUsing “mean”

Finally, we plotted meta-gene average profile with R scripts using the data.table and the ggplot2 package. We split genes into four groups using TPM values from RSEM and IS from cART-treated individuals studies as follow:

* HIV targeted genes
* Untranscribed genes (below 6 TPM)
* Transcribed genes
* Highly transcribed genes

Untranscribed genes were determined using the median of all genes below 6 TPM. Above this threshold, genes were split into two equally sized groups based on TPM values. The highest half was labelled "Highly transcribed", the lowest "Transcribed genes".

**ChIP-seq enrichment around HIV-1 Integration Site positions**

We then sought to compare ChIP-seq patterns between IS positions and randomly selected positions in genes with similar expression levels. To do so, we split all genes targeted by HIV in the cART studies into ten equally sized bins based on the TT-seq TPM values. We selected random genes not targeted by HIV and matched them to the same amount of HIV targeted genes from the same bin.

We use deeptools constructMatrix to construct matrices used to plot average profile around IS. We test the significance of differences using a Wilcoxon rank-sum test with the average signal of +/- 500 bp around IS versus average signal around random position.

**Hi-C library preparation**

Two million purified CD4+ T-cells were cross-linked with 1% formaldehyde solution quench then with the addition of 0.2M glycine. Cells were lysed with HiC lysis buffer (10mM Tris-HCl pH8.0 (Sigma-Aldrich), 10mM NaCl (Sigma-Aldrich), 0.2% Igepal CA63 (Sigma-Aldrich) and 1x of protease inhibitors (Roche)). DNA was digested using DpnII restriction enzyme. After biotin filing during DNA end repair, proximity ligation was carried out 4h at 18°C with 2000U T4 DNA ligase (NEB). After reverse-crosslinking, DNA was purified using ethanol precipitation and sheared to 300-500 bp fragments using Biorupter (Diagenode). DNA sizing was performed using two successive AMPure XP Beads (Beckman Coulter) purification to remove DNA shorter than 300 bp and bigger than 500 bp. Ligation fragments containing biotin were immobilized on MyOne Streptavidin T1 beads (ThermoFished Cat.N: 65602), end and A-tailed as described previously (Rao et al., 2014). NEXTflex adaptors (Bioo Scientific) were then ligated. Streptavidin beads were removed by heating DNA 98°C 10min to avoid potential interference during PCR amplification. Amplified DNA with adapted Illumina primer for 6 cycles. DNA was then size selected using AMPure XP beads with a ratio of 0.7:1 twice. The in situ HiC libraries were then quantified and analyzed on the Agilent Bioanalyzer 2100 with a DNA 6000 Pico and paired-end sequenced (50bp) on an Illumina HiSeq 2500 sequencer.

**Hi-C processing**

Hi-C data were processed using the Juicer pipeline (Durand et al., 2016) for visualisation and HicExplorer 3.7.1 for processing. In short, Hi-C reads were mapped to DnpII restriction fragments (hg19). Duplicates reads with MAPQ < 30, and intra-fragment reads were removed. Reads were then binned to generate contact matrices for all chromosomes at resolutions 5, 10, 25, 50 kb. Contact matrices were next normalized by reads depth, corrected using KR matrix balancing algorithm and we merge replicates.

Topologically associating domains (TADs) were identified using hicExplorer (Ramirez et al. 2018) at 25/50 kb resolution. Around 4500/3800 TADs were identified, and their median size was around 325/500 kb respectively. TADs could not be identified at lower resolutions due to matrix sparsity. TADs were then annotated according to TT-seq data: TADs were considered as active if the average TT-seq TPM (transcripts per kilobase million) was superior to 10, and were considered as inactive if not.

Enrichment analysis in tri-dimensional chromatin structure context

Enrichment of HIV IS in active TADs was assessed. For this purpose, IS were first mapped to genes. Then the enrichment of genes with IS in active TADs was compared to the enrichment of genes without IS in active TADs using the Fisher’s exact test. Similarly, the enrichment of genes with IS at loop anchors was compared to the enrichment of genes without IS at loop anchors using the Fisher’s exact test.

