

CRISPR screening at the single-cell level

Doctoral thesis at the Medical University of Vienna
for obtaining the academic degree

Doctor of Philosophy

Submitted by

Paul Datlinger, MSc

Supervisor:
Christoph Bock, PhD

CeMM Research Center for Molecular Medicine
of the Austrian Academy of Sciences

Lazarettgasse 14, AKH BT25.3
1090 Vienna, Austria

Vienna, 09/2019

DECLARATION

This doctoral thesis is written in a cumulative format and contains both published and unpublished results. The included manuscripts are listed below. Unless otherwise stated, the experiments and analyses described have been performed in the laboratory of Christoph Bock at the CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences in Vienna, Austria. The author of this thesis, Paul Datlinger, is first author on both included manuscripts and wrote the remainder of this thesis with input from the supervisor, Christoph Bock. Detailed author contribution statements approved by all authors are included in each manuscript in the section “Author Contributions”. Paul Datlinger and Christoph Bock declare competing financial interests in the form of licensing agreements for CROP-seq, and are inventors on a patent application relating to scifi-RNA-seq.

List of first-author publications:

Pooled CRISPR screening with single-cell transcriptome readout

Paul Datlinger, André F Rendeiro*, Christian Schmidl*, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar & Christoph Bock

Published in: Nature Methods 14: 297-301. 2017. doi: 10.1038/nmeth.4177

Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing

Paul Datlinger, André F Rendeiro*, Thorina Boenke*, Thomas Krausgruber, Daniele Barreca & Christoph Bock

In preparation (unpublished manuscript)

** These authors contributed equally / shared second authors*

TABLE OF CONTENTS

Declaration	ii
Table of contents	iii
List of figures and tables	vii
Abstract	viii
Zusammenfassung	ix
Publications arising from the thesis project	x
Abbreviations	xii
Acknowledgments	xiv
1. Introduction	1
1.1. Genome engineering and genetic screens	1
1.1.1. Profiling the genome, epigenome and transcriptome with sequencing technology	1
1.1.2. Genome engineering in the pre-CRISPR era	2
1.1.3. A prokaryotic immune system revolutionizes genome editing	3
1.1.4. Practical example: Single-gene knockouts with SpCas9	4
1.1.5. Notable alternatives to SpCas9	6
1.1.6. Extending the functionality of SpCas9 beyond cutting DNA	7
Turning SpCas9 into a versatile recruitment platform	7
Targeted gene activation (CRISPRa)	7
CRISPRa based on synergistic activators and RNA aptamers for recruitment	7
CRISPRa based on VP64 and signal amplification with the SunTag	8
Targeted inhibition of transcription (CRISPRi)	9
Epigenome editing	9
1.1.7. Pooled functional screens with CRISPR	10
Negative selection screens	11
Positive selection screens	12
Screens based on fluorescence-activated cell sorting (FACS)	13
Pooled CRISPR screens with single-cell RNA-seq readout	14
1.2. Implementing single-cell CRISPR screening	16
1.2.1. Single-cell RNA-seq platforms with sufficient throughput for pooled screening	16
Technologies based on the physical isolation of single cells	16
Nanowell plates	17
Microfluidic droplet generators	18

Technologies using single cells as the reaction compartment	20
Combinatorial indexing methods	20
Pooled optical screens	22
1.2.2. Platform-agnostic detection of gRNAs in single-cell transcriptomes	25
2. Results	28
2.1. Pooled CRISPR screening with single-cell transcriptome readout	28
2.2. Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing	51
3. Materials and Methods	71
3.1. Step-by-step protocol for CROP-seq (CRISPR droplet sequencing)	72
3.1.1. Pooled, amplification-free cloning of gRNA libraries	72
Validation of the CROPseq-Guide-Puro plasmid preparation	72
Preparation of CROPseq-Guide-Puro vector backbone	72
Plasmid digestion	72
UV-free gel extraction	73
Assembly of gRNA-encoding ssDNA oligonucleotides into vector backbone by Gibson Assembly	73
Design gRNAs	73
Gibson Assembly	74
Electroporation of Lucigen Endura cells	74
Preparations	74
Electroporation	75
Plasmid preparation	75
3.1.2. Next-generation sequencing of gRNA sequences for library quality control or pooled screens	75
3.1.3. Expected results for the pooled, amplification-free cloning of gRNA libraries	77
3.1.4. Lentivirus production and cell culture for CROP-seq screens	78
Production of lentiviral particles	78
Titration of lentivirus preparations (optional)	79
Transduction of suspension cells for pooled CROP-seq screens	80
Transduction of adherent cells for pooled CROP-seq screens	80
3.1.5. Single-cell RNA-seq based on Drop-seq	81
Droplet microfluidics	81
Prepare cells	81
Prepare beads	81
Drop-seq	81
Droplet breakage	82
Bead capture	82
Washing beads	83
Eluting beads	83

Concentrating beads	83
Preparation of cDNA library	83
Reverse transcription	83
Exonuclease I treatment	84
SMART PCR to enrich cDNA library	84
Purification, quality control and pooling of SMART-PCR enriched cDNA	86
Nextera XT library preparation	86
Tagmentation	86
Nextera XT PCR enrichment	87
Purification of the final library	87
Library quality control and next-generation sequencing	88
Quality control and dilution of the CROP-seq library	88
Next-generation sequencing on Illumina HiSeq 3000/4000 instruments	88
Next-generation sequencing on Illumina NextSeq 500 instruments	88
3.1.6. Expected results for single-cell RNA-seq by Drop-seq	89
3.1.7. Materials	90
Reagents required for pooled, amplification-free cloning of gRNA libraries	90
Reagents required for lentivirus production and cell culture for CROP-seq screens	91
Reagents required for single-cell RNA-seq based on Drop-seq	92
3.2. Step-by-step protocol for scifi-RNA-seq	93
3.2.1. Single-cell combinatorial fluidic indexing	93
Whole cell fixation for cell lines	93
Nuclei isolation for cell lines	94
Nuclei isolation with formaldehyde fixation for primary cells	94
Reverse transcription (round 1 indexing)	95
Pooling and cell counting	95
Thermocycling ligation (round 2 indexing)	96
Silane bead cleanup	97
SPRI cleanup	98
Template switching	98
3.2.2. cDNA enrichment, library preparation and next-generation sequencing	99
SPRI cleanup	99
cDNA enrichment	99
cDNA cleanup (0.8x, 0.6x) and quality control	99
Tagmentation with i7-only transposome	100
Library cleanup (0.7x, 0.8x) and quality control	101
Next-generation sequencing on the Illumina NovaSeq 6000 platform	102
Appendix A: Preparation of the custom i7 transposome and activity check	103
Anneal oligonucleotides	103

Load Tn5 transposase_____	103
Optional activity test of custom transposome in a qPCR assay _____	103
Appendix B: Expected results _____	106
Appendix C: Materials _____	107
Appendix D: Buffers for isolation of nuclei or whole cells without fixation _____	108
Appendix E: Buffers for isolation of nuclei with formaldehyde fixation and permeabilization ____	109
Appendix F: Buffers for transposome assembly and tagmentation _____	110
Appendix G: Oligonucleotides in tubes _____	111
Appendix H: Set of 384 barcoded reverse transcription primers with fixed bases for error correction _____	112
4. Discussion _____	116
4.1. Advantages of CROP-seq over competing technologies _____	116
4.2. The value of single-cell resolution in CRISPR screens _____	119
4.3. Current limitations of single-cell CRISPR screens _____	121
5. References _____	123
6. Curriculum Vitae _____	132

LIST OF FIGURES AND TABLES

Unless otherwise stated, figures and tables included in this thesis have been prepared by Paul Datlinger, the author of this thesis.

Figure 1	Layers of genomic and epigenomic information that can be profiled by next-generation sequencing.	1
Figure 2	Genome editing with the CRISPR/Cas9 system.	5
Figure 3	Three popular designs for targeted gene activation based on CRISPR.	8
Figure 4	Pooled CRISPR screening with positive or negative selection.	12
Figure 5	CRISPR screening at the single-cell level provides unique insights.	14
Figure 6	Technologies based on the isolation of single cells into miniaturized reaction compartments.	17
Figure 7	Single-cell combinatorial indexing methods.	21
Figure 8	Pooled optical CRISPR screens.	23
Figure 9	Technical solution to detect gRNAs directly from single-cell transcriptomes, while maintaining high gRNA expression and editing efficiencies.	26
Figure 10	Comparison of CROP-seq with alternative methods. <i>This figure was reprinted from (Wagner & Klein, 2017) with permission from the publisher.</i>	116
Figure 11	CROP-seq avoids lentiviral switching of gRNA to GBC associations. <i>Panel a was reprinted under the CC-BY-NC-ND 4.0 International license (Adamson et al., 2018). Panel b was modified and data was re-plotted as percent barcode recombination for clarity (Hill et al., 2018).</i>	117
Figure 12	CROP-seq has been adopted by labs world-wide.	119
Table 1	Comparison of current single-cell technology platforms and representative methods.	25

ABSTRACT

Revolutionary developments in DNA sequencing technology allow researchers to assay the genome, epigenome and transcriptome at unprecedented detail. Since the completion of the human reference genome, costs have dropped significantly - inspiring large-scale data collection efforts to catalogue mutations and aberrant gene-regulatory mechanisms in health and disease. However, research efforts remain mostly descriptive, and experimental validation is lagging behind due to a lack of suitable technologies. To realize the full potential of genome medicine will require tools that can introduce precise changes at the genomic and epigenomic level to establish causal links to diseases. This gap is increasingly filled by genome and epigenome editing and functional genetic screens, which are the focus of this work.

CRISPR-based genetic screens have the potential to revolutionize biological discovery and the interpretation of large datasets, yet current methods have inherent limitations. Pooled screens are highly scalable, but limited to simple readouts such as cell survival or sortable marker proteins. Screening in an arrayed format can provide a detailed molecular readout, but at a greatly reduced throughput. In this thesis, we introduce a novel paradigm for functional genetic screens. By combining pooled CRISPR screens with a single-cell RNA sequencing readout, we approach the scalability of pooled experiments with the versatile whole-transcriptome readout of arrayed assays. CRISPR screening at the single-cell level is a powerful approach that can localize the effects of a genetic perturbation in a heterogeneous mixture of cell types, or during a dynamic process such as differentiation. Our technology CROP-seq (for: CRISPR droplet sequencing) is based on a modified lentiviral vector that makes CRISPR guide RNAs detectable in single-cell transcriptomes while maintaining full compatibility with established CRISPR workflows. We applied our method to study T-cell receptor activation, demonstrating that CROP-seq can screen complex transcriptome signatures at unprecedented scale and detail.

While single-cell CRISPR screens promise to revolutionize functional genomics, the high cost and technical limitations of existing single-cell sequencing technologies prevent us from realizing their full potential. The second part of this thesis therefore explores single-cell combinatorial indexing methods to achieve ultra-high throughput, yet low-cost and versatile single-cell RNA sequencing. We introduce a new concept named scifi (single-cell combinatorial indexing with fluidic indexing), which combines combinatorial indexing with droplet microfluidics. Applying this idea to single-cell RNA sequencing (scifi-RNA-seq) allowed us to profile large numbers of single cells with a greatly simplified workflow and superior results.

ZUSAMMENFASSUNG

Revolutionäre Entwicklungen in der DNA-Sequenzierung ermöglichen es, das Genom, Epigenom und Transkriptom in bisher unerreichtem Detailgrad abzubilden. Seit der Fertigstellung des menschlichen Referenz-Genoms sind die Kosten so stark gesunken, dass umfassende Kataloge von Mutationen und abweichenden Genregulationsmechanismen für diverse Erkrankungen erstellt werden konnten. Leider bleiben diese Forschungsprojekte zumeist rein deskriptiv, da sich eine experimentelle Validierung der Resultate aufgrund unzureichender Technologien schwierig gestaltet. Um das volle Potential für die personalisierte Medizin ausschöpfen zu können, werden neue Ansätze benötigt, die präzise Veränderungen im Genom oder Epigenom ermöglichen, um so einen kausalen Zusammenhang zu einer Erkrankung festzustellen. Diese Lücke wird zunehmend durch Genom- und Epigenom-Editierung und Einzelzelltechnologien geschlossen, welche das Hauptthema dieser Arbeit darstellen.

Genetische Screens mit CRISPR revolutionieren derzeit die biologische Grundlagenforschung und die funktionale Interpretation genomischer Datensätze. Die geläufigen Methoden sind jedoch stark eingeschränkt. Gepoolte Screens erreichen zwar einen hohen Durchsatz, beschränken sich jedoch auf relativ einfache Endpunkte wie Zell-Proliferation oder die Expression von Markerproteinen. Im Gegensatz dazu ermöglichen Screens in Mikrotiterplatten die Generierung komplexer Datensätze, jedoch mit viel geringerem Durchsatz. In dieser Dissertation entwickeln wir ein neuartiges Konzept für genetische Screens. Wir erweitern gepoolte Screens durch RNA-Sequenzierung in Einzelzellen, wodurch es möglich wird, den hohen Durchsatz von gepoolten Screens mit dem Auslesen komplexer Genexpressions-Signaturen zu verbinden. CRISPR-Screens mit Einzelzell-Sequenzierung sind ein vielversprechender Ansatz, um die Effekte einer Genveränderung in unterschiedlichen Zelltypen eines komplexen Gewebes oder während eines dynamischen Prozesses wie der Zelldifferenzierung zu untersuchen. Unsere Technologie CROP-seq (abgekürzt für *CRISPR droplet sequencing*) basiert auf einem modifizierten lentiviralen Vektor, welcher CRISPR Guide RNAs in Einzelzell-Transkriptomen detektierbar macht, jedoch mit etablierten Protokollen für CRISPR Screens voll kompatibel bleibt. Durch Anwendung von CROP-seq auf T-Zellen konnten wir zeigen, dass sich die Methode auch für das Auslesen von komplexen Genexpressions-Signaturen wie der T-Zell-Rezeptor-Aktivierung eignet, die mit heute üblichen Verfahren für genetische Screens nur unzureichend untersucht werden können.

Obwohl CRISPR Screens auf Einzelzell-Ebene ein vielversprechendes neues Konzept darstellen, kann ihr volles Potential aufgrund der hohen Kosten und technischen Limitationen der Einzelzell-Sequenzierung bisher noch nicht annähernd ausgeschöpft werden. Der zweite Teil dieser Dissertation beschäftigt sich daher mit neuartigen Methoden zur RNA-Sequenzierung in Einzelzellen die einen sehr hohen Durchsatz bei möglichst geringen Kosten erreichen (Combinatorial Indexing). Wir stellen ein neuartiges Konzept vor, welches Combinatorial Indexing mit einem Mikrofluidik-Chip verbindet. Wir wenden diese neue Idee auf die RNA-Sequenzierung in Einzelzellen an (scifi-RNA-seq), wodurch wir sehr hohe Zellzahlen bei stark vereinfachtem Ablauf und besseren Resultaten erreichen konnten.

PUBLICATIONS ARISING FROM THE THESIS PROJECT

Datlinger P, Rendeiro AF*, Boenke T*, Krausgruber T, Barreca D, Bock C **Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing.** *In preparation (unpublished manuscript).*

Ferreira da Silva J, Salic S, Wiedner M, Datlinger P, Essletzbichler P, Hanzl A, Superti-Furga G, Bock C, Winter G, Loizou JI **Genome-scale CRISPR screens are efficient in non-homologous end-joining deficient cells.** *Submitted.*

Klughammer J*, Kiesel B*, Roetzer T, Fortelny N, Nemc A, Nenning KH, Furtner J, Sheffield NC, Datlinger P, Peter N, Nowosielski M, Augustin M, Mischkulnig M, Strobel T, Alpar D, Erguner B, Senekowitsch M, Moser P, Freyschlag CF, Kerschbaumer J, Thome C, Grams AE, Stockhammer G, Kitzwoegerer M, Oberndorfer S, Marhold F, Weis S, Trenkler J, Buchroithner J, Pichler J, Haybaeck J, Krassnig S, Mahdy Ali K, von Campe G, Payer F, Sherif C, Preiser J, Hauser T, Winkler PA, Kleindienst W, Wurtz F, Brandner-Kokalj T, Stultschnig M, Schweiger S, Dieckmann K, Preusser M, Langs G, Bauermann B, Knosp E, Widhalm G, Marosi C, Hainfellner JA, Woehrer A, Bock C (2018) **The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space.** *Nat Med* 24: 1611-1624

Datlinger P, Rendeiro AF*, Schmidl C*, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C (2017) **Pooled CRISPR screening with single-cell transcriptome readout.** *Nat Methods* 14: 297-301

Sheffield NC, Pierron G, Klughammer J, Datlinger P, Schonegger A, Schuster M, Hadler J, Surdez D, Guillemot D, Lapouble E, Freneaux P, Champigneulle J, Bouvier R, Walder D, Ambros IM, Hutter C, Sorz E, Amaral AT, de Alava E, Schallmoser K, Strunk D, Rinner B, Liegl-Atzwanger B, Huppertz B, Leithner A, de Pinieux G, Terrier P, Laurence V, Michon J, Ladenstein R, Holter W, Windhager R, Dirksen U, Ambros PF, Delattre O, Kovar H, Bock C, Tomazou EM (2017) **DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma.** *Nat Med* 23: 386-395

Bock C, Halbritter F, Carmona FJ, Tierling S, Datlinger P, Assenov Y, Berdasco M, Bergmann AK, Booher K, Busato F, Campan M, Dahl C, Dahmcke CM, Diep D, Fernández AF, Gerhauser C, Haake A, Heilmann K, Holcomb T, Hussmann D, Ito M, Kläver R, Kreutz M, Kulis M, Lopez V, Nair SS, Paul DS, Plongthongkum N, Qu W, Queirós AC, Reinicke F, Sauter G, Schlomm T, Statham A, Stirzaker C, Strogantsev R, Urdinguio RG, Walter K, Weichenhan D, Weisenberger DJ, Beck S, Clark SJ, Esteller M, Ferguson-Smith AC, Fraga MF, Guldborg P, Hansen LL, Laird PW, Martín-Subero JI, Nygren AOH, Peist R, Plass C, Shames DS, Siebert R, Sun X, Tost J, Walter J & Zhang K for the BLUEPRINT consortium (2016) **Quantitative comparison of DNA methylation assays for biomarker development and clinical applications.** *Nat Biotechnol* 34: 726-37

Klughammer J, Datlinger P, Printz D, Sheffield NC, Farlik M, Hadler J, Fritsch G, Bock C (2015) **Differential DNA Methylation Analysis without a Reference Genome**. Cell Rep 13: 2621-2633

Farlik M*, Sheffield NC*, Nuzzo A, Datlinger P, Schonegger A, Klughammer J, Bock C (2015) **Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics**. Cell Rep 10: 1386-97

Tomazou EM*, Sheffield NC*, Schmidl C, Schuster M, Schonegger A, Datlinger P, Kubicek S, Bock C, Kovar H (2015) **Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1**. Cell Rep 10: 1082-95

ABBREVIATIONS

10x Genomics Chromium. Commercial platform for single-cell RNA-seq using a microfluidic droplet generator

AAVs. Adeno-associated viruses

ATAC-seq. Assay for transposase-accessible chromatin using sequencing

BRCA1. Breast cancer type 1 susceptibility protein

Cas genes. CRISPR-associated genes, typically in the immediate vicinity of the CRISPR locus

Cas12a. also known as Cpf1, alternative nuclease for CRISPR-genome editing with the intrinsic ability to process the crRNA array, recognizing a short T-rich PAM and producing staggered cuts with 5' overhangs

Cas13a. Family of RNA-targeting CRISPR systems

CEL-Seq2. Plate-based single-cell RNA-seq technology

ChIP-seq. Chromatin immunoprecipitation sequencing

CjCas9. *Campylobacter jejuni* Cas9, a small variant useful for in vivo genome editing

C-NHEJ. Canonical non-homologous end joining

COSMIC. Catalogue of Somatic Mutations in Cancer

CRISPR. Clustered regularly interspaced short palindromic repeats

CRISPRa. CRISPR activation

CRISPRi. CRISPR inhibition without inducing DNA breaks

CRISP-seq. Method for single-cell CRISPR screening

CROP-seq. CRISPR droplet sequencing, a method for pooled CRISPR screens with single-cell transcriptome readout

crRNAs. CRISPR RNAs, small RNA species that mediates sequence-specific targeting of CRISPR systems

Cyto-Seq. Single-cell technology based on sub-nanoliter well plates

dCas9. Catalytically inactive (dead) Cas9

dCas9-VPR. System for CRISPR activation based on the activators VP64, p65 and Rta

DNA. Deoxyribonucleic acid

DNMT1. DNA Methyltransferase 1

DNMT3A. DNA methyltransferase 3A

dRNAs. Dead RNAs, gRNAs with shortened spacer sequence that still bind but do not cut

Drop-seq. Open-source single-cell RNA-seq technology based on a microfluidic droplet generator

evoCas9. High-fidelity variant of SpCas9

FACS. Fluorescence-activated cell sorting

FokI. Type IIS restriction endonuclease isolated from *Flavobacterium okeanokoites*

GBCs. Guide barcodes used by Perturb-seq, CRISP-seq and Mosaic-seq

GFP. Green fluorescent protein

gRNA. guide RNA, typically a chimeric fusion of crRNA and tracrRNA

GTEX. Genotype-Tissue Expression, large-scale collaborative project

HAP1. Near-haploid human cell line

HGP. Human Genome Project

Hi-C. Method for all vs. all chromosome conformation capture

hU6. Human U6 promoter, used for gRNA transcription by RNA polymerase III

HypaCas9. High-fidelity variant of SpCas9

IHEC. International Human Epigenome Consortium

indel. Insertion or deletion mutation

inDrop. Open-source single-cell RNA-seq technology based on a microfluidic droplet generator

I-SceI. Meganuclease isolated from *Saccharomyces cerevisiae*

KRAB. Krüppel-associated box, protein domain resulting in gene repression

LTR. Long terminal repeat, crucial part of the lentiviral genome

MARS-seq. Plate-based single-cell RNA-seq technology

MeCP2. Methyl CpG binding protein 2

MERFISH. multiplexed error-robust fluorescence in situ hybridization, for in situ transcriptome analysis

Microwell-Seq. Single-cell technology based on sub-nanoliter well plates

MMEJ. Microhomology-mediated end joining

MOI. Multiplicity of infection
 Mosaic-seq. Method for single-cell CRISPR screening
 mRNA. Messenger RNA
 NGS. Next-generation sequencing
 NmCas9. *Neisseria meningitidis* Cas9, a small variant useful for in vivo genome editing
 ORF. Open reading frame
 PAM. Protospacer adjacent motif
 PCR. Polymerase chain reaction
 PDMS. Polydimethylsiloxane
 Perturb-seq. Method for single-cell CRISPR screening
 RNA. Ribonucleic acid
 RNAi. RNA interference
 RNA-seq. RNA sequencing
 SAM. Synergistic activation mediator, a system for CRISPR activation
 scFv. Single-chain variable fragment antibody
 scifi. Single-combinatorial indexing with fluidic indexing
 scifi-RNA-seq. Ultra-high throughput method for scRNA-seq, based on scifi
 sci-RNA-seq. Combinatorial indexing technology for single-cell RNA-seq
 scRNA-seq. Single-cell RNA sequencing
 seqFISH. sequential fluorescence in situ hybridization, a method for in situ transcriptome analysis
 Seq-Well. Single-cell technology based on sub-nanoliter well plates
 Smart-Seq2. Plate-based single-cell RNA-seq technology
 SpCas9. CRISPR-associated gene 9 isolated from *Streptococcus pyogenes*, the most commonly used nuclease for CRISPR genome editing
 SPLIT-seq. Combinatorial indexing technology for single-cell RNA-seq
 STAMPs. Single-cell transcriptomes attached to microparticles
 SunTag system. System for CRISPR activation that uses signal amplification based on a small synthetic peptide paired with a single-chain variable fragment antibody
 TALENs. TALE nucleases consisting of a TALE protein fused to the DNA cutting domain of FokI
 TALEs. Transcription activator-like effectors
 TET1. Ten-Eleven Translocation 1, protein involved in active DNA demethylation
 tracrRNA. Trans-activating CRISPR RNA, that is essential for DNA cleavage by SpCas9

ACKNOWLEDGMENTS

First of all, I would like to thank my best friend and wife Niki. Without your love, support and endless patience, the completion of this thesis would not have been possible. But the zucchini cakes and our library sessions surely helped as well!

I'm extremely thankful to my family, especially my parents Geli & Petz for encouraging me to follow my interests in life (even though I ultimately did not become an astronaut or movie director), and for their advice and support. I would also like to thank Gerti & Kurt for their continuous support.

In particular, I would like to thank my supervisor Christoph Bock for the opportunity to work in his research group, his trust and the freedom to develop my own ideas. Thank you for the hours-long, enthusiastic discussions and for creating such a stimulating research environment. I'm grateful to all members of the Medical Epigenomics lab, in particular Christian Schmidl (for brutally honest feedback), André Rendeiro (for not kicking me out of his office most of the time), Thomas Krausgruber (for generously sharing his knowledge on T-cells and flow cytometry), Peter Traxler (for golden Drop-seq hands), Matthias Farlik (for the exciting time when the lab got started), Amelie Kuchler (for DNA methylation magic) and Thorina Boenke (for making long scifi-RNA-seq experiments fun). Furthermore, I would like to thank everyone at the CeMM Research Center for Molecular Medicine for contributing to its special atmosphere. It was real fun, but also a privilege to pursue my PhD studies surrounded by so many curious and clever people.

Special thanks go to my friends from the Bad Toro PhD generation: Adrián, Máté, Thea, Juliane, Bianca, Michel, Aniko and Anna for many unforgettable evenings, board game sessions and wonderful trips to Pamplona and Pécs. Thanks to you it never felt like work!

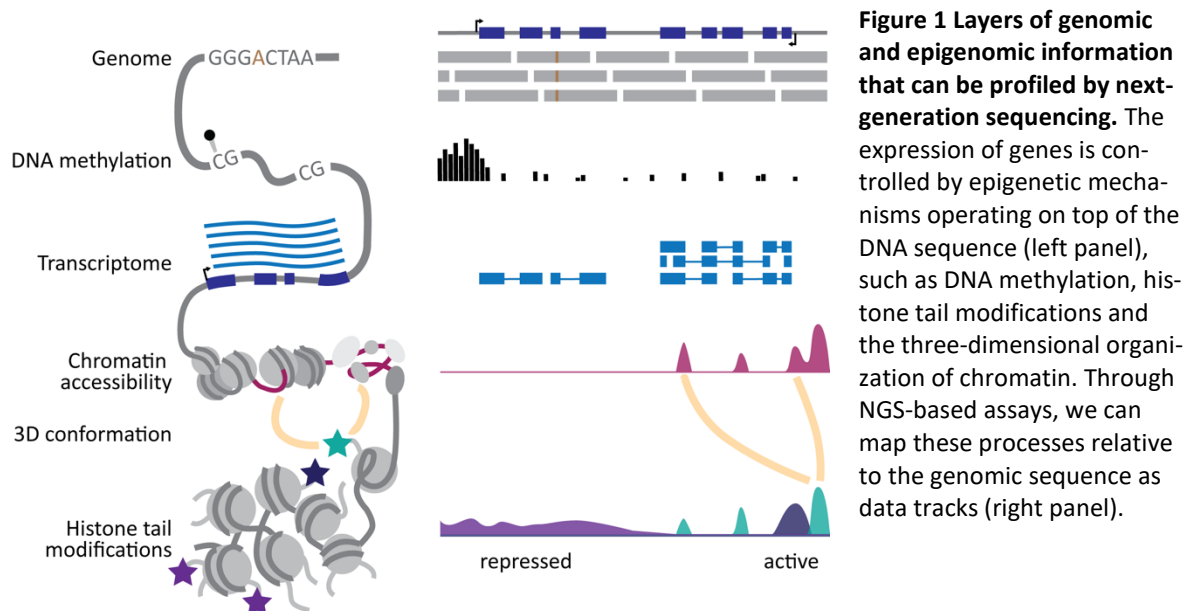
1. INTRODUCTION

1.1. Genome engineering and genetic screens

1.1.1. Profiling the genome, epigenome and transcriptome with sequencing technology

Sequencing the human genome was a major technological achievement and biomedical advance. The work was pioneered by a consortium of academic labs as part of the *Human Genome Project (HGP)* (Lander et al., 2001) and by the private company *Celera* (Venter et al., 2001), using technologies based on electrophoretic sequencing. Unfortunately, the medical impact of the draft sequences was sometimes exaggerated, and the human genome was advertised as the immediate solution to the genetics of many common diseases. In retrospect, this has generated some skepticism and disappointment (Wade, 2010): instead of quickly unlocking an arsenal of potent post-genomic drugs, the projects uncovered an enormous amount of genetic variation in humans that complicates the search for disease-causing DNA mutations.

The past 15 years since the completion of the *Human Genome Project* have been marked by revolutionary developments in DNA sequencing (Shendure et al., 2017). Electrophoretic sequencing has been superseded by *next-generation sequencing (NGS)* technology, a scalable approach resulting in an exponential growth of the sequencing capacity and greatly reduced costs. Whole-genome and exome sequencing enabled large international collaborations mapping pathogenic mutations, such as the *Catalogue of Somatic Mutations in Cancer, COSMIC* (Sondka et al., 2018), or natural genetic variation as exemplified by the *1000 Genomes Project* (1000 Genomes Project Consortium, 2015).



Meanwhile, NGS developed into a versatile platform technology with applications beyond just sequencing DNA (Shendure et al., 2017). By focusing on messenger RNA transcribed from active genes (RNA-seq (Cloonan et al., 2008)), a detailed readout of cellular states is obtained. NGS technologies also allow us to measure how genes are regulated by epigenetic mechanisms, which operate on top of the DNA sequence to define heritable cellular states. These modifications are set up during the

development of multicellular organisms and allow the formation of the plurality of cell types from a single zygote. Their highly context-specific distribution can be profiled genome-wide by employing specialized epigenomic assays. Most commonly, scientists measure DNA methylation (bisulfite sequencing (Bock et al., 2010)), chromatin accessibility (ATAC-seq (Buenrostro et al., 2013)), the three-dimensional conformation of chromatin (Hi-C (Rao et al., 2014)) or histone tail modifications (ChIP-seq (Schmidl et al., 2015)). Figure 1 provides an overview of how DNA is packaged into chromatin and highlights relevant layers of genetic and epigenetic information along with the respective data type.

Epigenomic technologies enabled data collection efforts such as *FANTOM* (Fantom Consortium, 2014), *IHEC/Blueprint* (Stunnenberg et al., 2016), *Roadmap Epigenomics* (Roadmap Epigenomics Consortium, 2015) and *GTEx* (GTEx Consortium, 2015). Through large-scale NGS-based projects, we have learned about natural genetic variation in humans, have created catalogues of somatic mutations in cancer, and have mapped gene-regulatory mechanisms at unprecedented detail. Our functional understanding of the human genome has progressed a lot since the early draft sequences. However, research efforts remain mostly descriptive and experimental validation is lagging behind due to a lack of suitable technologies. We have become experts at observing the genome and how it is regulated, but to realize the full potential of genome medicine requires tools that can introduce precise changes at the genomic and epigenomic level. This gap is increasingly filled by genome and epigenome editing and functional genetic screens, which are the focus of this work.

1.1.2. Genome engineering in the pre-CRISPR era

The first targeted genetic modifications of mammalian cells were achieved by homologous recombination (Capecchi, 1989). The method uses cloned DNA constructs containing an integration cassette that is flanked by homology arms derived from the target locus. Upon delivery to the target cells, intrinsic repair mechanisms incorporate the artificial DNA construct into the genome. While the process itself is inefficient (about 1 on-target event in 10^5 - 10^7 cells (Vasquez et al., 2001)), positive integration events can be selected, for instance by inclusion of an antibiotic resistance marker.

An important finding was that efficiency of homologous recombination could be greatly improved by introducing a DNA double-strand break at the target locus (Rouet et al., 1994). Initially, this was done with rare-cutting restriction enzymes called *meganucleases*. An example is I-SceI with its 18 base-long recognition sequence that does not naturally occur in the human or mouse genomes. Cell lines engineered by homologous recombination to include a single I-SceI target site in a safe-harbor locus were used for efficient genetic modifications, and also contributed to our understanding of DNA double-strand break repair. By introducing a DNA cut of known location and sequencing the target site, additional non-homologous repair mechanisms were discovered that cells can use to join DNA fragments in a non-templated fashion. While these mechanisms can rescue entire chromosomes, they come at the cost of small insertion and deletion mutations at the cut site (indels). This observation was key to the development of genome engineering technologies. Scientists reasoned that introducing a DNA double-strand break in the coding sequence of a gene would be sufficient to disrupt its reading frame by indel formation. Due to the triplet code, this leads to a frame-shift mutation in two thirds of the cases.

This finding motivated biotechnological efforts to engineer a targeted DNA endonuclease that combines sequence-specific DNA binding with DNA cutting. The restriction endonuclease *FokI* has favorable properties: It belongs to the type IIS family of restriction enzymes, which cut outside of their recognition sequence in a sequence-agnostic manner. Most importantly, binding and cutting by *FokI* are performed by two separate protein domains, so that the endonuclease activity could be isolated. Before CRISPR revolutionized genome engineering, specific DNA binding was achieved by combining domains from human zinc finger transcription factors or transcription activator-like effectors (TALEs) from *Xanthomonas* bacteria with the *FokI* nuclease domain. Both proteins are composed of short modules that bind either 3 bases (zinc fingers) or 1 base (TALEs) of DNA, and combinations of the modules can target almost any DNA sequence. By engineering a pair of zinc fingers or TALEs for the target gene and fusing them to the endonuclease domain of *FokI*, which is functional as a dimer, efficient cutting and indel formation could be achieved. Zinc fingers and TALE nucleases (TALENs) continue to be powerful tools for genome engineering, with clinical trials ongoing to achieve therapeutic genome editing. However, their binding specificity is encoded at the protein level and necessitates complex and time-consuming cloning procedures (Reyon et al., 2012).

1.1.3. A prokaryotic immune system revolutionizes genome editing

Truly groundbreaking developments in biomedicine are often enabled by new technologies that arise from seemingly unrelated research areas. The study of CRISPR/Cas systems and their repurposing as a versatile genome engineering platform is a particularly beautiful example of this phenomenon. CRISPR, short for *clustered regularly interspaced short palindromic repeats*, were independently discovered in 1987 and 1993 as an obscure DNA sequence in the genomes of bacteria and archaea (Ishino et al., 1987, Mojica et al., 1993). CRISPR loci were found to contain a series of near-palindromic repeats, separated by unique spacer sequences - but their function remained enigmatic until the advent of prokaryotic comparative genomics.

