

Differential Viral Accessibility (DIVA) identifies alterations in chromatin architecture through large-scale mapping of lentiviral integration sites

Richard T. Timms^{1†}, Iva A. Tchasovnikarova^{1,2†*} and Paul J. Lehner^{1*}

¹Department of Medicine, Cambridge Institute for Medical Research, Cambridge Biomedical Campus, Cambridge, CB2 0XY, UK

²Department of Molecular Biology, Massachusetts General Hospital, and Department of Genetics, Harvard Medical School, Boston, MA 02114, USA

*Correspondence should be addressed to I.A.T. (it257@cam.ac.uk) and P.J.L. (pjl30@cam.ac.uk).

†These authors contributed equally to the work.

SUMMARY

Alterations in chromatin structure play a major role in the epigenetic regulation of gene expression. Here we describe a step-by-step protocol for Differential Viral Accessibility (DIVA), a method for identifying changes in chromatin accessibility genome-wide. Commonly used methods to map accessible genomic loci have strong preferences towards detecting “open” chromatin found at regulatory regions, but are not well suited to study chromatin accessibility in gene bodies and intergenic regions. DIVA

overcomes this limitation, enabling a broader range of sites to be interrogated. Conceptually DIVA is similar to ATAC-seq in that it relies on the integration of exogenous DNA into the genome to map accessible chromatin, except that chromatin architecture is probed through mapping integration sites of exogenous lentiviruses. An isogenic pair of cell lines are transduced with a lentiviral vector, followed by PCR amplification and Illumina sequencing of virus-genome junctions; the resulting sequences define a set of unique lentiviral integration sites which are compared to determine whether genomic loci exhibit significantly altered accessibility between experimental versus control cells. Experienced researchers will take 6 days to generate lentiviral stocks and transduce the target cells, a further 5 days to prepare the Illumina sequencing libraries and a few hours to perform the bioinformatic analysis.

INTRODUCTION

As alterations in chromatin structure play a pivotal role in the epigenetic regulation of gene expression, the ability to examine chromatin structure is critical to understanding the molecular basis of gene regulation. A variety of approaches have been developed for probing chromatin accessibility, including DNase-seq¹, MNase-seq², FAIRE-seq³, NA-seq⁴, RED-seq⁵ and ATAC-seq⁶, but, whilst these methodologies effectively detect regions of accessible, “open” chromatin, they are

less suited to examining changes at “closed” genomic regions that adopt a more compact chromatin structure.

Development of DIVA.

We have recently characterized the Human Silencing Hub (HUSH) complex, along with its effector proteins SETDB1 and MORC2, as an epigenetic regulator of heterochromatin marked by the repressive H3K9me3 modification⁷⁻⁹. Motivated by a desire to examine potential dysregulation of heterochromatin architecture upon loss of HUSH function, we developed Differential Viral Integration (DIVA), a genome-wide method to profile chromatin accessibility in cultured cells⁹. Conceptually DIVA is similar to ATAC-seq, except that, rather than using transposon integration sites to map accessible chromatin, DIVA maps the integration sites of exogenous lentiviruses (**Fig. 1**). Whilst the Tn5 transposase used in ATAC-Seq exhibits a strong preference for integration near transcriptional start sites¹⁰, HIV-1-derived lentiviral vectors utilized in DIVA also target gene bodies and intergenic regions¹¹, thereby allowing the DIVA technique to interrogate the architecture of a broader range of loci.

We employed DIVA to compare the chromatin architecture of wild-type and MORC2 knockout HeLa cells, identifying a range of loci showing increased viral accessibility upon loss of MORC2⁹. In principle a similar genetic approach could be used to examine the effect of ablation or overexpression of any epigenetic regulator on chromatin architecture. Similarly, if chemical inhibitors targeting an epigenetic modifier of interest are available, DIVA could be employed to compare chromatin architecture in vehicle versus drug treated cells.

Overview of the procedure.

DIVA involves three phases (**Fig. 2**). In phase one, lentiviral particles are packaged in HEK-293T cells and used to infect the target cells at high multiplicity-of-infection (MOI) (**Fig. 2a**). Following integration of the lentiviruses into the genome of the target cells, the goal of phase two is to map the sites of lentiviral integration. Genomic DNA is extracted and digested, adaptors are ligated onto the ends of the fragments, and then fragments containing proviral DNA are enriched through a linear PCR reaction. The linear PCR products are subsequently captured and subjected to exponential PCR to generate libraries for Illumina sequencing (**Fig. 2b**). Finally, in phase three, the Illumina sequencing data is used to generate a set of unique lentiviral integration sites detected in the population; comparing the pattern of lentiviral integration between control and experimental samples reveals loci exhibiting differential accessibility (**Fig. 2c**).

Comparison with other methods.

When compared to alternative genome-wide high-throughput chromatin accessibility assays, a major advantage of DIVA is its ability to assay both accessible and inaccessible chromatin simultaneously due to the ability of lentiviruses to integrate across gene bodies and intergenic regions (**Fig. 3a**). In contrast, ATAC-seq and MNase-seq, plus the related techniques DNase-seq, NA-seq⁴ and RED-seq⁵, all employ enzymes which exhibit a strong preference for activity at accessible

chromatin regions found at transcriptional start sites (TSSs), and hence are less suited to the study of “closed”, genomic regions.

DIVA requires a similar sequencing depth to that of other chromatin accessibility assays⁶. 30 million uniquely mapped reads per sample are sufficient to identify a large number of unique lentiviral integration sites and provide coverage across a large mammalian genome; downsampling analysis of our data⁹ suggests that around 10 million uniquely mapped reads per condition would be the minimum sequencing depth at which meaningful conclusions could be drawn from a DIVA experiment (**Fig. 3b-d**). As ATAC-seq involves the purchase of an expensive Illumina kit and often requires deeper sequencing due to mitochondrial DNA contamination, DIVA is likely to be more cost-effective than a typical ATAC-seq experiment. Indeed, as the majority of the reagents required are inexpensive and the depth of Illumina sequencing required is modest, the overall cost of a DIVA experiment is relatively low.

However, DIVA is also subject to some technical limitations. DIVA requires a large number of cells (~6 million) to ensure sufficient coverage of the genome, whereas ATAC-seq can reveal regions of open chromatin using just 500 cells¹⁰, and has even been adapted to analyze single cells¹². Furthermore, in contrast to the simple and rapid sample preparation required for ATAC-seq, DIVA involves a more complex and time-consuming library generation protocol which also relies on the availability of a BSL-2 tissue culture room. Thus, whilst ATAC-seq retains clear advantages over DIVA for researchers interested only in interrogating sites of open

chromatin, DIVA may be better suited to examining changes in heterochromatic regions, gene bodies and intergenic regions.

Limitations of the method.

There are a number of limitations to the current version of the DIVA protocol. First, sufficient coverage of a large mammalian genome can only be obtained if millions of independent lentiviral integration sites are mapped. This relies on achieving a high multiplicity of infection (MOI) when transducing the target cells with the lentiviral vector, which may not be possible in cell lines or primary cells which are difficult to transduce. Second, DIVA may not be suitable for comparisons between non-isogenic cell lines, where it is difficult to control for differences in the abundance or activity of other factors besides changes in chromatin architecture that may influence viral integration preferences. For example, differences in the abundance of the LEDGF/p75 protein, a cellular tethering factor for HIV-1-based lentiviruses¹³, could result in altered patterns of viral integration that are not related to altered chromatin accessibility. Finally, DIVA may be less well suited to measuring decreases in chromatin accessibility: whereas increased accessibility is revealed by an increase in the number of viral integration sites detected at a given locus, decreased accessibility must be inferred from the absence of detection. However, provided that sufficient numbers of lentiviral integration sites are mapped across such regions, in principle a reproducible decrease in the number of lentiviral integration sites as measured by DIVA should be indicative of decreased chromatin accessibility.

Applications.

A critical part of the DIVA protocol is the mapping of the sites of integration of exogenous lentiviral DNA on a large scale. The linear amplification-mediated polymerase chain reaction (LAM-PCR) approach described here was adapted from that originally devised by the Brummelkamp laboratory¹⁴, and has been used extensively to map the integration sites of gene-trap retroviral vectors to analyze forward genetic screens in haploid human cell lines¹⁵⁻¹⁷. This technique is likely to be applicable in other experimental settings, and, employing appropriately redesigned primers, the LAM-PCR protocol used for DIVA could be readily adapted to locate any exogenously integrated DNA in large mammalian genomes.

Experimental design

Cell lines. DIVA was developed as a comparative method to reflect differences in chromatin accessibility between a control and treated cell line. Therefore we recommend that DIVA be performed in isogenic cell lines (step 8), as differences in the abundance or activity of cellular factors that affect lentiviral integration preferences could lead to altered integration patterns that are not indicative of changes in chromatin structure.

