PLOS ONE

RESEARCH ARTICLE

Frequent sgRNA-barcode recombination in single-cell perturbation assays

Shiqi Xie☯, Anne Cooley☯, Daniel Armendariz, Pei Zhou, Gary C. Hon*

Cecil H. and Ida Green Center for Reproductive Biology Sciences, Department of Obstetrics and Gynecology, University of Texas Southwestern Medical Center, Dallas, Texas, United States of America

☯ These authors contributed equally to this work.
* gary.hon@utsouthwestern.edu

Citation: Xie S, Cooley A, Armendariz D, Zhou P, Hon GC (2018) Frequent sgRNA-barcode recombination in single-cell perturbation assays. PLoS ONE 13(6): e0198635. https://doi.org/10.1371/journal.pone.0198635

Editor: Wenhui Hu, Lewis Katz School of Medicine at Temple University, UNITED STATES

Received: February 15, 2018

Accepted: May 22, 2018

Published: June 6, 2018

Data Availability Statement: Raw sequencing data is available through the NIH Sequence Read Archive (SRP132627).

Funding: This work is supported by the Cancer Prevention Research Institute of Texas (CPRIT) (www.cprit.state.tx.us, RR140023, G.C.H.), National Institute of General Medical Sciences (www.nigms.nih.gov, DP2GM128203, G.C.H.), the Department of Defense (cdmrp.army.mil/funding, PR172060, G.C.H.), the Welch Foundation (www.welch1.org, I-1926-20170325, G.C.H.), and the Green Center for Reproductive Biology. S.X. is an

Abstract

Simultaneously detecting CRISPR-based perturbations and induced transcriptional changes in the same cell is a powerful approach to unraveling genome function. Several lentiviral approaches have been developed, some of which rely on the detection of distally located genetic barcodes as an indirect proxy of sgRNA identity. Since barcodes are often several kilobases from their corresponding sgRNAs, viral recombination-mediated swapping of barcodes and sgRNAs is feasible. Using a self-circularization-based sgRNA-barcode library preparation protocol, we estimate the recombination rate to be ~50% and we trace this phenomenon to the pooled viral packaging step. Recombination is random, and decreases the signal-to-noise ratio of the assay. Our results suggest that alternative approaches can increase the throughput and sensitivity of single-cell perturbation assays.

Introduction

Recently, single-cell RNA sequencing (scRNA-seq) has been coupled with CRISPR-mediated perturbations, allowing functional assessment of genes (Perturb-seq, CRISP-seq, CROP-seq) [1–3] and enhancers (Mosaic-seq) [4] with a transcriptomic readout. All of these techniques deliver CRISPR components to cells through a lentiviral system, and each one has devised a unique strategy to detect sgRNAs through scRNA-Seq. Since the scRNA-seq strategies used are 3'-biased, most of these approaches insert a molecular barcode immediately before the poly (A) signal as an indirect proxy of sgRNA expression in each cell (Fig 1). Therefore, the accuracy and sensitivity of these approaches rely on pre-identification of sgRNA-barcode relationships and unambiguous recovery of barcode information in every cell assayed.

However, barcoding could introduce noise due to lentiviral recombination. Two viral genomes are packaged into each lentiviral / retroviral particle [5], and are non-covalently linked [6]. During viral genome replication, the reverse transcriptase can switch from one template to another when it synthesizes a DNA provirus from a dimeric RNA genome, and this process happens most frequently at homologous regions [7–9]. The frequency of recombination depends on the distance between the two regions, which has been estimated to be 2% every kilobase [7,10]. Thus, when libraries of distinct sgRNA-barcode viruses are packaged together in single-cell perturbation assays, template switching could lead to barcode

American Heart Association fellow (heart.org, 16POST29910007). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

recombination that randomly shuffles sgRNA/barcode linkages. This event would interfere with the accurate detection of sgRNAs. A similar concern has also been raised recently on lentivirus-based genetic screening technologies [11].