In 2005, bioinformatic analyses by three groups (Bolotin et al., 2005, Mojica et al., 2005, Pourcel et al., 2005) revealed that the spacer sequences in CRISPR loci were derived from viruses and plasmids encountered by the microbe, which suggested that CRISPR functions as an adaptive immune system. The first experimental confirmation of this hypothesis came from the analysis of CRISPR loci of *Streptococcus thermophilus* cultures used in dairy production (Barrangou et al., 2007). When wild-type strains were challenged with bacteriophages, resistant mutants emerged which had accumulated new spacer integrations in their CRISPR loci that were derived from phage DNA. Upon deletion of the newly acquired spacer sequences, the resistance was lost.

To uncover the enzymatic basis of the observed resistance mechanism, scientists started investigating *CRISPR-associated genes (Cas genes)* in the immediate vicinity of the CRISPR locus. It was found that the gene *Cas9*, with two prominent nuclease domains, was absolutely essential for acquired phage resistance (Barrangou et al., 2007). The discovery that *Cas9* cleaves DNA (Marraffini & Sontheimer, 2008), and not RNA, sparked the biotechnological interest to repurpose the system for genome editing.

Co-purification experiments revealed that *Cas9* interacts with small *CRISPR RNAs (crRNAs)*. Since their sequence corresponded to one spacer plus parts of the repeats, their likely role was to program the

nuclease in a sequence-specific manner. Nevertheless, crRNAs alone were not sufficient to reconstitute the CRISPR/Cas9 system *in vitro*. The last crucial bits of information came from next-generation sequencing experiments of small RNAs in *Streptococcus pyogenes*. It became clear that CRISPR loci are transcribed as long precursor RNAs, that are processed into smaller crRNAs. An abundant RNA species with 25 bases of near-perfect complementarity to crRNAs was identified, that was found to assist in their processing by RNase III and was essential for DNA cleavage (Deltcheva et al., 2011). It was named *trans-activating CRISPR RNA (tracrRNA)*. This made the CRISPR/Cas system of *S. pyogenes* the best understood at the time and the prime candidate for reconstituting the system *in vitro*.

In 2012, about 25 years after the initial observation of CRISPR loci, two groups reported efficient DNA cleavage (Gasiunas et al., 2012, Jinek et al., 2012) with four isolated components from *S. pyogenes*: the nuclease Cas9, RNase III, crRNA and tracrRNA. They demonstrated that the system can be reprogrammed by including custom spacers into the crRNA, and that shortening the spacer length to 20 nucleotides results in more efficient cutting. To simplify the workflow, crRNA and tracrRNA were fused into one chimeric *guide RNA (gRNA)* (Jinek et al., 2012), which has become the standard for most applications. The system proved surprisingly stable when expressed in other organisms, and within a year efficient genome editing in human and mouse cells was demonstrated (Cong et al., 2013).

These seminal papers established SpCas9 (for *S. pyogenes* Cas9) as an RNA-guided DNA endonuclease with great potential for genome engineering. In contrast to earlier technologies such as zinc fingers and TALENs, CRISPR does not require complex cloning strategies for protein re-engineering. The cutting specificity is determined by the short 20-nucleotide-long gRNA, which can be efficiently encoded with conventional or array-synthesized DNA oligonucleotides. Being both simple and scalable, CRISPR was quickly and universally adopted. The rapid development of CRISPR was also greatly facilitated by previous work on *RNA interference (RNAi)*. Many of the technical challenges of delivering and expressing small RNAs had already been overcome, such that scientists could select from a large collection of vector systems that needed only slight modifications.

1.1.4. Practical example: Single-gene knockouts with SpCas9

SpCas9 is a highly efficient DNA endonuclease that can be programmed with a small chimeric guide RNA to cut at virtually any genomic locus (Figure 2a). The gRNA consists of a constant *backbone* interacting with SpCas9 through pronounced secondary structures. It is preceded by a 20 nucleotide long *spacer* sequence for specific binding to the DNA target via direct base pairing (Ran et al., 2013). The target sequence is referred to as the *protospacer* and must be followed immediately by the *protospacer adjacent motif* or *PAM* (NGG is used for SpCas9, but this sequence varies between CRISPR systems). Of note, the PAM is not included in the gRNA, but rather constitutes an important mechanism of self-tolerance for CRISPR/Cas immune systems. It ensures that only invading DNA is cut, while the corresponding spacer in the CRISPR locus remains intact.

Wild type SpCas9 uses two nuclease domains to cut the complementary (HNH domain) and non-complementary (RuvC-like domain) DNA strands, resulting in a blunt DNA double-strand break exactly 3 bp upstream of the PAM (Jinek et al., 2012). The repair of these breaks by cell-intrinsic mechanisms proceeds with comparably slow kinetics, presumably because SpCas9 remains tightly bound to the target locus (Brinkman et al., 2018). The primary repair pathway leading to SpCas9-induced lesions is canonical non-homologous end joining (C-NHEJ), followed by microhomology-mediated end joining

(MMEJ) (Brinkman et al., 2018). Both mechanisms result in short insertion or deletion mutations with a random, but highly context-specific distribution. If the cut occurs inside an open-reading frame encoding for a protein, the repair outcome will shift the reading frame of the triplet code in roughly two thirds of the cases, resulting in missense and premature stop codons.

Figure 2b provides a real-world example of a single gene knockout achieved by targeting SpCas9 to exon 4 of the human DNMT1 gene. Sanger sequencing of clonally expanded cell lines identified a 20 bp deletion resulting in a premature stop codon immediately downstream. The truncated protein is missing the crucial DNA methyltransferase domain and constitutes a functional knockout.

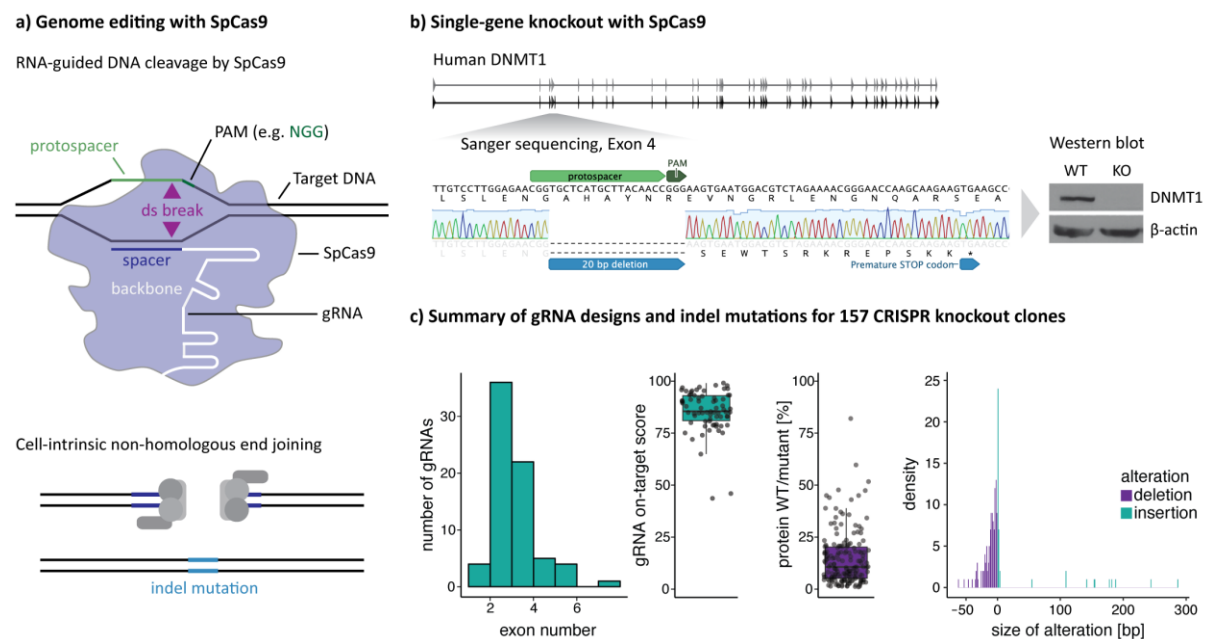


Figure 2 Genome editing with the CRISPR/Cas9 system. a) Schematic of SpCas9 in complex with the gRNA and target DNA. The induced double-strand break is repaired by non-homologous end joining, resulting in small insertion or deletion mutations. b) Example of a single-gene knockout achieved by targeting SpCas9 to exon 4 of the human DNMT1 gene. After clonal expansion, Sanger sequencing detected a 20 bp deletion resulting in a premature stop codon downstream. The truncated protein was not detected on a Western blot. c) Summary of the gRNA designs and editing outcomes for a total of 157 single knockout cell lines to illustrate the basic properties of editing with SpCas9.

Figure 2c summarizes gRNA designs and editing outcomes for a total of 157 knockout cell lines that I produced in the haploid cell line HAP1. This data illustrates the basic principles of genome engineering. When designing gRNA spacers to induce a gene knockout, it is important to have a good understanding of the target gene. Websites like Ensembl or Uniprot can provide an overview of the available isoforms and important protein domains before starting the design phase. Ideally, gRNAs should be placed in the earliest exon shared by all isoforms and should cut prior to key functional domains. In the presented dataset of 157 knockout lines, most genes could be targeted in exons 2-3 with highly active gRNAs (average on-target activity score (Doench et al., 2014) of 85.5). On average, 14.6% of the protein coding sequence remained intact after the frameshift, with some significant outliers. In such cases, it is important to ensure that key functional and structural domains are affected by the

frameshift. If this is not possible, functional protein domains can be targeted directly, because they are highly vulnerable to small insertions or deletions (Shi et al., 2015). Each of the genome-edited cell lines was clonally expanded, and the respective target sites were PCR-amplified from genomic DNA. Out of 157 indel mutations analyzed by Sanger sequencing, 133 (72%) were deletions and 44 (28%) were insertions. The size of deletions ranged from 1 to 60 base pairs. Indel mutations tended to be more variable in size: while the majority measured 1-3 base pairs, insertions of up to 300 bp were detected in rare cases.

1.1.5. Notable alternatives to SpCas9

Given the immense popularity of *Streptococcus pyogenes* Cas9 (*SpCas9*), the diversity of CRISPR/Cas systems (Koonin et al., 2017) is sometimes overlooked. As more of these prokaryotic immune systems are discovered through metagenomic data mining and functionally characterized, they will most certainly surprise us with many unexpected features. Here, I would like to provide a short overview of biomedically relevant alternatives to *SpCas9* and some of their most interesting applications.

Before turning to orthologous Cas proteins, I would like to highlight recent protein engineering efforts seeking to improve the properties of *SpCas9* itself. For instance, the available crystallographic data of *SpCas9* in complex with target DNA allowed the rational engineering of variants with greatly reduced off-target activity by mutating amino acids that mediated unspecific interactions to DNA (Kleinstiver et al., 2016, Slaymaker et al., 2016). More recently, the non-catalytic REC3 domain of *SpCas9* was shown to sense target complementarity as part of a proofreading mechanism – resulting in a new design called HypaCas9 (Chen et al., 2017). Similarly, screening of an *SpCas9* library with random mutations in the REC3 domain in yeast cells resulted in a highly specific variant called evoCas9 (Casini et al., 2018). Lastly, several groups have succeeded in altering and broadening the PAM recognition of *SpCas9*, thereby extending its targeting space (Hu et al., 2018, Nishimasu et al., 2018).

Cas12a (or Cpf1) is a single-protein CRISPR effector with several key properties (Zetsche et al., 2015). Most importantly, it has the intrinsic ability to process the crRNA array into mature crRNAs and does not require a tracrRNA. This greatly enhances its multiplexing capability (Zetsche et al., 2017). Cpf1 recognizes a short T-rich PAM and results in a staggered cut with a 5' overhang that is advantageous to introduce DNA into the genome via non-homology directed mechanisms.

Smaller Cas9 orthologues, such as NmCas9 (*Neisseria meningitidis*) and CjCas9 (*Campylobacter jejuni*) are useful for *in vivo* genome editing, because they are still within the packaging capacity of adeno-associated viruses (AAVs) (Cebrian-Serrano & Davies, 2017). This type of viral vector has displayed a good clinical efficacy and safety profile in animal models and human clinical trials (Liao et al., 2017). Both enzymes have more stringent PAM requirements than *SpCas9*, resulting in less off-target activity.

Cas13a (or C2C2) targets RNA instead of DNA. For some Cas13a orthologues, the recognition of the RNA target results in promiscuous RNase activity that can be detected with a quenched fluorescent RNA reporter. Combined with isothermal amplification and *in vitro* transcription methods, Cas13a has been used to develop a highly sensitive method for CRISPR-based RNA detection and diagnostics (Gootenberg et al., 2017), which successfully identified different strains of human pathogens. It has also been used to profile mutations in cell-free tumor DNA. More recently, an ortholog of Cas13a

avoiding collateral RNA degradation (LwaCas13a) has been used for targeted RNA knockdown and binding in mammalian cells (Abudayyeh et al., 2017).

1.1.6. Extending the functionality of SpCas9 beyond cutting DNA

Turning SpCas9 into a versatile recruitment platform

Within just a few years, CRISPR/Cas9 has developed into a rich biotechnology toolbox. This is thanks to numerous creative improvements by the scientific community that have extended its functions well beyond genome editing. Most methods use a catalytically inactive version of Cas9 (*dead Cas9* or *dCas9*) that still binds to the target locus and can serve as a versatile recruitment platform for effector proteins. dCas9 carries inactivating mutations in both the RuvC and HNH nuclease domains (Jinek et al., 2012) (most commonly D10A and H840A), but is otherwise identical to SpCas9. Alternative functions can be incorporated into the enzyme by fusion of effectors to dCas9, or by inclusion of small RNA aptamers into the gRNA that are recognized by specialized proteins fused to modifier proteins.

Another option to prevent SpCas9 from cutting is to shorten the gRNA spacer length to 14-15 bases (Dahlman et al., 2015). These *dead RNAs* (*dRNAs*) allow for efficient, but slightly more promiscuous DNA binding and can include aptamers for recruitment of effector proteins. dRNAs are a good option to combine DNA cutting with another function, and can be practical to use alternative functions in primary cells already expressing wild-type SpCas9, such as those obtained from SpCas9 mouse models (Chu et al., 2016, Platt et al., 2014). However, since dCas9 is more practical in most cases, it will be the focus of this section.

Targeted gene activation (CRISPRa)

Until recently, gain-of-function experiments were limited to the overexpression of cDNAs in the target cells. Cloning open-reading frames (ORFs) and optimizing their expression for a particular cell type is technically challenging and does not scale well. Despite these limitations, large collections of ORF libraries for gain-of-function screening have been established by academic consortia and commercial vendors (Sack et al., 2018). In the past few years, technologies for *CRISPR activation* (*CRISPRa*) have been developed, which promise a greatly simplified workflow compared to ORF-based methods. Here, I focus on the three most popular designs. All are based on dCas9 but differ in the choice of effector proteins and mode of recruitment.

CRISPRa based on synergistic activators and RNA aptamers for recruitment

Mammalian gene activation often employs synergistic factors that integrate input from multiple converging signaling pathways. Building on this basic concept of gene regulation, scientists set out to improve CRISPR gene activation by combining well-known activator proteins and protein modules with complementary modes of action. This resulted in two of the most popular designs for CRISPRa.

The dCas9-VPR system resulted from the systematic testing of 20 candidate effectors fused to dCas9 for their ability to activate a fluorescent reporter construct. This identified *VP64*, *p65* (transcription factor of the NF- κ B family) and *Rta* (transcriptional activator of *Epstein-Barr Virus*) as potent activators (Chavez et al., 2015). Through a series of systematic experiments, the three factors could be combined into a tripartite design called dCas9-VPR (for VP64, p65 and Rta). The dCas9-VPR system is the most

compact design for CRISPRa at the moment and the only efficient single-component system based on only a dCas9 fusion. This can be advantageous when the delivery of multi-component systems is challenging, such as in primary cells, or when the system has to be delivered as mRNA or protein.

The *Synergistic Activation Mediator (SAM)* system was developed based on the crystal structure of SpCas9 in complex with the gRNA and target DNA (Nishimasu et al., 2014). Importantly, two parts of the gRNA (tetra loop and stem loop 2) were found to protrude outside of the nuclease complex and can tolerate the insertion of small RNA aptamers (Konermann et al., 2015). RNA aptamers are often found in the RNA genomes of viruses, and interact with a specific protein partner. Specifically, the SAM system incorporates two hairpins from the MS2 bacteriophage into the gRNA, which can be recognized by a dimer of MS2 bacteriophage coat proteins (MCP).

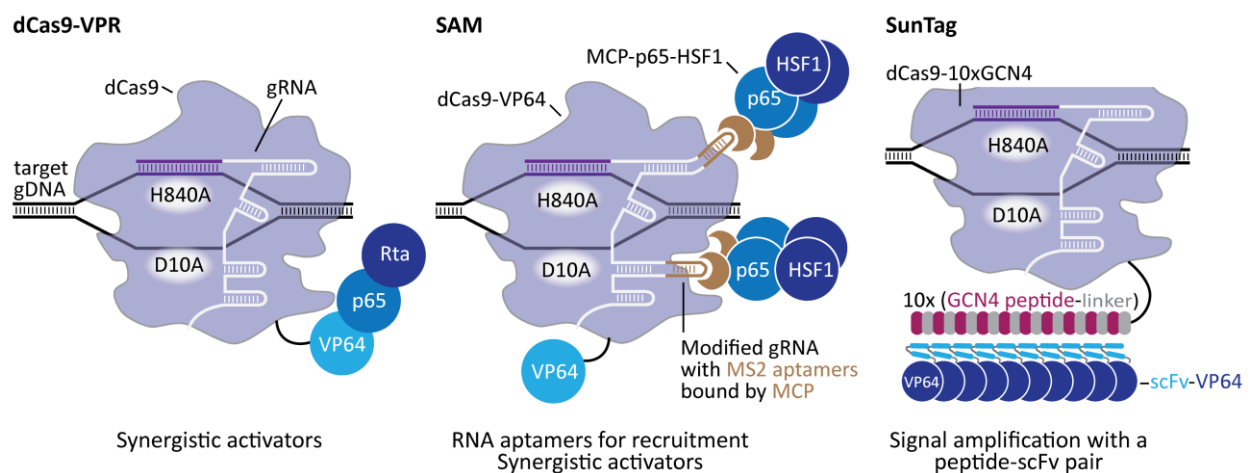


Figure 3 Three popular designs for targeted gene activation based on CRISPR. dCas9-VPR is based on a tripartite activator fusion of VP64, p65 and Rta. It is currently the only efficient design for CRISPRa using only a single effector protein, which can have advantages when the delivery of the components is challenging. The SAM system recruits additional activator proteins (p65 and HSF1) via MS2 RNA aptamers included in the gRNA. The SunTag system achieves efficient gene activation by signal amplification with a peptide-scFv antibody pair.

Efficient gene activation is achieved by fusing MCP to two effector proteins: p65 (as in dCas9-VPR) and HSF1, the activation domain from human heat-shock factor 1. In addition to recruiting four copies of MCP-p65-HSF1, the system also uses a dCas9-VP64 fusion. The original publication of the SAM system already introduced genome-wide activation screening for mediators of BRAF inhibitor resistance. The gRNA library for CRISPR activation was made available to the research community, which helps to explain its wide adoption.

CRISPRa based on VP64 and signal amplification with the SunTag

VP64 is a strong, synthetic transcriptional activation domain consisting of four tandem copies of the *Herpes Simplex Virus* protein VP16 that had already been successfully used for transcriptional activation with TALE proteins (Maeder et al., 2013b). Despite the optimized features of this module, early experiments with a simple dCas9-VP64 fusion targeted to gene promoters by single gRNAs proved ineffective (Maeder et al., 2013a, Perez-Pinera et al., 2013). In many cases the activation was below 2-fold and did not translate into a full biological response, as apparent when trying to reprogram fibroblasts to neuronal cells (Black et al., 2016). Scientists realized that this limitation can be overcome

by supplying several non-overlapping gRNAs for the same promoter, suggesting that efficient gene activation requires multiple copies of VP64. The SunTag system introduces an effective means of signal amplification based on a small synthetic peptide (GCN4) paired with a specific *single-chain variable fragment* (scFv) antibody (Tanenbaum et al., 2014). For CRISPR activation, an array of ten GCN4 peptides is C-terminally fused to dCas9, recruiting up to ten copies of GB1-scFv-VP64-sfGFP. The GB1 domain and super-folder GFP are required to enhance solubilization and folding of the antibody. This recruitment mode allows for efficient gene activation with a single gRNA, which is a prerequisite for CRISPRa-based gain-of-function screens.

Targeted inhibition of transcription (CRISPRi)

Certain situations call for a more controlled means of gene inactivation. In a gene therapy setting, it might be beneficial to repress a gene without DNA cleavage, to avoid undesired and unpredictable repair outcomes. Other experiments, such as functional screens, can improve their statistical power by perturbing all cells in a population in a homogenous way. CRISPR inhibition (CRISPRi) was first demonstrated in *Escherichia coli*, where the targeting of dCas9 to a promoter region is sufficient to disrupt the simple bacterial transcriptional machinery by steric hindrance (Qi et al., 2013). In mammalian cells, additional effector domains are required for efficient inhibition. The most commonly used repressive domain is KRAB (Gilbert et al., 2013) (for *Krüppel-associated box*) that naturally occurs in mammalian zinc finger transcription factors (Thakore et al., 2016). The recruitment of KRAB to a target promoter or distal regulatory element initiates heterochromatin formation by deposition of histone methylation and the removal of histone acetylation. More recently, it was described that the inclusion of MeCP2 into the repressive complex (to form dCas9-KRAB-MeCP2) can improve the robustness and efficiency of CRISPRi (Yeo et al., 2018). MeCP2 mediates the repressive effect of DNA methylation by binding to methylated CpG dinucleotides and compacting the chromatin.

Epigenome editing

Epigenetic mechanisms are the foundation for the development of multicellular organisms. In human development, they enable a single zygote with its defined genome to develop into the estimated two hundred, highly-specialized cell types of the human body – ultimately forming about 37 trillion cells (Bianconi et al., 2013). The epigenome describes the complete set of epigenetic modifications of a given cell, and defines its cellular identity. These mechanisms lead to heritable changes in gene expression without altering the underlying DNA sequence and can manifest as covalent modifications of DNA bases (e.g. DNA methylation), histone tail modifications, chromatin accessibility, or the 3-dimensional arrangement of chromatin - to name just a few of the diverse, multi-layered mechanisms. They are established by hundreds of proteins, functioning as writers, readers and erasers of a modification.

Through decades of research, we have understood that particular epigenetic features are characteristic for certain functional elements in the genome, such as promoters, gene bodies, enhancers, and repressed heterochromatin. These features have been mapped in great detail for both healthy and disease-derived cell types and tissues. Nevertheless, it is often hard to decide whether an epigenetic mark is causative for a particular phenotypic change, or just correlates with it. The situation is further complicated by the pronounced interactions between different layers of the epigenetic landscape. Ultimately, it will be important to define how epigenetic marks interact and depend on one another.

To address these complex questions, the research community has built a suite of *epigenome editing* tools that uses genome-targeting with CRISPR/Cas9 to direct epigenetic modifier proteins to a particular locus for functional epigenetic studies. The following paragraph summarizes the most promising approaches for epigenome editing that were introduced in recent years.

DNA methylation is well known for its negative regulatory effect on gene promoters, but can also affect transcription factor binding (Yin et al., 2017) and the activity of enhancers. A recently identified pathway for active DNA demethylation (Kriaucionis & Heintz, 2009, Tahiliani et al., 2009) was found to be mutated in many human cancers (Kohli & Zhang, 2013). To clarify the diverse functions of this epigenetic modification, its writer protein DNMT3A (Huang et al., 2017) and the eraser TET1 (Morita et al., 2016) have been targeted to dCas9 via the SunTag system, leading to measurable changes in the DNA methylation state and gene expression. While most epigenetic modifiers have the desired localized effect, DNMT3A displayed widespread off-target activity and will require further optimization (Galonska et al., 2018).

Gene promoters are the regions upstream of a gene's coding sequence where the transcriptional machinery assembles, and are marked by the histone modification H3K4 trimethylation when the promoter is active. By fusing the H3K4 methyltransferase PRDM9 to dCas9, moderate activation of a gene promoter could be achieved (Cano-Rodriguez et al., 2016). Of note, this process required artificial DNA demethylation by 5-azacytidine, which greatly limits the application of this epigenetic editing tool.

The functional assessment of enhancer regions is the most promising application of genome editing in the near future. By fusing the catalytic core domain of p300 to dCas9 (Hilton et al., 2015), H3K27 acetylation could be deposited at putative enhancer regions, resulting in the activation of interacting genes. The opposite was achieved with a dCas9-LSD1 construct, which successfully removed the typical H3K4 monomethylation (Kearns et al., 2015). Importantly, LSD1 was enhancer-specific and did not have an effect on the promoter region of the same gene.

Other prominent histone modifiers tested for epigenome editing include the mostly repressive H3K9 methyltransferases SUV39H1 and G9a, as well as EZH2 for H3K27 methylation. More exotic applications include the direct manipulation of the chromatin architecture by de novo loop formation (Morgan et al., 2017).

Epigenome editing tools mark a paradigm shift for epigenomics. Scientists are no longer limited to purely observing the epigenome, but can formulate testable hypotheses and address them with functional experiments. Before long, the descriptive bioinformatic analysis of epigenomic data will be followed by a series of epigenome editing experiments. If done at sufficiently high throughput, this can help to narrow down long lists of identified epigenomic alterations, prioritize those with a causal link to a disease and work on innovative epigenetic therapies. To fulfill this vision, we require large-scale screening methods, such as those introduced in the next section.

1.1.7. Pooled functional screens with CRISPR

The vast number of genetic and epigenetic variants identified by next-generation sequencing for which no causative link to the disease can be established calls for novel validation technologies. There is an increasing need for methods that can recapitulate a large collection of perturbations and evaluate their functional consequences. Pooled screens allow the comprehensive identification of genes

involved in biological processes, and the exhaustive functional analysis of gene variants. Consequently, they have accelerated and enriched basic biological discovery, but also helped to address urgent clinical problems. Pooled screening can take many forms, but generally involves four steps: (1) the introduction of targeted, traceable perturbations into a population of cells, (2) a strong means of selection for a functionally relevant sub-population, (3) an efficient readout of perturbations, (4) the identification of relevant perturbations by comparison of selected and unselected samples.

CRISPR-based genome and epigenome engineering technologies are ideally suited for pooled screening, because SpCas9 is targeted to DNA by a small gRNA. An effective means of introducing gRNAs into mammalian cells is to supply them as a DNA template under an RNA polymerase III promoter that ensures strong transcription with a defined start and end point (most commonly the human U6 promoter is used). Once transcribed inside the cell, gRNAs assemble with the SpCas9 effector complex and direct it to the target genomic locus specified by their 20 base-long spacer sequence. This short, specificity-encoding part of the gRNA is what makes the CRISPR system so incredibly scalable. Cloned with simple DNA oligonucleotides, it can be synthesized cost-effectively and at high throughput using array-based methods scaling to 1 million unique sequences (Kosuri & Church, 2014).

This has enabled the production of genome-wide gRNA libraries for CRISPR knockout, inhibition and activation that were designed to target every human or mouse gene with multiple gRNAs based on sophisticated design rules (Sanson et al., 2018). These libraries contain roughly 50,000 to 200,000 unique gRNA templates. Delivery by lentiviral particles at a low multiplicity of infection (MOI) ensures that each cell receives only a single gRNA. The DNA packaged into these vectors is integrated into the genome of the target cells, leaving a permanent record of the perturbation that cells retain throughout the selection procedure. The small size of the integrated gRNA spacers facilitates their detection by PCR and next-generation sequencing, such that they can be counted and compared between samples (Shalem et al., 2014, Wang et al., 2014). While the overall workflow is similar to pooled genomics screens with RNA interference (RNAi) (Whitehurst et al., 2007), CRISPR significantly outperforms RNAi by producing gene knockouts instead of just reducing the expression of the target (Evers et al., 2016, Morgens et al., 2016). Despite these advantages, pooled CRISPR screens still require a well-planned experimental design. At the heart of a successful screen is a selection method relevant to the biological question. The following section introduces key concepts for designing pooled CRISPR screens and includes examples from the recent research literature. It also introduces functional screening in single cells, a novel technology developed as part of this thesis.

Negative selection screens

When perturbed cells are cultured over an extended period of time, gRNAs targeting genes for proliferation and survival will drop out of the population. Negative selection screens require sophisticated statistical frameworks to distinguish potential hits from dropouts due to technical noise, because they are trying to prove that certain gRNAs are missing from the expanded cell population, compared to the initial population. While the concept of negative selection sounds straightforward, such screens are actually notoriously hard to design and perform well, requiring high coverage and many replicates.

Negative selection approaches have been used to identify genes and cellular processes that are essential for the survival of human cells (so-called *essentialomes*) (Blomen et al., 2015, Wang et al., 2015). About 1500 genes were mapped as essential across multiple cell lines, defining the minimum

functional requirements for a human cell. Most essential genes, however, appear to be context-specific (Hart et al., 2015). This could offer interesting therapeutic opportunities for cancer research. A number of international projects are collecting negative selection CRISPR screening data to uncover cancer and genotype-specific vulnerabilities that could eventually result in new, personalized therapies (e.g. DepMap, Project Achilles and PICKLES).

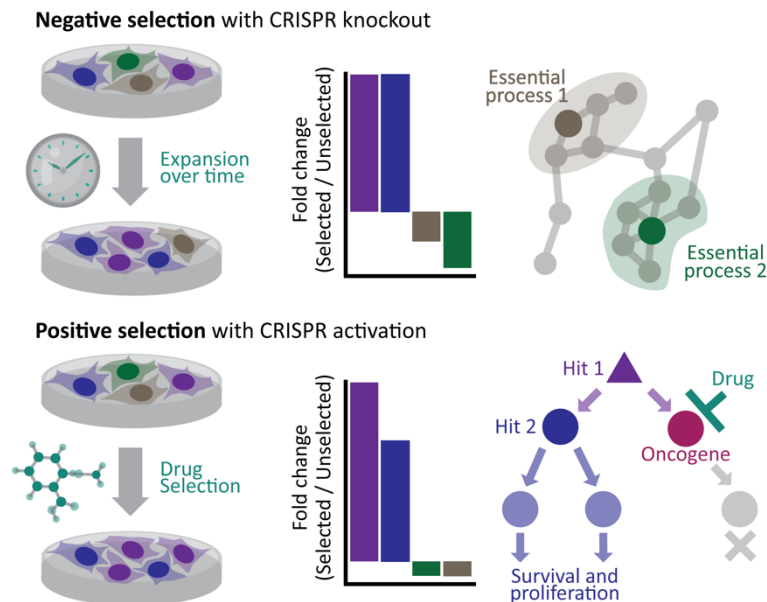


Figure 4 Pooled CRISPR screening with positive or negative selection. In its simplest form, negative selection involves the expansion of cells over an extended period of time. gRNAs targeting essential genes and processes will gradually drop out of the population. Positive selection identifies cells that survive a challenge. In the example in the lower panel, a population of genome-edited cells is treated with a drug blocking an essential oncogenic signaling pathway. Screening with CRISPR activation can identify potential resistance mechanisms (i.e. an alternative, parallel survival pathway).

If designed well, negative selection CRISPR screens can have an immediate impact on patient care. Recently, a screen using saturation genome editing (Findlay et al., 2014) on all 13 exons of the tumor suppressor BRCA1 in the haploid cell line HAP1 (where BRCA1 is essential), classified the pathogenicity of almost 4,000 single nucleotide variants (Findlay et al., 2018). Since germ-line mutations in BRCA1 can pre-dispose women to breast and ovarian cancer, this dataset is of tremendous value to interpret the risk of a patient's specific mutation when recommending surgery.

Positive selection screens

In positive selection screens, a population of perturbed cells is presented with a strong challenge, such as viral infection or a chemotherapeutic drug. While most cells will stop proliferating or die, some might carry resistance-conferring mutations and expand. Typically, resistant cells constitute a minor fraction of the final population, such that only a small portion of the initial gRNA library is detected. This makes positive selection screens technically more robust, especially when working with limiting cell numbers or in primary cells. Likewise, non-functional gRNAs are less likely to produce false-negative results, compared to negative selection designs.

This type of screen has been especially effective for identifying host factors required for viral infection. Amongst others, entry receptors for ebola virus (Carette et al., 2011) and adeno-associated virus (Pillay et al., 2016) as well as important host factors for infection with HIV (Park et al., 2017) and *Flaviviridae* (Marceau et al., 2016, Savidis et al., 2016) were identified (*Flaviviridae* include dengue virus, zika virus and hepatitis C virus).

Furthermore, positive selection screens combined with CRISPR activation can mimic mechanisms of acquired drug resistance in cancer. The specific inhibition of mutated oncogenes with small molecules is a promising therapeutic option, as exemplified by the drug vemurafenib targeting mutant BRAF in melanoma, but is often undermined by the upregulation of alternative oncogenic pathways (Bollag et al., 2012). The first genome-wide CRISPR activation screen ever performed (Konermann et al., 2015) introduced gain-of-function changes into the A375 melanoma cell line (harboring V600E, the most common BRAF mutation) treated with vemurafenib and uncovered many of the resistance mechanisms that had been described in patients. This goes to show that functional screening can identify resistance mechanisms quickly and comprehensively, which would take years if done on patient data.

Screens based on fluorescence-activated cell sorting (FACS)

In contrast to what the previous examples might suggest, functional screens are not limited to simple live/dead situations. To allow the investigation of more complex biological questions, relevant cell populations can be enriched by fluorescence-activated cell sorting (FACS). In FACS, intracellular or extracellular features of cells are labeled with fluorophore-conjugated antibodies or by fluorescent proteins. The cell suspension is processed with a cell sorter, where it flows through a vibrating nozzle to generate single-cell containing droplets. Upon excitation by lasers, labelled cells release a fluorescent signal that is filtered and detected. Droplets of desired properties are marked with an electric charge, deflected off the general path and isolated. Today's FACS instruments have reached a throughput that is sufficient for genome-wide screening, but the sorting is still quite time-consuming and labor-intensive, thus limiting the number and scale of screens that can be performed. Sorting complex marker combinations is possible, but most FACS-based CRISPR screens published to date have looked only at a single protein marker.

A particularly interesting example of a surface marker used for FACS-based screening is PD-L1, which is often expressed by tumor cells to evade the immune system. Presented on the cell surface, PD-L1 can inhibit tumor-infiltrating T-cells by engaging their negative co-receptor PD-1. By genome-wide CRISPR screening, two groups identified CMTM4 and CMTM6, two transmembrane proteins of previously unknown function that extend the half-life and surface presentation of PD-L1 (Burr et al., 2017, Mezzadra et al., 2017). This makes CMTM4 and CMTM6 inhibition an attractive target to render tumors more vulnerable to T-cell based cancer immunotherapy.