Choice of lentiviral vector. In principle any standard lentiviral vector could be used to perform a DIVA experiment (step 3). There are two particular advantages to the pHRSIN-P_{SFFV}-mCherry vector that we employed⁹: (1) expression of the mCherry fluorescent protein allows the transduction of the target cells to be easily monitored

by flow cytometry, and (2) the arrangement of NlaIII restriction sites in the vector is favorable for efficient LAM-PCR amplification of virus-genome junctions (**Box 1**).

Choice of restriction enzyme. This protocol uses a single restriction enzyme, NlaIII, to digest the genomic DNA isolated from the transduced target cells (step 16). NlaIII cleaves at its 4 bp recognition site 5'-CATG-3', which occurs every ~200 bp in the genome. A proportion of all the lentiviral integration sites in the transduced population will lie either too close or too far from an NlaIII cleavage site to be mapped, however; therefore, to increase the coverage obtained, the genomic DNA could also be digested (in separate reactions) with additional '4-cutter' restriction enzymes that are not affected by CpG methylation, such as MseI or MluCI. Adapting the protocol for additional enzyme(s) would require the design of modified adaptor oligonucleotides with the appropriate cohesive ends; in these cases this would require the removal of the CATG overhang from the 3' end of the NlaIII-BSA oligonucleotide and addition of a TA (MseI) or an AATT (MluCI) overhang to the 5' end of the TSA oligonucleotide (Table 1). Care also needs to be taken to ensure that the lentiviral vector used will not itself be cleaved by the restriction enzyme in a manner that would prevent PCR amplification of the virus-genome junction (see **Box 1** for further details).

Production of high-titer lentivirus. In order to gain sufficient coverage of the genome, millions of lentiviral integration sites must be mapped in each cell type. It is therefore critical that high-titer lentivirus is produced (steps 1-7). We recommend using low passage HEK 293T cells and endotoxin-free DNA vectors. To optimize the conditions for virus production, a lentiviral vector encoding a fluorescent protein can be used. HEK 293T cell seeding density, DNA:TransIT ratio, the number of cells to be transduced and the amount of viral supernatant used for the transduction will all

affect the transduction efficiency, and hence all of these parameters may need to be optimized to achieve maximum transduction efficiency for DIVA. Enhanced transduction efficiency can also be achieved by concentrating the viral supernatant, for example using Lenti-X Concentrator (Clontech, cat. no. 631231), or by performing the transduction in the presence of Polybrene at a final concentration of 8 µg/ml.

Sequencing depth required. We used a single run of an Illumina HiSeq 2500 machine in rapid run mode to generate the sequencing data for our DIVA experiment in HeLa cells⁹, which yielded approximately 50 million uniquely mapped reads for each condition. An analysis of the sequencing depth requirements for DIVA are presented in **Fig. 3b**. This analysis suggests that 30-50 million uniquely mapped reads are sufficient to map the majority of unique lentiviral integration sites present in a DIVA library, while 10 million uniquely mapped reads per condition should be considered the minimum required to generate meaningful data. We recommend single-end 50 bp reads to minimize the cost of the experiment (step 43).

Data analysis. This protocol uses the Seqmonk (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>) application to analyze the Illumina sequencing data (steps 44-70). Seqmonk is a powerful genome browser application that allows both the visualization and quantitation of mapped sequence data. Its graphical user interface is ideal for biologists with limited programming experience; experienced bioinformaticians, however, could readily automate these steps using standard tools (for example, the BedTools suite¹⁸).

Level of expertise required. Successful DIVA experiments can be performed by competent graduate students with relevant experience in cell culture and molecular

biology techniques and a basic level of Unix knowledge. As the LAM-PCR protocol to map viral integration sites is relatively challenging, however, we would not recommend it be attempted by users inexperienced in molecular biology techniques. The steps described in this protocol do not require any non-standard reagents or equipment, and hence could be adopted by any laboratory with access to a BSL-2 tissue culture room and a next-generation sequencing facility. Previous experience working with lentiviral vectors would be beneficial but not essential, as the technique could be easily mastered after a short induction.

MATERIALS

REAGENTS

- Target cell type of choice plus appropriate culture media. We have used HeLa cells (ATCC CCL-2), cultured in RPMI 1640 (see Reagent Setup).

CRITICAL DIVA relies on a comparison between two experimental conditions, and therefore a control and a treated cell line is required. For example, we used DIVA to compare wild-type versus mutant HeLa cells⁹.

! CAUTION The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.

- HEK293T cells (ATCC CRL-1573)

! CAUTION The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma.

- RPMI 1640 (Thermo Fisher Scientific, cat. no. 11875093)
- DMEM (Thermo Fisher Scientific, cat. no. 11965092)
- Opti-MEM I Reduced Serum Media (Thermo Fisher Scientific, cat. no. 31985070)
- Fetal bovine serum (HyClone, cat. no. SH30071.03)
- Penicillin-streptomycin (Thermo Fisher Scientific, cat. no. 15140122)
- Trypsin-EDTA (0.25%) (Thermo Fisher Scientific, cat. no. 25200056)
- Lentiviral transfer vector (Box 1). We have used the pHRSIN-P_{SFFV}-mCherry vector, which encodes mCherry from the spleen focus-forming virus (SFFV) long terminal repeat (LTR) promoter, allowing transduced cells to be fluorescently marked (Addgene, cat no. 74670).
- Lentiviral packaging plasmids pMD2.G and psPAX2 (Addgene, cat. no. 12259 and 12260)

- TransIT-293 transfection reagent (Mirus, cat. no. MIR 2700)
- Polybrene (Sigma-Aldrich, cat. no. H9268)
- PBS (Thermo Fisher Scientific, cat. no. 10010023)
- Genra Puregene Cell Kit (Qiagen, cat. no. 158767)
- NlaIII (NEB, cat. no. R0125L)
- Ultrapure water (Thermo Fisher Scientific, cat. no. 10977015)
- Agencourt AMPure XP beads (Beckman Coulter, cat. no. A63880)
- Oligonucleotide adaptors (Sigma-Aldrich; see **Table 1** for oligonucleotide sequences and Reagent Setup for preparation).
- 10x NEB buffer 2 (NEB, cat. no. B7002S)
- T4 DNA ligase (NEB, cat. no. M0202S)
- Biotinylated linear PCR primer (Sigma-Aldrich; see **Table 1** for oligonucleotide sequences)
- Accuprime Taq DNA polymerase (Thermo Fisher Scientific, cat. no. 12339016)
- dNTPs (NEB, cat. no. N0447L)
- Streptavidin Dynabeads M-280 (Thermo Fisher Scientific, cat. no. 11205D)
- Ethanol, Absolute (200 Proof), Molecular Biology Grade (Thermo Fisher Scientific, cat. no. BP2818500)
- 5 M NaCl (Thermo Fisher Scientific, cat. no. 24740011)

- 1 M Tris-HCl pH 7.5 (Thermo Fisher Scientific, cat. no. 15567027)
- 0.5 M EDTA (Thermo Fisher Scientific, cat. no. AM9260G)
- Tween-20 (Sigma-Aldrich, cat. no. P9416)
- Exponential PCR primers (Sigma-Aldrich; see **Table 1** for oligonucleotide sequences)
- Bioanalyzer High Sensitivity DNA Kit (Agilent, cat. no. 5067-4626)

TABLE 1 | Oligonucleotides required for DIVA.

Oligo name	Sequence 5' – 3'	Purpose	Step used
TSA	CCTATAGTGAGTCGTATTATATGAGTAGTACCATGGGAAC	Top strand adaptor	23
NlaIII-BSA	GTTCCCATGGTACTACTCATATAATACGACTCACTATAGGCATG	Bottom strand adaptor	23
Bio-Pack_F	[Biotin]-CTTTCGCTTTCAAGTCCCTGTTCG	Linear PCR primer	25
P5-HIV1_F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCTACACG ACGCTCTCCGATCTGATCTGGTCTAACCAGAGAG	Exponential PCR	
P7- Adaptor_R	CAAGCAGAAGACGGCATACGAGAXXXXXXGTGACTGGAGTTC AGACGTGTGCTCTCCGATCTGTTCCCATGGTACTACTCATATA ATACG	primers (X=index bases)	34
ISP	CTTGTCTTCGTTGGGAGTGAATTAGCCCTTCCA	Illumina sequencing primer	43