In active TADs, we assessed the enrichment of histone marks between TADs with IS (IS enriched domains) and TADs without IS (IS depleted domains). We used the Wilcoxon rank-sum test to assess the significance of the enrichment difference.

To assess the enrichment of IS at TAD borders, we computed an average profile of the number of IS that overlap TAD borders and the surrounding regions. Similarly, we computed average profiles of the number of IS that overlap enhancers (Andersson et al., 2014) and super-enhancers (Khan & Zhang, 2016) and the surrounding regions.

The enrichment of histone marks at enhancers/super-enhancers overlapping IS was compared between active and quiescent cells, depending on gene status (induced, repressed or unchanged). We used the Wilcoxon rank-sum test to assess the significance of the enrichment difference.

**Interaction between IS analysis in Hi-C**

To see if IS are more in contact in 3D context we split Hi-C normalized contact into ten equally sized groups (at 50 kb resolution) and count the proportion of IS contacting IS (or IS contacting others regions).

**APA analysis**

To test the specific interaction between IS-IS, IS-TSS and IS-enhancers/super-enhancers in elite controllers (De novo EC IS), healthy donors (De novo HD IS) and cART-treated studies (cART all IS), respectively, we use Hi-C matrix normalized at 50 kb resolution. Then we aggregate all local matrices by computing the mean of normalized contacts between all pairs that have a minimal distance of 1 Mb.

**APA differential boxplot**

We take the center of APA matrix and normalize by the bottom right of APA to have a factor of correction. We then apply this factor on all values of the matrices and performed a two-sided pairwise Wilcoxon rank-sum test. The enhancer annotation came from http://enhanceratlas.org/ (CD4+ ells).

**Construction of Promoter Capture Hi-C and Reciprocal Capture Hi-C network**

We used processed Promotor Capture data from Javierre et al., 2016 (<https://osf.io/u8tzp>). Briefly, data were mapped to GRCh37 using the HiCUP pipeline (<https://www.bioinformatics.babraham.ac.uk/projects/hicup/>), and interaction confidence scores were computed using the CHiCAGO pipeline (Cairns et al., 2016). We considered the set of interactions with high confidence scores (>=5) in this paper. We then constructed a network with significant interactions using igraph R package (version 1.1.2). Each node of the network represents HindIII fragments, and edges represent a significant interaction between two HindIII fragments. We constructed three networks. One for resting CD4 T cells (naCD4) specific interactions, one for Activated CD4 T cells (aCD4) specific interactions and one by merging all significant interactions between Resting and Activated CD4 T cells.

We used the R package igraph to analyze contact networks and visNetwork (v2.0.8) for visualization.

**Assortativity of HIV-1 IS**

We used the R data.table package to annotate HindIII fragments harbouring one or more IS (Boolean value) using foverlaps function. Assortativity was then computed on the constructed network using the R igraph package. To give significance to computed assortativity of IS, we randomly selected the same amount of nodes as those with IS and thus computed a “control” assortativity.

**Enrichment of IS in the network**

The majority of IS are found in what we termed a super-graph, so we decided to cluster the super-graph into communities using the *fastgreedy.community* function implemented in the igraph R package. This function uses a measure of modularity; it compares the number of intra versus inter-community connections to optimize clustering.

TPM values for each fragment were calculated using the mean TPM from 2 donors. We first computed the mean TPM of all genes, then selected genes that overlap each fragment. If there were multiple genes per fragment, we used their median values.

The size of graphs or communities corresponds to the total length of fragments (in bp). To measure the enrichment of IS we used a binomial law with the number of fragment with IS and size of graph/community (in bp). The probability is calculated as the total number of IS in captured fragments divided by the total length of captured fragments.

p(fragment with IS) = ∑ (length of fragment with IS)/∑ (length of all captured fragments)

**Transcriptional and epigenetic analysis of communities enriched with IS versus not enriched with IS**

Groups are based on enrichment in IS from the cART studies: “enriched” when p-value <= 0.05, "not enriched" when p-value > 0.05 & number of IS > 0, without IS when number of IS == 0.