Materials and methods

Cell lines and culture

K562 cells were cultured in IMDM Medium plus 10% FBS and pen/strep at 37˚C and 5% CO2. HEK293T cells were cultured in DMEM with 10% FBS and Pen/Strep. Both cells were acquired from ATCC (CCL-243 and CRL-3216).

Plasmids

The lenti-sgRNA(MS2)-puro plasmid (Addgene ID: 73795) was used for sgRNA expression. The 12-bp barcode region flanked by a BsrGI and an EcoRI cutsite was inserted into this plasmid by using overlap PCR and Gibson assembly. Specifically, a 108 bp oligo with 12 bp random oligo sequence was synthesized and amplified by PCR yielding double-stranded DNA. This fragment was then inserted into the linearized plasmid (cut with BsrGI and EcoRI) by Gibson assembly. After transformation, single clones were selected, and the barcode sequence of each clone was confirmed by Sanger sequencing. The insertion of sgRNAs was performed using BsmBI and T7 ligase, following the Golden Gate assembly protocol from the laboratory of Feng Zhang [12]. To minimize bacterial recombination, all the plasmids were transformed with Stellar Competent Cells (Clontech), and grown at 30˚C.

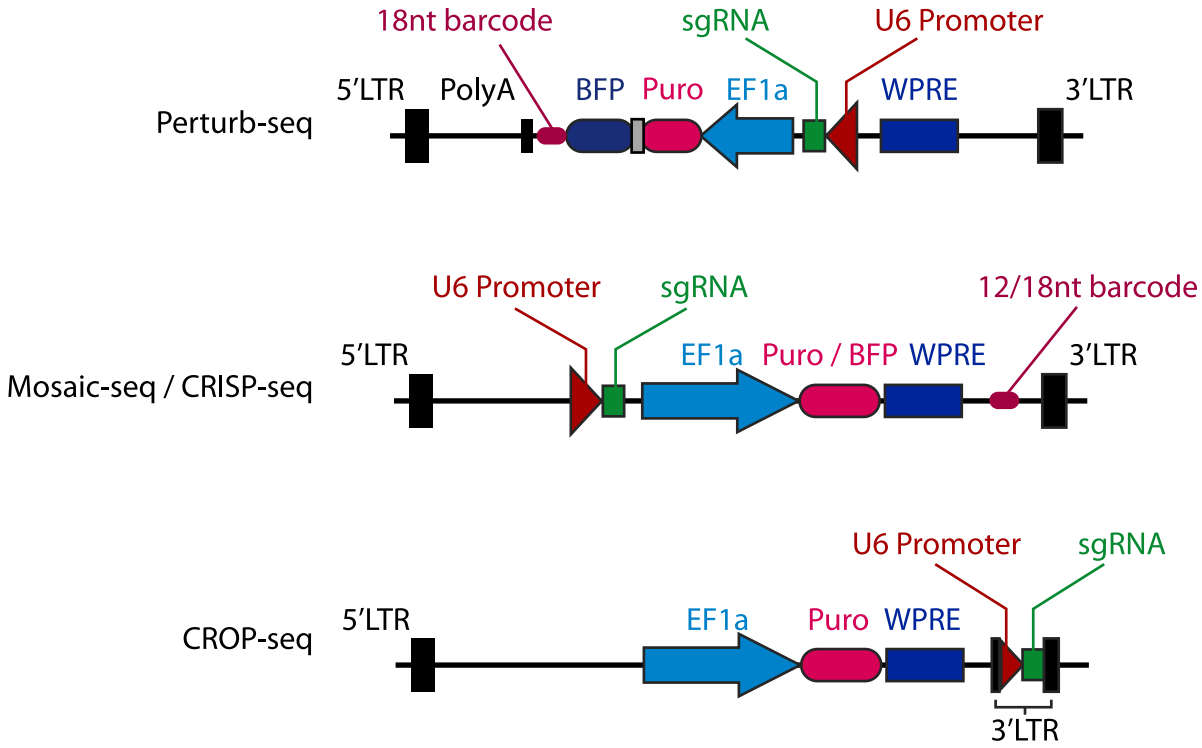

Fig 1. Vector structure of single-cell perturbation assays. The sgRNA barcode in Perturb-seq is part of the puromycin resistance gene / BFP transcript which is driven by core EF1α promoter (upper panel). Mosaic-seq and CRISP-seq share a similar design, in which the barcode is inserted immediately upstream of the lentiviral 3'LTR (middle panel). In CROP-seq, the sgRNA-expressing cassette is inserted into the 3'LTR, allowing direct detection of sgRNA sequences.