EQUIPMENT

- Syringe (VWR, cat. no. 309646)
- 0.45 micron syringe filters (Corning, cat. no. 431220)
- 5 ml polypropylene round-bottom tube for flow cytometry (Corning, cat. no. 352002)
- 6-well cell culture plates (for example, Thermo Fisher Scientific cat. no. 140675)
- Cell culture incubator (37°C, 5% CO₂, 90% humidified) (for example, Thermo Fisher Scientific cat. no. 51026332)
- Cell culture hood (for example, Thermo Fisher Scientific cat. no. 51023607)
- Centrifuge (for example, Eppendorf cat. no. 51026332)
- Flow cytometer (for example, BD LSRFortessa, BD Biosciences cat. no. 649225)
- Thermocycler (for example, C1000 Touch Thermal Cycler, Bio-Rad cat. no. 1851196)
- Nanodrop spectrophotometer (for example, Thermo Fisher Scientific cat. no. ND-2000)
- Orbital shaker (for example, Eppendorf cat. no. 5382000023)
- Microcentrifuge (for example, Fisher Scientific cat. no. 05-400-006)
- 0.2 ml PCR tubes (for example, Fisher Scientific cat. no. Fisher Scientific cat. no. 14-222-252)
- 1.5 ml microcentrifuge tubes (for example, Eppendorf cat. no. 022363204)
- Magnetic rack (for example, Thermo Fisher Scientific cat. no. 15567027)
- Bioanalyzer (Agilent, cat. no. G2939BA)

- Computer running Unix/Linux distribution with at least 4 GB of RAM and the following software installed:
- Fastqc (<https://www.bioinformatics.babraham.ac.uk/projects/fastqc/>)
- Fastx toolkit (http://hannonlab.cshl.edu/fastx_toolkit/)
- Samtools (<http://www.htslib.org/>)
- Bowtie 2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>)
- Seqmonk (<https://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>)

REAGENT SETUP

Cell culture medium. To make up complete DMEM and RPMI cell culture medium, in a cell culture hood add 50 ml of FBS (10% vol/vol final) and 5 ml of Penicillin-Streptomycin solution to 500 ml of cell culture medium and mix thoroughly by inverting. Media should be stored at 4°C and discarded if unused after approximately one month.

70% (vol/vol) ethanol. Add 0.3 volumes of ultrapure water to 0.7 volumes of stock 100% ethanol solution and mix by inverting. The solution can be kept tightly capped at room temperature (20°C-25°C).

Anneal adaptors. Dilute the 10x NEB buffer 2 two-fold by adding an equal volume of ultrapure water. Resuspend the TSA (Top Strand Adaptor; Table 1) and NlaIII-BSA (Bottom Strand Adaptor; Table 1) oligonucleotides in 5x NEB buffer 2 at a final

concentration of 100 μ M. Then add 50 μ l of each oligonucleotide to a PCR tube, mix by pipetting, and anneal by placing the tube in a thermocycler and running the following program:

Step	Denature	Anneal	Hold
1	95°C, 5 min		
2		Slow cool (0.1°C/s) to 4°C	
3			4°C

Annealed adaptors can then be stored indefinitely at -20°C.

PBS + 0.5% (wt/vol) BSA. Weigh out 0.25 g BSA, dispense into a 50 ml plastic tube, and add 50 ml PBS. Mix gently by inverting. Place in a water bath at 37°C for 30 min to dissolve fully. This solution should be made fresh on each occasion.

20% (vol/vol) Tween-20. Add 4 volumes of distilled water to 1 volume of Tween-20 and mix gently by inverting until the solution is homogeneous. The solution can then be stored for months at room temperature when protected from light.

2x B&W buffer. In a 500 ml bottle, add 200 ml 5 M NaCl, 5 ml 1M Tris-HCl pH 7.5 and 1 ml 0.5 M EDTA. Make up to a total volume of 500 ml with distilled water and mix by inverting. The solution can be stored for months at room temperature.

1x B&W buffer. In a 500 ml bottle, add 250 ml 2x B&W buffer and 250 ml distilled water and mix by inverting. The solution can be stored for months at room temperature.

2x B&W buffer + 0.2% (vol/vol) Tween-20. Pipette 500 μ l of 20% Tween-20 into 50 ml 2x B&W buffer and mix by inverting. The solution can be stored for months at room temperature.

1x B&W buffer + 0.1% (vol/vol) Tween-20. Pipette 250 μ l of 20% (vol/vol) Tween-20 into 50 ml 1x B&W buffer and mix by inverting. The solution can be stored for months at room temperature.

Software setup. Follow the online installation instructions to install Fastqc, the fastx toolkit and Bowtie 2 programs onto the linux computer. These need to be added to PATH environment variable in order to successfully execute the commands in the protocol; detailed instructions on how to do this can be found in the online documentation for each tool. Seqmonk can be installed on a computer running any operating system.

PROCEDURE

Generation of lentivirus stocks. • TIMING 4 days

1. After thawing a frozen vial of HEK 293T cells, allow 2-3 days for the cells to recover. Then trypsinize the stocks of HEK 293T cells, and seed 6×10^5 cells in each well of a 6-well plate in complete DMEM growth media. Grow overnight (~16 h-20 h) to achieve 70% confluency.

The following morning, perform a triple transfection of the HEK 293T cells to generate lentivirus. For each well of HEK 293T cells to be transfected, perform the following procedure (Steps 2-6) [AU: Editor has changed text. Please check if it is ok.]

2. Pipette 200 μ l Opti-MEM into a sterile 1.5 ml microcentrifuge tube. Then add 7.5 μ l TransIT-293 reagent and mix thoroughly by pipetting. Incubate at room temperature for 5 min.
3. During the incubation, pipette 50 μ l Opti-MEM into a second sterile 1.5 ml microcentrifuge tube. Then add the following amounts of plasmid DNA: 1 μ g lentiviral transfer vector, 0.7 μ g psPAX2 and 0.3 μ g pMD2.G. Mix by pipetting.
4. Add the TransIT-293/Opti-MEM mix from Step 2 to the DNA/Opti-MEM mix from Step 3 and mix gently but thoroughly by pipetting. Incubate at room temperature for 20 min.
5. During the incubation, aspirate media from HEK 293T cells from Step 1 and replace with 2 ml fresh DMEM per well. Pipette slowly onto the side of the well to minimize cell detachment.

6. Add the 250 μ l of DNA/TransIT-293/Opti-MEM mix from Step 4 dropwise to the HEK 293T cells. Rock the plate gently back and forth to evenly distribute transfection complexes. Return to 37°C incubator for 48 hours.

! CAUTION Users should familiarize themselves with their institutional guidelines for handling lentivirus. If appropriate, cells should be transferred to a BSL-2 tissue culture room at this point.

7. *Harvest viral supernatant.* Pipette media off of the cells and into a syringe attached to a 0.45 μ m filter. Insert the plunger and gently force the viral supernatant through the filter into a sterile collection tube.

■ **PAUSE POINT** Viral supernatants can be aliquoted in 1.5 ml microcentrifuge tubes and stored at -80°C for several months with minimal loss of activity.

Lentiviral Transduction. • TIMING 3 days

8. *Seed target cells for transduction.* For HeLa cells, seed 1×10^6 cells in each well of a 6-well plate in 2 ml complete RPMI growth medium (a total of 6×10^6 cells in 12 ml growth medium). Return to the 37°C incubator for 5 hours.

▲ **CRITICAL STEP** Both control and treated cells need be seeded to allow for parallel transduction. Continue to culture some wild-type target cells, as these will be required as a control sample when estimating transduction efficiency at Step 14.

9. *Transduce target cells with lentivirus.* Defrost the viral supernatant from Step 7 and add to target cells. For HeLa cells aspirate the 2 ml of growth media and replace it with 1.5 ml of viral supernatant.

10. (OPTIONAL) To maximize transduction efficiency, Polybrene (to a final concentration of 8 µg/ml) can be added with the virus and/or cells can be spun following addition of virus (800 x g, 60 min, 37°C).
11. Three hours later, aspirate and discard viral supernatant from the cells and replace with complete growth media. Return to 37°C incubator overnight.
12. *Repeat transduction of target cells.* To increase the overall multiplicity of infection achieved, 24 h after the initial transduction repeat the transduction process described in steps 9 - 11.
13. *Harvest transduced cells.* 24 h after the second transduction, trypsinize cells using 0.5 ml of Trypsin-EDTA per well, quench with an equal amount of RPMI media, pool and pipette into 50 ml conical tube. Pipette a few drops of the cell suspension into a round-bottom 5 ml polypropylene tube for flow cytometry analysis⁹ (Step 14). Then pellet the remainder by centrifugation (500 x g, 5 min, 4°C), wash the cells once by resuspending the pellet in 50 ml PBS and repeating the centrifugation (500 x g, 5 min, 4°C). Finally remove the supernatant and resuspend the pellet in 3 ml Cell Lysis Buffer from the Gentra Puregene Cell Kit.
 - **PAUSE POINT** According to the manufacturer, cell lysates can be stably stored at room temperature for at least two years.
14. *Verify transduction efficiency by flow cytometry.* If the lentiviral vector used expresses a fluorescent marker, then the transduction efficiency of the target cells can be assayed by flow cytometry. Setup the instrument using untransduced target cells, and then analyze the aliquot of transduced cells from Step 13. If high-level transduction has been achieved, then close to 100% of the cells should be positive for expression of the fluorescent protein.