The fluorescent signal for FACS-based CRISPR screens does not necessarily need to come from the cellular system itself. A recent screen in a macrophage cell line model used beads coupled to pH-sensitive and insensitive dyes to distinguish cell populations that had successfully completed phagocytosis and lysosome acidification (Sedlyarov et al., 2018). This example shows that multiple relevant populations can be isolated and screened simultaneously.

Ultimately, the number of parameters measured in FACS-based screens is limited by the availability of specific antibodies, fluorophores and their spectral overlap. Today's most sophisticated cell sorters employ up to six separate laser lines, each connected to an array of filters and detectors, and they run advanced compensation software to account for spectral overlap. Nevertheless, FACS is still limited to tens of parameters, and cannot comprehensively characterize the identity of a cell. Consequently, a FACS-based screen needs to describe a complex biological phenomenon with just a few sortable marker proteins.

Pooled CRISPR screens with single-cell RNA-seq readout

While FACS-based screening is well-suited for the relatively simple questions outlined above, many biological questions are at risk of being over-simplified when reduced to a few sortable marker proteins. This can be illustrated with a recent publication that addresses the differentiation of murine T helper cells into Th2 cells with a genome-wide genetic screen (Henriksson et al., 2019). The study uses mouse reporter stains with GFP placed under promoters responsive to key transcription factors of the Th2 lineage (Gata3, Irf4 and Xbp1) as well as antibody staining (Il4, Il13). By comparison of gRNAs in cells with high reporter levels to the dim population, the study identified positive and negative regulators of Th2 differentiation.

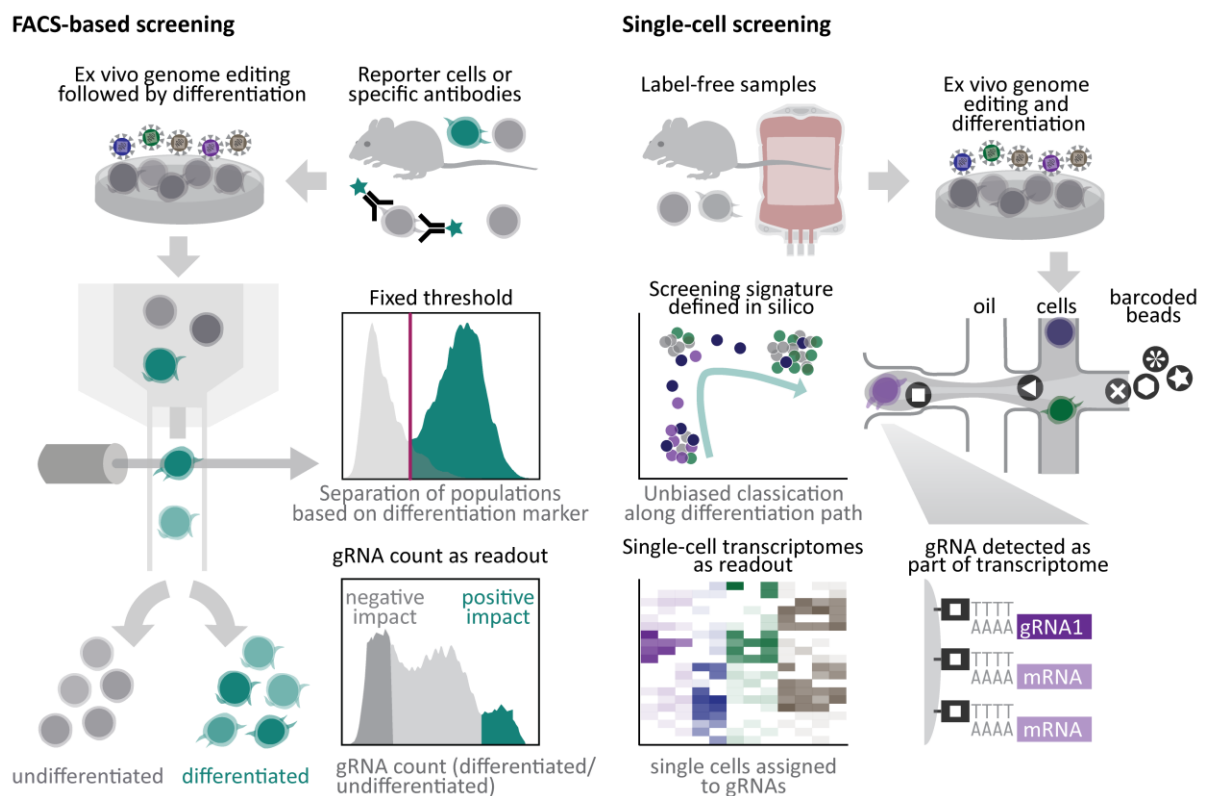


Figure 5 CRISPR screening at the single-cell level provides unique insights. FACS-based CRISPR screens require specific antibodies or reporter systems. Since they are limited to tens of parameters, they are also at risk of over-simplifying complex biological questions such as cellular differentiation. In contrast, single-cell based CRISPR screens can work on unlabeled samples, and measure the full transcriptomic response of each single cell in response to genetic perturbations. Thereby, they can assess the effect of a perturbation in different cell types, or along a developmental trajectory.

There are several limitations to this approach. First, experimenters are limited to well-studied systems where sortable markers exist, and rely on available mouse strains or antibodies. Crucially, they have to define a cutoff value for sorting which will have a great impact on the results, i.e. the GFP expression level that defines a true Th2 cell. Second, since gRNA sequences are the only readout of the screen, it requires extensive follow-up by additional methods such as bulk RNA-seq, ATAC-seq and ChIP-seq to understand the effects of a certain perturbation at the molecular level. Third, it would be interesting if perturbations could be studied over the time course of differentiation by classifying the fate of a cell

in an unbiased way. Single-cell technologies proved invaluable to answer this type of questions, most importantly single-cell RNA sequencing (scRNA-seq). We therefore set out to combine pooled CRISPR screening with a single-cell RNA-seq readout, resulting in a new method called *CRISPR droplet sequencing (CROP-seq)*. The next section will introduce scRNA-seq technologies with a particular focus on those that have a sufficient throughput for screening.

1.2. Implementing single-cell CRISPR screening

The previous section introduced single-cell CRISPR screens at the conceptual level. Here, I focus on implementing this concept using state-of-the-art technologies for single-cell RNA-seq (scRNA-seq). When combining CRISPR screening with single-cell analyses, choosing the right technology platform becomes crucial to ensure sufficient throughput, cost efficiency and data quality. This section is a review of current scRNA-seq technologies that are suitable for pooled screening, with a particular focus on their future potential. As an introduction to our publication on *CRISPR droplet sequencing (CROP-seq)*, I will also discuss how CRISPR gRNAs can be detected in a platform-agnostic manner to assign transcriptomes to particular gene perturbations. Along the way, I will discuss the future potential of single-cell genetic screens, and the technological advancements required to achieve their full potential. Leading up to our manuscript on *scifi-RNA-seq*, I will introduce a novel concept named *single-cell combinatorial indexing with fluidic indexing (scifi)* that we developed to generate single-cell transcriptomes at ultra-high throughput, with minimal costs.

1.2.1. Single-cell RNA-seq platforms with sufficient throughput for pooled screening

In pooled CRISPR screens, the gRNA representation is a key consideration for robust results with low technical variability. Typically, screens are done at 100 – 1000x coverage, referring to the number of cells that each individual gRNA is expressed in. But what are the throughput requirements for screening at the single-cell level? Most certainly, the answer will depend on the biological response profiled. The exact coverage requirements for single-cell screens remain to be determined empirically, but are likely lower than for standard screens due to the higher information content of the readout: single-cell transcriptomes instead of gRNA counts. When developing CROP-seq, we aimed to sequence 100 cells per gene perturbation, which is at the lower end of a typical pooled screen. This section discusses scRNA-seq methods with a suitable throughput for pooled screening: nanowell plates, microfluidic droplet generators, combinatorial indexing methods and pooled optical screening.

Technologies based on the physical isolation of single cells

The first whole transcriptome sequencing of single cells by next-generation sequencing was achieved in 2009 (Tang et al., 2009), applied to a handful of mouse oocytes and blastomeres. Since this seminal study, the sensitivity and precision of scRNA-seq have continuously improved, culminating in stable and widely-used protocols including Smart-Seq2 (Picelli et al., 2014), CEL-Seq2 (Hashimshony et al., 2016) and MARS-seq (Jaitin et al., 2014). All of these methods start with the physical isolation of single cells using simple glass capillaries or FACS. The library preparation is then performed in multiwell plates, a labor-intensive process that is wasteful in terms of the reagent volumes. In practice, microwell-based protocols are limited to a few hundred cells, and are therefore not attractive for pooled functional screening. Laboratory automation can increase the throughput to thousands of cells (Li et al., 2019), but the experimental design and cost do not scale well. Fortunately, the ever-increasing demand for higher throughput has been a major driver of technology development. Numerous innovative concepts and commercial platforms have emerged, revolutionizing the study of cellular identities and development (Shapiro et al., 2013) - even at the scale of whole organs and organisms (Rozenblatt-Rosen et al., 2017). Single-cell CRISPR screens are an exciting tool to introduce hypothesis-

driven perturbations into otherwise purely descriptive single-cell studies. The following section introduces the most promising approaches to maximize the number of cells analyzed in single-cell CRISPR screens.

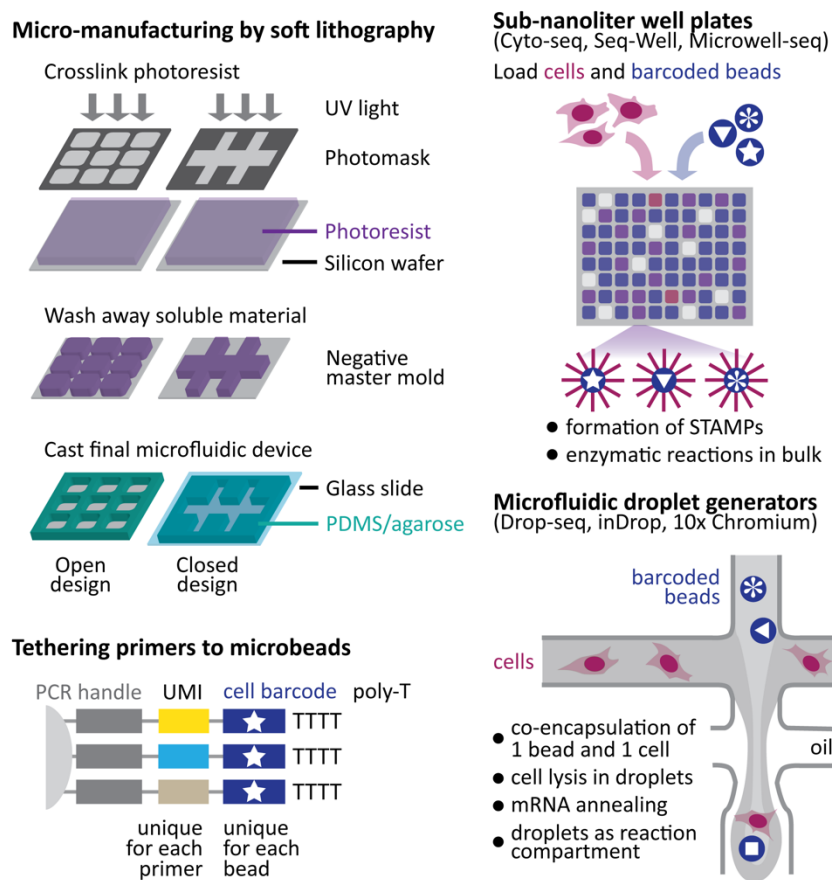


Figure 6 Technologies based on the isolation of single cells into miniaturized reaction compartments. High-throughput scRNA-seq is enabled by microfluidic devices produced by soft lithography and microbead-tethered reverse transcription primers (left). scRNA-seq in sub-nanoliter plates loads beads to near saturation, and cells at a limiting dilution. Wells containing both a bead and cell form *single-cell transcriptomes attached to microparticles (STAMPs, highlighted in purple)*, which are isolated and subjected to enzymatic reactions in bulk (top right). Droplet microfluidics allows the co-encapsulation of cells and barcoded microbeads in emulsion droplets that can be harnessed as reaction containers for cell lysis, mRNA capture and reverse transcription (bottom right).

Nanowell plates

The reaction volume of microwell plates is geared towards the human experimenter, but is inefficient for manipulating single cells. Unnecessarily large reagent volumes raise the cost per cell and dilute a single cell's transcriptome, thus resulting in a low efficiency of transcript capture and batch effects. In the search for a more suitable reaction compartment, scientists began to miniaturize the well size using soft lithography.

In this micromanufacturing method, a silicon wafer, which essentially acts as a base plate, is coated with a light-sensitive material called the photoresist. A photomask specific for the desired design is used to selectively cross-link the photoresist under UV-light. Material protected by the photomask remains soluble and is washed away with developer solution. This process shapes the photoresist into a negative master mold used to cast the final devices from materials such as PDMS or agarose. The single-cell protocols Cyto-Seq (Fan et al., 2015), Seq-Well (Gierahn et al., 2017) and Microwell-Seq (Han et al., 2018) use soft lithography to produce plates with reaction chambers in the nano- to picoliter range, only marginally larger than the cells to be analyzed. Given the right equipment, manufac-

turing such devices is relatively straightforward. The photomask is essentially an array of white rectangles on a black background, which correspond to the reaction chambers. This results in a master mold of micropillars that is used to cast plates for each experiment.

In scRNA-seq, individual cells are most efficiently labeled by including a DNA barcode into the reverse transcription primer. In traditional microwell-based protocols, soluble oligonucleotides with a particular *cell barcode* are pipetted directly into the reaction compartment, labeling all mRNAs the cell expresses with the same sequence. This is no longer possible for nano- and picowell plates, which are too small to be manipulated directly. The problem was elegantly solved by attaching the reverse transcription primers to carrier microbeads manufactured by split-pool DNA synthesis (Macosko et al., 2015) or primer extension (Fan et al., 2015, Klein et al., 2015) such that each bead carries primers sharing the same unique cell barcode. In a typical experiment, cells and beads are loaded onto the chip and settle into the nanowells by gravity. Cells are lysed due to the buffer composition, so that the released polyadenylated mRNA can anneal to the corresponding poly-T stretch on the bead-anchored primers. In all implementations published so far, the so-called *STAMPs* (Klein et al., 2015) (for *single-cell transcriptomes attached to microparticles*) are subjected to subsequent enzymatic reactions in bulk, after collecting them by centrifugation (Macosko et al., 2015), filtering (Datlinger et al., 2017) or with magnets (Han et al., 2018).

One of the main advantages of nanowell-based scRNA-seq over other methods is its portability. Aside from the PDMS or agarose array, the method does not require any specialized equipment, such that it can be applied almost anywhere. For instance, the 2017 publication describing Seq-Well (Gierahn et al., 2017) implemented scRNA-seq in a biosafety level 3 facility to profile primary human macrophages exposed to *Mycobacterium tuberculosis*. Nanowell-based methods also offer the highest cell capture efficiency of all methods, which is advantageous for scarce samples. Technically, this is achieved by selecting a well size that can accommodate only a single bead, and loading beads to saturation. This makes it more likely that a cell will encounter a microbead, resulting in a productive library preparation for up to 80% of the cells loaded (Gierahn et al., 2017).

Current nanowell-based methods reach a throughput of 10,000 cells per plate array, and can easily scale to 100,000 cells by running multiple plates in parallel or extending the chip design. This would make the system ideal for pooled single-cell screening. Unfortunately, at the time of writing this thesis, microwell-based scRNA-seq is still not widely adopted. This is in part due to a lack of commercially available chips and open source protocols. Another disadvantage is that the correct handling and sealing of the chip is critical to reduce cross-contamination between wells. Thus, nanowell-based methods are technically more challenging, require specific training and are harder to commercialize.

Microfluidic droplet generators

Soft lithography is not limited to open designs such as nanowell plates. When using PDMS material, the bottom side can be sealed by bonding it to a glass slide to realize complex closed channel designs. In their efforts to miniaturize the reaction compartments while increasing the throughput of scRNA-seq, scientists eventually turned to droplet microfluidics. These efforts have resulted in three popular microfluidic platforms for scRNA-seq: Drop-seq (Macosko et al., 2015), inDrop (Klein et al., 2015) and the commercially available 10x Genomics Chromium Controller (Zheng et al., 2017).

In a typical microfluidics-based scRNA-seq experiment, the aqueous cell suspension is run through a specially designed microfluidics chip, along with DNA-barcoded microbeads. Upon contact with a co-flowing stream of carrier oil, single cells and microbeads are co-encapsulated into nanoliter-sized emulsion droplets that are sufficiently stable to be used as reaction chambers. Depending on the method used, hundreds (Klein et al., 2015) to thousands (Macosko et al., 2015) of droplets are formed per second. As in nanowell-based scRNA-seq, the buffer composition ensures efficient cell lysis, releasing transcribed mRNAs into the reaction compartment where they anneal to the complementary poly-T stretch on the bead-tethered oligonucleotides. The Drop-seq protocol is similar to the previously introduced nanowell-based methods in that the reverse transcription is performed in bulk after breaking the emulsion droplets – with some potential for cross-contamination. Other protocols such as inDrop and the 10x Genomics Chromium system avoid this issue by encapsulating the reverse transcription reagents along with cells and beads, such that the reaction can take place inside the emulsion droplets. To improve transcript capture efficiency, these protocols also release the primers from their carrier beads by exposure to reducing conditions (Zheng et al., 2017) or UV light (Klein et al., 2015).

The biggest advantage of droplet microfluidics for scRNA-seq is the availability of both open-source protocols that enable technology development (Klein et al., 2015, Macosko et al., 2015), and reliable commercial platforms for large-scale projects (Zheng et al., 2017). Labs interested in establishing open-source protocols can outsource challenging tasks such as the manufacturing of beads and microfluidic devices to specialized companies. The throughput of the method is highly scalable, by increasing the run time or through parallelization of the encapsulation process into multiple channels (Zheng et al., 2017). Recent work based on Drop-seq has generated single-cell transcriptomes for 690,000 cells from 9 regions of the adult mouse brain (Saunders et al., 2018), an impressive demonstration of what microfluidic scRNA-seq is capable of.

The cell capture efficiency is variable between methods and depends mostly on the chip design as well as the material of the barcoded microbeads. Drop-seq has the least sophisticated bead design: the brittle resin material tends to fragment over the course of a run, and does not allow the packing of beads for efficient encapsulation. Instead, beads are supplied at a limiting concentration, which results in a lower rate of productive library preparation events. While the unique split-pool DNA synthesis approach utilized in Drop-seq generates the largest pool of cell barcodes, the primers also suffer from substantial synthesis errors that have to be corrected *in silico*. inDrop and the commercial 10x Genomics Chromium platform both use deformable hydrogel beads, which allows their synchronized encapsulation in the microfluidic channel with fill rates of up to 80%. This means that precious samples can be captured with up to 50% productive library preparation events.

Based on these advantages, microfluidic scRNA-seq is currently the ideal choice for pooled CRISPR screens at the single-cell level. Commercial platforms can profile up to 100,000 cells in a single run, albeit at very high costs. Through parallelization in the chip design, samples can be processed swiftly, resulting in high-quality data with minimal batch effects. Labs seeking to maximize the number of cells profiled in their screens can establish low-cost workflows based on available open-source protocols and chip designs, such as the Drop-seq workflow we used to first demonstrate CROP-seq. Nevertheless, single-cell CRISPR screens still have a long way to go before we can realize their full potential. For example, a genome-wide single-cell screen targeting the roughly 20,000 (Willyard, 2018) protein-coding human genes at 100x coverage would require 2 million single-cell transcriptomes. This ambitious

goal is beyond the capabilities of both nanowell-based and microfluidic-based scRNA-seq, and will require next-generation approaches such as those outlined below.

Technologies using single cells as the reaction compartment

The methods introduced in this section constitute a paradigm shift in single-cell genomics and hold great potential for single-cell CRISPR screening. They no longer try to physically isolate single cells into ever smaller reaction chambers. Instead, these protocols use gentle permeabilization and harness intact cells or nuclei as the reaction compartment. Below, I discuss the concept of combinatorial indexing, promising exceptionally high throughput and multi-omics measurements, and I will examine the even more early-stage idea of optical CRISPR screening that combines genetic perturbations with a microscopy readout of protein localization or cell-cell interactions. Finally, this section will introduce our method *scifi-RNA-seq*, which combines combinatorial indexing with droplet microfluidics.

Combinatorial indexing methods

Combinatorial indexing is not restricted to scRNA-seq, but refers to a more general concept for single-cell analyses. This experimental design has broad applicability, and it has been successfully applied to ATAC-seq (Cusanovich et al., 2015), whole-genome sequencing (Vitak et al., 2017), DNA methylation analysis (Mulqueen et al., 2018) and chromosome conformation capture by Hi-C (Ramani et al., 2017). In combinatorial indexing, permeabilized cells or nuclei go through multiple rounds of split-pool barcoding on multiwell plates. In the process, each cell travels its own path through the plate positions, resulting in a unique composite cell barcode. Importantly, thousands of cells share a reaction compartment during a barcoding step, but are pooled and randomly re-distributed for the next round of indexing.

In designing a combinatorial indexing method, one first identifies steps in a library preparation that can incorporate oligonucleotides into the DNA fragments. For a typical RNA-seq protocol, the most obvious steps are the reverse transcription, tagmentation and PCR enrichment. For instance, the method sci-RNA-seq (Cao et al., 2017) incorporates barcodes into the reverse transcription primer, transposome, and PCR primers, thus making a relatively standard RNA-seq protocol compatible with three rounds of combinatorial barcoding.

In developing a combinatorial indexing method, the main challenge lies in keeping cells or nuclei intact throughout the multi-step process, which often requires to find a substitute for harsh reaction buffers and incubation temperatures while maintaining the efficiency of the enzymatic reaction. Since the inactivation of enzymes by heating or cleaning the sample is not possible, unwanted enzymatic activities have to be kept in check by alternative means. It is also challenging to devise an elegant library structure that can read all barcodes cost-effectively, and to design barcodes with sufficient room for error correction. The original protocols for sci-RNA-seq and SPLIT-seq have demonstrated that combinatorial indexing is feasible for scRNA-seq, but will require further optimization to be broadly useful for single-cell CRISPR screening. Below, I highlight some of the technical challenges that have yet to be overcome, in order to make effective use of these promising new technologies.

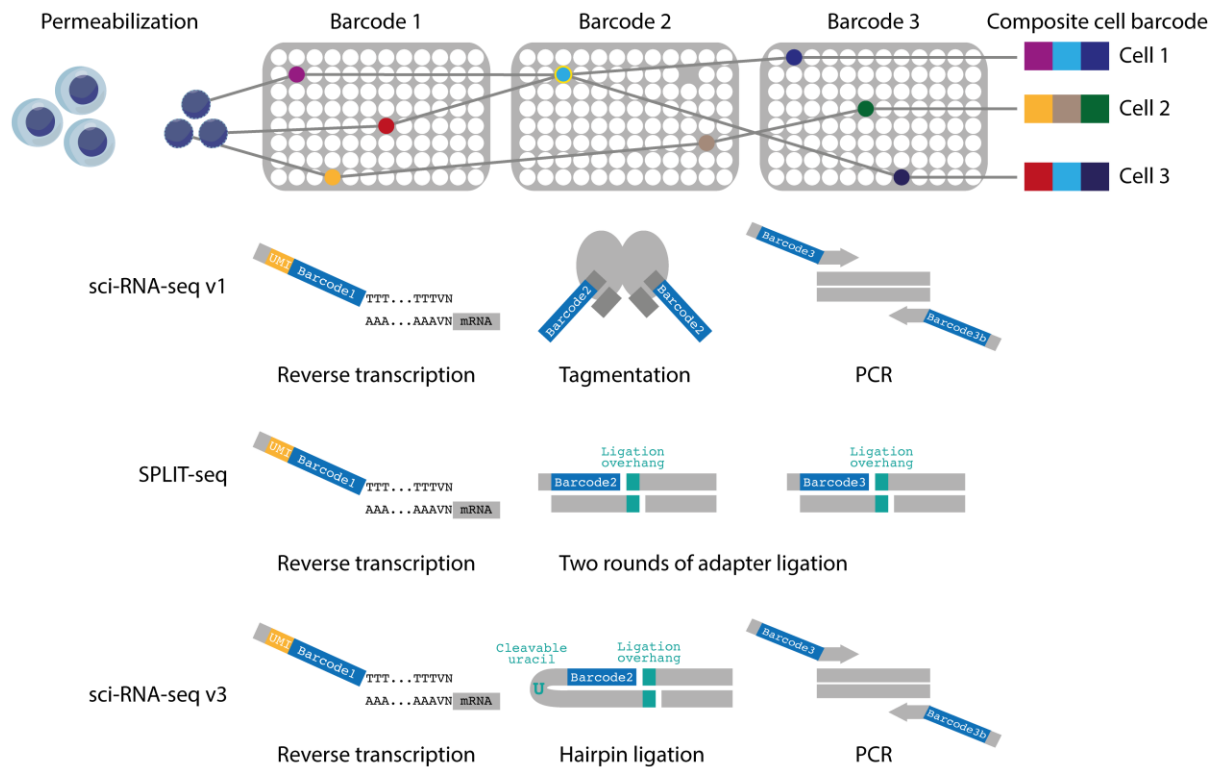


Figure 7 Single-cell combinatorial indexing methods. Combinatorial indexing does not isolate single cells. Instead, permeabilized cells or nuclei go through a series of barcoding steps, resulting in a composite cell barcode. Importantly, in each step hundreds to thousands of cells share a reaction compartment, but are mixed before being re-distributed for the next indexing reaction. This ensures that individual cells take a unique path through the plate positions. Both sci-RNA-seq and SPLIT-seq introduce the round 1 barcode via the reverse transcription. sci-RNA-seq introduces barcode 2 by tagmentation and barcode 3 by PCR. SPLIT-seq employs sequential adapter ligation. In sci-RNA-seq v3, the tagmentation step has been replaced by a hairpin ligation at a lower temperature where nuclei are more stable.

The requirement to run incubations on permeabilized cells or nuclei is destructive and results in very low recovery rates. This precludes the analysis of rare samples, but might be tolerated in a screening setting where enough cells are available as input. The problem is particularly apparent in the original sci-RNA-seq protocol, where the tagmentation step leads to pronounced sample loss and inferior single-cell transcriptomes. Realistically, the protocol was limited to two levels of barcoding and did not provide real advantages over existing methods unless impractical numbers of barcodes were used. The tagmentation step in sci-RNA-seq also prevented the PCR-enrichment of gRNAs from the transcriptome pool for efficient assignment of cells to perturbations, since this barcode would be lost during a gRNA-specific PCR. Recently, both issues have been addressed in an updated version of the protocol (Cao et al., 2019) (sci-RNA-seq v3). Replacing the tagmentation step with a ligation to a hairpin adapter, the method is now less destructive and provides an efficient and cost-effective way for introducing the second index.

While SPLIT-seq employed low-temperature incubations from the beginning and is compatible with the enrichment of CRISPR gRNAs, it is impractical due to its inefficient barcode design. In the final library, barcodes are separated by short sequences required for adapter annealing and ligation. During

next-generation sequencing, these uninformative regions have to be sequenced along with the composite cell barcode. This requires more expensive reagents, at a time where sequencing cost is often the most limiting factor.

At the moment, combinatorial indexing scRNA-seq is not directly applicable to pooled functional screening. The comparably low number of detected genes per cell or nucleus limits the efficiency of gRNA detection, thereby complicating the assignment of cells to their respective perturbations. Furthermore, three rounds of combinatorial labeling with substantial numbers of barcodes are work-intensive and error-prone. With *scifi-RNA-seq* we propose an alternative approach that combines combinatorial indexing with droplet microfluidics (the method is described in more detail in the corresponding manuscript in section 2.2 of this thesis). Ultimately, optimized protocols for PCR-based gRNA enrichment will be required. Once this challenge has been solved, combinatorial indexing could become the most attractive technology for single-cell screening. While challenging to design and optimize, combinatorial indexing protocols have advantages over conventional methods that make them especially interesting for single-cell CRISPR screens. Most importantly, the number of cells profiled per experiment grows exponentially with the number of barcoding rounds. Assuming a cell collision rate of 10%, three rounds of barcoding in 384-well plates would allow the analysis of over one million cells, which would be very hard to achieve with conventional methods such as microwell plates and droplet generators and could enable genome-wide screens in single cells in the near future. For *scifi-RNA-seq*, the second combinatorial index is introduced via a microfluidic droplet generator, such that the same barcode complexity is reached with only two rounds of labeling.

Moreover, all combinatorial indexing protocols perform at least some of the steps in multiwell plates. This is familiar territory for molecular biologists without microfluidics experience who want to adapt their favorite protocols to combinatorial indexing. Since the material is accessible at all times, combinatorial indexing also offers interesting opportunities for the development of multi-omics protocols, such as the recently published combination of scRNA-seq and scATAC-seq (Cao et al., 2018). I expect that combinatorial indexing and scifi methods will be quickly adopted and improved upon, and will enable the first truly large genetic screens in single cells.

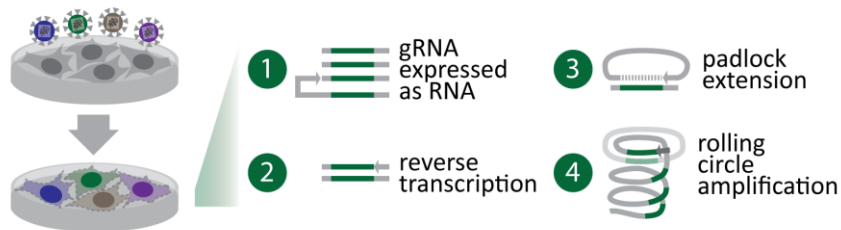
Pooled optical screens

Pooled CRISPR screens with single-cell transcriptome readout can achieve a high-dimensional readout of cell states, but also come with limitations. All scRNA-seq methods introduced so far require the dissociation of tissues into a single-cell suspension, for loading into a nanowell plate, microfluidic droplet generator or as input for a combinatorial indexing experiment. Thereby, information on the native location of cells within the tissue and cell-to-cell interactions is lost. Even though new methods can measure the abundance of cell surface proteins in parallel to the single-cell transcriptome (Peterson et al., 2017, Shahi et al., 2017, Stoeckius et al., 2017), they have not been adapted for intracellular proteins and, by design, cannot pinpoint the exact subcellular localization. Many biological phenotypes however, have an important spatial or temporal component. Microscopic imaging is ideally suited to address these questions, but has so far been limited to arrayed screening formats.

Optical CRISPR screens are still a very new and experimental concept, but proof-of-principle has already been provided in a recent preprint (Feldman et al., 2018). This method is directly compatible

with our single-cell CRISPR screening method CROP-seq, where the gRNA is included in a highly expressed messenger RNA. Similar to combinatorial indexing, cells are permeabilized and subjected to reverse transcription, which is followed by targeted in situ sequencing of the gRNAs. The sequencing uses off-the-shelf reagents from the Illumina MiSeq kit and a standard automated microscope equipped with filters to read the fluorescence intensities for the four bases. After completing the sequencing-by-synthesis reaction, sequencing clusters are linked to the nearest nucleus, and bases and the respective perturbation are called. This process is followed by immunofluorescence imaging to assess complex spatial phenotypes. In the preprint describing pooled optical screening, 8 million single cells were imaged, out of which 3 million could be confidently assigned to one of 3,063 gRNAs targeting 952 genes. In this screen, a spatial readout of p65 translocation to the nucleus allowed the identification of regulators of NFkB signaling.

Cell permeabilization and library preparation



gRNA in situ sequencing with microscopy readout

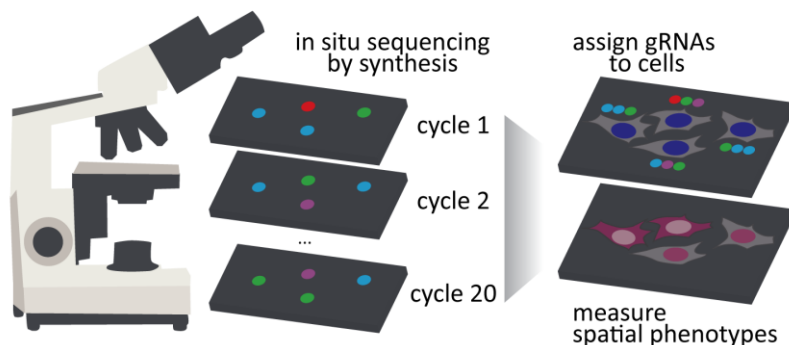


Figure 8 Pooled optical CRISPR screens

Cells are permeabilized and expressed gRNA molecules are reverse transcribed, followed by signal amplification through padlock extension and rolling circle amplification. A cell's gRNA is detected by in situ sequencing, using sequencing by synthesis reagents. Based on the gRNA reads, each cell is assigned a genetic perturbation. In addition, complex phenotypes can be visualized, such as the subcellular localization of proteins, or cell-cell interactions.

The imaging readout alone offers very interesting opportunities that are highly complementary to pooled scRNA-seq screens. Nevertheless, even richer single-cell phenotyping could be achieved by combining optical screening with an imaging-based transcriptome readout such as MERFISH (Chen et al., 2015) or seqFISH (Lubeck et al., 2014, Shah et al., 2016). In these methods, fixed cells are subjected to multiple rounds of fluorescence in situ hybridization targeting cellular RNAs. While the exact technical solution differs, both methods identify individual transcripts by a combination of fluorescently labelled probes, read sequentially to yield a temporal transcript barcode. To illustrate the performance of these methods, a recent report using MERFISH profiled 130 RNA species in a total of 100,000 cells within 45 hours (Moffitt et al., 2016b). seqFISH has a comparable throughput, with 249 genes detected in 16,958 cells (Shah et al., 2016). By design, imaging-based transcriptomics is not limited to mRNA, but can be readily adapted to detect non-coding RNAs or intronic sequences (Shah et al., 2018). More recently, refined clearing methods have allowed the application of MERFISH and seqFISH to tissue slices (Moffitt et al., 2016a) and three-dimensional organs (Shah et al., 2016), thereby greatly expanding their potential.