https://doi.org/10.1371/journal.pone.0198635.g001

Virus packaging, titration and infection

For virus packaging, 293T cells were seeded in a 6-cm dish ($3X10^6$ cells) one day before transfection. The indicated viral plasmid(s) were co-transfected with lentiviral packaging plasmids pMD2.G and psPAX2 (Addgene ID 12259 and 12260) with 4:2:3 ratio by using linear polyethylenimine (PEI). Twelve hours after transfection, media was changed to fresh DMEM with 10% FBS plus Pen/Strep. Seventy-two hours after transfection, virus-containing media was collected, passed through a 45 μm filter, and aliquoted into 1.5ml tubes. Viruses were stored in -80C before infection or titration. Virus were then titrated and used for infection based on the methods described previously [4]. For infection of K562 cells, $2X10^5$ cells (in 500μl medium, with 8ng/μl polybrene) were used. After mixing with the indicated amount of virus stock, the cells were centrifuged at 1000g for 1 hour at 37C and then returned to the incubator. The media was changed with fresh media containing 1μg/μl puromycin in the following day. The cells were selected for 7 days with media refreshed every two days and then collected for genomic DNA extraction and downstream library preparation.

Construction of sequencing libraries

Library construction was performed as previously described [4], with some modifications. Briefly, a 3kb amplicon flanked by the sgRNA and barcode sequences was amplified from plasmids or genomic DNA extracts. Then the fragment was self-circularized, and a second round of PCR was performed to yield a 400bp fragment with sgRNA and barcode adjacent to each other. The detailed protocol is available through potocols.io (dx.doi.org/10.17504/protocols.io.pufdntn).

Analysis

The Illumina NextSeq500 bcl files were de-multiplexed by using bcl2fastq (Illumina). Then the fastq files of two reads were combined and the reads with any base under quality score 10 were discarded. Then the sgRNA sequences and barcode sequences were extracted and compared with the known list, allowing 2 base-pair mismatches. The total reads per barcode-sgRNA pair were summarized and used to plot the figures.

Results

To systematically measure the noise introduced by viral recombination during Mosaic-seq, we individually cloned 20 unique sgRNAs into backbones with known barcode sequences (Fig 2A). We then monitored how pooling the samples at the transformation, viral packaging, or viral infection steps affected sgRNA-barcode recombination. To directly measure sgRNA-barcode pairs in each sample, we constructed deep sequencing libraries on plasmid pools and genomic DNA extracts. However, this problem is complicated by the large distance (~3-kb in Mosaic-seq) separating each sgRNA to its barcode. Our strategy involves PCR amplification of ~3kb sgRNA-barcode amplicons followed by a self-circularization step, which reduces the sgRNA/barcode distance to a sequenceable distance of ~400-bp (Fig 2B).

Since self-circularization is mediated by ligation, noise could be introduced by this method to assess recombination rates. To quantify this noise, we first examined Plasmid Library 2 (PL2), in which every sgRNA-barcode plasmid was constructed, transformed and extracted separately. We observe that 74.0% of reads (median) for each barcode is correctly linked to its known sgRNA pair, while the remaining 26.0% of reads are randomly linked to other sgRNAs (Fig 3A). This random collision rate correlates with the total abundance of each