? TROUBLESHOOTING

LAM-PCR amplification of virus-genome junctions. • TIMING 5 days

15. *Extract genomic DNA from cell lysates.* Follow the Genra Puregene Cell Kit manual protocol for genomic DNA extraction from cultured cell lysates from Step 13. Resuspend the final DNA pellet in 500 μl DNA hydration solution.

■ **PAUSE POINT** Genomic DNA can be stored stably for years at 4°C.

16. *Digest genomic DNA with NlaIII.* Measure the DNA concentration using a Nanodrop spectrophotometer. Take half of the DNA out for restriction enzyme digestion and keep the other half at 4°C as a back-up. For each 10 μg of DNA to be digested, set up the following reaction in a 1.5 ml microcentrifuge tube:

Component (add in this order)	Amount	Final concentration
Genomic DNA	10 μg	0.1 $\mu\text{g } \mu\text{l}^{-1}$
10x NEB CutSmart buffer	10 μl	1x
ddH ₂ O	Up to 90 μl	
NlaIII (10 units μl^{-1})	10 μl	1 unit μl^{-1}
Total volume	100 μl	

For example, if in total 200 μg of DNA was extracted, digest 10 μg of DNA in 10 separate digestion reactions. Incubate all tubes at 37°C overnight.

? TROUBLESHOOTING

17. *Purify and concentrate the digested DNA using a spin column.* Follow the Qiagen PCR purification kit protocol. To avoid exceeding the binding capacity

of the columns, perform a separate purification for each 10 µg of digested DNA. Elute the DNA from each column in 30 µl of ultrapure water, and pool all of the eluted material in a fresh 1.5 ml microcentrifuge tube. Alternatively, a standard phenol/chloroform extraction followed by ethanol precipitation¹⁹ could be used to purify all of the DNA in a single process.

■ **PAUSE POINT** Digested DNA can be stored at -20°C.

18. *Size selection to enrich for large DNA fragments that contain virus-genome junctions.* Vortex the stock solution of Agencourt AMPure XP beads thoroughly. Add 0.5 volumes of beads to the pooled digested DNA solution from Step 17 and mix well by pipetting. Incubate for 5 min at room temperature.
19. Place the tube on the magnetic stand and wait until the beads are collected by the magnet on the back of the tube. Then remove the supernatant and discard.
20. Add 500 µl of 70% (vol/vol) ethanol to wash the beads. Aspirate, and discard, and repeat the wash step. Allow the beads to air-dry for a few minutes.
21. Remove the tube from the magnet. Add 100 µl of ultrapure water and pipette to resuspend the beads. Incubate for 5 min at room temperature.
22. Place the tube back on the magnet. Wait for the magnet to collect the beads on the back of the tube (~ 1 min). Collect the supernatant containing the eluted DNA into a fresh 1.5 ml microcentrifuge tube, and measure the concentration of the sample using a nanodrop spectrophotometer. Typical recovery would be about one-third of input DNA.

23. *Ligate annealed adaptors to the size-selected DNA fragments.* For each 3 µg of DNA, set up the following reaction in a 0.2 ml PCR tube:

Component (add in this order)	Amount	Final concentration
Size-selected, NlaIII-digested genomic DNA	3 µg	0.06 µg µl ⁻¹
Annealed adaptor (50 µM)	5 µl	5 µM
10x NEB T4 Ligase buffer	5 µl	1x
ddH ₂ O	Up to 48 µl	
T4 DNA ligase (400 units µl ⁻¹)	2 µl	16 units µl ⁻¹
Total volume	50 µl	

Incubate at 16°C overnight in a PCR instrument.

24. *Clean-up the ligated DNA using Agencourt Ampure XP beads.* Pool all of the ligation reactions in a 1.5 ml microcentrifuge tube, add 0.7 volumes of beads and follow the procedure described in steps 18-22.

▲ **CRITICAL STEP** This clean-up step removes un-ligated adaptors from the sample, which otherwise decrease the efficiency of subsequent PCR steps.

■ **PAUSE POINT** Ligated DNA can be stored indefinitely at -20°C.

25. *Amplify virus-genome junctions by linear PCR.* Measure the concentration of adaptor-ligated DNA using a Nanodrop spectrophotometer; the total amount of DNA should be roughly equivalent to the total amount of input DNA. To amplify the material, for each 2 µg of adaptor-ligated DNA, make up the following mix:

Component (add in this order)	Amount	Final concentration
Adaptor-ligated genomic DNA	2 µg	0.04 µg µl ⁻¹
Accuprime buffer II (10x)	5 µl	1x
Bio-Pack_F (Table 1; 0.25 µM)	1 µl	5 nM
ddH ₂ O	Up to 49.6 µl	
Accuprime Taq polymerase (2 U µl ⁻¹)	0.4 µl	0.016 U µl ⁻¹
Total volume	50 µl	

Run the following thermocycler program:

Step	Denature	Anneal	Extend	Hold
1	94°C, 5 min			
2-200	94°C, 60 s	55°C, 45 s	68°C, 90 s	
201			68°C, 7 min	
202				4°C

■ **PAUSE POINT** PCR products can be stored indefinitely at -20°C.

? TROUBLESHOOTING

26. *Capture biotinylated PCR products onto streptavidin-coated magnetic beads.*

Pool all of the linear PCR reactions into a single 1.5 ml microcentrifuge tube.

Estimate the total volume using a pipette.

27. *Prepare streptavidin beads.* Vortex the tube of beads thoroughly, and then

pipette 5 µl of beads per PCR reaction (ie. 40 µl beads total for 8 PCR

- reactions) into a 1.5 ml microcentrifuge tube. Block the beads by collecting the beads on the magnet for approximately 1 min. Aspirate the supernatant, discard and resuspending the beads in PBS + 0.5% (wt/vol) BSA.
28. Incubate for 30 min at room temperature, inverting the tube occasionally to prevent the beads from sinking.
29. Collect the beads on the magnet for approximately 1 minute, wash once with 1 ml 1 x B&W buffer, and discard. Then resuspend the beads in 2 x B&W buffer + 0.2% (vol/vol) Tween-20; the volume of buffer should be half that of the volume of the pooled PCR products
30. Add the pooled PCR products from Step 26 to the beads from Step 29, and mix thoroughly by pipetting. Incubate for 2 h at room temperature, shaking at 1100 rpm in an orbital shaker to prevent the beads from sinking.
31. To wash unbound material from the streptavidin beads, place the tube on the magnet to collect the beads, and then remove the supernatant. (Store the initial supernatant at -20°C as a back-up; this could be re-amplified in a further linear PCR reaction later if necessary).
32. Add 1 ml of 1x B&W + 0.1% (vol/vol) Tween-20 to the beads, resuspend by pipetting, and return the tube to the orbital shaker at 1100 rpm for 10 min.
33. Wash the beads thoroughly by repeating steps 31 and 32 ten times.
- **PAUSE POINT** Beads can be stored in 1x B&W + 0.1% Tween-20 at 4°C for several weeks.
34. *Generate Illumina sequencing libraries through an on-bead exponential PCR reaction.* Prepare the PCR mastermix on ice. In total, perform the same number of independent exponential PCR reactions as the number of linear

PCR reactions performed in step 25. The following mix details the components for one reaction:

Component (add in this order)	Amount (μl)	Final concentration
Accuprime buffer II (10x)	5	1x
P5-HIV1_F (Table 1; 100 μM)	0.25	0.5 μM
P7-Adaptor_R (Table 1; 100 μM)	0.25	0.5 μM
ddH ₂ O	Up to 49.4	
Accuprime Taq polymerase (2 U μl^{-1})	0.6	0.024 U μl^{-1}
Total volume	50	

For example, if 8 separate linear PCR reactions were performed in step 25, multiply each of the values above by 8 to make up an 8x mastermix of 400 μl total volume.

35. Collect the streptavidin beads from Step 33 on the magnet, and perform one final wash with PBS. Then collect the beads on the magnet, aspirate the PBS, and resuspend the beads in all of the PCR mastermix. Distribute 50 μl of the bead/PCR mix into each tube of an 8-strip PCR tube.

36. Run the following thermocycler program:

Step	Denature	Anneal	Extend	Hold
1	94°C, 2 min			
2-11	94°C, 30 s	55°C, 30 s	68°C, 90 s	
12			68°C, 7 min	
13				4°C

37. Carefully remove the tubes from the PCR machine and place on ice. Without disturbing the pellet of beads, remove 15 µl of the supernatant from each of the reactions and pool into a separate 1.5 ml tube kept on ice. This is the 10 cycle sample.

38. Return the remaining PCR reaction mix to the thermocycler for a further four cycles of PCR:

Step	Denature	Anneal	Extend	Hold
1	94°C, 2 min			
2-5	94°C, 30 s	55°C, 30 s	68°C, 90 s	
6			68°C, 7 min	
7				4°C

39. Again, carefully remove 15 µl of the supernatant from each of the reactions and pool into a separate 1.5 ml tube kept on ice. This is the 14 cycle sample.