Since both protocols implement barcoding correction schemes to address signal dropouts and misidentifications, they can theoretically scale to the whole transcriptome (Chen et al., 2015). Nevertheless, the imaging time and overlap of fluorescent signal are limiting the number of RNAs and cells that can be profiled. Visualizing more transcripts by imaging inevitably results in a substantial overlap of the fluorescent signal from individual transcripts. This problem has been partially addressed by expansion microscopy (Wang et al., 2018), but the resulting increase in imaging time was still a major limitation. Recently, the problem was resolved by a novel pseudocolour-encoding scheme that dilutes mRNA molecules into 60 separate images, such that mRNA dots can be localized below the diffraction limit and combined into one super resolution image. This encoding scheme has been implemented in the improved method seqFISH+ (Eng et al., 2019), which can image mRNAs for 10,000 genes in single cells while reducing the imaging time eightfold. seqFISH+ has been applied to the cortex, subventricular zone and olfactory bulb of the mouse brain, which highlights its broad utility for the generation of spatial cell atlases. While the number of single cells profiled with MERFISH and seqFISH seems low compared to what sequencing-based technologies can achieve, many biological applications can benefit from in situ data, and future single-cell studies will likely combine both data types (Zhu et al., 2018).

Optical screens, with or without transcriptome readout, provide a complementary perspective for many screening problems. However, it is hard to imagine them being as widely adopted as pooled scRNA-seq screens due to the challenging experimental setup, long imaging times and relatively specialized data analysis requirements. Nevertheless, they might be preferable for certain biological questions where a clear spatial or temporal phenotype can be defined.

1.2.2. Platform-agnostic detection of gRNAs in single-cell transcriptomes

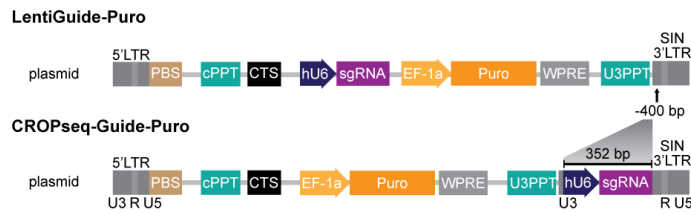
As the previous section has demonstrated, single-cell genomics is a dynamic research field marked by continuous technological innovation. Within just one decade, the experimental throughput for scRNA-seq has increased from a few selected oocytes and blastomeres (Tang et al., 2009) to over 2 million single nuclei derived from the developing mouse embryo (Cao et al., 2019). This rapid development has been possible due to several disruptive new concepts, which were already introduced in the previous chapter and are summarized below (Table 1). Microfluidic droplet generators, such as the 10x Genomics Chromium controller, have enabled the wide adoption of scRNA-seq – an assay now applied routinely in laboratories around the globe. In the recently developed combinatorial indexing methods for scRNA-seq, the limiting factor for experimental throughput is no longer the library preparation, but the cost of next-generation sequencing.

Single-cell CRISPR screens can provide a unique perspective, even with just a few selected perturbations (Jaitin et al., 2016). Large exploratory screens however, inevitably rely on large cell numbers and benefit greatly from supporting both current and emerging technologies. Therefore, a key design consideration for CROP-seq was to remain platform independent and future-proof. A mechanism common to all scRNA-seq technologies is to enrich for polyadenylated mRNA by reverse transcription with oligo-dT primers. The process of polyadenylation is dependent on and coupled to transcription by RNA polymerase II. CRISPR gRNAs have important structural requirements, and are therefore transcribed from RNA polymerase III promoters such as hU6, which allow precise control over the transcript's start and end point. Thus, gRNAs lack a poly-A tail and cannot be detected by standard scRNA-seq technologies.

	Multiwell plates	Nanowell plates	Microfluidic droplet generators	Combinatorial indexing
Representative methods	Smart-Seq2 CEL-Seq2 MARS-seq	Cyto-Seq Seq-Well Microwell-Seq	Drop-seq inDrop 10x Chromium	sci-RNA-seq SPLIT-seq
Cells per run	100 - 1,000 10,000 (automation)	10,000 - 100,000	10,000 - 100,000	100,000 - 2,000,000
Cell capture efficiency	80%	80%	10% (Drop-seq) 50% (inDrop, 10x Chromium)	8%
Cost per cell [\$]	10	0.02	0.06 (inDrop, Drop-seq) 0.1 - 1 (10x Chromium)	0.01
Suitable for single-cell screening	not suitable	yes	yes	yes
Perturbations at 100x coverage	1 - 10	100 - 1,000	100 - 1,000	1,000 - 20,000 (genome-wide)
Requires physical isolation of cell	yes	yes	yes	no
Reverse transcription	compartment	bulk	bulk (Drop-seq) compartment (inDrop, 10x Chromium)	compartment
Advantages	cell capture efficiency, transcriptome quality	cell capture efficiency, portability	open-source and commercial platforms available, widely adopted	ultra-high throughput multi-omics measurements, encourages development of custom protocols
Disadvantages	cost per cell workload	lack of open-source protocols and chip designs	protocol development requires expertise in microfluidics and bead manufacturing	not suitable for low input samples

Table 1 Comparison of current single-cell technology platforms and representative methods. Key parameters for single-cell CRISPR screening are the experimental throughput and cost (Cells per run, Cost per cell). Since most screens start with large cell numbers, low cell capture efficiencies can be tolerated.

a) CROP-seq uses a modified lentiviral vector



b) Detailed mechanism for gRNA duplication during the transduction of target cells

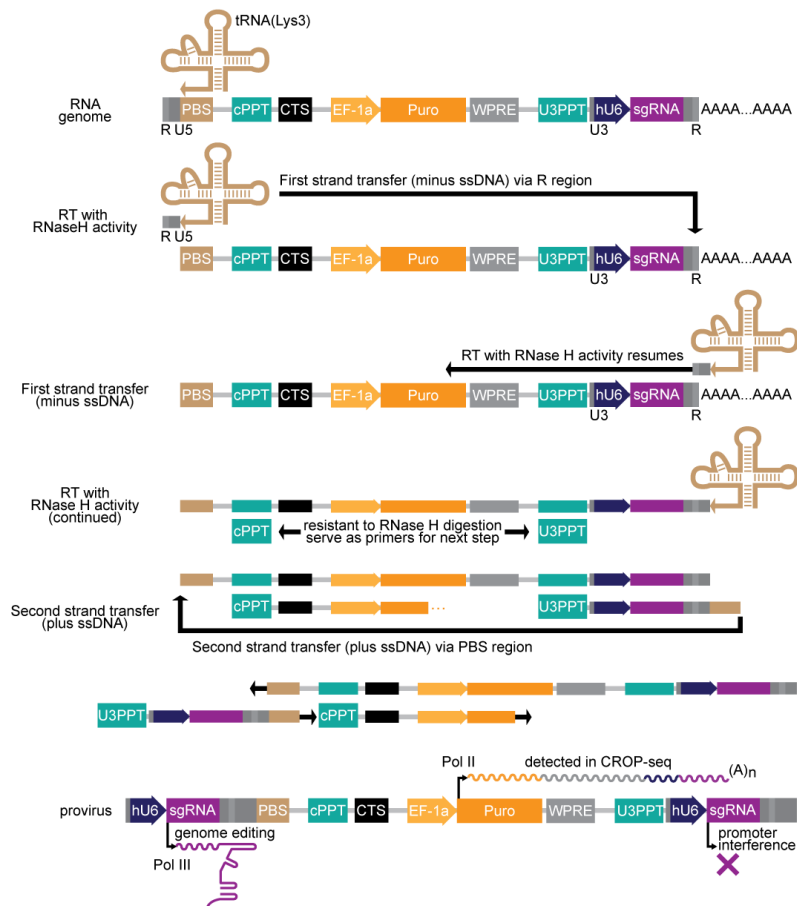


Figure 9 Technical solution to detect gRNAs directly from single-cell transcriptomes, while maintaining high gRNA expression and editing efficiencies.

a) CROP-seq uses a modified lentiviral vector that places the gRNA expression cassette inside the lentiviral 3' long terminal repeat (LTR), inside a highly expressed selectable marker for puromycin resistance. Driven by a polymerase II promoter, this transcript is effectively polyadenylated and detected in standard single-cell RNA-seq. b) During transduction of the target cell, the gRNA cassette gets copied to the 5' LTR by lentiviral mechanisms. While the original gRNA cassette is repressed by promoter interference, the duplicate copy is highly expressed and results in efficient genome editing. Upon transduction of the target cell, the single-stranded lentiviral RNA genome is reverse transcribed. Synthesis of the minus strand DNA is initiated from a tRNA molecule originating from the producer cells. Once the end of the genome is reached, the ssDNA gets transferred to the 3' end via the homologous R region, where the reverse transcription resumes. The RNase activity of the reverse transcriptase degrades the RNA template except for two purine-rich sequences, the central polypurine tract (cPPT) and a polypurine tract towards the 3' LTR (U3PPT). These two RNA sequences serve as primers for

the plus ssDNA synthesis step. The plus ssDNA gets transferred to the 3' end via the homologous PBS region. This is the step that ultimately leads to the duplication of the gRNA cassette, and allows CROP-seq to combine direct gRNA detection with efficient editing. Abbreviations: LTR long terminal repeat, PBS primer binding site, cPPT central polypurine tract, CTS central termination sequence, hU6 human U6 promoter for defined transcription of gRNAs, EF-1a RNA polymerase II promoter leading to A-tailing, Puro puromycin antibiotic resistance marker, WPRE Woodchuck Hepatitis Virus Posttranscriptional Regulatory Element, U3PPT polypurine tract before the 3' LTR

When designing CROP-seq, we decided against modifying an existing scRNA-seq protocol for gRNA detection, which might restrict the method to a particular technology platform. Instead, CROP-seq solves the problem of making gRNAs detectable in scRNA-seq data at the vector level.

We designed a lentiviral vector that places the gRNA expression cassette inside a highly expressed, polyadenylated transcript encoding the antibiotic resistance marker puromycin (Figure 9a). Thereby, the gRNA is effectively polyadenylated and should be detected irrespective of the scRNA-seq protocol used. However, transcriptional interference between the promoters for the antibiotic resistance gene

and gRNA can be expected to reduce gRNA expression and the efficiency of genome editing. To avoid this issue, we decided to exploit lentivirus biology. We placed the gRNA cassette into a specific location in the 3' long terminal repeat (LTR) of the lentiviral vector, which gets copied over to the 5' LTR during lentiviral reverse transcription. Thus, we obtain one copy of the gRNA expression cassette in the 3' LTR that is detected in scRNA-seq and one additional copy in the upstream 5' LTR that is unaffected by promoter interference and results in efficient editing.

The molecular mechanism by which the gRNA cassette gets duplicated is shown in detail in Figure 9b. During lentivirus production in HEK293T producer cells, the CROP-seq construct is transcribed from the plasmid to form the RNA genome of the lentiviral vector, starting at the R region of the 5' LTR and extending into the R region of the 3' LTR. The transcript contains both a 5' cap structure and poly-A tail, and is packaged into lentiviral particles along with a host tRNA molecule binding to the *primer binding site (PBS)* region. Upon transduction of the target cells, the tRNA serves as a primer for the reverse transcription of the minus ssDNA strand. Once the end of the genome is reached, the ssDNA gets transferred to the 3' end via the homologous R region, where the reverse transcription resumes. Of note, the reverse transcriptase has RNase activity, which degrades most of the RNA template except two purine-rich sequences, the central polypurine tract (cPPT) and a polypurine tract towards the 3' LTR (U3PPT). These two RNA sequences serve as primers for the plus ssDNA synthesis step. Again, once the end of the genome is reached, the plus ssDNA gets transferred to the 3' end via the homologous *PBS* region. This is the step that ultimately leads to the duplication of the gRNA cassette, and allows CROP-seq to combine direct gRNA detection with efficient editing.

2. RESULTS

2.1. Pooled CRISPR screening with single-cell transcriptome readout

Next-generation sequencing technology allows researchers to profile disease states at unprecedented detail, but the functional interpretation of the vast amounts of data and subsequent translation into clinical practice is challenging. Functional genetic screens with CRISPR/Cas9 are a promising tool for genomic medicine, as they can help to identify causal variants from the large set of alterations typically found in patients.

The aim of this thesis was to develop a new screening paradigm for functional genomics that combines the scalability of pooled screens with the detailed molecular readout of arrayed assays. In the following paper (Datlinger et al. 2017 Nature Methods), we developed and validated a groundbreaking method for CRISPR screens at the single-cell level. Our method CROP-seq (“CRISPR droplet sequencing”) uses a modified lentiviral vector to make CRISPR guide RNAs detectable from single-cell RNA sequencing data. We also developed a bioinformatic framework to assign CRISPR perturbations to their respective single-cell transcriptomes. Finally, we apply CROP-seq to study T-cell receptor activation in the human Jurkat T-cell line. This application demonstrates that CROP-seq is suitable for addressing complex biological phenomena such as the dynamics of a signaling pathway. Furthermore, our assay can derive relevant screening signatures from unmodified control cells, which provides a self-calibrating readout that does not require any prior knowledge about the biological system.

Pooled CRISPR screening with single-cell transcriptome readout

Paul Datlinger¹, André F Rendeiro^{1,4}, Christian Schmidl^{1,4}, Thomas Krausgruber¹, Peter Traxler¹, Johanna Klughammer¹, Linda C Schuster¹, Amelie Kuchler¹, Donat Alpar¹ & Christoph Bock¹⁻³

CRISPR-based genetic screens are accelerating biological discovery, but current methods have inherent limitations. Widely used pooled screens are restricted to simple readouts including cell proliferation and sortable marker proteins. Arrayed screens allow for comprehensive molecular readouts such as transcriptome profiling, but at much lower throughput. Here we combine pooled CRISPR screening with single-cell RNA sequencing into a broadly applicable workflow, directly linking guide RNA expression to transcriptome responses in thousands of individual cells. Our method for CRISPR droplet sequencing (CROP-seq) enables pooled CRISPR screens with single-cell transcriptome resolution, which will facilitate high-throughput functional dissection of complex regulatory mechanisms and heterogeneous cell populations.

Pooled CRISPR screening is a powerful and widely used method for identifying genes involved in biological mechanisms such as cell proliferation^{1,2}, drug resistance³, and viral infection⁴. Cells are transduced in bulk with a library of guide RNA (gRNA)-encoding vectors, and the distribution of gRNAs is measured before and after applying a selective challenge (Fig. 1a). Pooled CRISPR screens work well for mechanisms that affect cell survival and proliferation, and they can be extended to measure the activity of individual genes (e.g., by using engineered reporter cell lines). However, they do not support complex molecular readouts such as transcriptome profiling, which is one of the most comprehensive and informative measures of cellular response⁵. Arrayed CRISPR screens, in which only one gene is targeted at a time, make it possible to use RNA-seq as the readout⁶. But this comes at the cost of much lower throughput because cells transduced with individual gRNAs need to be physically separated (Fig. 1b).

Here we propose a third and complementary screening paradigm, single-cell CRISPR screens, based on the idea that gRNAs and their induced cellular responses are already compartmentalized within single cells. Detecting each cell's gRNA along with its single-cell transcriptome would thus provide gene expression signatures for individual gene knockouts in a complex pool of cells (Fig. 1c). Implementing this concept, our CROP-seq method combines four

key components: a gRNA vector that makes individual gRNAs detectable in single-cell RNA-seq experiments, a high-throughput assay for single-cell RNA-seq⁷, a computational pipeline for assigning single-cell transcriptomes to gRNAs, and a bioinformatic method for analyzing and interpreting gRNA-induced transcriptional profiles. CROP-seq thereby enables pooled screens with single-cell transcriptome readout, providing a scalable method for dissecting complex signaling pathways and other biological mechanisms that are not easily reduced to a single selectable marker (Fig. 1d).

RESULTS

Direct detection of gRNAs from single-cell transcriptome data

CRISPR gRNAs are typically transcribed by RNA polymerase III from a human U6 promoter (hU6). They lack a polyadenylated (poly-A) tail and are not detectable with current single-cell RNA-seq assays. We thus re-engineered a popular construct for pooled CRISPR screening (LentiGuide-Puro)⁸ to include the gRNA in a polyadenylated mRNA transcript (Fig. 1e, Supplementary Data, and Supplementary Fig. 1). Because LentiGuide-Puro was originally derived from a lentiviral vector rendered self-inactivating by a 400-bp deletion of key promoter elements, we hypothesized that this position within the 3' long terminal repeat (LTR) would tolerate the insertion of the similarly sized hU6-gRNA cassette. At this position, the gRNA becomes part of the puromycin-resistance mRNA transcribed by RNA polymerase II and detectable by RNA-seq protocols that use poly-A enrichment. Functional gRNAs continue to be expressed from the hU6 promoter. In addition, the entire hU6-gRNA cassette is copied to the 5' LTR during reverse transcription and integration of the virus (Supplementary Fig. 1). This results in a second copy upstream of the possibly interfering EF-1a promoter—a design reminiscent of an early shRNA expression vector⁹. CROP-seq thereby solves the challenge of detecting gRNAs in single-cell transcriptomes at the vector level, which makes it compatible with various single-cell RNA-seq assays and with widely used cloning protocols for pooled screening.

Having assembled our CROPseq-Guide-Puro plasmid from four PCR products using the ligase cycling reaction¹⁰ (Supplementary Fig. 1 and Supplementary Table 1), we extensively validated its

¹CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria. ²Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria. ³Max Planck Institute for Informatics, Saarland Informatics Campus, Saarbrücken, Germany. ⁴These authors contributed equally to this work. Correspondence should be addressed to C.B. (cbock@cemm.oew.ac.at).

RECEIVED 11 OCTOBER 2016; ACCEPTED 10 JANUARY 2017; PUBLISHED ONLINE 18 JANUARY 2017; DOI:10.1038/NMETH.4177

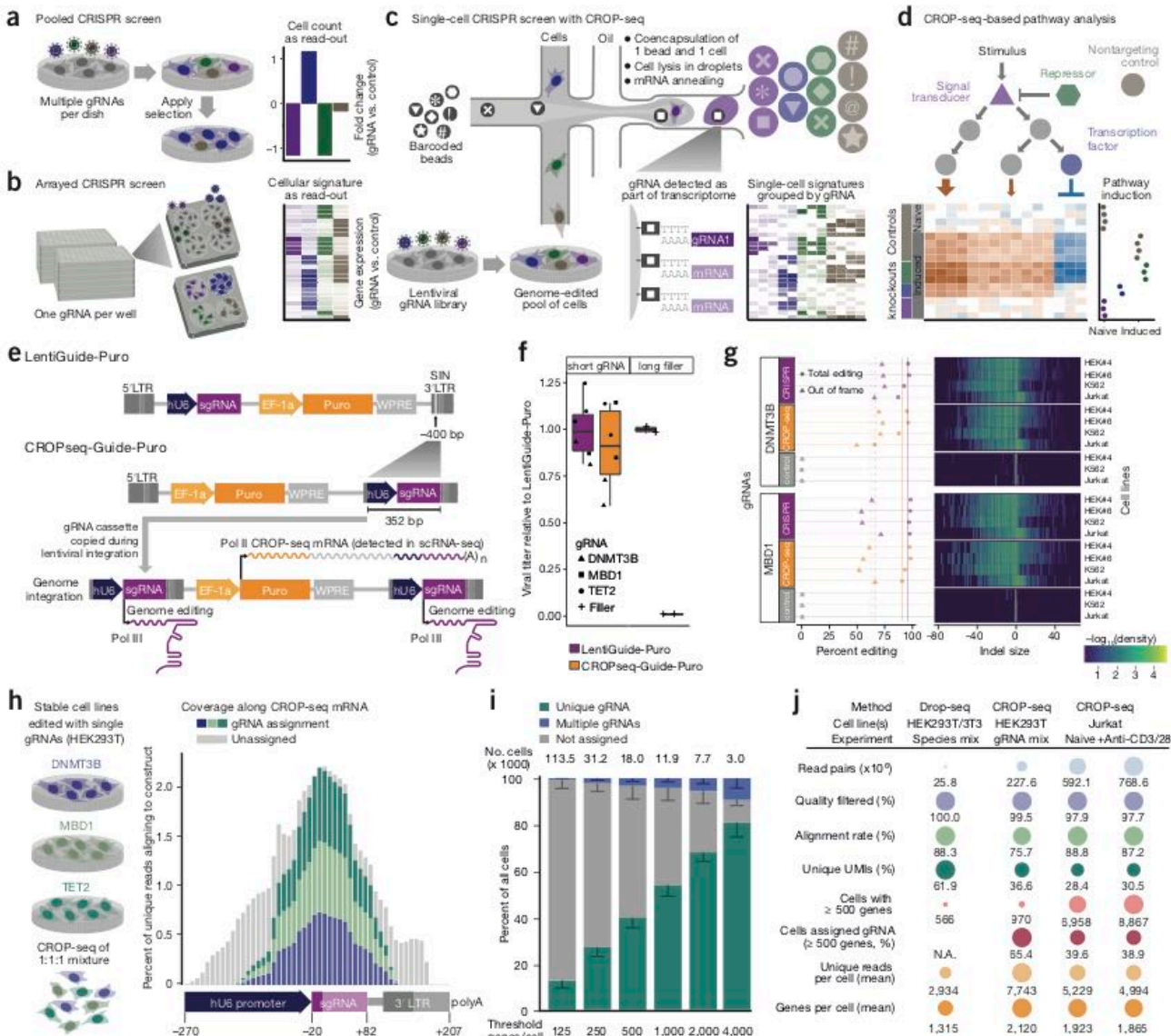


Figure 1 | CROP-seq enables pooled CRISPR screening with single-cell transcriptome readout. **(a)** Pooled screens detect changes in gRNA abundance among bulk populations of cells, which limits them to simple readouts based on cell frequencies. **(b)** Arrayed screens support complex readouts such as transcriptome profiling, but cells transduced with different gRNAs have to be physically separated. **(c)** CROP-seq uses droplet-based single-cell RNA-seq to profile each cell's transcriptome together with the expressed gRNA, and knockout signatures are derived by averaging across cells that express gRNAs for the same target gene. **(d)** Data analysis identifies pathway signature genes and quantifies the effect of specific gRNAs on these signatures. **(e)** The CROP-seq lentiviral construct includes a gRNA cassette within the 3' long terminal repeat (LTR), which is duplicated during viral integration. It expresses an RNA polymerase III transcript for genome editing and a polyadenylated RNA polymerase II transcript detected by single-cell RNA-seq. **(f)** Cloning the hU6-gRNA cassette into the 3' LTR to generate CROPseq-Guide-Puro does not compromise lentiviral function for gRNAs. In contrast, 1,885 bp of filler DNA result in a 98-fold reduction of the viral titer. **(g)** Genome editing efficiencies and indel signatures are highly similar between LentiGuide-Puro and CROPseq-Guide-Puro. **(h)** CROP-seq can detect gRNAs from single-cell transcriptomes. **(i)** The rate of successful gRNA assignments is associated with single-cell transcriptome quality, expressed as the number of detected genes per cell. Most cells were assigned to one gRNA, except for a small fraction of cell doublets. Error bars, 95% CI. **(j)** Performance statistics across all CROP-seq experiments.

performance and suitability for CRISPR screening. The hU6–gRNA insertion into the 3' LTR had no adverse effects on the formation of lentiviral particles nor on puromycin resistance of transduced cells. We produced virus from LentiGuide-Puro and CROPseq-Guide-Puro containing one of three validated gRNAs (targeting *DNMT3B*, *MBD1*, or *TET2*) or a longer filler construct that was expected to interfere with viral activity, and we observed comparable functional

titers for all tested gRNA sequences (**Fig. 1f**). Genome editing efficiencies were benchmarked in three Cas9-expressing cell lines (K562, Jurkat, and two HEK293T clones) using two gRNAs and two complementary technologies—the T7 endonuclease assay¹¹ (**Supplementary Fig. 2**) and next-generation sequencing of PCR amplicons (**Fig. 1g**). Genome editing was highly efficient for both constructs, averaging at 95.5% (LentiGuide-Puro) and 90.5%

(CROPseq-Guide-Puro), with the expected two thirds (66.2% and 62.4%) of molecules carrying out-of-frame insertions and deletions. For each locus and cell line, we also observed characteristic editing signatures, which were highly similar between both constructs (Fig. 1g). CROPseq-Guide-Puro thus provides excellent results both in terms of the quantity and quality of induced genome editing events.

Next, we tested whether single-cell RNA-seq could detect expressed gRNAs in cells transduced with CROPseq-Guide-Puro (Fig. 1h and Supplementary Fig. 3). We generated three knockout cell lines from HEK293T cells, targeting *DNMT3B*, *MBD1*, and *TET2* with validated gRNAs. By performing the transductions independently, we ensured that each cell received a unique gRNA—even in the case of multiple infection events. When we applied Drop-seq⁷ on a 1:1:1 mixture of the three cell lines, the resulting single-cell transcriptomes provided high sequencing coverage for the gRNAs, and we were able to assign single cells to gRNAs at the expected ratios (Fig. 1h).

Combining all CROP-seq data from this study, we assessed the confidence of our gRNA assignments (Fig. 1i), which depended on the number of detected genes per cell. For example, 38.7% of cells with at least 500 and 78.9% of cells with at least 4,000 detected genes were uniquely assigned. Few cells matched more than one gRNA (e.g., 2.7% for a threshold of 500 detected genes), although this rate increased with the detected number of genes per cell (e.g., 9.8% for a threshold of 4,000). This rate increase was most likely due to rare cell doublets (Supplementary Fig. 3d) that release twice the amount of RNA, resulting in more detected genes and multiple gRNA assignments. Furthermore, we excluded any cells that were assigned to multiple gRNAs from the downstream analysis, and CROP-seq is robust toward potentially undetected doublets because it combines data across all single cells assigned to the same gRNA.

Single-cell CRISPR screening for T-cell-receptor induction

Having established and validated CROP-seq as a method for single-cell CRISPR screens (Fig. 1j) provides detailed performance statistics across all 12 CROP-seq experiments), we tested our method in a proof-of-concept screen of T-cell receptor (TCR) activation in Jurkat cells (Fig. 2a). We designed a gRNA library for six high-level regulators of TCR signaling and 23 transcription factors, targeting each gene with three distinct gRNAs (Supplementary Table 2). We also included 20 nontargeting gRNAs as negative controls and 9 gRNAs for essential genes² as positive controls. Jurkat cells that stably express Cas9 were transduced with lentivirus produced from this CROPseq-Guide-Puro gRNA library, and genome-edited cells were selected with puromycin. At day 10 post-transduction, the surviving pool of cells was serum starved, split, and subjected to either TCR stimulation via anti-CD3 and anti-CD28 antibodies or to continued starvation; and both cell populations were analyzed with CROP-seq.

Comparing gRNA assignments based on CROP-seq with gRNA counts obtained by sequencing the plasmid library, we observed a consistent depletion of positive controls, confirming efficient editing (Fig. 2b and Supplementary Fig. 4). The gRNAs in our TCR library showed more diverse patterns; gRNAs targeting *ETS1*, *RUNX1*, and *GATA3* were depleted, suggesting that the encoded transcription factors are essential in Jurkat cells. In contrast, most gRNAs behaved similarly to the negative controls, indicating that

they did not have strong antiproliferative effects. At the transcriptome level, pairwise distances between gRNAs targeting the same gene were smaller than those between gRNAs targeting different genes (Supplementary Fig. 5a,b), and transcriptome profiles assigned to specific gRNAs and target genes were distinguishable from those of cells expressing nontargeting gRNAs (Supplementary Fig. 5c,d).

We established a transcriptome signature of TCR induction directly from the CROP-seq data by applying dimensionality reduction to single-cell RNA-seq profiles (Supplementary Fig. 6a–c). Principal component analysis for cells grouped by gRNA target genes separated naive and TCR-induced Jurkat cells along the first principal component, which defined a TCR induction signature of 165 genes (Fig. 2c and Supplementary Fig. 6d,e). This signature was enriched for genes known to be relevant in TCR signaling (Fig. 2c and Supplementary Fig. 6f).

Aggregating single-cell RNA-seq profiles by their assigned target genes and gRNAs (Fig. 2d and Supplementary Fig. 7a,b), we found that induced cells expressing gRNAs for *LCK* and *ZAP70* were similar to the naive group, indicating that knockout of these key mediators of TCR signaling interfered with TCR induction. For a more quantitative picture, we positioned each cell on a gradient defined by naive and induced cells that expressed nontargeting control gRNAs (Fig. 2e and Supplementary Fig. 7c–e). This analysis was based on 5,798 high-quality single-cell transcriptome profiles with unique gRNA assignments and a median of 80 cells per targeted gene; while 1,320 control cells with nontargeting gRNAs were used as reference.

Validation by bulk RNA-seq and flow cytometry

For validation, we performed an arrayed screen using 48 independent lentiviral constructs, targeting 20 genes with two gRNAs and including eight nontargeting controls (Supplementary Table 3). We measured TCR induction by bulk RNA-seq (87 RNA-seq libraries; Fig. 2f and Supplementary Fig. 8a,b) and flow cytometry (96 flow cytometry profiles; Fig. 2g). High correlations were observed between bulk RNA-seq and CROP-seq data, both for inferred TCR signature intensities and for global expression profiles (Fig. 2f and Supplementary Fig. 8c,d). Flow cytometry for surface proteins that were correlated with our TCR signature (CD69 and CD82) or previously reported as markers of T-cell activation (CD25, CD38, CD154, and PD-1) further validated the CROP-seq results (Fig. 2g and Supplementary Fig. 9).

The arrayed screen also confirmed the results for individual genes that CROP-seq identified as important for TCR signaling in Jurkat cells. Depletion of the kinases *LCK* and *ZAP70* and the adaptor protein *LAT* had a strong negative effect on the TCR activation signature, consistent with their established role in TCR signal propagation and amplification¹² (Fig. 2h,i and Supplementary Fig. 8e). Moreover, we observed that targeting known negative regulators of TCR signaling (such as *PTPN6*, *PTPN11*, and *EGR3*) slightly increased the TCR activation signature. While the focus of this screen was to validate CROP-seq as a method, we also identified a few transcription factors that may warrant further investigation of their potential functional roles during TCR induction (Fig. 2d,e and Supplementary Fig. 7).

To assess how well CROP-seq might scale to larger (and eventually genome-wide) gRNA libraries, we performed downsampling analysis and identified the minimum number of cells that

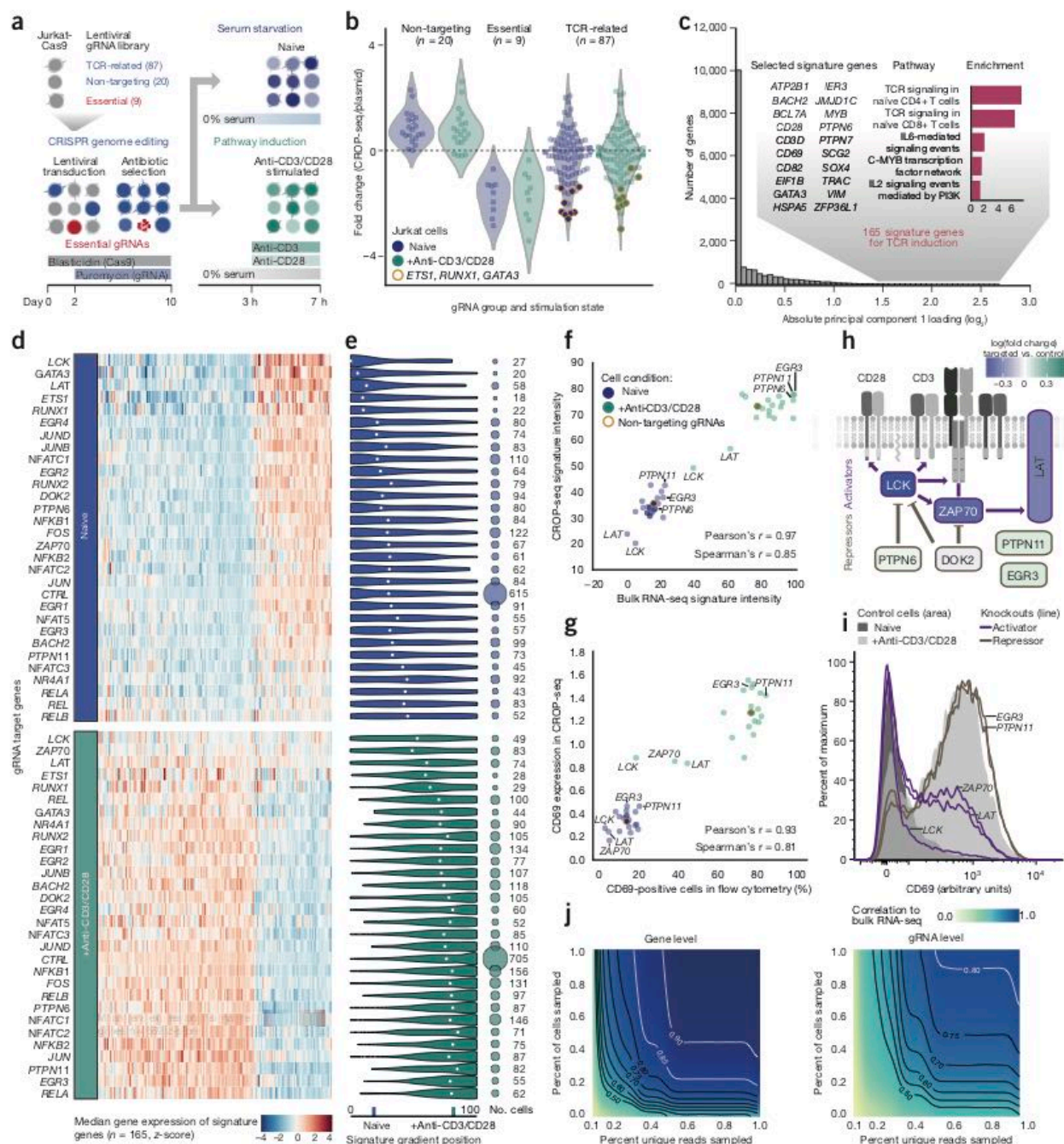


Figure 2 | CROP-seq analysis of T cell receptor signaling. (a) Experimental design of a single-cell CRISPR screen for T cell receptor (TCR) pathway induction. **(b)** Fold change of gRNA abundance between cell assignments from CROP-seq and gRNA counts from plasmid library sequencing. Values were normalized to the total of assigned cells or reads, respectively. **(c)** Inference of pathway signature from CROP-seq data. Single-cell transcriptomes were aggregated by gRNA target genes, and principal component analysis separated naive and anti-CD3/CD28-stimulated cells. Genes with absolute loading values for principal component 1 that exceeded the 99th percentile were included in the TCR induction signature ($n = 165$). The signature was enriched for genes with a known role in TCR signaling (inset). **(d)** Median relative expression (column z-score) across the 165 pathway signature genes (columns), aggregating cells that express gRNAs targeting the same gene (rows). **(e)** Distribution of signature intensity across single cells (left) and number of cells (right) for each gRNA target gene. The median is indicated with a white dot. **(f)** Gene signature concordance between CROP-seq and bulk RNA-seq in an arrayed validation screen. Known positive and negative regulators of the TCR pathway are highlighted. **(g)** Concordance of the CD69 marker of TCR induction between CROP-seq and an arrayed validation screen with flow cytometry readout. **(h)** Changes in TCR pathway induction detected by CROP-seq mapped onto a schematic of the T-cell receptor with key downstream regulators. **(i)** CD69 marker levels in control cells and knockouts for important TCR activators or repressors. **(j)** Robustness of CROP-seq signatures in a downsampling analysis at the gene and gRNA levels, evaluated against bulk RNA-seq data.

resulted in concordant TCR signatures regarding the TCR signature intensities of the targeted genes (Fig. 2j). We found that 870

single cells (15% of the data set) yielded results similar to those of the entire screen. However, the exact number of cells needed for

future screens will also depend on the strength of the investigated biological signal.