For TT-seq data we used the median of fragment TPMs overlapping genes by communities. For ChIP-seq data, we compute the normalized mean of the signal by fragment and use mean of this value for each community.

**Bootstrapping IS present in A/B compartments**

IS, including “cART IS untreat”, “cART IS short treatment”, “cART IS long treatment”, “Long term EC IS”, “De novo EC IS” and “De novo HD IS” were overlapped with A/B compartments matrices defined in primary CD4 T cells ([GSE138767](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138767)) (Yang et al., 2020) by using BEDTools (Quinlan and Hall, 2010) with the command intersect and the opinions -wa and -wb. 200 IS in each IS category were randomly sampling with replacement; the process was repeated 10,000 times. Statistical tests were performed with R with default options.

**Bootstrapping IS present in satellite DNA, SINES and LINES**

IS, including “cART IS untreat”, “cART IS short treatment”, “cART IS long treatment”, “Long term EC IS”, “De novo EC IS” and “De novo HD IS” were overlapped with human hg38 genomic repeats (<https://www.repeatmasker.org/species/hg.html>) released by RepeatMasker (Smit et al.) by using BEDTools (Quinlan and Hall, 2010) with the command intersect and the opinions -wa and -wb. 200 IS in each IS category were randomly sampling with replacement; the process was repeated 10,000 times. Statistical tests were performed with R with default options.

**Code availability**

Code for bootstrapping IS present in A/B compartments and in satellite DNA, SINES and LINES is available at <https://github.com/HCAngelC>.

**Data availability**

The raw sequencing data (B-HIVE outputs) generated during this study are available from Gene Expression Omnibus (XXX). A/B compartments data ([GSE138767](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE138767)) were downloaded from Gene Expression Omnibus. Human hg38 genome repeats were downloaded from RepeatMasker (<https://www.repeatmasker.org/species/hg.html>).

**References:**

Anders, S., Pyl, P. T., & Huber, W. (2015). HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)*, *31*(2), 166–169. <https://doi.org/10.1093/bioinformatics/btu638>

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., … Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493), 455–461. <https://doi.org/10.1038/nature12787>

Burren, O. S., Rubio García, A., Javierre, B.-M., Rainbow, D. B., Cairns, J., Cooper, N. J., … Wallace, C. (2017). Chromosome contacts in activated T cells identify autoimmune disease candidate genes. *Genome Biology*, *18*(1), 165. <https://doi.org/10.1186/s13059-017-1285-0>

Cairns, J., Freire-Pritchett, P., Wingett, S. W., Várnai, C., Dimond, A., Plagnol, V., … Ryba, T. (2016). CHiCAGO: robust detection of DNA looping interactions in Capture Hi-C data. *Genome Biology*, *17*(1), 127. <https://doi.org/10.1186/s13059-016-0992-2>

Chen, H.-C., Martinez, J.P., Zorita, E., Meyerhans, A., & Filion, G.J. (2017). Position effects influence HIV latency reversal. *Nature Structural & Molecular Biology*, 24, 47–54. <https://doi.org/10.1038/nsmb.3328>

Chen, H., Zorita, E., & Filion, G.J. (2018). Using Barcoded HIV Ensembles (B‐HIVE) for Single Provirus Transcriptomics. *Current Protocols in Molecular Biology* <https://doi.org/10.1002/cpmb.56>

Derrien, T., Estellé, J., Marco Sola, S., Knowles, D. G., Raineri, E., Guigó, R., & Ribeca, P. (2012). Fast computation and applications of genome mappability. *PloS One*, *7*(1), e30377. <https://doi.org/10.1371/journal.pone.0030377>

Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., … Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*, *29*(1), 15–21. <https://doi.org/10.1093/bioinformatics/bts635>