Then return the tubes to the thermocycler for a final 4 cycles of PCR:

Step	Denature	Anneal	Extend	Hold
1	94°C, 2 min			
2-5	94°C, 30 s	55°C, 30 s	68°C, 90 s	
6			68°C, 7 min	
7				4°C

40. Remove all of the remaining volume (including the beads) from the PCR reactions and pool into a separate 1.5 ml tube kept on ice. To remove the

beads, place the tube on the magnet for 2 min and transfer the supernatant into a fresh tube. This is the 18 cycle sample.

41. *Perform a final clean-up of the Illumina libraries.* For each of the samples (10 cycles, 14 cycles and 18 cycles), perform an Agencourt AMPure XP bead clean-up as described in steps 18-22, using 0.8 volumes of beads. Elute in 20 µl of ultrapure water.

■ **PAUSE POINT** Sequencing libraries can be stored indefinitely at -20°C.

42. *Validate the Illumina sequencing libraries using an Agilent Bioanalyzer.* Run 1 µl of each of the sequencing libraries on a DNA High Sensitivity Chip. The library electropherogram should reveal a broad size distribution between ~240 bp and ~600 bp, with minimal adaptor dimer contamination at ~120 bp.

? TROUBLESHOOTING

43. *Submit optimal sample for Illumina sequencing.* Using the Bioanalyzer software, quantify the concentration of the DNA of the correct size (between ~240 bp and ~600 bp). Choose the sample which has been subjected to the lowest number of PCR cycles, yet still exceeds the minimum amount required by your sequencing facility. Submit the final libraries for Illumina sequencing, along with a custom sequencing primer that binds to the extreme end of the lentiviral 5'LTR (ISP, in our case; see Table 1). We generated 50 bp single-end sequence reads using an Illumina HiSeq2500 machine.

? TROUBLESHOOTING

■ **PAUSE POINT** Analysis of sequencing data can be performed at any time.

Mapping of Illumina reads and identification of genomic loci exhibiting differential accessibility. • **TIMING** <1 day

44. Copy the output fastq sequencing files generated by the Illumina sequencing run to the hard drive of the Linux computer.

45. Run fastqc to check the run. Navigate to the folder containing the sequencing files, open a terminal window, and type the following command after the \$ sign:

```
$ fastqc inputfile.fastq
```

(where inputfile is the name of the file to be analyzed).

? TROUBLESHOOTING

46. Perform an adaptor trim to remove adaptor sequences from the 3' ends of reads. Type the following command:

```
$ fastx_clipper -a CCTATAGTGAGTCGTA -l 18 -v -i inputfile.fastq -o  
outputfile_trimmed.fastq
```

(where inputfile is the name of the file to be analyzed and outputfile_trimmed is the desired name of the output fastq file containing the trimmed reads).

47. Map to the human genome using Bowtie 2. Type the following command:

```
$ bowtie2 -t -x GRCh38 -U inputfile_trimmed -S outputfile.sam
```

(where GRCh38 is the basename of the Bowtie2 index, inputfile_trimmed is the file of trimmed reads from step 46, and outputfile is the desired name of the output sam file containing the mapped reads).

48. To save disk space, compress the output sam file to a bam file by typing:

```
$ samtools view -b outputfile.sam > outputfile.bam
```

49. *Import the resulting bam file into SeqMonk.* The following can be performed on any operating system. Navigate to the folder containing the unzipped

SeqMonk installation files, and launch the program by running the SeqMonk executable file.

50. From the File menu, select New Project. From the resulting dialog box, click the 'Import Genome From Server' button and select the genome build to which the data was mapped (GRCh38 in this example). Then click on 'Start New Project' to initialize the project.

51. To load the data, from the 'File' tab select 'Import Data' and then click on 'BAM/SAM'. Select the location of the relevant sam or bam file(s) from Step 47 or Step 48 on your computer, and then click 'OK'.

52. Specify import options: set 'Remove duplicate reads' to 'Yes' and 'Min mapping quality' to '20'. Click 'OK'.

53. *Collapse the sequencing reads into a set of unique sites of lentiviral integration.* Using Seqmonk's import function, from the 'File' tab, choose 'Import Data' and then 'Visible Data Stores'.

54. Set 'Extend reads by (bp)' to '-1000', and then click 'Import'. This creates a new data track for each sample where only the first base of each read (i.e. the point of viral integration) is displayed.

55. Remove the original data tracks: click on the 'Change Data Tracks' icon, highlight the original data files in the 'Displayed Tracks' window, and click 'Remove' button followed by the 'OK button'.

56. *Segment the genome into a set of equally-sized intervals and quantitate the number of unique lentiviral integration sites in each window.* From the 'Data'

- tab, choose 'Define Probes' and highlight 'Running Window Generator' from the 'Probe Generator Options' list.
57. Set 'Probe Size (bp)' to '10000' and the 'Step Size (bp)' to '5000'. Then click the 'Create Probes' button. This divides the genome into a set of 10 kb intervals which overlap by 5 kb.
58. To count the number of integration sites in each window, highlight 'Read Count Quantitation' from the 'Quantitation Options' list. Ensure that the 'Correct for total read count' and 'Log Transform Count' boxes are checked, and click 'Quantitate'.
59. *Visualize data as a scatter plot and allow normalization.* From the 'Plots' tab, select 'Scatter Plot'.
60. Select the two relevant datasets to compare, for example wild-type on the x-axis and mutant on the y-axis. Each dot on the resulting plot represents the quantitation for one 10 kb genomic interval. Dots lying on the diagonal represent intervals containing equal frequencies of viral integration in both conditions.
61. The majority of the dots should lie along the diagonal. Because the data has been collapsed into a set of unique integration sites, skewing of the data away from the diagonal may be observed at the top end of the distribution. If this is observed, normalize the data by clicking on the 'Data' tab, select 'Quantitate Existing Probes', and then highlight 'Match Distribution Quantitation' from the list of 'Quantitation Options'. Select the two datasets to normalize on the 'Data Stores to Normalise' list on the right, and then click the 'Quantitate' button.

Re-display the scatter plot to verify that the normalization has been successfully applied.

62. *Identify loci exhibiting significant changes in lentiviral accessibility.* From the 'Filtering' tab, choose 'Filter by Statistical Test', then 'Intensity Difference' and 'Individual Probes'.

63. In the resulting dialog box, select the relevant datasets to compare. The control (eg. wild-type) dataset should be selected in the first list called 'From Data Store / Group', and the experimental (eg. mutant) dataset should be selected in the second list called 'To Data Store / Group'. Maintain the default parameters on the right, and click 'Run Filter'.

64. To visualize the loci exhibiting significant changes in lentiviral accessibility re-create the scatter plot by following the instructions in steps 59 and 60.

65. Then click on the 'Highlight Sublists' button in the top right corner. From the 'Available Lists' list, highlight the entry that begins 'Intensity Difference $p < 0.05$ ' and click 'Add' followed by 'OK'. The scatter plot now shows loci exhibiting significantly altered accessibility in the experimental versus the control datasets colored in blue.

66. Export the scatter plot by clicking on the 'Save Image' button.

67. Finally, a summary of the data can be exported for further analysis. On the list on the left-hand side of the home screen, under 'Probe Lists' click on the icon next to 'All Probes' to expand the display and highlight the list of significant loci by clicking on it.

68. Under the 'Reports' tab, select 'Annotated Probe Report'.
69. In the resulting dialog box, under 'Annotate with' select 'overlapping' from the left-hand menu and 'gene' from the right-hand menu, and then click 'OK'.
70. The 'Annotated Probe Report' can then be saved and further manipulated in standard programs such as Microsoft Excel.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
14	Low efficiency transduction observed by flow cytometry	Inefficient viral packaging in HEK 293T cells	Thaw a fresh aliquot of HEK 293T cells and verify that they are free of <i>Mycoplasma</i> infection. Passage the cells every 2 days and do not allow them to become over-confluent. Prepare fresh plasmid stocks of the lentiviral vectors.
14	Low efficiency transduction observed by flow cytometry	Target cells refractory to lentiviral transduction	Verify that high-titre lentiviral stocks have been generated by performing a parallel transduction of a cell line that is readily transducible, for example HEK 293T or HeLa.
16	Inefficient digestion of genomic DNA	Addition of too much DNA to the digestion reaction	If the genomic DNA samples are “stringy”, dilute them further in DNA hydration solution prior to measuring their concentration. Take duplicate nanodrop readings of the genomic DNA samples and

			verify that the readings are similar.
25	Low sample volumes after the linear PCR	Evaporation during the PCR reaction	Avoid placing samples in the outer tubes of the PCR strip. Instead add 200 µl water to outer tubes. Ensure PCR strips are tightly capped.
42	No peaks detected in the Bioanalyzer trace	Could result from failure at any step of LAM-PCR protocol	Efficient NlaIII digestion of genomic DNA can be verified by agarose gel electrophoresis of the digested DNA; this should generate a smear with a mean fragment size ~200bp. Efficient adaptor ligation can be verified by a simple PCR reaction using the adaptor-ligated DNA (from Step 24) as the template with a single primer that will anneal to the adaptor ligated on to both ends of the fragments (eg. P7-Adaptor_R; Table 1). Using standard cycling conditions (eg. 30 cycles of those described in Step 36), the product should

appear as a bright smear should when observed by agarose gel electrophoresis. If both of these steps have worked correctly, then repeat the LAM-PCR, bead annealing and exponential PCR steps of the protocol.