DISCUSSION

We have presented proof of concept for single-cell CRISPR screening as an approach that combines key strengths of pooled and arrayed screens, and we validated CROP-seq as a flexible and straightforward method for performing such screens. We showed that gene signatures can be derived directly from single-cell data using the transcriptomes of cells with nontargeting gRNAs, which makes the assay self-calibrating and independent of any prior knowledge about relevant marker genes.

While this paper was in review¹³, two alternative methods—Perturb-seq and CRISP-seq—were published that use transcribed barcodes for single-cell CRISPR screening^{14–16}. Compared with these methods, CROP-seq has the key advantage that it reads the gRNA directly, which dramatically simplifies single-cell CRISPR screening with large gRNA libraries and makes CROP-seq fully compatible with standard cloning protocols for pooled CRISPR screens.

We expect CROP-seq to be broadly relevant for studying biological mechanisms that are difficult to reduce to a simple read-out needed for classical pooled screens. Applied to heterogeneous cell populations and *in vivo* tissue (e.g., in combination with Cas9-expressing mice), CROP-seq can localize cell-type-specific changes in complex organs and cellular differentiation hierarchies. Given the increasing throughput of single-cell transcriptomics¹⁷ and the advent of single-cell multi-omics technology¹⁸, CROP-seq has the potential to provide comprehensive characterization of large CRISPR libraries and constitutes a powerful method for dissecting cellular regulation at scale.

METHODS

Methods, including statements of data availability and any associated accession codes and references, are available in the [online version of the paper](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We would like to thank J. Bigenzahn, A. Fauster, and M. Owusu (CeMM) for providing Cas9-expressing cell lines; M. Farlik for contributing to the Drop-seq setup; F. Müller and J. Menche for bioinformatic discussions; N. Winhofer for feedback on the illustrations; the Biomedical Sequencing Facility at CeMM for assistance with next-generation sequencing; and all members of the Bock lab for

their help and advice. C.S. is supported by a Feodor Lynen Fellowship of the Alexander von Humboldt Foundation. C.B. is supported by a New Frontiers Group award of the Austrian Academy of Sciences and by an ERC Starting Grant (European Union's Horizon 2020 research and innovation programme, grant agreement no. 679146).

AUTHOR CONTRIBUTIONS

P.D., A.F.R., C.S., and C.B. conceptualized the project; P.D. designed and developed CROP-seq; P.D., C.S., P.T., and L.C.S. conducted CROP-seq experiments; D.A. optimized sequencing protocols; P.D., T.K., L.C.S., and A.K. performed the arrayed validation screen; A.F.R. and J.K. developed software; P.D., A.F.R., C.S., and T.K. analyzed data; P.D. and A.F.R. visualized data; P.D., A.F.R., C.S., and C.B. wrote the original draft; T.K., P.T., L.C.S., A.K., and D.A. reviewed the draft; and C.B. supervised the project.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

1. Blomen, V.A. *et al.* Gene essentiality and synthetic lethality in haploid human cells. *Science* **350**, 1092–1096 (2015).
2. Wang, T. *et al.* Identification and characterization of essential genes in the human genome. *Science* **350**, 1096–1101 (2015).
3. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84–87 (2014).
4. Marceau, C.D. *et al.* Genetic dissection of *Flaviviridae* host factors through genome-scale CRISPR screens. *Nature* **535**, 159–163 (2016).
5. Lamb, J. The Connectivity Map: a new tool for biomedical research. *Nat. Rev. Cancer* **7**, 54–60 (2007).
6. Gapp, B.V. *et al.* Parallel reverse genetic screening in mutant human cells using transcriptomics. *Mol. Syst. Biol.* **12**, 879 (2016).
7. Macosko, E.Z. *et al.* Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell* **161**, 1202–1214 (2015).
8. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat. Methods* **11**, 783–784 (2014).
9. Tiscornia, G., Singer, O. & Verma, I.M. Design and cloning of an shRNA into a lentiviral silencing vector: version A. *CSH Protoc.* **1**, pdb.prot5009 (2008).
10. de Kok, S. *et al.* Rapid and reliable DNA assembly via ligase cycling reaction. *ACS Synth. Biol.* **3**, 97–106 (2014).
11. Guschin, D.Y. *et al.* A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).
12. Brownlie, R.J. & Zamoyska, R. T cell receptor signalling networks: branched, diversified and bounded. *Nat. Rev. Immunol.* **13**, 257–269 (2013).
13. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome read-out. Preprint at <http://biorxiv.org/content/early/2016/10/27/083774> (2016).
14. Adamson, B. *et al.* A multiplexed single-cell CRISPR screening platform enables systematic dissection of the unfolded protein response. *Cell* **167**, 1867–1882.e21 (2016).
15. Dixit, A. *et al.* Perturb-Seq: dissecting molecular circuits with scalable single-cell RNA profiling of pooled genetic screens. *Cell* **167**, 1853–1866.e17 (2016).
16. Jaitin, D.A. *et al.* Dissecting immune circuits by linking CRISPR-pooled screens with single-cell RNA-seq. *Cell* **167**, 1883–1896.e15 (2016).
17. Zheng, G.X.Y. *et al.* Massively parallel digital transcriptional profiling of single cells. Preprint at <http://biorxiv.org/content/early/2016/07/26/065912> (2016).
18. Bock, C., Farlik, M. & Sheffield, N.C. Multi-omics of single cells: strategies and applications. *Trends Biotechnol.* **34**, 605–608 (2016).

ONLINE METHODS

Step-by-step protocol. A step-by-step experimental protocol for CROP-seq is included as **Supplementary Protocol**. Future updates of the protocol will be shared via <http://crop-seq.computational-epigenetics.org>.

Cloning and validation of the CROPseq-Guide-Puro plasmid. To clone the CROPseq-Guide-Puro plasmid, we amplified four PCR products from LentiGuide-Puro (Addgene 52963) and assembled them in different order using the ligase cycling reaction (LCR)¹⁰. Bridge oligonucleotides for LCR were designed such that each side had a melting temperature of 70 °C. The sequences of all primer pairs, the respective annealing temperatures, and the overlapping bridge oligonucleotides are provided in **Supplementary Table 1**. Plasmid fragments were amplified by mixing 12.5 µl Q5 Hot Start High-Fidelity 2× Master Mix (NEB cat. no. M0494L), 1.25 µl of 10 µM FWD primer, 1.25 µl of 10 µM REV primer, 10 ng of plasmid template, and water to 25 µl, and incubating as follows: 98 °C for 30 s, 25× (98 °C for 10 s, Ta for 30 s, 72 °C for 1 min 30 s), 72 °C for 2 min, hold at 10 °C. PCR products were cleaned by a 2.0× AMPure XP bead cleanup (Beckman Coulter cat. no. A63880), measured in a Qubit dsDNA HS assay (Invitrogen cat. no. Q32854), and diluted to 20 nM. Next, we mixed 3.125 µl of the four diluted PCR products, added 1.5 µl of 10× Ampligase buffer and 1 µl of DpnI (20 U/µl, NEB cat. no. R0176S), and digested at 37 °C for 1 h without heat inactivation. LCR was started by mixing 15 µl of DpnI-digested PCR preparations, 1 µl of 10× Ampligase buffer, 1.5 µl of Ampligase enzyme (5 U/µl, Epicentre cat. no. A32250), 2.25 µl 5 M betaine (Sigma Aldrich cat. no. B0300-1VL), 2 µl DMSO (Sigma Aldrich cat. no. D9170-1VL), 0.5 µl each of the four bridge oligonucleotides (1.5 µM in water), and 1.25 µl of water. The reaction was incubated for 2 min at 94 °C, 50× (94 °C for 10 s, 55 °C for 30 s, 66 °C for 1 min), storage at 4 °C. 2.5 µl of the LCR assembly were transformed into 25 µl of NEB Stable *Escherichia coli* (NEB cat. no. C3040H) according to the manufacturer's high-efficiency protocol. We screened for correctly assembled clones by colony PCR and further validated the assembly by restriction digestion with either SphI or SapI as well as Sanger sequencing. Moreover, to validate the duplication of the hU6-gRNA cassette, genomic DNA from cells infected with LentiGuide-Puro lentivirus, CROPseq-Guide-Puro lentivirus, or the CROPseq-Guide-Puro plasmid pool were PCR amplified with the primers hU6_FWD1 and hU6_REV2 (**Supplementary Table 1**), which are oriented in opposite directions. In this PCR, productive amplification occurs only when the gRNA cassette is duplicated during lentiviral reverse transcription and integration, or when amplifying from a circular plasmid.

Cloning of individual gRNAs into the CROPseq-Guide-Puro plasmid. gRNA cassettes were annealed from two oligonucleotides (top: 5'-CACCG(N)₂₀-3' bottom: 5'-AAAC(N)₂₀-3') by combining 1 µl of each 100 µM oligonucleotide with 1 µl of 10× T4 ligation buffer (NEB cat. no. B0202S), 6.5 µl of water, and 0.5 µl of T4 polynucleotide kinase (NEB cat. no. M0201S), incubating as follows: 37 °C for 30 min (oligonucleotide phosphorylation), 95 °C for 5 min, then ramping from 90 °C to 25 °C at 5 °C/min. Plasmid backbone was prepared by digesting 1 µg of CROPseq-Guide-Puro with 10 units of BsmBI (NEB cat. no. R0580L) in a volume of 30 µl 1× NEB buffer 3.1, incubating for 1 h at 55 °C. To dephosphorylate the digested plasmid, we added 2 µl of shrimp alkaline phosphatase

(rSAP, NEB cat. no. M0371L), incubating for 1 h at 37 °C followed by heat inactivation (both BsmBI and rSAP) for 20 min at 80 °C. Ligation reactions were set up as follows: 1.6 µl of rSAP reaction, 1 µl gRNA cassette (diluted 1:200 in water), 5 µl 2× Quick Ligase buffer, 2.4 µl water, and 1 µl Quick Ligase (NEB cat. no. M2200S), incubated at 25 °C for 15 min. The ligation reaction was chemically transformed into NEB Stable *E. coli* (NEB cat. no. C3040H) following the manufacturer's high-efficiency protocol.

Amplification-free cloning of pooled gRNA libraries into the CROPseq-Guide-Puro plasmid. Vector backbone was prepared by digesting 5 µg of CROPseq-Guide-Puro with 20 units of BsmBI (NEB cat. no. R0580L) in a total volume of 25 µl 1× NEB buffer 3.1, incubating for 1 h at 55 °C and inactivating the restriction enzyme for 20 min at 80 °C. The 8,333 bp fragment was purified using the SNAP UV-free gel purification kit (Invitrogen cat. no. K200025). gRNAs were synthesized by Sigma Aldrich as 74 base oligonucleotides with 18 and 35 bases of homology to the hU6 promoter and guide RNA backbone, respectively. Oligonucleotides were diluted to 100 µM and pooled in equal amounts. The oligonucleotide pool was further diluted 1:1,000 for a final molarity of 100 nM for assembly into the vector backbone. gRNA libraries were cloned by Gibson's isothermal assembly: 11 fmoles (56.7 ng) of CROPseq-Guide-Puro backbone and 200 fmoles of pooled ssDNA oligonucleotides (2 µl of the 100 nM solution) were combined with 10 µl of NEBuilder HiFi DNA assembly master mix (NEB cat. no. E2621S) and water to 20 µl. After 1 h of incubation at 50 °C, reactions were desalted by filter dialysis (Merck cat. no. VMWP04700), and 10 µl of the reaction was electroporated into 25 µl of Lucigen Endura *E. coli* cells (Lucigen cat. no. 60242-2) using prechilled 1 mm electroporation cuvettes (BioRad cat. no. 1652089) in a BioRad GenePulser I machine set to 25 µF, 200 Ω, and 1.5 kV. Within seconds after the pulse, 1 ml of 37 °C Recovery Medium (Lucigen) was added and bacteria were grown in round-bottom tubes for 1 h at 37 °C while shaking at 180 r.p.m. Then, 1 ml of the bacterial culture was plated on a 25 × 25 cm bioassay plate containing LB medium (Miller) with 100 µg/ml carbenicillin. Plates were incubated at 32 °C for 22 h, then LB medium was added and cells were scraped off the plate. Bacterial cells were pelleted by 15 min of centrifugation at 5,000 RCF at 4 °C, and plasmid DNA was extracted with Qiagen's EndoFree Plasmid Mega kit (Qiagen cat. no. 12381). Library coverage was estimated by counting the number of bacterial colonies on a 1:1,000 dilution plate. The T-cell receptor library was cloned at 853× coverage.

Lentivirus production for single gRNAs and pooled CROP-seq screens. HEK293T cells were plated onto six-well plates at 1.2 million cells per well in 2 ml of lentivirus packaging medium (Opti-MEM I (Gibco cat. no. 51985-034), 5% FCS (Sigma), 200 µM sodium pyruvate (Gibco cat. no. 11360-070), no antibiotics). The next morning, cells were transfected with lipofectamine 3000 (Invitrogen cat. no. L3000015) using 1.7 µg of CROPseq-Guide-Puro (containing single gRNAs or libraries) or LentiCas9-Blast (Addgene 52962), and 0.9 µg each of the three packaging plasmids pMDLg/pRRE (Addgene 12251), pRSV-Rev (Addgene 12253), and pMD2.G (Addgene 12259). The medium was exchanged for fresh lentivirus packaging medium 6 h after the transfection. Supernatant containing viral particles was harvested at 24 h and 48 h, pooled, and passed through a 0.45 µm filter to remove cells.

Further concentration of viral particles was not required. Virus was stored at -80°C . One vial was subsequently thawed for estimating the virus titer.

Lentivirus titration for single gRNAs. HEK293T cells were seeded onto 24-well plates at 50,000 cells per well in 500 μl of culture medium (DMEM (Gibco cat. no. 10569010), 10% FCS, no antibiotics) and grown overnight to reach 30% to 50% confluence. The next day, medium was exchanged for 450 μl per well of fresh culture medium (prepared as described above) containing 8 $\mu\text{g}/\text{ml}$ polybrene (Sigma Aldrich cat. no. H9268-5G), which was also used to dilute the viral stock. Lentivirus preparations were thawed from storage at -80°C and titrated in a 1:5 dilution series ranging over ten wells (1:5 to 1:9,765,625). Each dilution was tested in duplicate by adding 50 μl per well to the 24-well plate. At least two wells per plate served as untransduced controls. 24 h after the transduction, the medium was exchanged for 500 μl per well of complete culture medium (DMEM (Gibco cat. no. 10569010), 10% FCS, penicillin–streptomycin). Starting at 48 h post-transduction, the medium was exchanged for selection medium every 2–3 d. Depending on the construct, the selection medium for HEK293T cells contained either 22.5 $\mu\text{g}/\text{ml}$ of blasticidin (Invivogen cat. no. ant-bl-5) or 2.25 $\mu\text{g}/\text{ml}$ of puromycin (Fisher Scientific cat. no. A1113803). As soon as all cells in the untransduced controls had died, medium was removed and the plate was washed once with $1\times$ PBS. Resistant colonies were stained for 15 min in a solution of 1% (w/v) crystal violet (Sigma Aldrich cat. no. 61135-25G) in a 10% solution of ethanol in water. After washing the plate several more times with $1\times$ PBS, colonies were counted, and the virus titer (transducing units per ml) was calculated as (no. of resistant colonies $\times$ dilution factor) $\times$ (1,000 $\mu\text{l}/50 \mu\text{l}$).

Lentiviral transduction with gRNA libraries or single gRNAs. For suspension cell lines (Jurkat), cells were seeded in six-well plates at 5 million cells per well in 2 ml of complete culture medium (RPMI (Gibco cat. no. 21875-034) with penicillin–streptomycin and 10% FCS (Sigma)) containing 8 $\mu\text{g}/\text{ml}$ polybrene (Sigma Aldrich cat. no. H9268-5G) and blasticidin (Invivogen cat. no. ant-bl-5) to maintain selection for LentiCas9-Blast. Wells 1 to 5 were transduced with 250 or 50 μl of the virus stock; or 50 μl of a 1:3, 1:6, or 1:12 dilution. Well 6 served as the untransduced control. Immediately after addition of the virus, cells were centrifuged at 37°C and 1,200 RCF for 45 min, followed by overnight incubation at 37°C , 5% CO_2 . After 24 h, cells were pelleted in 15 ml tubes, taken up in 30 ml of fresh complete culture medium containing blasticidin and transferred to T75 culture flasks. At 48 h post-transduction, selection with puromycin (Fisher Scientific cat. no. A1113803) was started. A flask with about 30% surviving cells was chosen and grown for 10 d to allow for efficient genome editing while renewing the selective medium (containing blasticidin and puromycin) every 2–3 d. Adherent cells (HEK293T) were plated at 5 million cells per 15 cm dish in complete culture medium (DMEM (Gibco cat. no. 41965-039), 10% FCS, penicillin–streptomycin), preparing six plates per library. After 24 h, the medium was replaced by complete culture medium containing 8 $\mu\text{g}/\text{ml}$ polybrene, and different amounts of virus were added to the plates: 250 or 50 μl of the virus stock; or 50 μl of a 1:3, 1:6, or 1:12 dilution. The sixth plate remained untransduced. After 24 h, the virus-containing medium was replaced by fresh complete culture medium. Puromycin

selection was started at 48 h post-transduction, and 2 d later a plate with about 30% surviving cells was selected and grown under blasticidin and puromycin selection for 10 d.

Preparation of lentivirus for the arrayed validation screen. We cloned 48 individual CROPseq-Guide-Puro plasmids, targeting 20 genes with two gRNAs each and including eight nontargeting controls. All plasmids were verified by Sanger sequencing with the primer 5'-TTGGGCACTGACAATTCCTG-3'. HEK293T cells were seeded onto a 96-well plate at 50,000 cells per well in 200 μl lentivirus packaging medium (Opti-MEM I (Gibco cat. no. 51985-034), 5% FCS (Sigma), 200 μM sodium pyruvate (Gibco cat. no. 11360-070), no antibiotics) and incubated overnight at 37°C and 5% CO_2 . Virus production was initiated by transfection with Lipofectamine 3000 reagents (ThermoFisher cat. no. L3000015): transfection mix A (per reaction: 25 μl Opti-MEM I, 0.6 μl P3000 Enhancer Reagent, 90 ng each of the three packaging plasmids pMDLg/pRRE (Addgene 12251), pRSV-Rev (Addgene 12253), and pMD2.G (Addgene 12259)) was distributed in a 96-well plate, and 170 ng of the respective CROPseq-Guide-Puro transfer plasmid were added. Lipid complexes were formed by addition of transfection mix B (per reaction: 25 μl Opti-MEM I, 0.7 μl Lipofectamine 3000 Reagent) and incubating for 20 min at room temperature. 100 μl of medium was removed from the HEK293T cells, then 50 μl of the lipid complexes was added. After 6 h, the medium was exchanged for fresh lentivirus production medium. Viral supernatant was collected at 24 h and 48 h, combined in a sterile deepwell plate (Corning cat. no. 07-200-700), and filtered through a sterile 0.45 μm filter plate (Merck cat. no. MSHVS4510) to remove cells.

Production of single gene knockout cell lines for the arrayed validation screen. Jurkat cells stably expressing Cas9 (after transduction with LentiCas9-Blast, Addgene 52962) were seeded onto a 96-well plate at 250,000 cells per well in 100 μl of complete culture medium (RPMI (Gibco cat. no. 21875-034), 10% FCS (Sigma), penicillin–streptomycin) containing 25 $\mu\text{g}/\text{ml}$ blasticidin (Invivogen cat. no. ant-bl-5) to maintain the selection pressure for Cas9 and 8 $\mu\text{g}/\text{ml}$ polybrene (Sigma Aldrich cat. no. H9268-5G) to facilitate the interaction of viral particles with the cells. To transduce the cells, between 2.5 and 10 μl of the prepared lentiviral particles were added with a multichannel pipette, and the plate was centrifuged at 37°C and 1,200 RCF for 45 min, followed by overnight incubation at 37°C and 5% CO_2 . After 24 h, cells were pelleted by centrifugation at 300 RCF for 5 min, and the medium was exchanged to select for transduced cells expressing the gRNA construct (RPMI, 10% FCS, penicillin–streptomycin, 25 $\mu\text{g}/\text{ml}$ blasticidin, 2 $\mu\text{g}/\text{ml}$ puromycin (Fisher Scientific cat. no. A1113803)). 48 h post-transduction, the virus concentration yielding about 30% transduced cells was chosen for further expansion. Cells were subjected to continued selection over the course of 10 d, moving to 24-well and eventually to 6-well plates.

T7 endonuclease assay. For easy-to-resolve cleavage products, PCR primers were designed such that the *DNMT3B* and *MBD1* gRNA targets are located off center of the amplicons (sequences and annealing temperatures are provided in **Supplementary Table 1**). PCR reactions were set up in a reaction volume of 50 μl : 25 μl of Q5 Hot Start High-Fidelity 2 $\times$ Master Mix (NEB cat. no. M0494L), 2.5 μl 10 μM FWD primer, 2.5 μl 10 μM REV primer, 100 ng of

genomic DNA, and water, and incubated as follows: 98 °C for 30 s, 35× (98 °C for 5 s, Ta for 10 s, 72 °C for 20 s), 72 °C for 2 min and storage at 10 °C. PCR products were purified by a 2.0× AMPure XP bead clean-up (Beckman Coulter cat. no. A63880) and measured in a Qubit HS assay (Invitrogen cat. no. Q32854). 200 ng of each PCR product were taken up in 19 µl of 1× NEB buffer 2 and subjected to denaturation and reannealing: 95 °C for 5 min, 95 to 85 °C at −2 °C/s, 85 to 25 °C at −0.1 °C/s, hold at 4 °C. Mismatched DNA duplexes were then digested by addition of 1 µl T7 endonuclease I (NEB cat. no. M0302) followed by incubation at 37 °C for 15 min. The reaction was stopped with 1.5 µl of 0.25 M EDTA (Invitrogen cat. no. 15575020), and 1 µl was analyzed on an Agilent High Sensitivity DNA chip (Agilent cat. no. 5067-4626), yielding the molarity of digested and undigested fragments from which the editing efficiencies were calculated as: $\text{average}(\text{digested1}, \text{digested2}) / \text{sum}(\text{average}(\text{digested1}, \text{digested2}), \text{undigested})$.

Indel analysis by next-generation sequencing. Amplicons were designed to be smaller than 500 bp, with the gRNA target site at the center. PCR was performed as described for the T7 endonuclease assay. PCR products were purified by a 2.0× AMPure XP bead cleanup (Beckman Coulter cat. no. A63880) and used as input for Nextera XT (Illumina cat. no. 15032350) library preparations done according to the standard protocol. Libraries were sequenced on the Illumina MiSeq platform using a 150-cycle v3 flow cell with dual indexing. The machine was set to read lengths of 159 (read1) + 8 (i7) + 8 (i5) bases. To analyze the data, we first defined two 10 bp long ‘anchor’ sequences on both sides of the gRNA at a fixed distance of 30 bp. We then extracted reads spanning the gRNA target site from the BAM file via a grep operation for the pattern “<anchor_left>.*<anchor_right>” on the BAM file, using the -o option to return only the matching part of the sequence. For unedited fragments, this sequence is exactly 100 bp long: 10 bp anchor_left + 30 bp + 20 bp (gRNA) + 30 bp + 10 bp anchor_right. The size of insertions and deletions was calculated as the deviation from the unedited fragment length, summarized, and plotted.

Plasmid library complexity and gRNA dynamics. The gRNA expression cassette was amplified from either the plasmid library or the genomic DNA of cells at day 10 post-transduction. All functional sequences required for compatibility with Illumina sequencing were introduced in a single PCR reaction with specially designed primers (**Supplementary Table 1**). We included 8 bp long i7 barcodes and a stagger sequence to increase library complexity, resulting in a set of primer variants that can be freely combined. Purified PCR products were pooled and sequenced on an Illumina MiSeq machine, using a 150-cycle v3 flowcell with a read configuration of 159 (read1) + 8 (i7) + 8 (i5). Read1 covers the stagger sequence (variable length), the FWD primer binding site (23 bases), the gRNA target (20 bases), and the backbone (82 bases). gRNA read counts were normalized by the total number of sequenced reads. The number of cells assigned to a specific gRNA in CROP-seq (regardless of the number of genes per cell) was normalized to the total number of cells assigned to any gRNA. Fold change values were calculated between normalized cell counts from the CROP-seq screen and the normalized gRNA counts from the plasmid library.

Anti-CD3/CD28 stimulation of the T-cell receptor pathway in Jurkat cells. Jurkat cells were serum starved for 3 h before stimulation. They were stimulated with 1 µg/ml anti-CD3 (eBioscience cat. no. 16-0037) and 1 µg/ml anti-CD28 antibody (eBioscience cat. no. 16-0289-81) for 4 h, while the unstimulated (naive) control was subjected to continued starvation. Samples collected for both conditions were subjected to CROP-seq.

Anti-CD3/CD28 stimulation of single gene knockout cells for the arrayed validation screen. Jurkat cells that were knockouts for single genes were grown in six-well plates and split into two equal parts of 900 µl in a sterile deepwell storage plate (Corning cat. no. 07-200-700). The cells were washed once with 800 µl of starvation medium (RPMI (Gibco cat. no. 21875-034), 0% FCS, penicillin–streptomycin) and then resuspended in 600 µl of starvation medium. After 3 h of serum starvation, one part was stimulated by addition of 1 µg/ml anti-CD3 (eBioscience cat. no. 16-0037) and 1 µg/ml anti-CD28 antibody (eBioscience cat. no. 16-0289-81) diluted in serum-free medium, while the other part was subjected to continued starvation. After another 4 h, we proceeded with extraction of genomic DNA and RNA (Qiagen cat. no. 80311) as well as flow cytometry.

Flow cytometry for the arrayed validation screen. Following treatment, anti-CD3/CD28 stimulated and unstimulated Jurkat T cells were transferred to 96-well V-bottom plates and washed twice with PBS containing 0.1% BSA and 5 mM EDTA (PBS + BSA + EDTA). Cells were then incubated with anti-CD16/CD32 (clone 93, Biolegend) to prevent nonspecific binding. Single-cell suspensions were stained with antibodies against CD25 (clone BC96), CD38 (clone HB-7), CD69 (clone FN50), CD82 (clone ASL-24), CD154 (clone 24-31), PD-1 (clone EH12.2H7), and Zombie NIR fixable viability dye (all from Biolegend) for 30 min at 4 °C. Cells were washed twice with PBS + BSA + EDTA and fixed in Biolegend fixation/permeabilization buffer for 1 h at room temperature followed by permeabilization using True-Nuclear Transcription Factor buffer (Biolegend) for 45 min in the presence of antibody against Ki-67 (clone Ki-67). Cells were washed twice with permeabilization buffer, acquired with a LSRFortessa (BD) cell analyzer, and data analysis was performed with FlowJo (Tree Star) software.

Drop-seq protocol for highly multiplexed single-cell RNA-seq in microfluidic droplets. Adherent cells were detached using Trypsin–EDTA (Gibco cat. no. 25300-054) following standard cell culture practices. Cells were collected by centrifugation at 300 RCF for 5 min, washed once with PBS–0.01% BSA (freshly prepared on the day of the run), and resuspended in 1 ml of PBS–0.01% BSA. Cells were filtered through a 40 µm cell strainer to obtain a single-cell suspension, and cell numbers were counted using a CASY device. Single cells were then coencapsulated with barcoded beads (ChemGenes cat. no. Macosko-2011-10) using an Aquapel-coated PDMS microfluidic device (FlowJEM) connected to syringe pumps (kdScientific) via polyethylene tubing with an inner diameter of 0.38 mm (Scicominc cat. no. BB31695-PE/2). Cells were supplied in PBS–0.01% BSA at a concentration of 220 cells/µl, while barcoded beads were resuspended in Drop-seq lysis buffer at a concentration of 150 beads/µl. The flow rates for cells and beads were set to 1.6 ml/h, while droplet generation oil (BioRad cat. no. 1864006) was provided at 8 ml/h. During the run, the barcoded bead solution was mixed by magnetic stirring with a mixing disk

set to 1 jump/s. A typical run lasted between 35 and 40 min. In case of multiple runs per day, droplets were temporarily stored at 4 °C and processed together. Our most important modification to the Drop-seq protocol was an alternative way to break droplets, which recovered beads much more efficiently than in the original publication of Drop-seq⁷. After removing as much oil below the droplet layer as possible, we added 30 ml of 6× SSC buffer (Promega cat. no. V4261) and 1 ml of Perfluorooctanol (Sigma Aldrich cat. no. 370533-25G) and shook the tube forcefully six times to break the droplets. Based on their large diameter, beads were then collected by syringe filtering the solution through a 0.22 µm filter unit (Merck cat. no. SLGV033RS), washing 2× with 20 ml of 6× SSC buffer and eluting by turning the filter upside down and rinsing it with 3× 10 ml of 6× SSC buffer. Beads were collected by centrifugation at 1,250 RCF for 2 min, setting the brake speed to 50%. After washing a second time with 10 ml 6× SSC buffer, the pellet was taken up in 200 µl of 5× RT buffer and transferred to a 1.5 ml tube. Reverse transcription and exonuclease I treatment were performed as described in the original publication⁷, and the number of barcoded beads was estimated using a Fuchs–Rosenthal counting chamber (mixing the bead suspension with 6× DNA loading dye). Depending on the performance of the experiment, we prepared up to 24 PCR reactions per Drop-seq run, adding 4,400 beads (~220 cells) per PCR and enriching the cDNA for 4 + 10 cycles, using the already described reagents. We then prepared Drop-seq libraries using the Nextera XT kit (Illumina cat. no. 15032350), starting from 1 ng of cDNA pooled in equal amounts from all PCR reactions for a given run. We typically required an additional ten enrichment cycles using the Illumina Nextera XT i7 primers along with the Drop-seq New-P5 SMART-PCR hybrid oligonucleotide. The slightly increased cDNA input typically resulted in an average size distribution of about 575 bp. After quality control, libraries were sequenced with paired-end SBS chemistry on Illumina HiSeq 3000/4000 instruments. Drop-seq Custom Read1 Primer was spiked into the HP10 primer solution, located in column 11 of the cBot Reagent Plate at 1 µM final concentration. High sequence complexity needed for optimal base calling performance was achieved by adding 20–30% PhiX as spike in. Cluster generation and Read1 primer hybridization were completed using the Illumina cBot protocol “HiSeq_3000_4000_HD_Exclusion_Amp_v1.0.” Alternatively, libraries were sequenced on an Illumina NextSeq 500 instrument using the 75 cycle High Output v2 Kit (Illumina cat. no. FC-404-2005). We loaded 1.8 pM library and provided Drop-seq Custom Read1 Primer at 0.3 µM in position 7 of the reagent cartridge. On NextSeq machines, we sequenced without PhiX spike-in, using a read configuration of 20 bases (Read1), 8 bases (Index1), and 64 bases (Read2).

Preprocessing of single-cell sequencing data. Single-cell sequencing data were processed using the Drop-seq Tools v1.12 software⁷. Briefly, each transcriptome Read2 was tagged with the cell barcode (bases 1 to 12) and unique molecular identifier (UMI) barcode (bases 13 to 20) obtained from Read1, trimmed for sequencing adapters and poly-A sequences, and aligned using STAR v2.4.0 (ref. 19) to a concatenation of the mouse and human genomes (for the species-mixing experiment) or to the human reference genome assembly (Ensembl GRCh38 release), which was extended with artificial chromosomes that represent the CROPseq-Guide-Puro plasmid construct (250 bp common U6 promoter sequence, one gRNA sequence per artificial chromosome (20 bp), and the

remaining 260 bp downstream plasmid backbone until the poly-A consensus sequence). Cell barcodes were corrected for possible bead synthesis errors, allowing the removal of up to four bases using the DetectBeadSynthesisErrors tool from the Drop-seq Tools v1.12 software. Reads aligning to exons were tagged with the respective gene name, and counts of UMI-deduplicated reads per gene within each cell were used to build a digital gene expression matrix comprising all cells with transcript counts for at least 500 genes. This matrix was converted to transcripts per million and log transformed for further analysis. We excluded weakly expressed genes and genes coding from ribosomal proteins from the analysis given that their widespread detection may be caused by solubilization of their mRNAs as noted previously⁷. For the assignment of gRNAs to cells, we quantified the overlap of UMI-deduplicated reads to the specific gRNA sequence within the CROPseq-Guide-Puro plasmid chromosomes and assigned the most abundant gRNA to the respective cell. For the analysis shown in **Figure 1i**, cells were classified as uniquely assigned if the dominant gRNA was three times more abundant than the sum of all other gRNAs; they were classified as containing multiple gRNAs where this difference was smaller than three; and they were classified as unassigned where no reads were overlapping with the gRNA sequence. For the Jurkat screen, we calculated the purity for each cell as the fraction of unique reads for the most abundant detected gRNA over the sum of all unique gRNA reads in the same cell, and we retained only cells with >0.9 purity for further analysis.

Preprocessing of bulk RNA sequencing data. Reads were trimmed with Trimmomatic²⁰ and aligned to the GRCh38 assembly of the human genome using Bowtie1²¹ with the following parameters: -q -p 6 -a -m 100 -minins 0 -maxins 5000 -fr -sam -chunkmbs 200. Duplicate reads were removed with Picard's MarkDuplicates utility using standard parameters, followed by transcript quantification with BitSeq²² using the Markov chain Monte Carlo method and standard parameters. To obtain gene-level quantifications, we assigned to each gene the expression value of its most highly expressed transcript, calculated transcripts per million, and log transformed the expression matrix.

Transcriptome signature analysis. We first investigated the concordance between gRNAs across the whole transcriptome by measuring pairwise distances (L2-norm metric) between gRNAs. We used the median expression of single cells with the same gRNA assigned and assessed the significance of differences between the distributions of gRNAs targeting the same gene and gRNAs targeting different genes using the Mann–Whitney U test. Significant perturbations by a gRNA were assessed by grouping cell transcriptomes in the same manner (median) and comparing each gRNA with the respective stimulation condition, assessing significance using the Mann–Whitney U test. For gene-level quantifications, we combined the gRNA *P* values obtained from the previous step with Fisher's method.