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3*(1), 95–98. <https://doi.org/10.1016/j.cels.2016.07.002>

Einkauf, K. B., Osborn, M. R., Gao, C., Sun, W., Sun, X., Lian, X., Parsons, E. M., Gladkov, G. T., Seiger, K. W., Blackmer, J. E., Jiang, C., Yukl, S. A., Rosenberg, E. S., Yu, X. G., & Lichterfeld, M. (2022). Parallel analysis of transcription, integration, and sequence of single HIV-1 proviruses. *Cell*, *185*(2), 266–282.e15. <https://doi.org/10.1016/j.cell.2021.12.011>

Javierre, B. M., Burren, O. S., Wilder, S. P., Kreuzhuber, R., Hill, S. M., Sewitz, S., … Fraser, P. (2016). Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters. *Cell*, *167*(5), 1369–1384.e19. <https://doi.org/10.1016/j.cell.2016.09.037>

Jiang, C., Lian, X., Gao, C., Sun, X., Einkauf, K. B., Chevalier, J. M., Chen, S. M. Y., Hua, S., Rhee, B., Chang, K., Blackmer, J. E., Osborn, M., Peluso, M. J., Hoh, R., Somsouk, M., Milush, J., Bertagnolli, L. N., Sweet, S. E., Varriale, J. A., … Yu, X. G. (2020). Distinct viral reservoirs in individuals with spontaneous control of HIV-1. *Nature*, *585*(7824), 261–267. <https://doi.org/10.1038/s41586-020-2651-8>

Khan, A., & Zhang, X. (2016). dbSUPER: a database of super-enhancers in mouse and human genome. *Nucleic Acids Research*, *44*(D1), D164-71. <https://doi.org/10.1093/nar/gkv1002>

Li, B., & Dewey, C. N. (2011). RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics*, *12*(1), 323. <https://doi.org/10.1186/1471-2105-12-323>

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*, *25*(14), 1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>

Liao, Y., Gordon K. S., Wei S. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. <https://doi.org/10.1093/bioinformatics/btt656>

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics , 26(6), 841–842 <https://doi.org/10.1093/bioinformatics/btq033>

Rajkumar, A. P., Qvist, P., Lazarus, R., Lescai, F., Ju, J., Nyegaard, M., … Christensen, J. H. (2015). Experimental validation of methods for differential gene expression analysis and sample pooling in RNA-seq. *BMC Genomics*, *16*, 548. <https://doi.org/10.1186/s12864-015-1767-y>

Ramachandran, P., Palidwor, G. A., Porter, C. J., & Perkins, T. J. (2013). MaSC: mappability-sensitive cross-correlation for estimating mean fragment length of single-end short-read sequencing data. *Bioinformatics (Oxford, England)*, *29*(4), 444–450. <https://doi.org/10.1093/bioinformatics/btt001>

Ramírez, F., Ryan, D. P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A. S., … Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research*, *44*(W1), W160–W165. <https://doi.org/10.1093/nar/gkw257>

Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., Aiden, E.L. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*,*159*(7):1665-80. <https://doi.org/10.1016/j.cell.2014.11.021>

Schwalb, B., Michel, M., Zacher, B., Fruhauf, K., Demel, C., Tresch, A., … Cramer, P. (2016). TT-seq maps the human transient transcriptome. *Science*, *352*(6290), 1225–1228. <https://doi.org/10.1126/science.aad9841>

Smit, A.F.A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. 2013-2015 <http://www.repeatmasker.org>

Yang, J., McGovern, A., Martin, P., Duffus, K., Ge, X., Zarrineh, P., Morris, A. P., Adamson, A., Fraser, P., Rattray, M., & Eyre, S. (2020). Analysis of chromatin organization and gene expression in T cells identifies functional genes for rheumatoid arthritis. [https://doi.org/10.1101/827923](http://www.apple.com/tw)

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., … Liu, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9), R137. <https://doi.org/10.1186/gb-2008-9-9-r137>