43 Failed Illumina sequencing run Low complexity

Correct library preparation should result in a complex library with approximately even representation of each base across the read. Problems with low complexity indicate an excessive amount of adaptor artefacts. In this case it would be advisable to repeat the LAM-PCR procedure, verifying efficient NlaIII digestion by agarose gel electrophoresis and adaptor ligation by PCR prior to the LAM-PCR steps.

- **TIMING**

Steps 1-7, generation of lentivirus stocks: 3 days

Steps 8-14, lentiviral transduction of target cells: 3 days

Steps 15-43, LAM-PCR amplification of virus-genome junctions: 5 days

Steps 44-70, Bioinformatic analysis of Illumina sequencing data: <1 day

ANTICIPATED RESULTS

Transduction of target cells. We recommend using a lentiviral vector which encodes a fluorescent marker, which enables the transduction of the target cells to be assayed by flow cytometry (Step 14). If high efficiency transduction of the target cells has been achieved, then ~100% of the cells should be positive. An example of successful transduction of wild-type and mutant HeLa cells with the pHRSIN-P_{SFFV}-mCherry lentiviral vector is shown in **Fig. 4a**.

LAM-PCR preparation of Illumina sequencing libraries. For HeLa cells, in step 15 the total amount of DNA extracted from the entire 6-well plate of transduced cells should be approximately 200 µg. Following successful digestion with NlaIII, the vast majority of the genomic DNA should be fragmented to < 1 kb and can be visualized as a smear by agarose gel electrophoresis (**Fig. 4b**). After size-selection for large DNA fragments in steps 18-22, the amount of DNA recovered should be approximately 20-30% of the total input digested DNA. Successful LAM-PCR amplification of virus-genome junction can be assayed through Agilent Bioanalyzer analysis of the final Illumina library; we would expect the Bioanalyzer trace in step 42 to look similar to that shown in **Fig. 4c**, with the trace starting at ~220 bp, peaking at ~240 bp, and then gradually tailing off by ~600 bp.

Bioinformatic analysis of Illumina sequencing data. Any abnormalities with the sequencing data should be revealed by Fastqc at step 45. Efficient amplification of virus-genome junctions should result in approximately even representation of each of the four bases across the sequence reads; strong bias in base representation is indicative of adaptor artefacts.

Based on downsampling analysis of our DIVA Illumina sequencing data⁹ (Figure 3), we found that approximately 10 million uniquely mapped reads per condition would have been sufficient to draw meaningful conclusions from the experiment. Therefore, for high coverage of a large mammalian genome, we recommend that >10 million uniquely mapped reads in each of the control and treated samples should be obtained at steps 43-47. Useful data may still be derived from samples with lower coverage, however, although an increased number of loci exhibiting a significant change in accessibility would likely be uncovered as more uniquely mapped reads are generated (Figure 3d).

The overall aim of this protocol is to identify genomic loci exhibiting altered chromatin accessibility between control and treated cells. A scatterplot (steps 59 and 60) summarizing the results of a DIVA experiment is shown in **Figure 4d**. At the completion of a successful experiment the researcher should have derived a list of genomic sites (tabulated in steps 67-70) which show significantly altered accessibility between the two cell types.

REFERENCES

1. Boyle, A. P. *et al.* High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–22 (2008).
2. Schones, D. E. *et al.* Dynamic Regulation of Nucleosome Positioning in the Human Genome. *Cell* **132**, 887–898 (2008).
3. Giresi, P. G., Kim, J., McDaniell, R. M., Iyer, V. R. & Lieb, J. D. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* **17**, 877–85 (2007).
4. Gargiulo, G. *et al.* NA-Seq: a discovery tool for the analysis of chromatin structure and dynamics during differentiation. *Dev. Cell* **16**, 466–81 (2009).
5. Chen, P. B., Zhu, L. J., Hainer, S. J., McCannell, K. N. & Fazio, T. G. Unbiased chromatin accessibility profiling by RED-seq uncovers unique features of nucleosome variants in vivo. *BMC Genomics* **15**, 1104 (2014).
6. Tsompana, M. & Buck, M. J. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* **7**, 33 (2014).
7. Tchasovnikarova, I. A. *et al.* Epigenetic silencing by the HUSH complex mediates position-effect variegation in human cells. *Science (80-)*. **348**, 1481–5 (2015).
8. Timms, R. T., Tchasovnikarova, I. A. & Lehner, P. J. Position-effect variegation revisited: HUSHing up heterochromatin in human cells. *Bioessays* **38**, 333–43 (2016).
9. Tchasovnikarova, I. A. *et al.* Hyperactivation of HUSH complex function by Charcot–Marie–Tooth disease mutation in MORC2. *Nat. Genet.* **49**, 1035–1044 (2017).
10. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–8 (2013).
11. Kvaratskhelia, M., Sharma, A., Larue, R. C., Serrao, E. & Engelman, A. Molecular mechanisms of retroviral integration site selection. *Nucleic Acids Res.* **42**, 10209–25 (2014).
12. Buenrostro, J. D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486–490 (2015).
13. Debyser, Z., Christ, F., De Rijck, J. & Gijssbers, R. Host factors for retroviral

- integration site selection. *Trends Biochem. Sci.* **40**, 108–16 (2015).
14. Carette, J. E. *et al.* Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat. Biotechnol.* **29**, 542–6 (2011).
 15. Jae, L. T. *et al.* Virus entry. Lassa virus entry requires a trigger-induced receptor switch. *Science* **344**, 1506–10 (2014).
 16. Blomen, V. A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science (80-.)*. **350**, 1092–6 (2015).
 17. Timms, R. T. *et al.* Genetic dissection of mammalian ERAD through comparative haploid and CRISPR forward genetic screens. *Nat. Commun.* **7**, 11786 (2016).
 18. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).
 19. Sambrook, J. & Russell, D. W. Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc.* **2006**, pdb.prot4455 (2006).
 20. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330 (2015).

Related links**Key reference using this protocol**

Tchasovnikarova, I. et al. Nat. Genet. 49, 1035–1044 (2017) <https://doi.org/10.1038/ng.3878>

ACKNOWLEDGEMENTS

We thank S. Andrews for assistance with data analysis using SeqMonk. This work was supported by the Wellcome Trust, through a Principal Research Fellowship to P.J.L. (101835/Z/13/Z), a Sir Henry Wellcome Fellowship to R.T.T. (201387/Z/16/Z) and a Ph.D studentship to I.A.T.. I.A.T. is a Damon Runyon Fellow supported by the Damon Runyon Cancer Research Foundation (DRG-2277-16). The CIMR is in receipt of a Wellcome Trust strategic award.

AUTHOR CONTRIBUTIONS

R.T.T., I.A.T. and P.J.L. conceived the method. I.A.T. and R.T.T. performed all of the experiments, and, together with P.J.L., analysed the data and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

FIGURE LEGENDS

Figure 1 | Conceptual overview of DIVA. DIVA examines the integration preferences of exogenous lentiviruses to infer changes in chromatin accessibility between control (wild-type) and treated (mutant) cells.

Figure 2 | Workflow of a DIVA experiment.

(a) Lentiviral transduction of target cells. Lentivirus is generated through the triple transfection of HEK-293T cells, and the viral supernatant is used to transduce the matched pair of control and treated target cells.

(b) Mapping of lentiviral integration sites. Genomic DNA harvested from the transduced cells is enzymatically digested and adaptor ligated, followed by linear PCR amplification of virus-genome junctions. Biotinylated linear PCR products are captured on streptavidin beads which are then used as the template for a final exponential PCR to generate a library for Illumina sequencing.

(c) Bioinformatic analysis of Illumina sequence data. Illumina reads are mapped to the reference genome to define a set of unique lentiviral integration sites in each cell type (control vs treated); comparing the pattern of integration sites in control (wild-type) versus treated (mutant) cells reveals genomic loci exhibiting altered chromatin architecture.

Figure 3 | Genome coverage and sequencing depth required for DIVA.