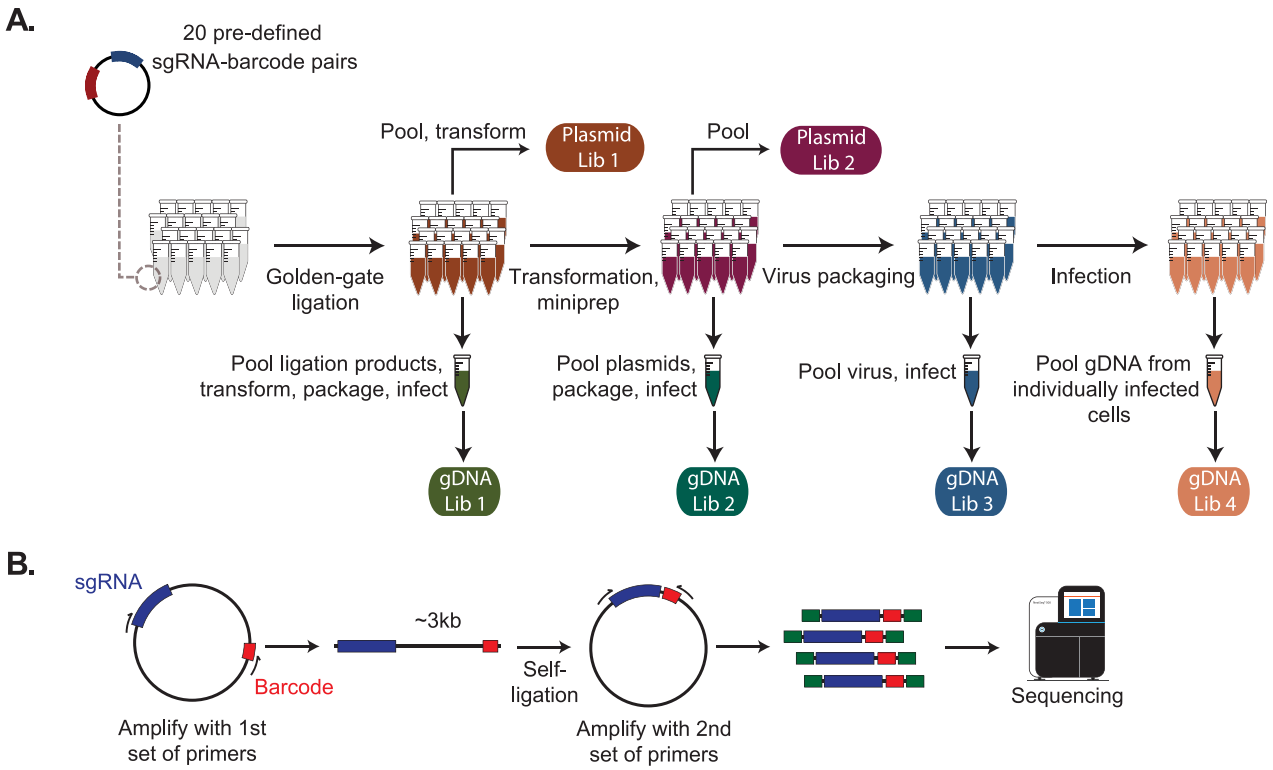

Fig 2. Schematic representation of the experimental design. (A) 20 sgRNAs were inserted into 20 sgRNA backbones with distinct barcodes by Golden Gate assembly. Then the samples were pooled at different steps of the procedure and sequencing libraries were constructed from either plasmids or genomic DNA extracts of infected K562 cells (see Methods). In total, we constructed two libraries from plasmids and four from the genomic DNA samples. (B) Schematic representation of the library construction procedure.

https://doi.org/10.1371/journal.pone.0198635.g002

sgRNA in the library. As PL2 plasmids were independently processed, sgRNA-barcode recombination should be negligible. Therefore, 26.0% noise we observed is likely derived from our ligation-mediated method for detecting recombination.

Then, we examined recombination after pooled bacterial transformation (PL1). We also observed a median of 79.1% of reads exhibited correct sgRNA-barcode linkages, suggesting that pooled transformation does not significantly contribute to recombination in a library of sgRNA-barcode plasmids.