Sparse transcriptome signatures were used to position single cells sharing a putative perturbation (here, gene knockouts) within a multidimensional trajectory of cell states (here, measured activation of the TCR pathway in Jurkat cells). This method is reminiscent of a method for the assessment of cross-contamination between cell types in single cells²³ or a recently described approach for monitoring cell fate conversion²⁴. We identified genes associated with

the TCR response (signature genes), selected cells assigned to a nontargeting control gRNA as representatives of the unperturbed stimulation (control cells), produced *in silico* mixtures of gene expression profiles of increasingly more stimulated cells based on the control cells from both conditions (mix profiles), and identified the mixed profile that best matched the expression profile of each single cell (signature position).

To determine a TCR-specific activation signature, we aggregated single cells by their assigned gRNA target genes and performed principal component analysis on all genes. This analysis detected a clear separation of genes by the TCR-activating anti-CD3/CD28 stimulation condition in principal component 1 (**Supplementary Fig. 6c**). Signature genes were defined as having an absolute loading higher than the 99th percentile for this principal component. Bioinformatic analysis of the signature gene function was performed with the gene set enrichment analysis tool Enrichr²⁵, and the retrieved combined score ($\log(P \text{ value}) \times z\text{-score}$) was displayed.

Based on the median expression level of cells assigned to nontargeting gRNAs, we constructed for the signature genes a matrix Z of stepwise weighted linear mixtures of expression profiles between the unstimulated and stimulated conditions, which is given by:

$$Z_{i..n,j..m} = \frac{\mu_{a,i} \times j + \mu_{b,i} \times (m-j)}{m} \quad (1)$$

where μ_a and μ_b are vectors with mean expression values of cells in the unstimulated and stimulated conditions, respectively; and i and j are indices of the signature genes (n) and the number of desired mixtures m . We selected $m = 100$; and to account for cell-to-cell variability as well as overshooting changes in the signature genes compared with the mean of the control cells, we generated an extended linear space of length m with boundaries -20 and 120 . For $m = 100$, this is equivalent to μ_a and μ_b being placed at index $j = 20$ and $j = 80$ of the Z matrix, respectively.

We positioned each single cell in the matrix by retrieving the arg-max of the Pearson correlation of each cell to each mixture in the matrix and visualized the relationship between the correlation arg-max and the mean number of unique reads per cell, the minimum and $(1 - \text{maximum})$ correlation to ensure these variables were not

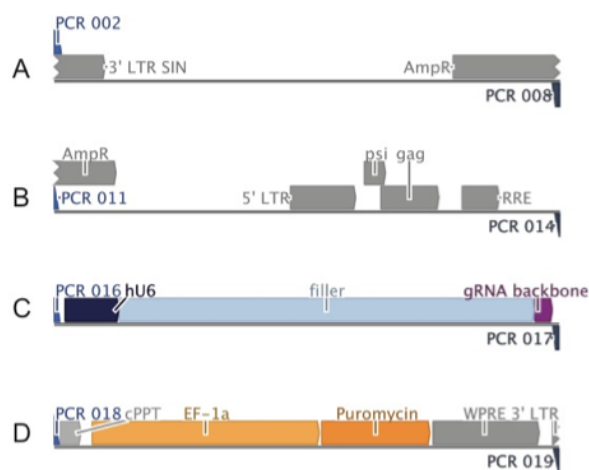
confounding the signature positioning. Grouping cells by their gRNA assignment, we visualized the distribution of cell signature positions per group and calculated the mean group signature position for groups with more than ten cells. We then calculated the log fold deviation of each group of cells to the group of control cells within the respective stimulation condition to determine a deviation relative to genetically unperturbed cells.

The above procedure was repeated for each bulk RNA-seq sample on the same set of genes, but with the μ_a and μ_b of the signature matrix as the medians of the eight nontargeting control bulk cell lines. Signature values for gRNAs or genes obtained from the CROP-seq data were compared with those derived from bulk RNA-seq if available, and the Pearson and Spearman correlations were calculated. To assess the robustness of signature positions derived from the CROP-seq data, we subsampled 100 fractions of the data set and compared the Pearson correlation of the gRNA or gene aggregated signatures with those provided by the bulk RNA-seq as reference at each fraction for 100 iterations with random sampling.

Data availability statement. The CROPseq-Guide-Puro plasmid is available from Addgene (86708). Raw and processed CROP-seq data are provided via the NCBI GEO database (GSE92872). The source code of the computational analyses is included as **Supplementary Software** and maintained in an open Github repository (<https://github.com/epigen/crop-seq>).

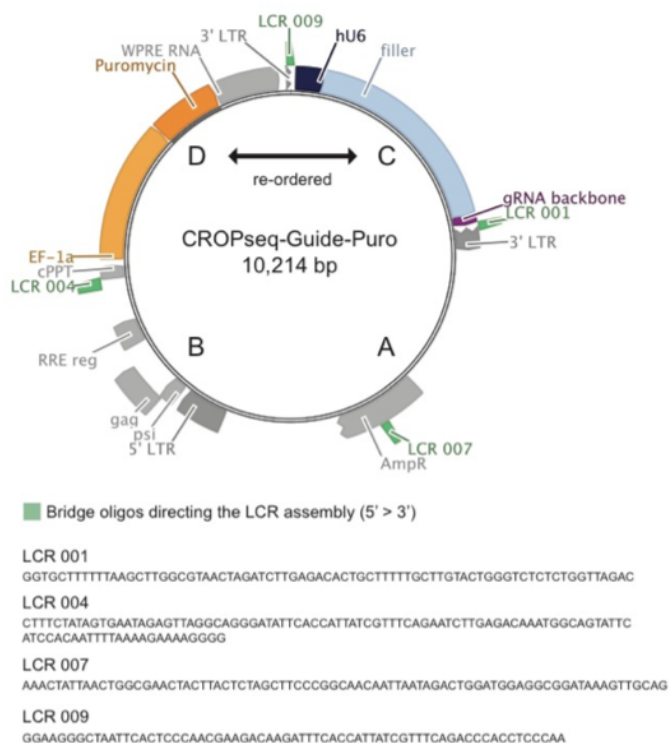
19. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
20. Bolger, A.M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
21. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
22. Glaus, P., Honkela, A. & Rattray, M. Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics* **28**, 1721–1728 (2012).
23. Li, J. *et al.* Single-cell transcriptomes reveal characteristic features of human pancreatic islet cell types. *EMBO Rep.* **17**, 178–187 (2016).
24. Treutlein, B. *et al.* Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* **534**, 391–395 (2016).
25. Kuleshov, M.V. *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).

a Step 1 - Amplification of PCR products from LentiGuide-Puro

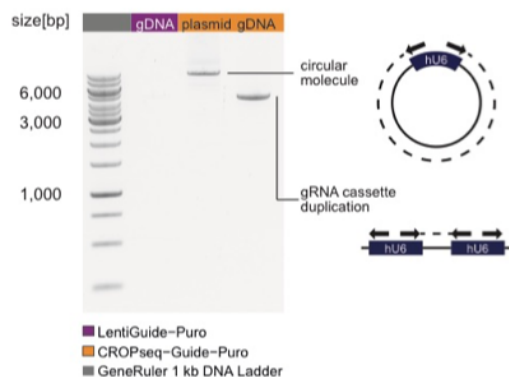


Product	PCR primers
A	FWD 002 pCTGCTTTTGGCTGTACTGGTCTCTCTG REV 008 pCCGGAAGCTAGAGTAAGTAGTTCGCCAG
B	FWD 011 pCAACAATTAATAGACTGGATGGAGCGG REV 014 pTCTGAAACGATAATGGTGAATATCCCTGC
C	FWD 016 pTTCACCATTATCGTTTCAGACCCACC REV 017 pTGTCTCAAGATCTAGTTACGCCAAGC
D	FWD 018 pATCTTGAGACAAATGGCAGTATTCATCCAC REV 019 pATCTTGCTCTCGTTGGAGTGAATTAGCC

b Step 2 - Assembly of CROPseq-Guide-Puro from PCR products via the ligase cycling reaction (LCR)



c Validation of gRNA cassette duplication after viral integration

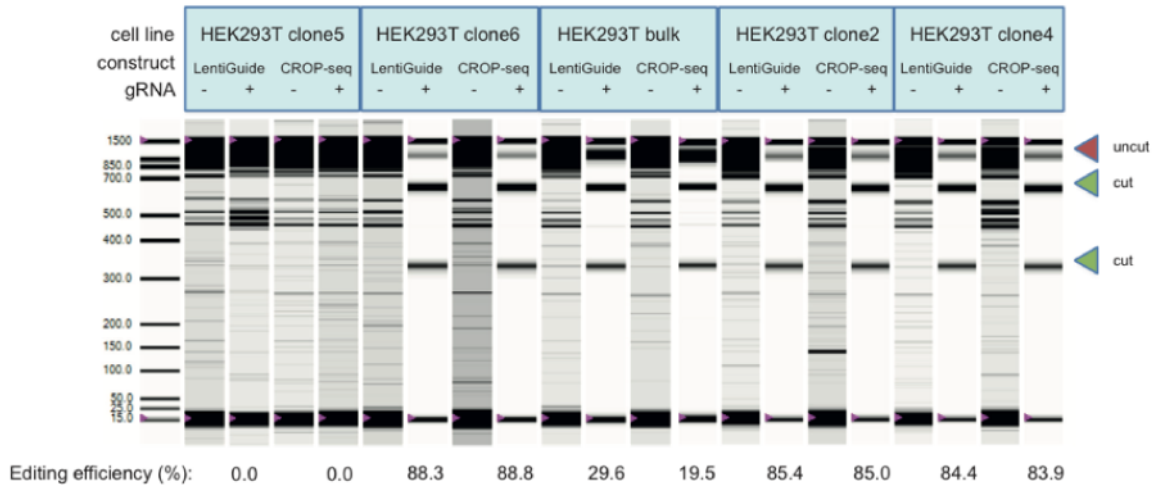


Supplementary Figure 1

Cloning and validation of the CROPseq-Guide-Puro plasmid

a) As starting point for preparing the CROPseq-Guide-Puro plasmid, we amplified four PCR products (A, B, C, and D) from the LentiGuide-Puro plasmid with the indicated primer pairs. **b)** CROPseq-Guide-Puro was constructed from these four amplicons using the ligase cycling reaction (LCR). The assembly was directed by four overlapping bridge oligonucleotides to flip the order of parts C and D. This rearrangement places the hu6-gRNA cassette into the 3' LTR, downstream of the EF-1a puromycin marker. **c)** To validate the duplication of the hu6-gRNA cassette during lentiviral integration, we performed PCRs with primers that bind to the hu6 promoter but face in opposite directions. Productive amplification can occur only when amplifying from a circular plasmid or following duplication of the cassette during viral integration. As templates, we used gDNA from LentiGuide-Puro transduced cells (lane 1, resulting in no amplification), a plasmid preparation of CROPseq-Guide-Puro (lane 2), or gDNA from CROPseq-Guide-Puro transduced cells (lane 3).

a T7 endonuclease assays for bulk as well as clonally expanded HEK293T cells



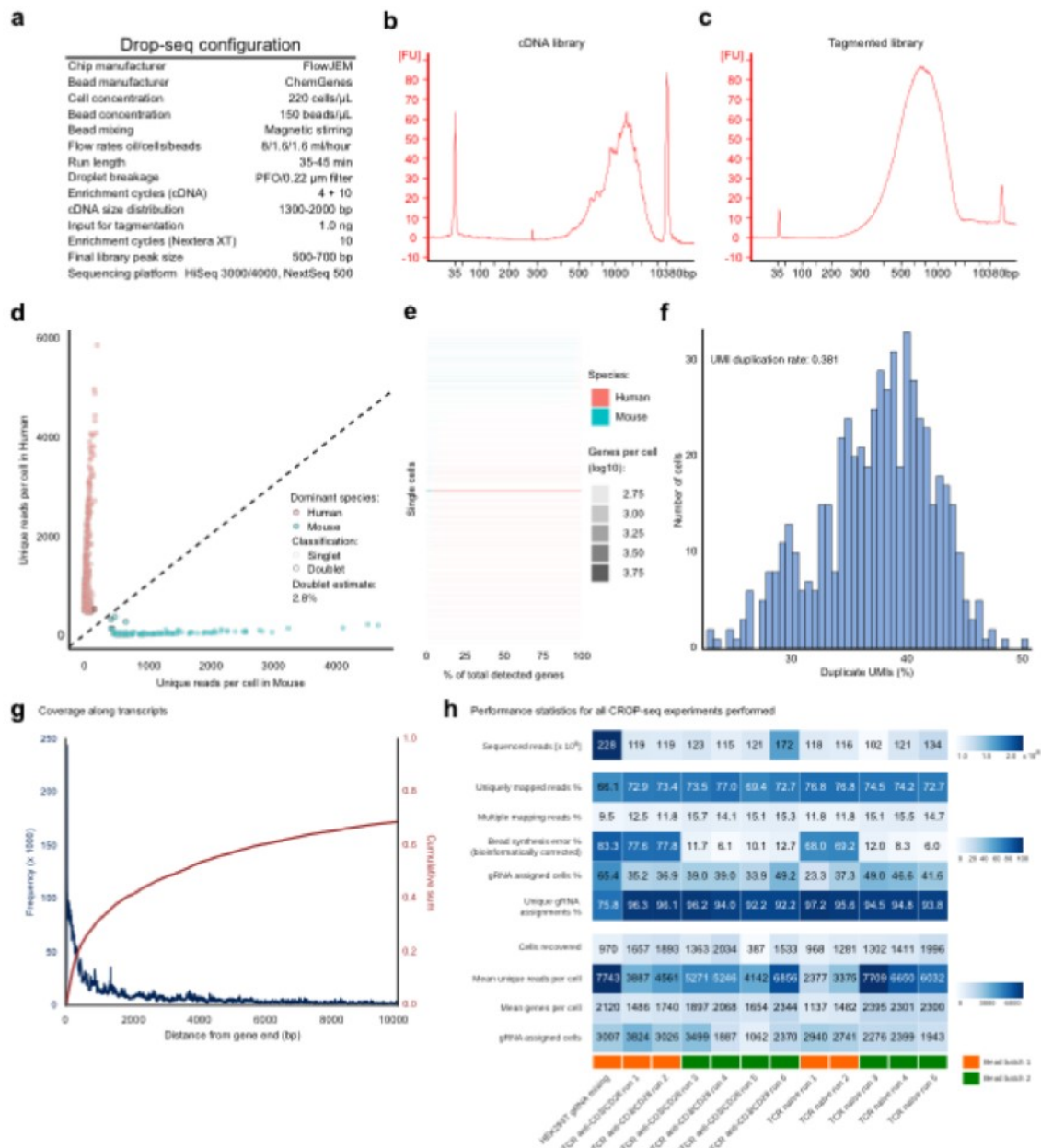
b Summary table of all T7 endonuclease assays performed

sample_ID	cell_line	viral_construct	gene	undigested [nM]	digested 1 [nM]	digested 2 [nM]	editing_efficiency[%]
1	HEK clone5	LentiGuide-Puro	MBD1	9.26	0	0	0.0
2	HEK clone5	CROPseq-Guide-Puro	MBD1	9.52	0	0	0.0
3	HEK clone6	LentiGuide-Puro	MBD1	1.01	7.26	8.02	88.3
4	HEK clone6	CROPseq-Guide-Puro	MBD1	1.07	8.05	8.98	88.8
5	HEK bulk	LentiGuide-Puro	MBD1	7.25	2.85	3.24	29.6
6	HEK bulk	CROPseq-Guide-Puro	MBD1	7.79	1.85	1.93	19.5
7	HEK clone2	LentiGuide-Puro	MBD1	1.44	8.22	8.68	85.4
8	HEK clone2	CROPseq-Guide-Puro	MBD1	1.54	8.36	9.16	85.0
9	HEK clone4	LentiGuide-Puro	MBD1	1.58	8.3	8.76	84.4
10	HEK clone4	CROPseq-Guide-Puro	MBD1	1.69	8.55	9.09	83.9
11	HEK clone5	LentiGuide-Puro	DNMT3B	10.54	0	0	0.0
12	HEK clone5	CROPseq-Guide-Puro	DNMT3B	8.81	0	0	0.0
13	HEK clone6	LentiGuide-Puro	DNMT3B	0.92	12.37	9.71	92.3
14	HEK clone6	CROPseq-Guide-Puro	DNMT3B	0.76	11.06	8.2	92.7
15	HEK bulk	LentiGuide-Puro	DNMT3B	6.37	2.61	2.13	27.1
16	HEK bulk	CROPseq-Guide-Puro	DNMT3B	6.48	2.64	2.25	27.4
17	HEK clone2	LentiGuide-Puro	DNMT3B	1.24	14.26	11.64	91.3
18	HEK clone2	CROPseq-Guide-Puro	DNMT3B	1.17	14.25	11.2	91.6
19	HEK clone4	LentiGuide-Puro	DNMT3B	0.97	12.97	9.97	92.2
20	HEK clone4	CROPseq-Guide-Puro	DNMT3B	0.91	12.37	9.45	92.3
21	K562 bulk	LentiGuide-Puro	MBD1	0.7	6.61	6.88	90.6
22	K562 bulk	CROPseq-Guide-Puro	MBD1	0.63	6.31	6.12	90.8
23	Jurkat bulk	LentiGuide-Puro	MBD1	1.75	7.31	8.33	81.7
24	Jurkat bulk	CROPseq-Guide-Puro	MBD1	1.8	7.57	8.87	82.0
25	KBM7 bulk	LentiGuide-Puro	MBD1	0.98	6.89	7.32	87.9
26	KBM7 bulk	CROPseq-Guide-Puro	MBD1	2.55	11.7	10.72	81.5
27	K562 bulk	LentiGuide-Puro	DNMT3B	0.89	10.38	8.65	91.4
28	K562 bulk	CROPseq-Guide-Puro	DNMT3B	0.83	10.59	8.54	92.0
29	Jurkat bulk	LentiGuide-Puro	DNMT3B	1.24	11.66	9.75	89.6
30	Jurkat bulk	CROPseq-Guide-Puro	DNMT3B	1.86	10.63	9.53	84.4
31	KBM7 bulk	LentiGuide-Puro	DNMT3B	1.65	6.84	6.4	80.0
32	KBM7 bulk	CROPseq-Guide-Puro	DNMT3B	2.32	8.34	7.4	77.2

Supplementary Figure 2

Genome editing efficiencies of LentiGuide-Puro and CROPseq-Guide-Puro based on the T7 endonuclease assay

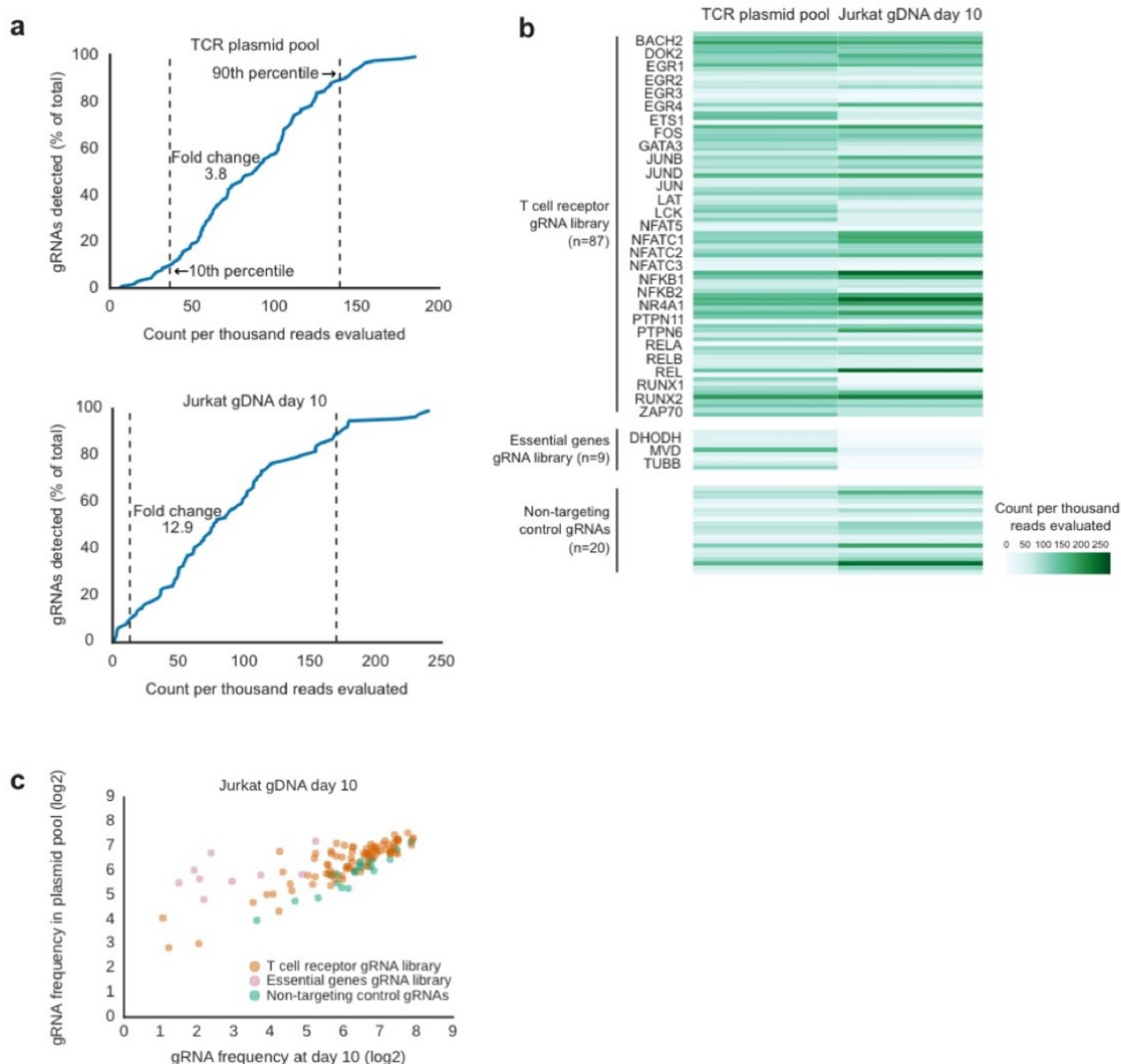
a) Clonally expanded HEK293T cell lines as well as a HEK293T bulk population were transduced with LentiGuide-Puro (LentiGuide) or CROPseq-Guide-Puro (CROP-seq) vectors containing a gRNA targeting the MBD1 locus (+) or targeting a different locus (-). Genome editing efficiencies for MBD1 were measured using the T7 endonuclease assay, which indicated highly similar performance between the two vectors. HEK293T clone 5 did not show any genome editing and was not used for further experiments. **b)** Table summarizing genome editing efficiencies for four cell lines (HEK293T, K562, Jurkat, KBM7) and two gRNAs (MBD1, DNMT3B).



Supplementary Figure 3

Configuration and validation of the droplet-based assay for single-cell transcriptome profiling

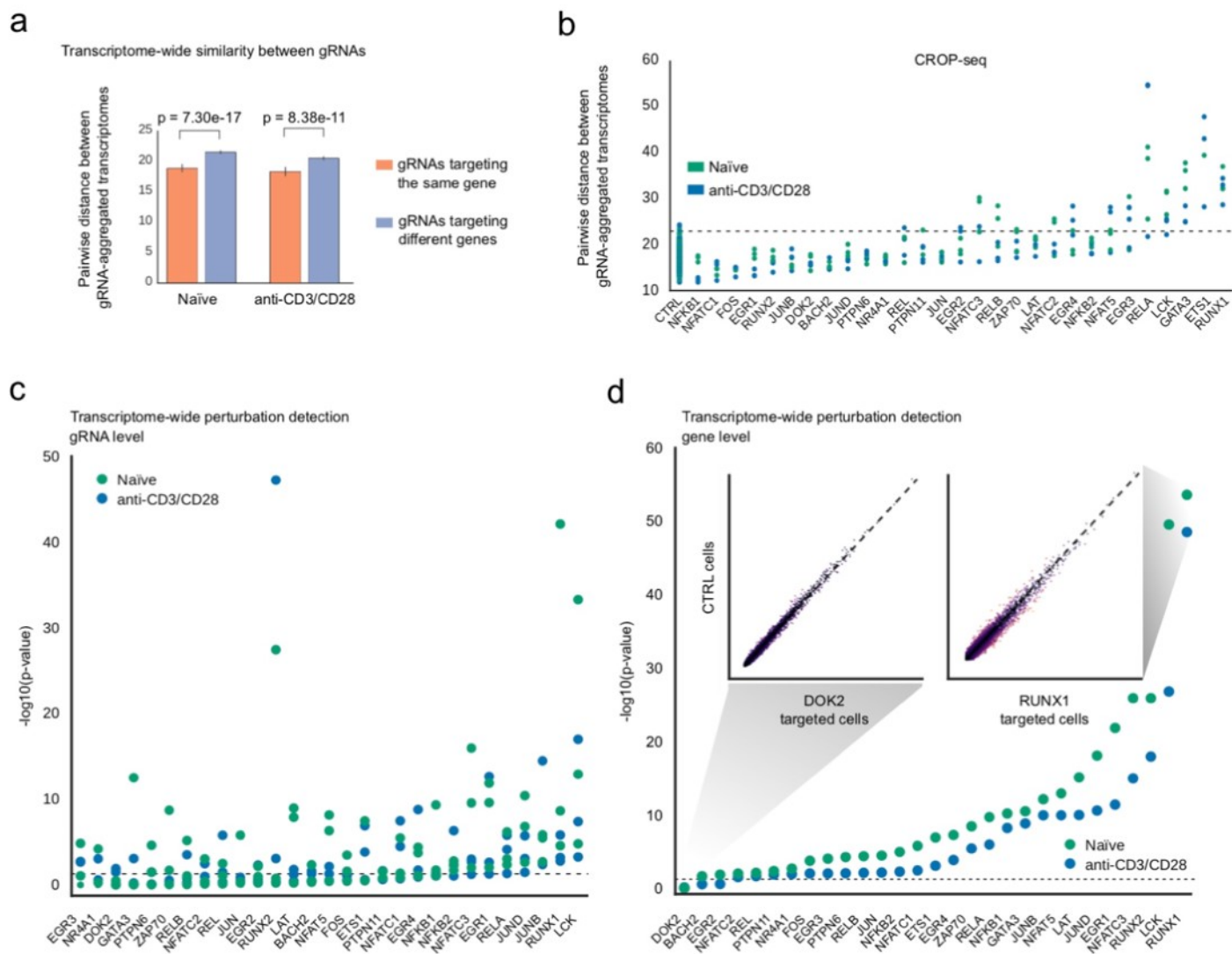
a) Setup of the Drop-seq workflow used as part of CROP-seq. **b)** Bioanalyzer trace of a typical cDNA library prepared with CROP-seq. **c)** Electropherogram of a sequencing-ready CROP-seq library after tagmentation. **d)** Doublet estimates based on a HEK293T (human) / 3T3 (mouse) mixing experiment across all detected cells and transcripts (without filtering). **e)** Percent of detected genes aligning to the human and mouse transcriptomes (filtered for cells with more than 500 detected genes). **f)** PCR duplication rates based on unique molecular identifiers (UMIs) in the HEK293T (human) / 3T3 (mouse) mixing experiment. **g)** Distribution of the distance of read mapping positions to the 3' end of gene models (blue line) and their cumulative sum (red line). **h)** Detailed performance statistics for twelve CROP-seq experiments. Green and orange labels indicate different batches of Drop-seq beads, where batch 1 suffered from production problems affecting the cell barcodes, which have been bioinformatically corrected to improve the data quality of the affected samples.



Supplementary Figure 4

Validation of the T cell receptor gRNA library and gRNA dynamics

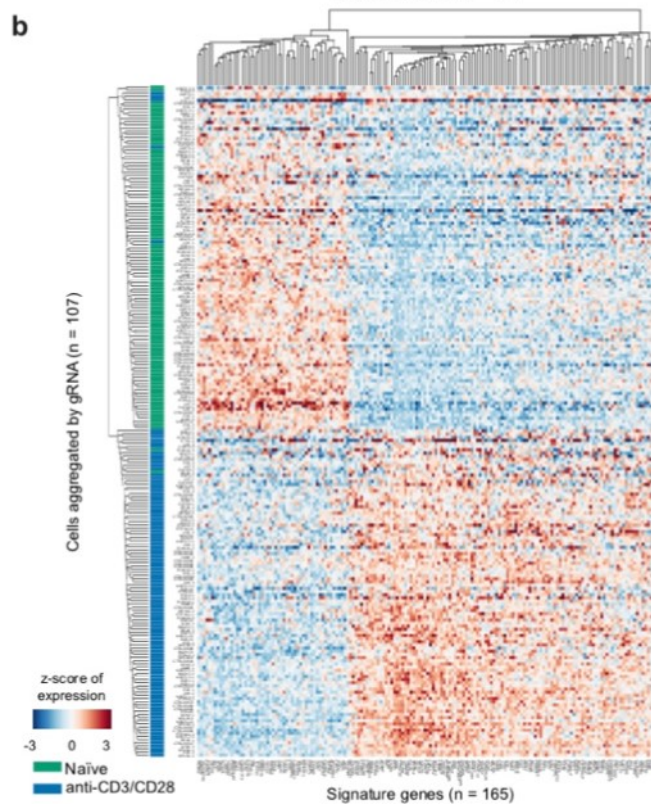
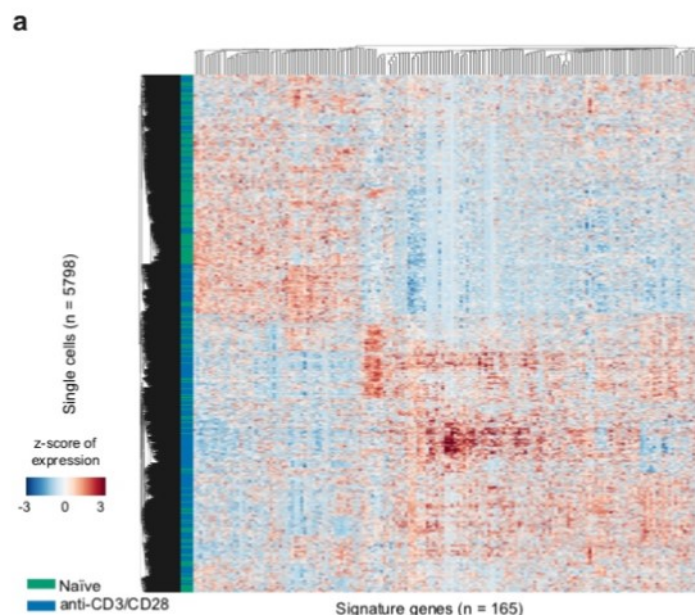
a) gRNA representation in the T cell receptor (TCR) gRNA library, assessed by amplicon sequencing of the plasmid pool (top) and the gDNA of Jurkat cells at day 10 post transduction with CROPseq-Guide-Puro (bottom), both displayed as cumulative distribution plots. The fold change between the 10th and 90th percentile is highlighted as a measure of library imbalance, which expectedly increases upon transduction. **b)** Abundance of each gRNA shown as a heatmap. **c)** Scatterplots of gRNA abundance from amplicon libraries at day 10 versus the original plasmid library. Frequencies of detected gRNAs have been normalized to the evaluated reads in each experiment.