(a) Comparison of the genomic coverage obtained with DIVA and ATAC-seq. The bar chart indicates the percentage of either ATAC-seq peaks (blue) or unique lentiviral integration sites identified by DIVA (orange) that occur across 15 types of chromatin. Compared to ATAC-seq, lentiviral integration sites are depleted from transcriptional start sites and enhancers, and are instead enriched in gene bodies and intergenic regions. ATAC-seq peaks were from¹⁰, DIVA lentiviral integration sites were from⁹, and the chromatin state classifications were from²⁰.

(b-d) Sequencing depth required for DIVA. **(b)** Downsampling analysis of Illumina sequencing data for wild-type HeLa cells⁹. Whilst the proportion of uniquely mapped reads scales linearly with the total number of sequencing reads generated, the number of unique integration sites detected plateaus rapidly. Thus, further sequencing of these DIVA libraries would have yielded few additional unique integration sites. **(c)** The maximal correlation between wild-type and mutant samples was essentially achieved with 10 million uniquely mapped reads per condition, and did not improve further with the addition of more reads. **(d)** Quantification of the number of genomic loci exhibiting a significant change in accessibility as a function of sequencing depth. The number of significant loci continues to increase as the number of mapped reads increases; in our experiment, however, similar biological

conclusions could have been drawn with 10 million uniquely mapped reads per cell type.

Figure 4 | Anticipated results of a DIVA experiment.

(a) Successful lentiviral transduction of target cells. Transduction of wild-type and mutant cell lines with high-titer virus should result in ~100% infection; this can be easily monitored by flow cytometry if the virus used expresses a fluorescent protein.

(b) Efficient digestion of genomic DNA with NlaIII as visualized by 2% agarose gel electrophoresis. Molecular weight markers are shown in lanes 1, 3 and 5. A broad smear should be observed with fragments ranging from ~100 bp to ~2 kb in both cell types.

(c) Validation of successful Illumina library generation. Using the pHRSIN-P_{SFFV}-mCherry lentiviral vector with NlaIII digestion, Agilent Bioanalyzer analysis of the final exponential PCR product should reveal a trace starting at ~220 bp, peaking at ~240 bp, and then gradually tailing off by ~600 bp. A similar profile should be observed in both cell types.

(d,e) Results of a successful DIVA experiment. Each point on the scatterplot **(d)** indicates the correlation between the number of unique lentiviral integrations detected in the wild-type cells versus the mutant cells for a given 10 kb genomic locus. Loci exhibiting a significant change in accessibility between the wild-type and

mutant samples are highlighted in blue. Raw data for one locus of interest (*ZFN772*) is depicted in **(e)**, comparing the coverage achieved using DIVA and ATAC-seq. ATAC-seq transposon integrations are concentrated at the transcriptional start site, whereas DIVA lentiviral integration sites are found predominantly across the body of the gene. A greater number of unique integration sites are detected across the body of the gene in the mutant sample, indicating an increase in chromatin accessibility in the mutant cell line.

Box 1 | Size-selection of large DNA fragments containing virus-genome junctions.

During steps 18-22 of the protocol, a size-selection is performed to enrich for large DNA fragments following *Nla*III digestion of the genomic DNA. The diagram (inset) shows the location of *Nla*III recognition sites in the pHRSIN- P_{SFFV} -mCherry lentiviral vector when integrated into genome, plus the primer binding sites for PCR amplification of the virus-genome junction. The distribution of *Nla*III sites is such that a genomic DNA fragment containing the 5'LTR must be at least ~1.8 kb; therefore, selecting for large DNA fragments enriches for the fragments of interest that contain proviral DNA. The distribution of restriction sites in the proviral DNA is therefore an important consideration should the protocol be adapted to use alternative lentiviral vectors or restriction enzymes. Abbreviations are as follows: LTR, long terminal repeat; WPRE, Woodchuck Posttranscriptional Regulatory Element; P_{SFFV} , spleen focus-forming virus promoter.