Next, we examined sgRNA-barcode pairs after viral integration into the human genome. At the four stages of Golden Gate ligation, transformation, viral packaging, and infection, samples were pooled, and sgRNA-barcode sequencing libraries were constructed on genomic DNA (Fig 2A). Two genomic DNA libraries, in which sgRNA-barcode lentiviruses were individually packaged (GL3-4), maintain the correct sgRNA-barcode linkages (median of 83.6% and 84.4%, respectively) (Fig 3A), which is comparable to the plasmid libraries PL1-2.

In contrast, genomic DNA libraries in which plasmid libraries were pooled prior to viral packaging (GL1 and GL2), exhibited significant sgRNA-barcode recombination. The most abundant sgRNA of each barcode occupies less than half of the reads (median of 42.2% and 41.3%, respectively), which is greater than a 50% loss compared to GL3-4. Recombination is random, and none of the incorrect sgRNA-barcode pairs are dominant over the expected pairs (Fig 3B). These results suggest that, using a strategy in which sgRNAs are separated from barcodes by several kilobases, recombination will be frequent if plasmid libraries are pooled prior to viral packaging.

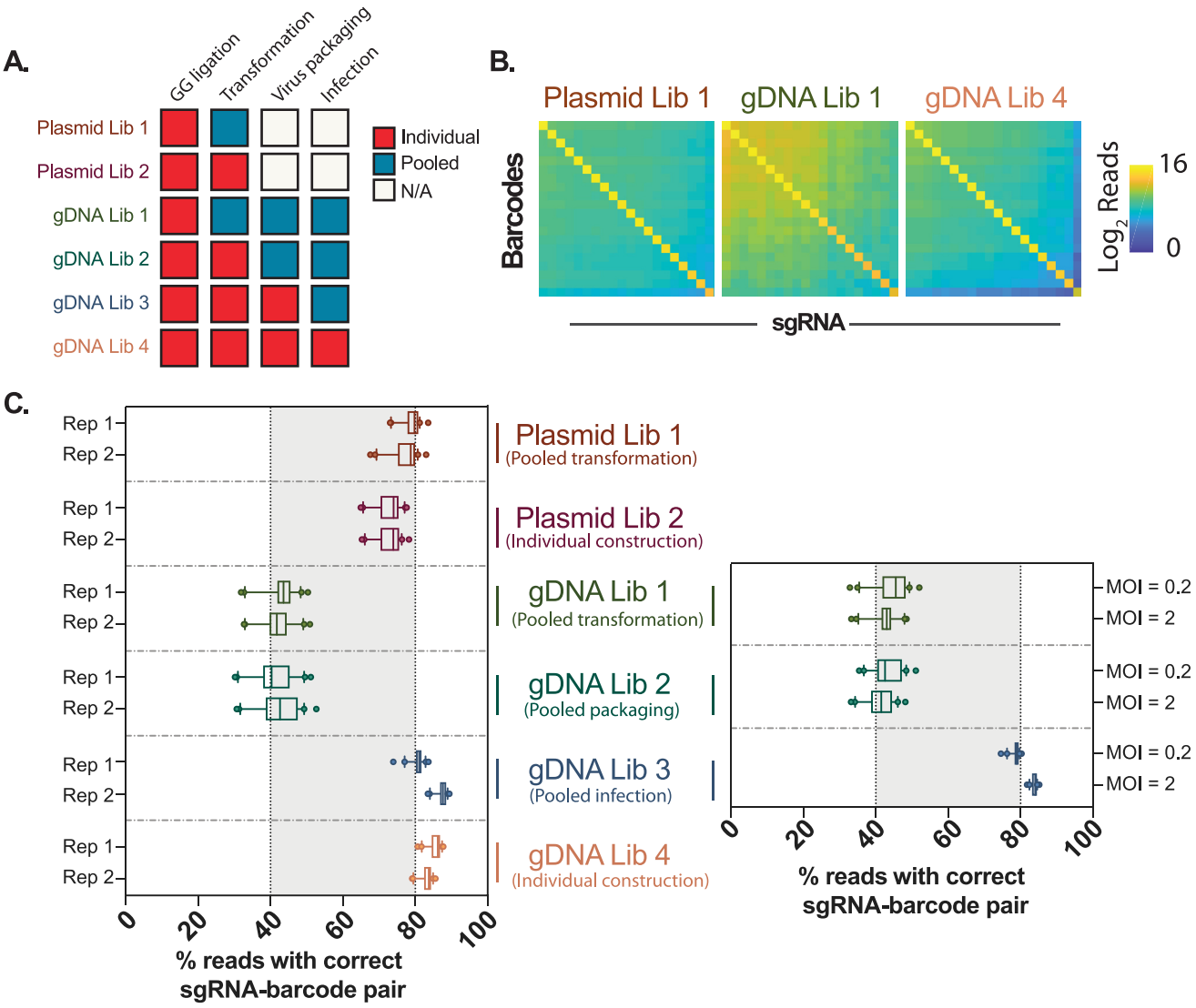

Fig 3. Barcode shuffling during multiplexed Mosaic-seq library preparation. (A) Summary of the sample conditions. (B) Read distribution of three representative libraries. For each barcode, we only observed one dominant sgRNA sequence, which is always the expected sgRNA. (C) The percentage of reads for the most abundant sgRNA for each barcode are plotted in the boxplot. Whiskers represent the 10th and 90th percentiles, and the dots represent outliers. Virus infection in the left panel were performed by using the same volume of virus stock (MOI varies from 1-2.4); right panel shows an independent experiment with high and low MOI, using the same virus packaged in the left panel.

https://doi.org/10.1371/journal.pone.0198635.g003

To further test whether recombination depends on viral titer, we infected cells at high and low multiplicity of infection (MOI = 2 and MOI = 0.2). Based on Poisson statistics, >90% of antibiotic-selected cells are expected to be infected by exactly one virus at MOI = 0.2, which we hypothesize could reduce observed recombination rates compared to cells infected at high MOI. However, we observed no significant difference in recombination between the high and low MOI samples (Fig 3A), suggesting that the observed sgRNA-barcode shuffling is not due to recombination between multiple viruses infecting a single cell.