Supplementary Figure 5

Similarities and differences in the transcriptome response at the gRNA and gene level

a) Mean and standard deviation of pairwise distances (L2-norm) between CROP-seq transcriptomes for pairs of gRNAs that target the same gene (orange) or different genes (blue). Statistical significance was assessed with the Mann-Whitney U test. **b)** Pairwise distances as in panel a, shown separately for gRNAs targeting specific genes and for naïve as well as anti-CD3/CD28 stimulated cells. The dotted line indicates the 99th percentile of the distribution of distances between non-targeting gRNAs. **c)** Statistical significance of the transcriptome-wide effect induced by gRNAs targeting specific genes relative to cells with non-targeting gRNAs (based on the Mann-Whitney U test). The dotted line indicates a p-value of 0.01. **d)** As in panel c, but aggregated at the gene level by combining p-values using Fisher's method. Insets show scatterplots for the expression of two example genes with low (left) and high (right) systematic effects on the transcriptome compared to cells expressing non-targeting control gRNAs in the same stimulation condition (y-axis).



c

1) Creating matrix of synthetic transcriptomes between cell states

$$Z_{(n \times m)} = \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{pmatrix} \cdot j + \begin{pmatrix} g_1 \\ g_2 \\ \vdots \\ g_m \end{pmatrix} \cdot (m - j)$$

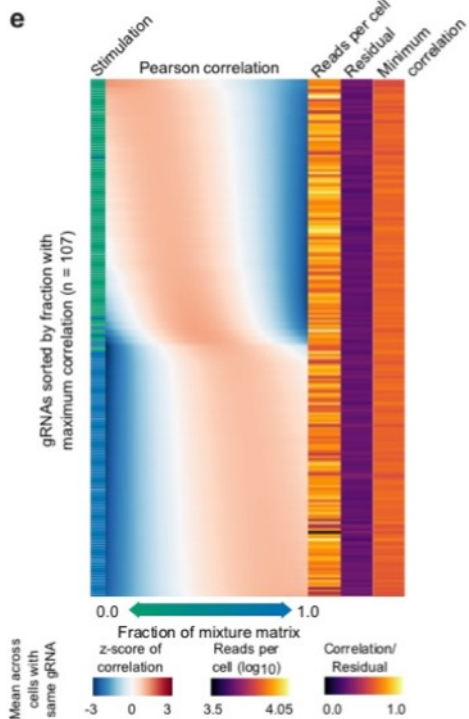
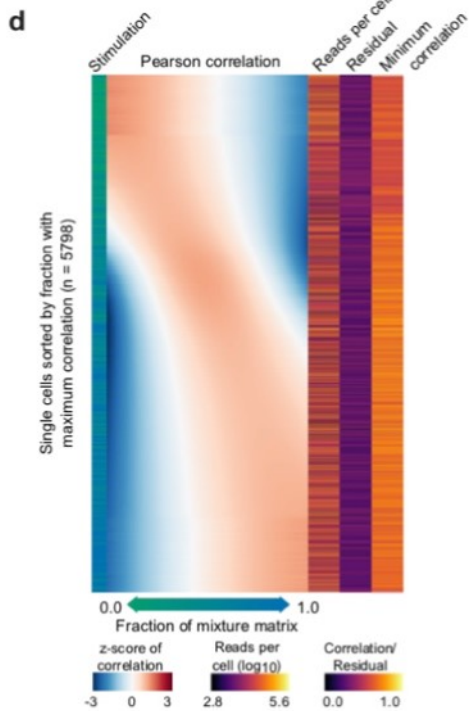
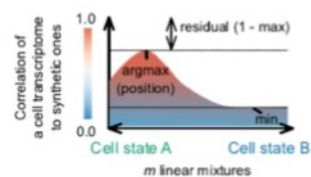
n signature genes (i, j)

m linear mixtures of transcriptome profiles (j, m)

2) Positioning cells in the gradient of cell state mixtures

$$C_{i,p} = \underset{1..K}{\operatorname{argmax}} \operatorname{Pearson}(E_{i,p}, Z_p)$$

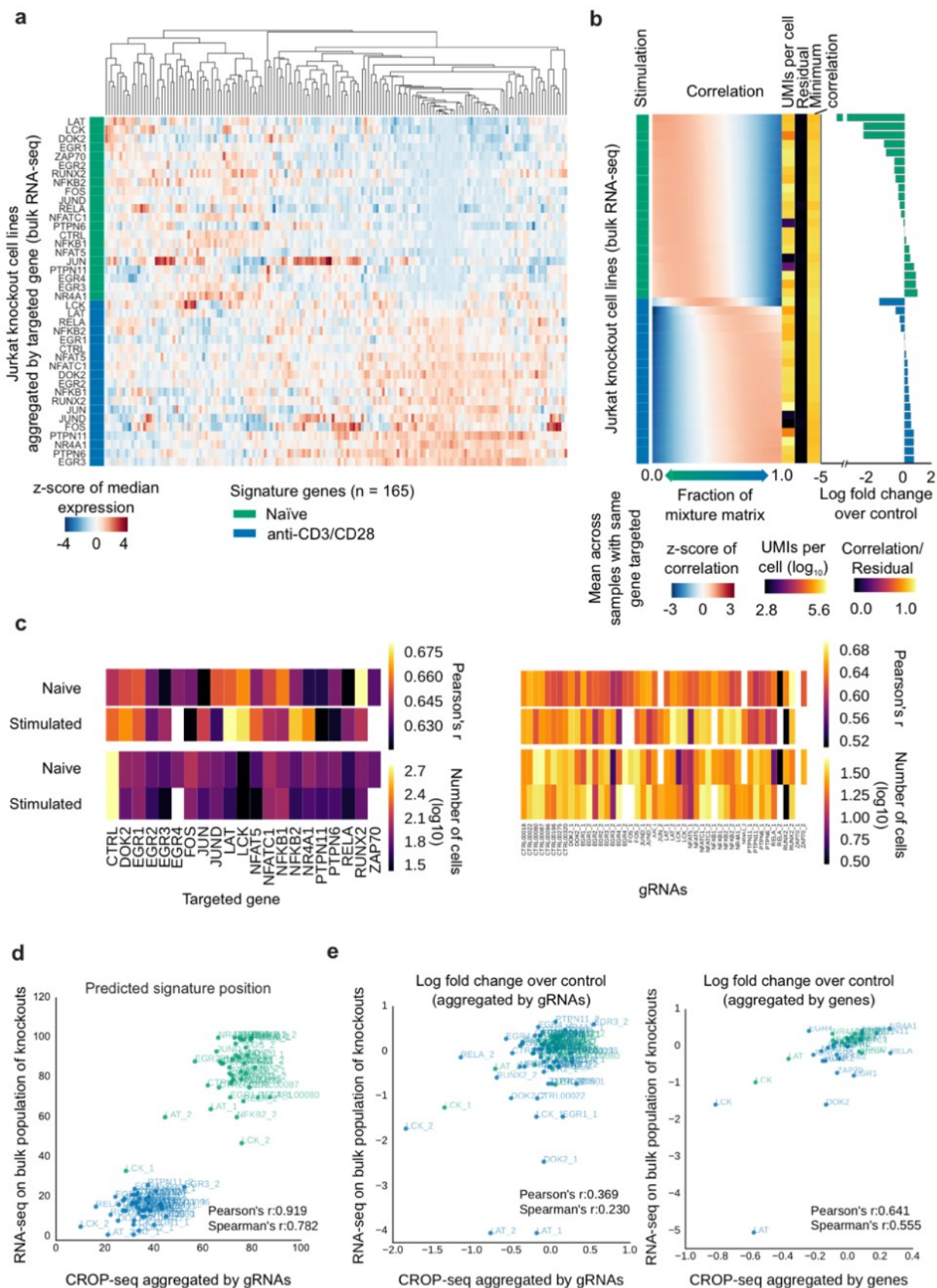
for p cells (i, j, p) with E as an expression matrix



Supplementary Figure 7

Positioning cells and target genes on a spectrum defined by naïve and stimulated cell states

a) Hierarchical clustering of single-cell transcriptomes with unambiguously assigned gRNA ($n = 5,798$) based on all genes included in the TCR activation signature. Clustering for cells and genes used the Pearson correlation, and the z-score of expression is displayed along with the stimulation state for each cell (left column). **b)** Hierarchical clustering of median gene expression values aggregated across cells expressing gRNAs for the same target gene ($n = 107$). **c)** Analytical procedure for assigning each cell to a specific position on a spectrum defined by the CROP-seq transcriptomes of non-targeted cells in the naïve and the anti-CD3/CD28 stimulated cell state. In a first step (left), a matrix of synthetic transcriptome signatures (Z) is built by linear combination of the transcriptomes in the two defining cell states (μ_A, μ_B). In a second step (right), the position of the Z matrix that shows the maximum Pearson correlation with the transcriptome of the cells (E matrix) is taken as the cell's position along the spectrum of cell states. **d)** Correlation (row-wise z-score) of single-cell transcriptomes with a matrix comprising synthetic mixtures of transcriptome profiles between the median of non-targeted cells in both conditions (values close to one reflect similarity with anti-CD3/CD28 stimulated cells). Additional data on the transcriptome quality for each cell (unique reads per cell) and the overall correlation performance are shown as columns and reveal no relationship with the inferred position of the corresponding cells. All cells were ordered by the inferred signature value (position with maximum correlation), rather than being clustered as in panel a. **e)** Same as in panel c, but for cells grouped by gRNA target genes.



Supplementary Figure 8

Bulk RNA-seq analysis for the arrayed validation screen

a) Hierarchical clustering of the median expression of TCR activation signature genes, based on bulk RNA-seq data aggregated across gRNAs for the same target gene. Clustering of rows and columns used the Pearson correlation, and z-scores of expression are shown.

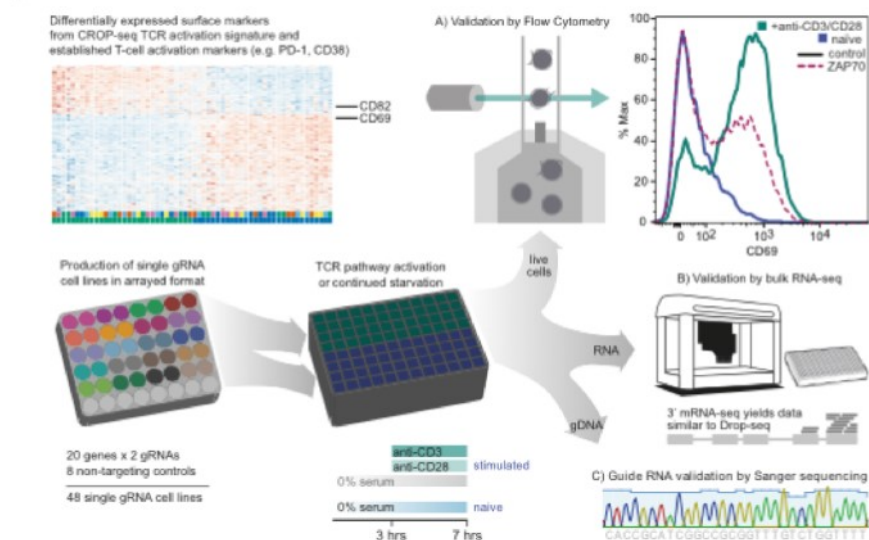
b) Correlation (row-wise z-score) of bulk RNA-seq transcriptomes with a matrix comprising synthetic mixtures of transcriptome profiles between the median of non-targeted cells in both conditions (values close to one reflect similarity with anti-CD3/CD28 stimulated cells). Additional metrics are shown as columns (center) and reveal no relationship with the inferred position of the corresponding samples. The effect that perturbing each target gene had on the TCR activation signature based on the bulk RNA-seq data was assessed in comparison to the control group (barplot on the right).

c) Correlation (top) of median expression levels based on CROP-seq aggregating across target genes (left) or gRNAs (right) compared to the respective bulk RNA-seq libraries across all TCR activation signature genes. The corresponding number of cells in each group is shown at the bottom. Empty values reflect lack of matching bulk RNA-seq libraries due to failed samples in the bulk RNA-seq.

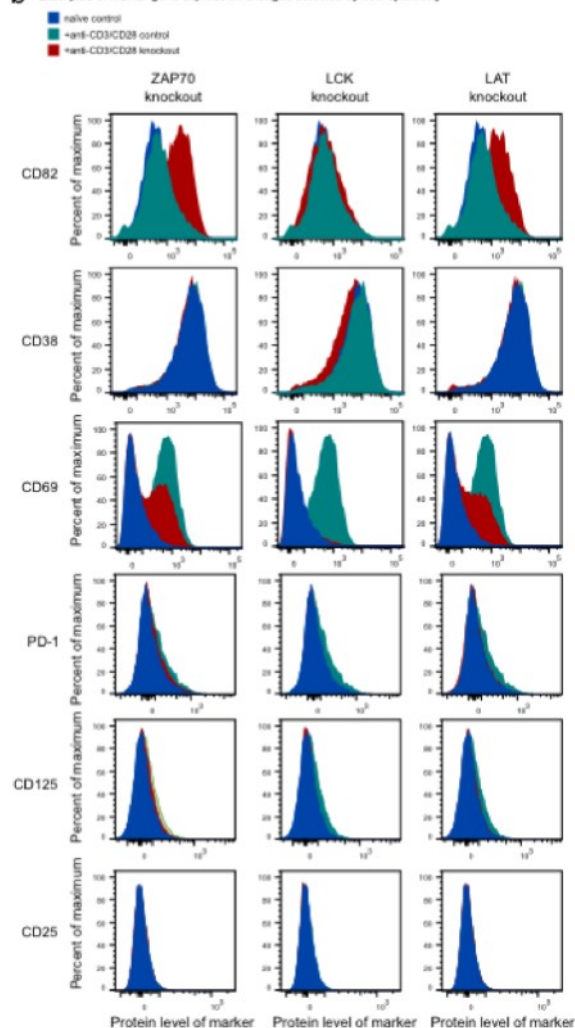
d) Comparison of the signatures inferred from CROP-seq (x-axis) with those derived from bulk RNA-seq data (y-axis) across all shared gRNAs.

e) Comparison of the relative impact of each gRNA (left) or target gene (right) on the TCR activation signature relative to the corresponding control for CROP-seq (x-axis) and bulk RNA-seq (y-axis).

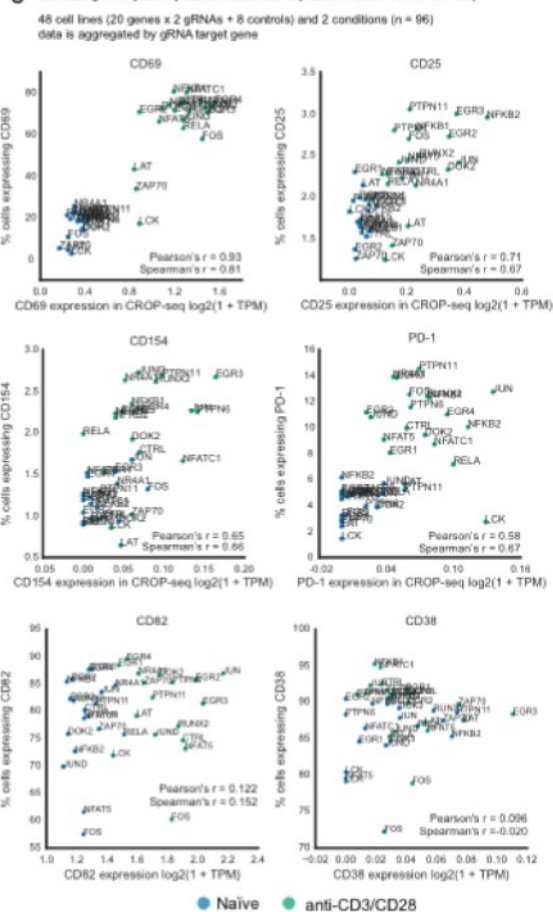
a Technical and biological validation of CROP-seq via arrayed screening with two complementary read-outs



b Examples of marker gene expression changes detected by flow cytometry



c Correlating flow cytometry marker levels with expression values from CROP-seq



Supplementary Figure 9

Flow cytometry analysis for the arrayed validation screen

a) Experimental strategy of the arrayed validation screen. For this screen, 48 CROPseq-Guide-Puro constructs were individually cloned, targeting 20 genes with two gRNAs each and including eight non-targeting controls. Lentivirus production and transductions were performed in 96-well plates, and cells were expanded for 10 days under puromycin and blasticidin selection. Cells were then split into two parts, serum starved for three hours, and subjected to either anti-CD3/CD28-stimulation or continuous starvation for another four hours. The resulting cell populations were validated by Sanger sequencing of the corresponding gRNAs. For validation of the CROP-seq signature, bulk RNA-seq was performed using a 3' enrichment protocol, yielding data similar to Drop-seq (n = 87 RNA-seq libraries). As a complementary single-cell and protein-based read-out, flow cytometry (n = 96 samples) was performed for surface markers enriched in the TCR induction signature derived from CROP-seq (CD69, CD82) or previously reported as markers of T cell activation (CD25, CD38, CD154, PD-1). **b)** Examples of marker expression changes for TCR pathway activators identified by CROP-seq (ZAP70, LCK, LAT). **c)** Scatterplots comparing protein levels for TCR induction markers to RNA expression values obtained by CROP-seq.

2.2. Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing

CRISPR screens at the single-cell level (e.g. by CROP-seq) are a powerful tool to localize the effect of a perturbation in a heterologous mixture of cell types, or during a dynamic process such as cellular differentiation. Due to the efficiency of single-cell profiling, they can be performed on rare populations of primary cells, instead of relying on artificial cell line models. Methods such as CROP-seq can remove the need for an optimized screening readout, and provide versatile whole-transcriptome data that can be used to develop relevant screening signatures without prior knowledge about the biological system. However, due to the high cost of single-cell analyses and next-generation sequencing, we cannot yet realize the full potential of single-cell CRISPR screens.

The following manuscript is the result of our exploration of single-cell combinatorial indexing methods to achieve ultra-high throughput, yet low-cost single-cell RNA sequencing. When comparing published assays in our lab, we found that they require a substantial experimental workload and result in relatively low-quality transcriptomes. We thus developed a new concept named scifi (single-cell combinatorial indexing with fluidic indexing), which combines combinatorial indexing with droplet microfluidics. This allows us to profile even higher cell numbers with a greatly simplified workflow and superior results.

Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing

Paul Datlinger¹, André F. Rendeiro^{1*}, Thorina Boenke^{1*}, Thomas Krausgruber¹, Daniele Barreca¹, Christoph Bock^{1,2}

¹ CeMM Research Center for Molecular Medicine of the Austrian Academy of Sciences, Vienna, Austria

² Department of Laboratory Medicine, Medical University of Vienna, Vienna, Austria

* These authors contributed equally to this work

Correspondence: Christoph Bock (cbock@cemm.oeaw.ac.at)

Keywords: Single-cell RNA-seq, transcriptome profiling, combinatorial indexing, droplet microfluidics, next-generation sequencing, high-throughput biology

Abstract

Cell atlas projects and single-cell CRISPR screens hit the limits of current single-cell RNA-seq technology, as they require the profiling of millions of individual cells. To satisfy these enormous throughput requirements, we developed the “single-cell combinatorial indexing with fluidic indexing” (scifi) method and applied it to single-cell RNA sequencing. This resulted in the scifi-RNA-seq assay, which is both massively scalable and easy-to-use. Our method combines single-step combinatorial indexing of entire transcriptomes with subsequent single-cell RNA-seq using droplet microfluidics. The indexing step allows us to load multiple cells per droplet, which increases the throughput of microfluidic single-cell RNA-seq up to 50-fold. Compared to multi-round combinatorial indexing, scifi-RNA-seq provides a simpler, faster, and more efficient workflow, thereby enabling massive-scale RNA-seq experiments for a broad range of applications.

Introduction

Microfluidic droplet generators are currently the most popular single-cell technology platform. Through their speed, ease of use, and consistent data quality, they have enabled the widespread adoption of single-cell RNA-seq (scRNA-seq)¹⁻³. In a typical droplet-based scRNA-seq experiment, the cell suspension is loaded onto a microfluidic chip, together with uniquely barcoded microbeads, reverse transcription reagents, and carrier oil (**Fig. 1a**). When aqueous and oil phases are combined at controlled flow rates, emulsion droplets co-encapsulate individual cells with individual microbeads. Cells are lysed inside the droplets, and RNA molecules anneal to bead-tethered primers. On each bead, all primers carry the same cell barcode and thus uniquely identify transcripts derived from the same cell. Emulsion droplets are then used as reaction compartments for the reverse transcription reaction. Afterwards, droplets are broken, and the remaining steps of the library preparation are performed in bulk.

In droplet-based scRNA-seq, two cells that enter the same droplet will have their transcriptomes labelled with the exact same cell barcode. As the result, transcripts from these two cells are indistinguishable in the downstream analysis, creating in a cell doublet. To minimize this issue, the cell suspension is loaded into the microfluidic device at a limiting concentration. This “workaround” leads to a major conceptual inefficiency: Most emulsion droplets are fully functional (they contain both barcoded microbeads and reverse transcription reagents), but do not receive a cell and thus do not result in usable transcriptome data (**Fig. 1a**). Thus, microfluidic scRNA-seq uses its microfluidic reaction compartments very inefficiently, while being cost-prohibitive for large-scale studies.

To overcome this conceptual limitation of microfluidic scRNA-seq and to unlock its full potential for massive-scale RNA-seq experiments, we drew inspiration from recently developed methods for combinatorial indexing^{4,5}. We hypothesized that pre-indexing entire single-cell transcriptomes prior to droplet generation could dramatically

increase the efficiency and yield of existing microfluidic droplet generators (**Fig. 1b**). With this setup, a pool of permeabilized cells or nuclei with pre-indexed transcriptomes can be encapsulated at much higher fill rates, aiming for multiple cells per droplet. Groups of cells contained in the same droplet will then share the same microfluidic (round2) barcode, but their single-cell transcriptomes can still be distinguished by the pre-indexing (round1) barcode. Importantly, this concept is superior to cell hashing with DNA-labelled antibodies⁶ or lipids⁷, which encode a cell's identity by a single index sequence but do not barcode the entire transcriptome. Therefore, cell hashing can identify and exclude doublets from the downstream analysis, but cannot resolve their transcriptomes. To distinguish our new concept from conventional combinatorial indexing on multiwell plates, but still pay tribute to the elegant design of the original methods, we named it "single-cell combinatorial indexing with fluidic indexing" (scifi).

Results

Our method is fundamentally based on the assumption that microfluidic droplet generators can tolerate substantial overloading while maintaining a stable droplet emulsion with the desired random loading distribution. We successfully validated this assumption for the popular Chromium platform commercialized by 10x Genomics^{3, 8} (**Fig. 1d-f**), on which we focused our technology development due to its large user base and scalability through parallelization into channels. Beyond this specific platform, the scifi method can be easily adopted to boost the throughput of other microfluidic platforms including Drop-seq¹ and inDrop², and it would also be relatively straightforward to implement for sub-nanoliter well plate assays such as Cyto-seq⁹, Seq-Well¹⁰, and Microwell-seq¹¹.

To test the capacity for overloading of the Chromium system, we prepared a nuclei suspension from human Jurkat cells and loaded it onto the system at different concentrations while omitting the lysis reagents. Without lysis reagents, the nuclei that have been loaded into the emulsion droplets stay intact and can be counted under a light microscope (**Fig. 1d** and **Supplementary Fig. 2a-b**). We first assessed the maximum recommended loading concentration for the Chromium device, which is 15,300 nuclei per microfluidic channel. Based on 609 representative droplet images, we found that only 16.4% of droplets contained one or more nuclei (average of 0.2 nuclei per droplet). Importantly, massive overloading did not clog the microfluidic channels and still resulted in a stable droplet emulsion (**Supplementary Fig. 2c**). As we scaled the loading concentration to 191,000; 383,000; 765,000 and 1,530,000 nuclei per channel, droplet fill rate and the number of nuclei per droplet increased in a controlled fashion (**Fig. 1e-f**). When loading 1.53 million nuclei per channel, which constitutes ~100-fold overloading, we increased the droplet fill rate to 95.5% with an average of 9.6 nuclei per droplet (628 droplet images evaluated).

Loading multiple cells or nuclei into the same droplet is a promising approach only in combination with pre-indexing of the entire transcriptome, which would allow us to assign the RNA reads from a single droplet to the individual cells that were contained within it. To develop a suitable pre-indexing strategy, we modeled the droplet fill rate with a zero-inflated Poisson distribution (**Supplementary Fig. 2d-e**), and we calculated the expected percentage of doublets for defined numbers of round1 barcodes as a function of the droplet fill rate (calculations for 96, 384, or 1,536 barcodes are shown in **Fig. 1g**). This analysis demonstrated that scifi-RNA-seq can distinguish large numbers of single cells with relatively few round1 indices, which has important practical advantages (including easier handling, higher throughput, and lower setup cost). We independently confirmed this result by Monte Carlo simulations of the scifi process (**Supplementary Fig. 2f**). Due to the high complexity of the microfluidic round2 barcode (i.e., 737,000 indices when using off-the-shelf reagents), two-round scifi-RNA-seq already exceeds the barcoding capacity of much more tedious, damaging, three-round combinatorial indexing (**Fig. 1h**).

The method design for scifi-RNA-seq is outlined in **Fig. 1c**, and shown with detailed oligonucleotide sequences in **Supplementary Fig. 1**. Similar to sci-RNA-seq^{4, 12} and SPLIT-seq⁵, permeabilized cells or nuclei are indexed with barcoded oligo-dT primers by reverse transcription on multiwell plates. This step labels the entire transcriptome

with a round1 index specific to each well, and it introduces a unique molecular identifier (UMI) as well as a primer binding site for next-generation sequencing (PBS). To introduce a second defined end, scifi-RNA-seq employs a reverse transcriptase that adds untemplated C bases to the cDNA upon reaching the transcript end. After the reverse transcription reaction, cells or nuclei containing pre-indexed cDNA are pooled, washed, filtered, and counted. Next, cells or nuclei are run on the Chromium system at a high droplet fill rate leading to multiple cells per droplet, along with commercially available gel beads. Oligonucleotides carrying the microfluidic round2 barcode are ligated to the cDNA via the 5'-phosphate group of the reverse transcription primer, directed by a complementary bridge oligo. Efficient ligation is achieved through repeated cycling between denaturation and ligation temperatures with a thermostable ligase. After breaking the droplet emulsion, the 3'-end of the cDNA library is extended by template switching, so that the cDNA can be amplified. Finally, double-stranded cDNA is tagged with a custom i7-only transposome and PCR-enriched, where further sample-specific barcodes can optionally be introduced. **Fig. 1c** shows the read structure for the Illumina NovaSeq 6000 platform, which provides adequate sequencing yield for large scifi-RNA-seq libraries.

In addition to the overloading of microfluidic droplet generators (which we validated above), the second major feasibility concern of our method was whether pre-indexed cells and nuclei would withstand the pressures inside microfluidic droplet generators, after having been exposed to damaging high-temperature incubations and buffers. To demonstrate the feasibility of our method, we processed nuclei derived from a 1:1 mixture of human and mouse cell lines (Jurkat and 3T3, respectively) with the entire scifi-RNA-seq protocol. We loaded 15,300, 383,000 or 765,000 nuclei into single microfluidic channels of the Chromium system. In **Fig. 2a**, we plot all detected, ranked barcodes against the number of sequenced reads per barcode on logarithmic scales, to identify the characteristic inflection point that separates nuclei from noise. The results indicate that scifi-RNA-seq can recover input nuclei with high efficiency, even when overloading the device up to 50-fold.

This proof-of-concept dataset also allowed us to validate our pre-indexing strategy for single-cell transcriptome deconvolution. To that end, we compared human-mouse cell doublets using – versus not using – the pre-indexing information in the form of the round1 barcode. The effect of pre-indexing was striking: We observed that almost all droplets (as identified by round2 barcodes) contained cells of both species, which was expected for overloading at a rate of 383,000 nuclei per channel (**Fig. 2b**, left panel); in contrast, deconvolution based on pre-indexing (considering round1 barcodes) accurately resolved these cell doublets (**Fig. 2b**, right panel). As expected, this was seen only when the droplet generator was overloaded, whereas the microfluidic round2 barcode alone was sufficient for minimizing cell doublets at a standard loading rate of 15,300 nuclei per channel. The composite barcode measures the exact number of individual cells (round1 barcodes) inside each droplet (round2 barcode). Reassuringly, nuclei loading distributions based on sequencing data (**Fig. 2c**) agreed with those derived from microscopic images (**Fig. 1f**).

Having validated the conceptual and practical feasibility of scifi-RNA-seq, we next investigated and optimized the efficiency of our method. To this end, we processed five human cell lines with variable cell size and transcript content (HEK293T, K562, Jurkat, NALM-6, HL-60). During the reverse transcription, each cell line was assigned a specific set of round1 indices. scifi-RNA-seq performed well on four of the cell lines (**Fig. 2d**). Due to low quality transcriptomes, we excluded HL-60 from the further analysis. **Fig. 2e** shows the number of unique molecular identifiers (UMIs) per nucleus as a function of the sequencing coverage, demonstrating that scifi-RNA-seq yields complex single-cell transcriptomes. Our pilot experiment had relatively low read duplication rates, as apparent when plotting the fraction of unique reads as a function of sequencing depth (**Fig. 2f**). To assess whether the amount of reagents loaded into each droplet are indeed sufficient for barcoding multiple cells per droplet, we plotted both quality metrics against the number of nuclei per droplet (**Fig. 2g**). We did not observe any deterioration for overloaded droplets, confirming that reagents were not a limiting factor.

130 Using a principal component analysis on the dataset, we found that most of the variation was explained by the cell
131 type (**Fig. 2h** and **Supplementary Fig. 4a-c**). The round1 indices corresponding to each cell line mapped to defined
132 clusters on the PCA projection, indicating that there is little cross-contamination during the pre-indexing. Through
133 gene set enrichment analysis, we could clearly verify the identity of each cell line (**Supplementary Fig. 4d**), and
134 we identified cell line specific transcriptomic signatures (**Fig. 2i** highlights the top 75 specific genes per cell line).

135

136 Discussion

137 We have developed “single-cell combinatorial indexing with fluidic indexing” (scifi) as a concept for single-cell
138 analyses that provides very high cell throughput while retaining a straightforward, user-friendly workflow that is
139 readily applicable to a broad range of biological questions. We demonstrated this combination of combinatorial
140 pre-indexing with microfluidics indexing for single-cell RNA-seq, giving rise to the scifi-RNA-seq assay, for
141 which we developed and optimized a detailed protocol (provided here as a **Supplementary Protocol**).

142 The presented scifi-RNA-seq assay combines the scalability of combinatorial indexing protocols with the simplic-
143 ity and efficiency of droplet generators. While this initial demonstration of scifi-RNA-seq focused on the widely
144 available 10x Genomics Chromium platform, scifi-RNA-seq can in principle be implemented on any microfluidic
145 system as well as for high-throughput single-cell RNA-seq assays that use sub-nanoliter well plates.

146 Our scifi concept is not limited to scRNA-seq, but can be used to boost the throughput of any single-cell assay
147 compatible with combinatorial indexing, such as single-cell ATAC-seq¹³, single-cell whole genome sequencing¹⁴,
148 DNA-methylation analysis¹⁵, and chromatin conformation capture by Hi-C¹⁶. Due to the high complexity of the
149 microfluidic round2 barcode, 2-level *scifi* indexing outperforms the more complex 3-level combinatorial indexing.

150 We expect that *scifi-RNA-seq* will be enabling for large-scale single-cell atlas projects¹⁷. For example, the large
151 number of cells profiled can enhance the detection of rare cell types and helps resolve developmental trajectories
152 at high resolution¹⁸. Furthermore, the round1 indexing provides an effective means for sample multiplexing. Com-
153 bined with efficient PCR-based guide RNA enrichment, scifi-RNA-seq could also become the method of choice
154 for single-cell CRISPR screens (e.g., using CROP-seq¹⁹, Perturb-seq^{20, 21}, or CRISP-seq²²), where very high cell
155 numbers are needed for large-scale genetic screens with transcriptome readout.

156

157 Online Methods

158 A step-by-step experimental protocol for scifi-RNA-seq is provided as a **Supplementary Protocol**. Future updates
159 of the protocol will be shared via a Supplementary Website (<http://scifi-rna-seq.computational-epigenetics.org>),
160 which will become available online upon journal publication.

161

162 *Assembly of the custom i7-only transposome*

163 Oligonucleotides Tn5-top_ME and Tn5-bottom_Read2N were synthesized by Sigma Aldrich (sequences are pro-
164 vided in **Supplementary Table 1**) and reconstituted in EB buffer (Qiagen cat. no. 19086) at 100 μ M. We mixed
165 22.5 μ l of each oligonucleotide and 5 μ l of 10x Oligonucleotide Annealing Buffer (10 mM Tris-HCl (Sigma cat.
166 no. T2944-100ML), 50 mM NaCl (Sigma cat. no. S5150-1L), 1 mM EDTA (Invitrogen cat. no. AM9260G)) and
167 annealed them in a thermocycler: 95 °C for 3 min, 70 °C for 3 min, ramp to 25 °C at 2 °C per minute. The annealing
168 reaction was then diluted by addition of 180 μ l of water. At this point, the diluted oligonucleotide cassette can be
169 aliquoted and frozen for future transposome assemblies. To load the Tn5 transposase, we mixed 20 μ l of diluted
170 oligonucleotide cassette from the previous step with 20 μ l of 100% glycerol (Sigma cat. no. G5516-100ML) and

171 10 µl of EZ-Tn5 Transposase (Lucigen cat. no. TNP92110), and incubated for 30 min at 25 °C in a thermocycler.
172 The resulting 50 µl of assembled transposome can be stored at -20 °C for at least one month.

173

174 *Activity check for the custom i7-only transposome by qPCR*

175 Tagmented DNA flanked by two Illumina i7 adapters is suppressed in PCR reactions due to competition between
176 intramolecular annealing and primer binding. Our custom i7-only transposome is therefore tested in a negative
177 qPCR assay as described previously²³. Briefly, a defined PCR product is subjected to one tagmentation reaction
178 and one no-enzyme control reaction. Both samples are then re-amplified with the same primers in a qPCR reaction.
179 Since the tagmentation fragments the PCR product, the corresponding reaction should yield higher Ct values. The
180 tagmentation efficiency can then be calculated from the shift of Ct values:

181 $\text{Tagmentation efficiency [\%]} = 100 / [2^{(\text{average Ct tagmentation} - \text{average Ct no-enzyme control})}]$

182 *Generation of the PCR product:* Oligonucleotides pUC19-FWD and pUC19-REV were synthesized by Sigma Al-
183 drich (sequences are provided in **Supplementary Table 1**) and reconstituted in EB buffer (Qiagen cat. no. 19086)
184 at 100 µM. Next, a 1,961 bp PCR product was generated by mixing 128.7 µl of water, 33 µl of 50 pg/µl pUC19
185 plasmid (NEB cat. no. N3041S), 1.65 µl each of primers pUC19-FWD and pUC19-REV (100 µM) combined with
186 165 µl of 2x Q5 HotStart High-Fidelity Master Mix (NEB cat. no. M0494L). The resulting 6.6x master mix was
187 distributed into a tube strip (six reactions of 50 µl) and amplified in a thermocycler: 98 °C for 30 s; 31x (98 °C for
188 10 s, 68 °C for 30 s, 72 °C for 1 min), 72 °C for 2 min, storage at 12 °C. To each 50 µl PCR reaction, we added
189 6.25 µl of 10x CutSmart Buffer and 6.25 µl of DpnI (NEB cat. no. R0176L) and incubated at 37 °C for 1 hour to
190 digest the PCR template plasmid. The six PCR reactions were pooled and cleaned with the QiaQuick PCR Purifi-
191 cation Kit (Qiagen cat. no. 28106) using two columns and eluting with 30 µl of EB buffer per column. Eluates
192 were pooled, and the purity of the PCR fragment was checked on a 1% agarose gel containing ethidium bromide.
193 We then measured the concentration of dsDNA with a Qubit HS assay (Thermo Fisher Scientific cat. no. Q32854),
194 and diluted the PCR product to 25 ng/µl with EB buffer.

195 *Tagmentation:* Tagmentation reactions were set up by mixing 2 µl of 25 ng/µl pUC19 PCR product from the
196 previous step, 7 µl of ATAC Buffer (10x Genomics cat. no. 2000122), and either 6 µl of custom i7-only transpo-
197 some (tagmentation reaction) or 6 µl of water (no-enzyme control reaction). After 60 min of incubation at 37 °C,
198 the Tn5 enzyme was stripped from the DNA by addition of 1.75 µl of 1% SDS solution (Sigma cat. no. 71736-
199 100ML) followed by incubation at 70 °C for 10 min. The two reactions were diluted 1/100 with EB buffer, and
200 qPCR reactions were set up in triplicates: 2 µl of 1/100-diluted reaction, 10 µl of 2x GoTaq qPCR Master Mix
201 (Promega cat. no. A600A), 0.1 µl each of 100 µM pUC19-FWD and pUC19-REV primers and 7.8 µl of water.
202 qPCR reactions were incubated as follows: 95 °C for 2 min, 40x (95 °C for 30 s, 68 °C for 30 s, 72 °C for 2 min
203 and plate read).

204

205 *Preparation of permeabilized whole cells from human and mouse cell lines*

206 5 million cells were washed with 10 ml of ice-cold 1x PBS (Gibco cat. no. 14190-094, centrifugation: 300 rcf, 5
207 min, 4 °C) and fixed in 5 ml of ice-cold methanol (Fisher Scientific cat. no. M/4000/17) at -20 °C for 10 min. After
208 two additional washes (centrifugation: 300 rcf, 5 min, 4 °C) with 5 ml of ice-cold PBS-BSA-SUPERase (1x PBS
209 supplemented with 1% w/v BSA (Sigma cat. no. A8806-5) and 1% v/v SUPERase-In RNase Inhibitor (Thermo
210 Fisher Scientific cat. no. AM2696)) permeabilized cells were resuspended in 200 µl of ice-cold PBS-BSA-
211 SUPERase, and filtered through a cell strainer (40 µM or 70 µM depending on the cell size). We then used 5 µl of

212 the sample for cell counting in duplicates on a CASY device (Schärfe System), and diluted to 5,000 cells per μl
213 with ice-cold PBS-BSA-SUPERase. We immediately proceeded with the reverse transcription step.