Figure 1

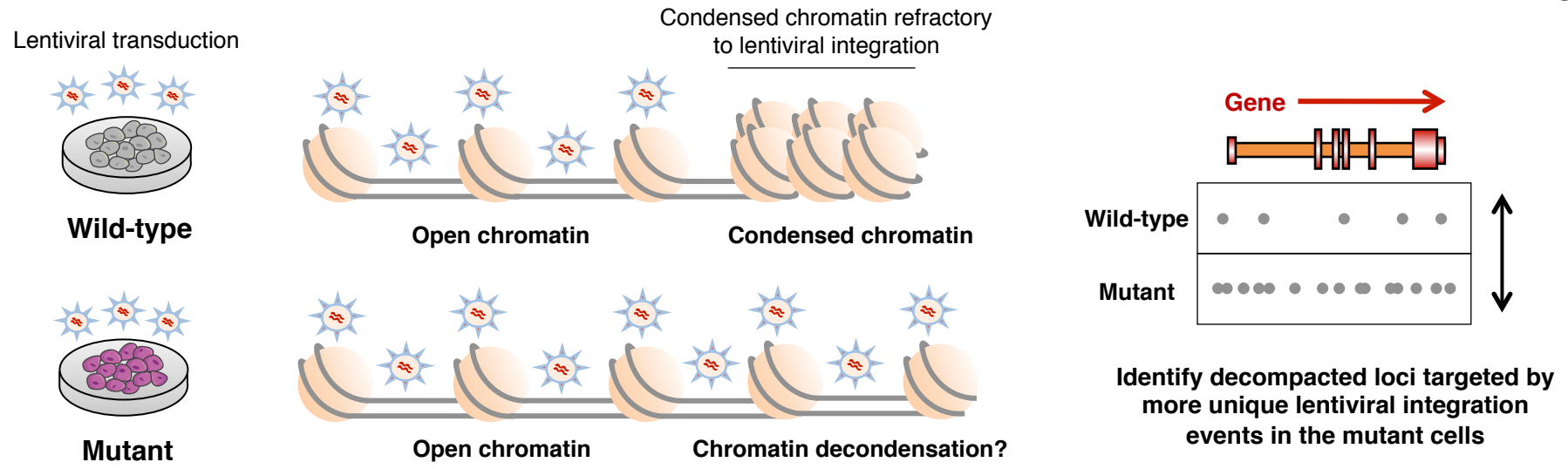


Figure 2

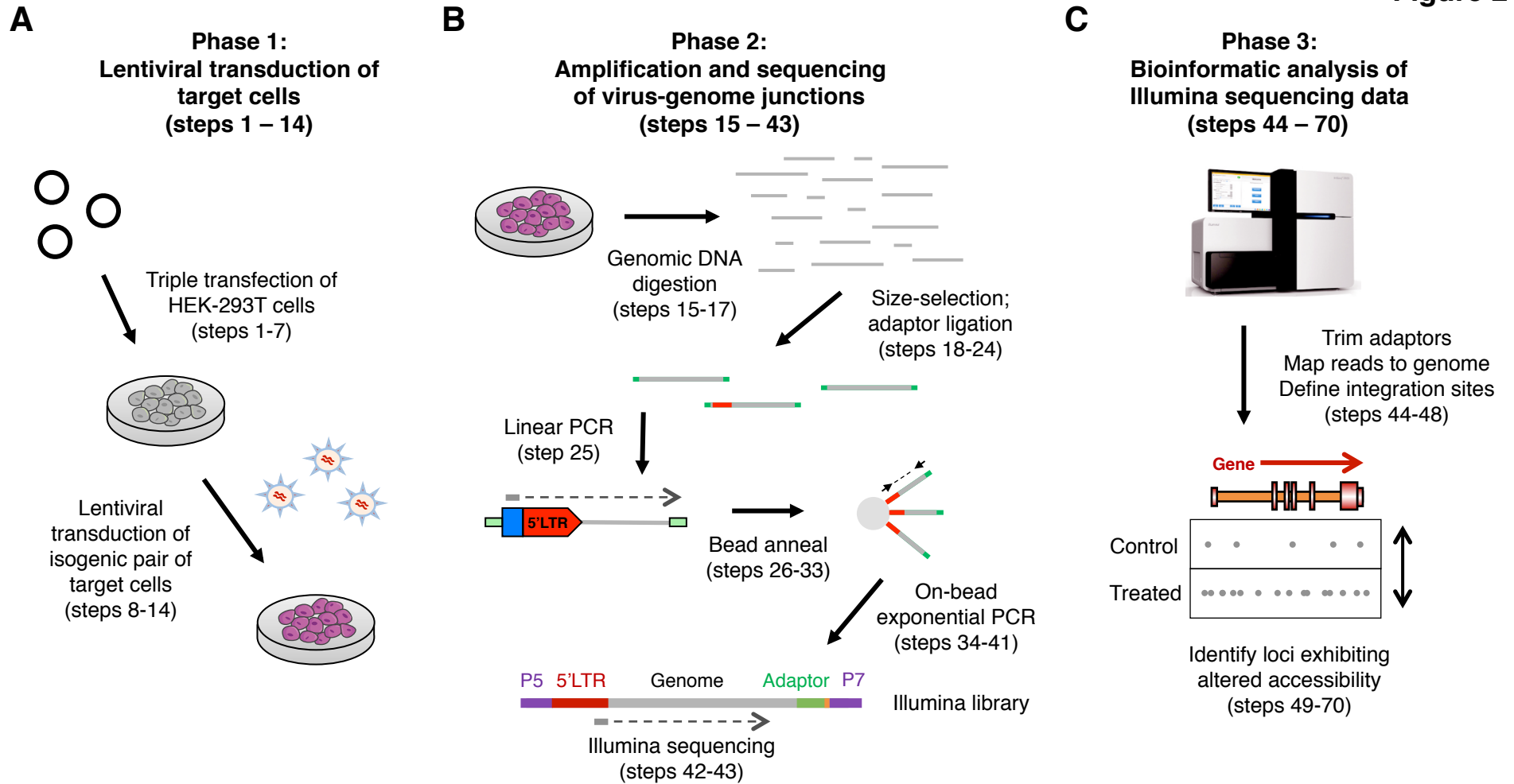


Figure 3

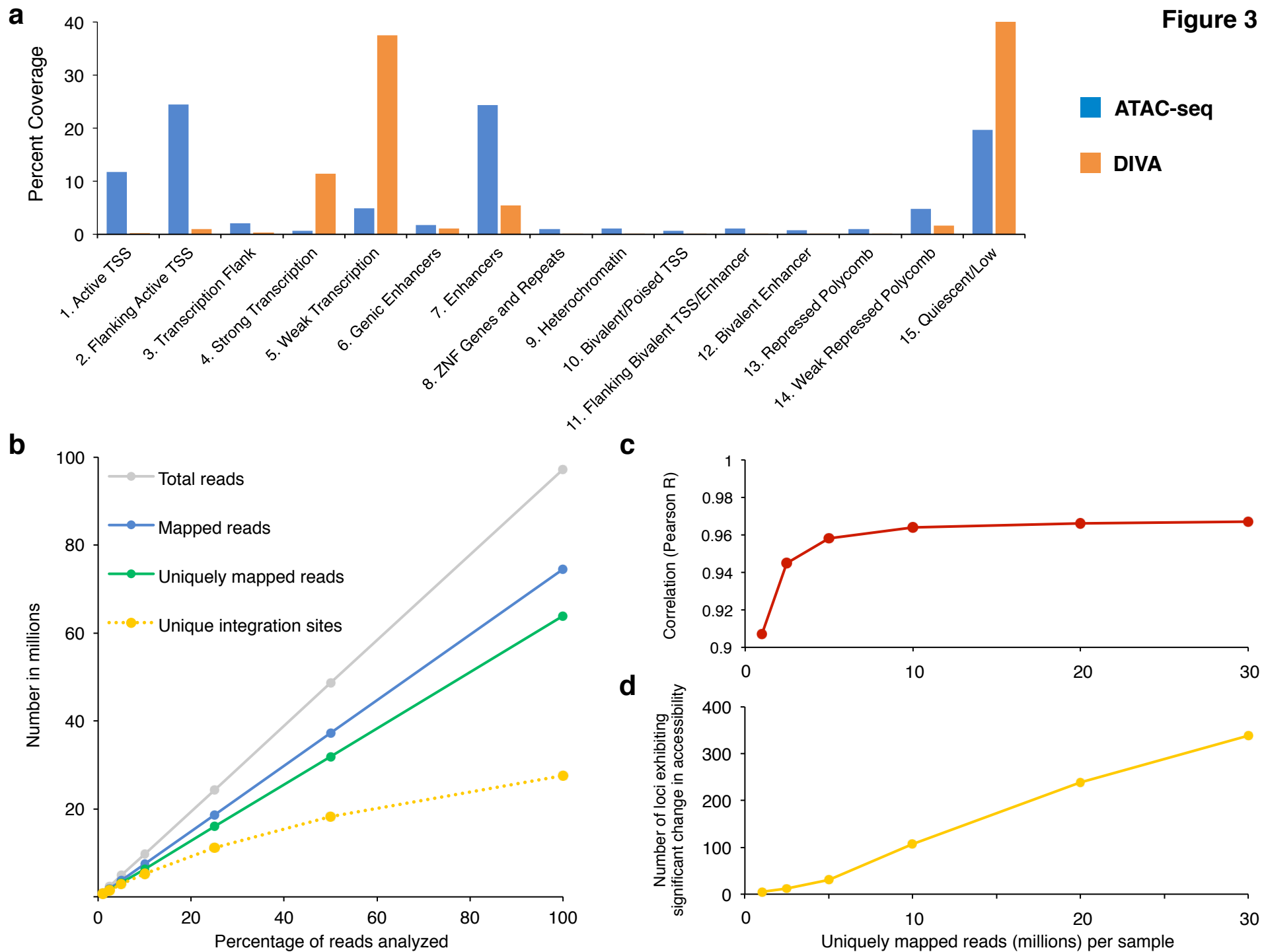
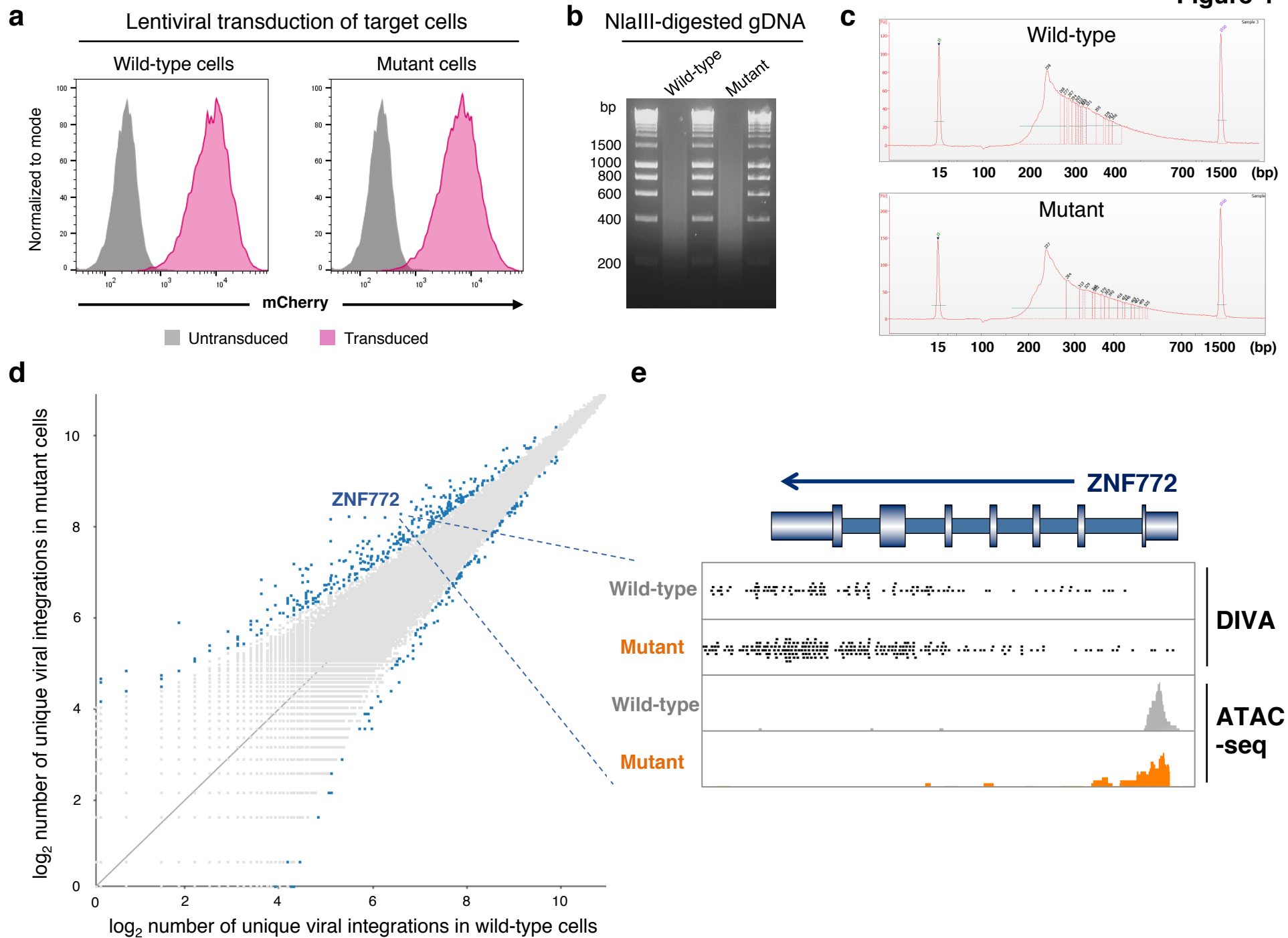


Figure 4

Box 1

Integrated pHR SIN-PSFFV-mCherry