Discussion and conclusions

Here we used a self-ligation-based method to assess the recombination between sgRNAs and barcodes during Mosaic-seq. While our method has a relatively high baseline level of noise

(~20%), our data confirms sgRNA-barcode recombination during pooled preparation of Mosaic-seq libraries. Recombination is random and accounts for ~50% of reads. While this noise is unlikely to create false positive hits, it does reduce the overall signal-to-noise of the assay, which we expect will decrease sensitivity. We postulate that similar recombination events will exist in other methods that rely on lentiviral/retroviral delivery systems that are coupled to indirect detection of DNA-based barcodes.

We observed that recombination only occurs when libraries are pooled before the virus production step, independent of the viral titer used during infection. This suggests that recombination predominantly occurs between two viral genomes packaged into the same virion, but not between distinct virus infecting the same cell. Thus, at low throughput, this problem can be overcome by constructing and packaging each virus separately. However, for large scale library preparation, the CROP-seq sgRNA plasmid is an improved solution [3]. In CROP-seq, the sgRNA cassette is inserted into the 3'LTR of the virus, which becomes part of the puromycin-resistance mRNA transcribed by EF1α promoter. Therefore, the sgRNA can be directly detected by scRNA-seq without the use of indirect barcodes. Moreover, CROP-seq dramatically simplifies the construction of large-scale sgRNA libraries since barcodes do not need to be constructed. By reducing sgRNA/barcode recombination, the sensitivity of single-cell perturbation assays could increase substantially. During preparation of this manuscript, similar observations have also been independently reported [13–15]. We believe that these improvements will significantly expand the application of single-cell perturbation assays, enabling the construction of large-scale libraries to systematically perturb and unravel transcriptional regulation from systems perspective.

Supporting information

S1 Table. Primers used in this study.
(DOCX)

S1 File. Summary of sgRNA-barcode read count.
(GZ)

Acknowledgments

We thank Vijay Ramani for his helpful insights into the sgRNA/barcode recombination problem, and we thank all the members in Hon lab for insightful discussion. We acknowledge the BioHPC computational infrastructure at UT Southwestern for providing HPC and storage resources that have contributed to the research results reported within this paper. We also acknowledge UT Southwestern's McDermott Center for providing next-generation sequencing services for this work.

Author Contributions

Conceptualization: Shiqi Xie, Gary C. Hon.

Formal analysis: Shiqi Xie.

Funding acquisition: Shiqi Xie, Gary C. Hon.

Investigation: Shiqi Xie, Anne Cooley, Daniel Armendariz, Pei Zhou.

Methodology: Shiqi Xie.

Project administration: Gary C. Hon.

Supervision: Shiqi Xie, Gary C. Hon.

Writing – original draft: Shiqi Xie.

Writing – review & editing: Shiqi Xie, Gary C. Hon.

References

1. Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. Cell. 2016; 167: 1853–1866. e17. https://doi.org/10.1016/j.cell.2016.11.038 PMID: 27984732
2. Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. Cell. Elsevier; 2016; 167: 1883–1896. e15.
3. Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, et al. Pooled CRISPR screening with single-cell transcriptome readout. Nat Methods. 2017; 14: 297–301. https://doi.org/10.1038/nmeth.4177 PMID: 28099430
4. Xie S, Duan J, Li B, Zhou P, Hon GC. Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. Mol Cell. 2017; 66: 285–299.e5. https://doi.org/10.1016/j.molcel.2017.03.007 PMID: 28416141
5. Nikolaitchik OA, Dilley KA, Fu W, Gorelick RJ, Tai S-HS, Soheilian F, et al. Dimeric RNA Recognition Regulates HIV-1 Genome Packaging. PLoS Pathog. Public Library of Science; 2013; 9: e1003249.
6. Paillart J-C, Shehu-Xhilaga M, Marquet R, Mak J. Dimerization of retroviral RNA genomes: an inseparable pair. Nat Rev Microbiol. Nature Publishing Group; 2004; 2: 461.
7. Hu WS, Temin HM. Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. Proc Natl Acad Sci U S A. 1990; 87: 1556–1560. PMID: 2304918
8. Zhang J, Temin HM. Retrovirus recombination depends on the length of sequence identity and is not error prone. J Virol. 1994; 68: 2409–2414. PMID: 7511170
9. Peliska JA, Benkovic SJ. Mechanism of DNA strand transfer reactions catalyzed by HIV-1 reverse transcriptase. Science. American Association for the Advancement of Science; 1992; 258: 1112–1118.
10. Schlub TE, Smyth RP, Grimm AJ, Mak J, Davenport MP. Accurately Measuring Recombination between Closely Related HIV-1 Genomes. PLoS Comput Biol. Public Library of Science; 2010; 6: e1000766.
11. Sack LM, Davoli T, Xu Q, Li MZ, Elledge SJ. Sources of Error in Mammalian Genetic Screens. G3. 2016; 6: 2781–2790. https://doi.org/10.1534/g3.116.030973 PMID: 27402361
12. Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, et al. Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature. Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.; 2014; 517: 583.
13. Feldman D, Singh A, Garrity AJ, Blainey PC. Lentiviral co-packaging mitigates the effects of intermolecular recombination and multiple integrations in pooled genetic screens [Internet]. bioRxiv. 2018. p. 262121. https://doi.org/10.1101/262121
14. Adamson B, Norman TM, Jost M, Weissman JS. Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens [Internet]. bioRxiv. 2018. p. 298349. https://doi.org/10.1101/298349
15. Hill AJ, McFaline-Figueroa JL, Starita LM, Gasperini MJ, Matreyek KA, Packer J, et al. On the design of CRISPR-based single-cell molecular screens. Nat Methods. 2018; 15: 271–274. https://doi.org/10.1038/nmeth.4604 PMID: 29457792