214

215 *Preparation of fresh nuclei from human and mouse cell lines*

216 5 million cells were washed with 10 ml of ice-cold 1x PBS (Gibco cat. no. 14190-094, 300 rcf, 5 min, 4 °C). Nuclei
217 were prepared by resuspending cells in 500 μl of ice-cold Nuclei Preparation Buffer (10 mM Tris-HCl pH 7.5
218 (Sigma cat. no. T2944-100ML), 10 mM NaCl (Sigma cat. no. S5150-1L), 3 mM MgCl₂ (Ambion cat. no.
219 AM9530G), 1% w/v BSA (Sigma cat. no. A8806-5), 1% v/v SUPERase-In RNase Inhibitor (Thermo Fisher Sci-
220 entific cat. no. AM2696), 0.1% v/v Tween-20 (Sigma cat. no. P7949-500ML), 0.1% v/v IGEPAL CA-630 (Sigma
221 cat. no. I8896-50ML), 0.01% v/v Digitonin (Promega cat. no. G944A)), followed by 5 min of incubation on ice.
222 Lysis of the plasma membrane was stopped by adding 5 ml of ice-cold Nuclei Wash Buffer (10 mM Tris-HCl pH
223 7.5, 10 mM NaCl, 3 mM MgCl₂, 1% w/v BSA, 1% v/v SUPERase-In RNase Inhibitor, 0.1% v/v Tween-20).
224 Nuclei were collected by centrifugation (500 rcf, 5 min, 4 °C), resuspended in 200 μl of ice-cold PBS-BSA-
225 SUPERase (1xPBS supplemented with 1% w/v BSA and 1% v/v SUPERase-In RNase Inhibitor (20 U/ μl , Thermo
226 Fisher Scientific cat. no. AM2696)) and filtered through a cell strainer (40 μm or 70 μm depending on the cell
227 size). We then used 5 μl of the sample for cell counting in duplicates on a CASY device (Schärfe System), and
228 diluted to 5,000 cells per μl with ice-cold PBS-BSA-SUPERase. We immediately proceeded with the reverse tran-
229 scription step.

230

231 *Preparation of nuclei from primary cells with formaldehyde fixation and permeabilization*

232 5 million primary cells were washed with 10 ml of ice-cold 1x PBS (Gibco cat. no. 14190-094, centrifugation: 300
233 rcf, 5 min, 4 °C). Nuclei were prepared by resuspending cells in 500 μl of ice-cold Nuclei Preparation Buffer
234 without Digitonin and without Tween-20 (10 mM Tris-HCl pH 7.5 (Sigma cat. no. T2944-100ML), 10 mM NaCl
235 (Sigma cat. no. S5150-1L), 3 mM MgCl₂ (Ambion cat. no. AM9530G), 1% w/v BSA (Sigma cat. no. A8806-5),
236 1% v/v SUPERase-In RNase Inhibitor (Thermo Fisher Scientific cat. no. AM2696), 0.1% v/v IGEPAL CA-630
237 (Sigma cat. no. I8896-50ML)), followed by 5 min of incubation on ice. Lysis of the plasma membrane was stopped
238 by addition of 5 ml of Nuclei Wash Buffer without Tween-20 (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 3 mM
239 MgCl₂, 1% w/v BSA, 1% v/v SUPERase-In RNase Inhibitor). Nuclei were collected by centrifugation (500 rcf, 5
240 min, 4 °C), and fixed in 5 ml of ice-cold 1x PBS containing 4% Formaldehyde (Thermo Fisher Scientific cat. no.
241 28908) for 15 min on ice. Fixed nuclei were collected (500 rcf, 5 min, 4 °C), the pellet was resuspended in 1.5 ml
242 of ice-cold Nuclei Wash Buffer without Tween-20 and transferred to a 1.5 ml tube. After one more wash with 1.5
243 ml of ice-cold Nuclei Wash Buffer without Tween-20 (500 rcf, 5 min, 4 °C), fixed nuclei were resuspended in 200
244 μl of Nuclei Wash Buffer without Tween-20, snap-frozen in liquid nitrogen and stored at -80 °C.

245 For processing with scifi-RNA-seq, frozen samples were thawed in a 37 °C water bath for exactly 1 min, and
246 immediately placed on ice. Following centrifugation (500 rcf, 5 min, 4 °C), fixed nuclei were resuspended in 250
247 μl of ice-cold Permeabilization Buffer (10 mM Tris-HCl, 10 mM NaCl, 3 mM MgCl₂, 1% w/v BSA, 1% v/v
248 SUPERase-In RNase Inhibitor, 0.01% v/v Digitonin (Promega cat. no. G944A), 0.1% v/v Tween-20 (Sigma cat.
249 no P7949-500ML)). After 5 min of incubation in ice, 250 μl of Nuclei Wash Buffer without Tween-20 were added
250 per sample, and nuclei were collected (500 rcf, 5 min, 4 °C). After one more wash with 250 μl of Nuclei Wash
251 Buffer without Tween-20, nuclei were taken up in 100 μl of 1x PBS containing 1% w/v BSA and 1% v/v
252 SUPERase-In RNase Inhibitor. We used 5 μl for cell counting in duplicates on a CASY device (Schärfe Systems),
253 and diluted to 5,000 cells per μl with PBS-BSA-SUPERase. We immediately proceeded with the reverse transcrip-
254 tion step.

255

256 *Testing the nuclei loading capacity of the Chromium system*

257 Human Jurkat cells (clone E6-1) were cultured in RPMI medium (Gibco cat. no. 21875-034) supplemented with
258 10% FCS (Sigma) and penicillin-streptomycin (Gibco cat. no. 15140122). Fresh nuclei were isolated as described
259 above. Next, we prepared samples of 15.3k, 191k, 383k, 765k and 1.53M nuclei, added 1.5 µl of Reducing Agent
260 B (10x Genomics cat. no. 2000087) and 1x Nuclei Buffer (10x Genomics cat. no. 2000153) to a total volume of
261 80 µl. This buffer does not contain detergents, hence the nuclei remain intact during the microfluidic run and can
262 be visualized inside the emulsion droplets with a standard light microscope. At the same time, Reducing Agent B
263 dissolves the Gel Beads, which might otherwise obstruct the view. The microfluidic chip (Single Cell E Chip, 10x
264 Genomics 2000121) was loaded as follows: 75 µl of nuclei sample at the indicated loading concentrations into inlet
265 1, 40 µl of Single Cell ATAC Gel Beads (10x Genomics cat. no. 2000132) into inlet 2, and 240 µl of Partitioning
266 Oil (10x Genomics cat. no. 220088) into inlet 3. To image the resulting droplets, 15 µl of Partitioning Oil were
267 pipetted onto a glass slide, followed by 5 µl of emulsion droplets, and images were taken at 10x magnification.
268 Nuclei were counted for an average of 653 droplets.

269

270 *Measuring the bead fill rate of the Chromium system*

271 To measure the bead fill rate, we loaded the Single Cell E Chip (10x Genomics 2000121) with 80 µl of 1x Nuclei
272 Buffer (10x Genomics cat. no. 2000153) into inlet 1, 40 µl of Single Cell ATAC Gel Beads (10x Genomics cat.
273 no. 2000132) into inlet 2, and 240 µl of Partitioning Oil (10x Genomics cat. no. 220088) into inlet 3. By omitting
274 Reducing Agent B, we ensured that Gel Beads remain intact throughout the microfluidic run, such that they can be
275 visualized inside the emulsion droplets using a standard light microscope. Our fill rate calculations are based on a
276 total of 1,265 droplets.

277

278 *Detailed scifi-RNA-seq protocol description*

279 *Reverse Transcription:* 384 indexed reverse transcription primers were synthesized by Sigma Aldrich and shipped
280 at 100 µM in EB Buffer in 96-well plates (sequences are provided in **Supplementary Table 1**). 96-well plates with
281 barcoded oligo-dT primers were prepared prior to the experiment and stored at -20 °C (1 µl of 25 µM per well).
282 10,000 permeabilized cells or nuclei (2 µl of a 5,000/µl suspension) were added to the pre-dispensed primers and
283 well assignments were recorded. The plate was incubated for 5 min at 55 °C (to resolve RNA secondary structures),
284 then placed immediately on ice (to prevent their re-formation). Per well, a mix of 3 µl nuclease-free water, 2 µl 5x
285 Reverse Transcription Buffer, 0.5 µl of 100 mM DTT (Sigma cat. no. 646563-10x.5ML), 0.5 µl of 10 mM dNTPs
286 (Invitrogen cat. no. 18427-088), 0.5 µl of RNaseOUT RNase inhibitor (40 U/ml, Invitrogen cat. no. 10777019),
287 and 0.5 µl of Maxima H Minus Reverse Transcriptase (200 U/ml, Thermo Fisher Scientific cat. no. EP0753) were
288 added. The reverse transcription was incubated as follows: (heated lid set to 60 °C), 50 °C for 10 min, 3 Cycles of
289 {8 °C for 12 sec, 15 °C for 45 sec, 20 °C for 45 sec, 30 °C for 30 sec, 42 °C for 2 min, 50 °C for 3 min}, 50 °C for
290 5 min, store at 4 °C.

291 *Cell/Nuclei recovery and pooling:* Processed cells/nuclei were recovered from the plates and pooled in one 15 ml
292 tube per plate on ice. Wells were washed with ice-cold 1xPBS-1%BSA, which was transferred to the same tube
293 for maximum recovery. The volume was topped up to 15 ml with 1xPBS-1%BSA, and nuclei were collected (500
294 rcf, 5 min, 4 °C). The resulting pellet was resuspended in 1.0 ml of 1x Ampligase Reaction Buffer (Lucigen
295 #A0102K), filtered through a cell strainer (40 µm or 70 µm depending on the cell/nuclei size) into a 1.5 ml tube

296 and centrifuged (500 rcf, 5 min, 4 °C). The supernatant was removed almost completely, and the tube was centri-
297 futed briefly (500 rcf, 30 s, 4 °C) to collect the remaining liquid at the bottom of the tube. Typically, this resulted
298 in ~10 µl of a highly concentrated suspension, which was diluted 1:200 with 1x Ampligase Buffer and counted in
299 a Fuchs Rosenthal counting chamber (Incyto cat. no. DHC-F01). The desired number of cells/nuclei was brought
300 to a volume of 15 µl with 1x Ampligase Reaction Buffer (Lucigen #A0102K).

301 *Microfluidic thermoligation barcoding:* Unused channels in the Chromium Chip E (10x Genomics cat. no.
302 2000121) were filled with 75 µl (inlet 1), 40 µl (inlet 2) or 240 µl (inlet 3) of 50% glycerol solution (Sigma cat.
303 no. G5516-100ML). Right before loading the chip, a mix of 47.4 µl nuclease-free water, 11.5 µl of 10x Ampligase
304 Reaction Buffer (Lucigen #A0102K), 2.3 µl of 100 U/µl Ampligase (Lucigen #A0102K), 1.5 µl of Reducing Agent
305 B (10x Genomics cat. no. 2000087) and 2.3 µl of Bridge Oligo (100 µM) was added per sample. The microfluidic
306 chip was loaded with 75 µl of cells/nuclei in thermoligation mix (inlet 1), 40 µl of Single Cell ATAC Gel Beads
307 (inlet 2, 10x Genomics cat. no. 2000132) and 240 µl of Partitioning Oil (inlet 3, 10x Genomics cat. no. 220088)
308 and run on the 10x Genomics Chromium controller. For thermoligation barcoding the droplet emulsion was incu-
309 bated as follows: (heated lid set to 105 °C, volume set to 100 µl), 12x (98 °C for 30 s, 59 °C for 2 min), storage at
310 15 °C. The emulsion was broken by addition of 125 µl Recovery Agent (10x Genomics cat. no. 220016) and 125
311 µl of the pink oil phase were removed by pipetting. The remaining sample was mixed with 200 µl of Dynabead
312 Cleanup Master Mix (per reaction: 182 µl Cleanup Buffer (10x Genomics cat. no. 2000088), 8 µl Dynabeads
313 MyOne Silane (Thermo Fisher Scientific cat. no. 37002D), 5 µl Reducing Agent B (10x Genomics cat. no.
314 2000087), 5 µl of nuclease-free water). After 10 min of incubation at room temperature, samples were washed
315 twice with 200 µl of freshly prepared 80% ethanol (Merck cat. no. 603-002-00-5) and eluted in 40.5 µl of EB
316 Buffer (Qiagen cat. no. 19086) containing 0.1% Tween (Sigma cat. no. P7949-500ML) and 1% v/v Reducing
317 Agent B. Bead clumps were sheared with a 10 µl pipette or needle. 40 µl of the sample were transferred to a fresh
318 tube strip and subjected to a 1.0x cleanup with SPRIselect beads (Beckman Coulter cat. no. B23318), eluting in 22
319 µl of EB Buffer.

320 *Template switching:* 20 µl of sample from the previous step were mixed with 10 µl of 5x Reverse Transcription
321 Buffer, 10 µl of Ficoll PM-400 (20%, Sigma cat. no. F5415-50ML), 5 µl of 10 mM dNTPs (Invitrogen cat. no.
322 18427-088), 1.25 µl of Recombinant Ribonuclease Inhibitor (Takara cat. no. 2313A), 1.25 µl of Template Switch-
323 ing Oligo (100 µM) and 2.5 µl of Maxima H Minus Reverse Transcriptase (200 U/ml, Thermo Fisher Scientific
324 cat. no. EP0753). The template switching reaction was incubated for 30 min at 25 °C, 90 min at 42 °C, storage at
325 4 °C and cleaned with a 1.0x SPRI cleanup, eluting in 17 µl of EB buffer.

326 *cDNA enrichment:* 15 µl of the above sample were mixed with 33 µl of nuclease-free water, 50 µl of NEBNext
327 High Fidelity 2x Master Mix (NEB cat. no. M0541S), 0.5 µl of 100 µM Partial P5 primer, 0.5 µl of 100 µM TSO
328 Enrichment Primer and 1 µl of 100x SYBR Green in DMSO (Life Technologies cat. no. S7563). cDNA was am-
329 plified in a thermocycler: 98 °C for 30 sec, Cycle until fluorescent signal >2000 RFU {98 °C for 20 sec, 65 °C for
330 30 sec, 72 °C for 3 min}, 72 °C for 5 min in another thermocycler, storage at 4 °C. cDNA was cleaned by one 0.8x
331 SPRI cleanup followed by a 0.6x SPRI cleanup, quantified with a Qubit HS assay (ThermoFisher Scientific cat.
332 no. Q32854) and 1.5 ng were checked on a Bioanalyzer High-Sensitivity DNA chip (Agilent cat. no. 5067-4626
333 and 5067-4627).

334 *Tagmentation:* To achieve the maximum library complexity, the entire sample was processed in multiple tagmen-
335 tation reactions with 1 ng input. cDNA was diluted to 0.2 ng/µl with nuclease-free water and 5 µl (1 ng) per reaction
336 were distributed into a 96-well plate on ice. A mix of 11.25 µl nuclease-free water, 5 µl of 5x Tn5 Reaction Buffer
337 (50 mM TAPS (Sigma cat. no. T9659-100G), 25 mM MgCl₂ (Ambion cat. no. AM9530G), pH adjusted to 8.5,
338 sterile-filtrated), 2.5 µl of Dimethylformamide (Sigma cat. no. D4551-250ML) and 1.25 µl of freshly diluted i7-
339 only transposome (prepared as described above and diluted 1:4.5 in Tn5 Dilution Buffer (50 mM Tris-HCl pH 7.5
340 (Sigma cat. no. T2944-100ML), 100 mM NaCl (Sigma cat. no. S5150-1L), 0.1 mM EDTA (Invitrogen cat. no.

AM9260G), 50% glycerol (Sigma cat. no. G5516-100ML), 0.1% Triton-X100 (Sigma cat. no. X100-100ML), 1 mM freshly added DTT (Sigma cat. no. 646563-10x.5ML)) was added. Reactions were incubated for 10 min at 55 °C, then cooled for 1 min on ice. The enzyme was inactivated by addition of 2.5 µl 10x Stop Solution (included with Lucigen cat. no. TNP92110) for 5 min at room temperature. Next, the volume was brought to 50 µl and the fragmented cDNA was purified with a 1.0x SPRI cleanup, eluting in 17 µl of EB buffer (Qiagen cat. no. 1014609).

Library enrichment: 15 µl of tagged cDNA were mixed with 5 µl of 10 µM barcoded P7 primer. Per reaction, a mix of 28.5 µl nuclease-free water, 50 µl NEBNext High Fidelity 2x Master Mix (NEB cat. no. M0541S), 1 µl of 100x SYBR Green in DMSO (Life Technologies cat. no. S7563) and 0.5 µl of 100 µM Partial-P5 primer were added. Reactions were incubated in a qPCR cycler: 72 °C for 3 min (for end fill-in after tagmentation), 98 °C for 30 sec, Cycle {98 °C for 10 sec, 65 °C for 30 sec, 72 °C for 30 sec, plate read}. During the run, we monitored the fluorescence signal and removed samples from the thermocycler when they reached >4000 RFU. To complete unfinished PCR products, we incubated for 5 min at 72 °C in another thermocycler. Libraries were cleaned with a 0.7x SPRI cleanup with AMPure XP beads. All wells with the same P7 barcode were pooled, and subjected to a 0.8x SPRI cleanup, eluting in 0.2 bead volumes of EB buffer (Qiagen cat. no. 1014609). The final library concentration was measured with the Qubit High Sensitivity DNA assay (Thermo Fisher Scientific cat. no. Q32854), and 1.5 ng were run on a Bioanalyzer High Sensitivity DNA chip (Agilent cat. no. 5067-4626).

Next-generation sequencing: Libraries were diluted to 2.0 nM with EB buffer (Qiagen cat. no. 1014609) containing 0.1% Tween-20 (Sigma cat. no. P7949-500ML), and sequenced on the Illumina NovaSeq 6000 platform with standard sequencing primers and a read structure of 21 bases (Read 1), 8 bases (Index 1, i7), 16 bases (Index 2, i5), 78 bases (Read 2). Depending on the expected number of single-cell transcriptomes, NovaSeq 6000 SP (Illumina cat. no. 20027464), S1 (Illumina cat. no. 20012865) or S2 (Illumina cat. no. 20012862) reagent kits were used.

Theoretical expectation of collisions for scifi-RNA-seq experiments

To obtain an expectation of cell doublets per compartment (droplet and/or well), we used the birthday problem as described by others^{4, 13}, where the probability of collision of n cells in c compartments is defined as

$$p(n, c) \approx 1 - \left(\frac{c-1}{c}\right)^{\frac{n(n-1)}{2}}$$

and the amount of cells per compartment c at a collision rate p is given by the reverse problem

$$n(p, c) \approx \sqrt{2c \cdot \ln\left(\frac{1}{1-p}\right)}.$$

In this framework, combining pre-indexing with a microfluidic droplet generator is equivalent to the expectation of placing cells in a number of compartments that is the product of the number of pre-indexing compartments and the number of droplets with beads in the microfluidic device. These estimates provide a useful upper bound, but are an overestimate due to high-order collisions expected for high numbers of cells or nuclei.

Monte Carlo experiments on microfluidic device loading

As a complementary approach that makes no assumptions about specific distributions, we performed the following Monte Carlo experiments: We generated two vectors of size n for which we randomly sampled integers from the range of combinatorial indexing compartments $r1$ and microfluidics barcodes $r2$. We then calculated the fraction of non-one occurrences of the same $r1$ integer in the $r2$ compartments. This was repeated for a range of input cell numbers (n), $r1$ and $r2$ compartments and 100 iterations (i).

380

381 *Modeling of microfluidic device loading*

382 Both the naive combinatorial expectations and Monte Carlo simulations lack the consideration of biological dou-
383 blets (i.e., cell clumps) and should be seen as an optimal upper bound. To gain more accurate insights into the
384 dynamics of loading the Chromium system, we loaded nuclei without lysis reagents, so that they could be visual-
385 ized inside the emulsion droplets under a light microscope (as described above).

386 Since we observed a fraction of droplets containing no cells even at high cell loading concentrations, we specified
387 a Bayesian model where the observed number of cells per droplet is explained by a zero-inflated Poisson distribu-
388 tion of latent parameters λ for mean and variance and ψ for the zero-inflated component. We imposed a prior on λ
389 as the observed mean cells per droplet and estimated a posterior distribution of parameters by Markov Chain Monte
390 Carlo sampling with the NUTS sampler for 5,000 draws after 10,000 tuning draws using PyMC3²⁴.

391 To estimate doublet rates for scifi-RNA-seq experiments with various rounds of pre-indexing, we made the as-
392 sumption that the doublet rate of cells is independent of the pre-indexing. This has the practical advantage that the
393 number of input cells for a scifi-RNA-seq experiment regarding doublet rates can be seen as the number of real
394 input cells divided by the number of unique pre-indexing barcodes. We therefore interpolated the estimated model
395 parameters across the range of cell loading concentrations. Using those parameters for the practical number of
396 input cells calculated the 1 - probability mass function (PMF) of the zero-inflated Poisson distribution for more
397 than one cell per droplet as the doublet rate.

398 While considering observational data from the microfluidic device, these predictions are still optimistic as they are
399 based on cells as the countable unit, whereas in the real scifi-RNA-seq experiment the countable units are transcript
400 molecules and cells are deduced as a proxy.

401

402 *Processing of scifi-RNA-seq data*

403 For fast parallel processing of raw sequence data, we demultiplexed the raw sequence calls based on the sample
404 and round1 barcode, allowing up to 3 mismatches; and demultiplexed files were written in unaligned single-read
405 BAM format, with dedicated tags for the round2 barcode (“r2 tag”) and UMI (“RX”).

406 The STAR aligner (version 2.7.0e) was used to map the demultiplexed unaligned BAM files with the following
407 parameters: “--readFilesCommand samtools view -h --readFileType SAM SE”. In addition, we allowed trimming
408 of poly-A stretches by passing the “--clip3pAdapterSeq AAAAAA” parameters and obtained all alignments via the
409 following paramters: “--outFilterScoreMinOverLread 0 --outFilterMatchNminOverLread 0 --outFilterMatchNmin
410 0”. We set the output of STAR to aligned BAM using the “--outSAMtype BAM Unsorted” parameters.

411 To quantify gene expression, we use the featureCounts tool from the Subread package (version v1.6.2) to add an
412 “XT” tag to the aligned BAM file with the respective gene using the “-g gene_id” parameter. Only alignments with
413 quality above 30 were tagged (“-q 30” parameter). We quantify alignments in an unstranded manner (“-s 0” pa-
414 rameter) for either only exons “-t exon” or whole gene body “-t gene” since nuclei can contain high levels of
415 unspliced transcripts. We then used Pysam (version 0.15.2) to filter out alignments from low-quality or unmapped
416 reads, multi-mappers, secondary or supplementary alignments. Effectively, we kept only reads which aligned
417 uniquely to genes with quality higher than 30.

418 To generate a sparse matrix of gene expression, we considered a unique molecule count as a sequenced molecule
419 with unique read name, round2 barcode, UMI, associated gene and mapping position in the chromosome. We then
420 created a Scipy (version 1.3.0) compressed sparse row matrix of the form (UMI count, cell barcode, gene) and
421 initialized an Anndata (version 0.6.21) object to save it as h5ad format. We used Python 3.7.2 for all analysis.

422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459

Analysis of scifi-RNA-seq data

We used Scanpy (version 1.4.3) for further analysis of the resulting single-cell RNA-seq data. The Anndata object was filtered for cell barcodes containing between 100 and 80,000 UMI counts and genes with at least 200 UMI counts. To perform dimensionality reduction, we selected variable genes with the *seurat* flavor. For the *scanpy.pp.highly_variable_genes* function, *normalized_dispersion* was set to a minimum value of 1, and *mean_expression* to a maximum of 1 on a log scale. Gene expression values were scaled, centered and principal component analysis (PCA) was performed using highly variable genes as input and the *arpack* solver. To discover genes associated with each cell line, the normalized and log transformed expression values were used as input for the *scanpy.tl.rank_genes_groups* function using the *t-test_overstim_var* test for the top 2000 genes per group. For these differential gene sets between cell lines, we performed gene set enrichment analysis using the Enrichr API for the *ARCHS4_Cell-lines* database.

Data availability

All data will be available through the Supplementary Website (<http://scifi-rna-seq.computational-epigenetics.org>), which will become available online upon journal publication. Single-cell RNA-seq data will also be openly available from the NCBI GEO database.

Code availability

The analysis source code underlying the final version of the paper will be provided on the Supplementary Website (<http://scifi-rna-seq.computational-epigenetics.org>) and a Github repository (<https://github.com/epigen/scifiRNA-seq>).

Acknowledgements

We also thank the Biomedical Sequencing Facility at CeMM for assistance with next generation sequencing and all members of the Bock lab for their help and advice. C.B. is supported by a New Frontiers Group award of the Austrian Academy of Sciences and by an ERC Starting Grant (European Union’s Horizon 2020 research and innovation programme, grant agreement n° 679146).

Author contributions

P.D. conceived the project and developed the method; P.D. and T.B. optimized the protocol and conducted the experiments; T.B. performed next-generation sequencing; A.F.R performed computational modeling and analyzed the data; P.D. and A.F.R visualized the results and made the figures; T.K. performed flow cytometry; D.B. performed data pre-processing; P.D., A.F.R., and C.B. wrote the manuscript with contributions from all authors; C.B. supervised the project.

Competing financial interests

P.D. and C.B. are inventors on a patent application relating to scifi-RNA-seq.

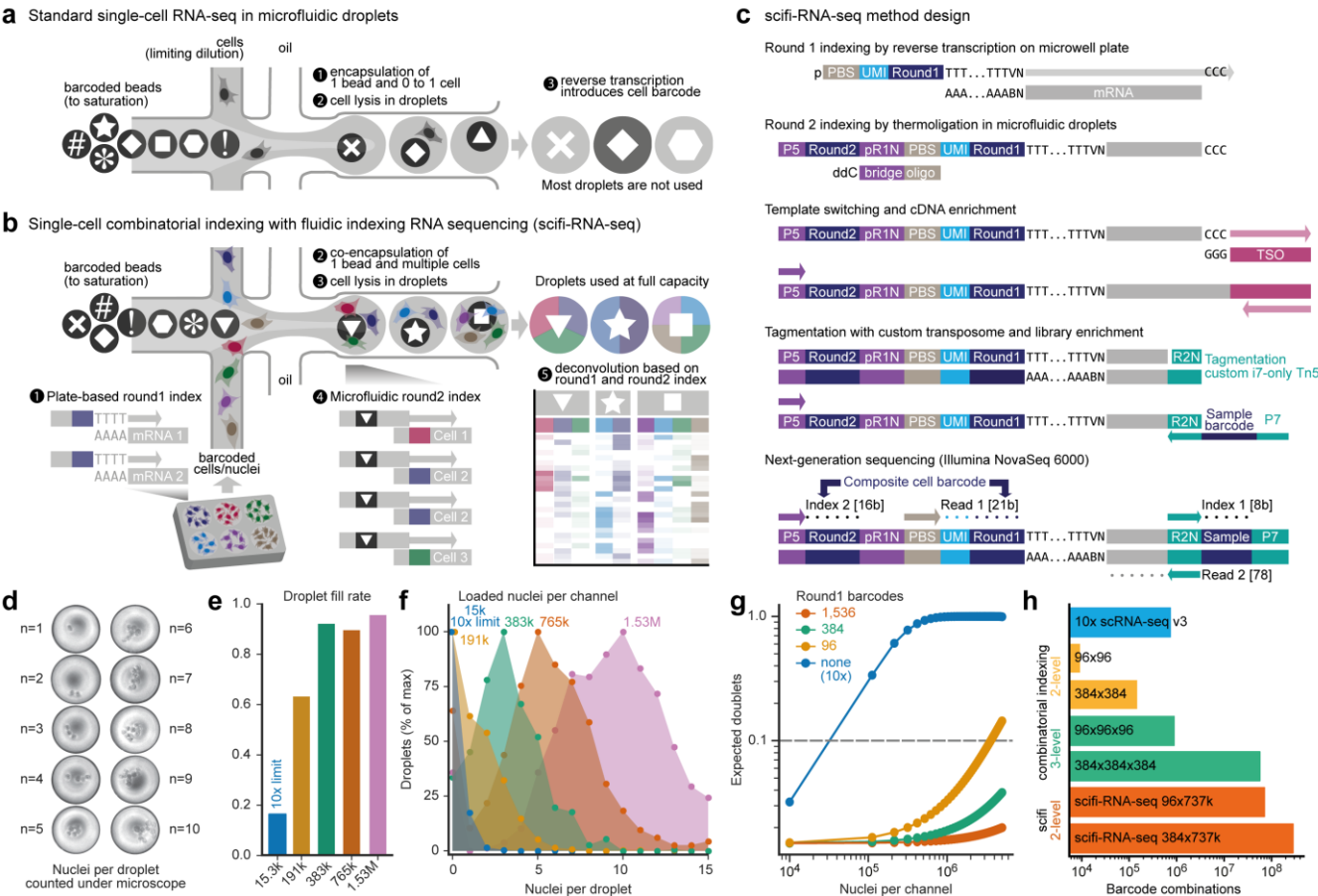


Figure 1: Single-cell combinatorial fluidic indexing (scifi) combines pre-indexing of entire transcriptomes with droplet-based single-cell RNA-seq

a) Standard droplet-based scRNA-seq using a microfluidic droplet generator is highly inefficient in its use of the droplets. Most droplets contain both a barcoded microbead and the reverse transcription reagents (and are thus fully functional), but never receive a cell. Furthermore, the reagents within a droplet are sufficient to barcode more than one cell. **b)** scifi-RNA-seq unlocks the full potential of microfluidic droplet generators. Prior to the microfluidic run, entire transcriptomes are pre-indexed by reverse transcription inside permeabilized cells or nuclei. The differentially barcoded pool is loaded at fill rates of multiple cells/nuclei per droplet. Cells/nuclei inside the same emulsion droplet are labelled with an identical microfluidic (round2) barcode, but can still be distinguished via their transcriptome (round1) index. **c)** Detailed method design for scifi-RNA-seq. **d)** By omitting the lysis reagents, intact nuclei can be imaged inside emulsion droplets, confirming the feasibility of overloading microfluidic droplet generators. Representative droplets containing between 1 and 10 nuclei are shown. **e)** Overloading boosts the percentage of droplets filled with nuclei from 16.4% (10x Genomics maximum) to 95.5% (100-fold overloading using 1.53 million nuclei per channel). **f)** Overloading causes the average number of nuclei per droplet to increase in a controlled fashion while maintaining the desired random loading distribution. **g)** Expected doublet rate as a function of the cell/nuclei loading concentration per channel for defined sets of round1 barcodes. The cell/nuclei fill rate was modelled as a zero-inflated Poisson distribution. **h)** Due to the high number of microfluidic round2 barcodes, 2-level scifi exceeds the barcode combinations of 3-level combinatorial indexing.

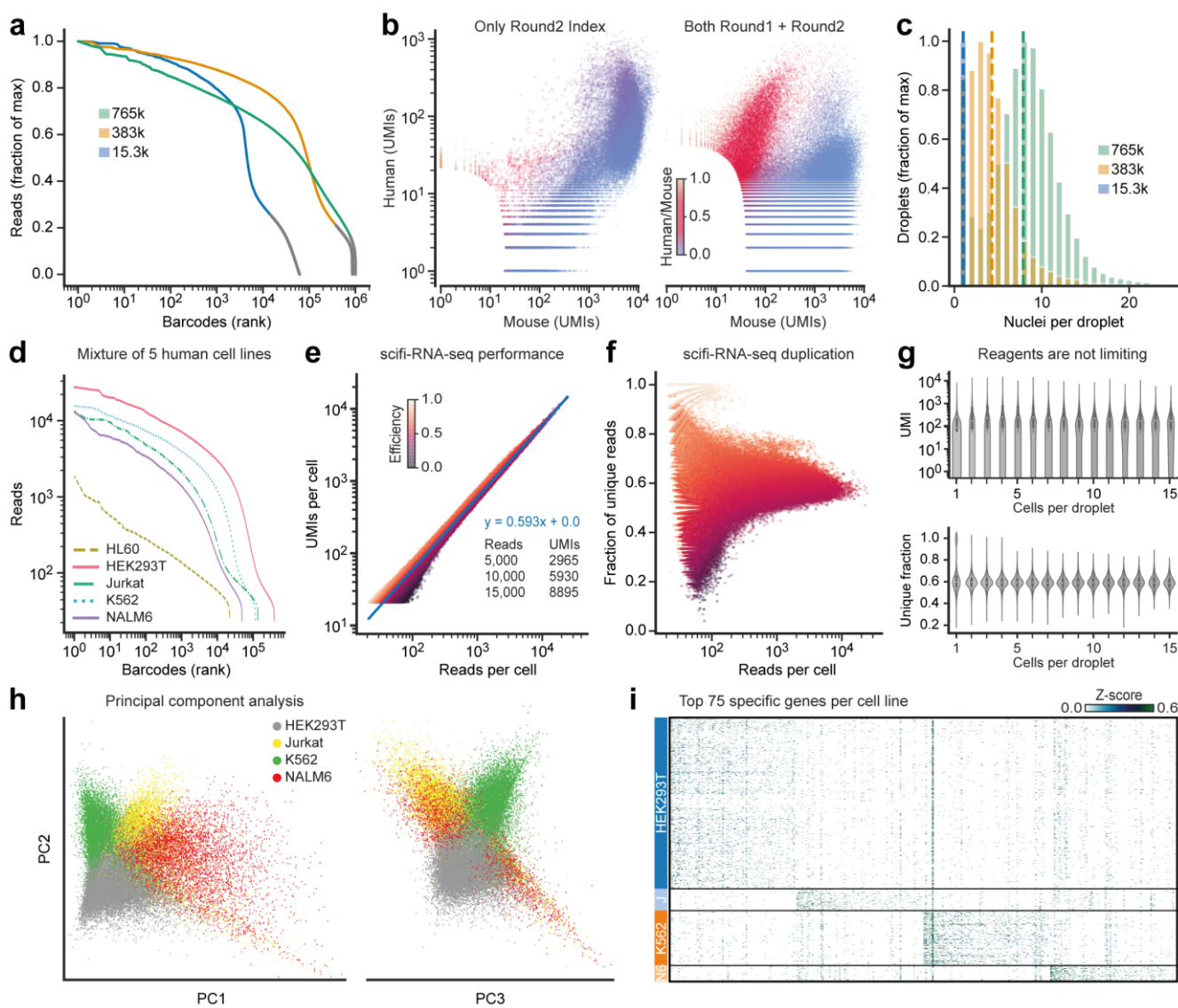


Figure 2: scifi-RNA-seq yields high quality single-cell transcriptomes

483 **a)** A 1:1 mixture of human and mouse cells (Jurkat and 3T3, respectively) was processed with scifi-RNA-seq.
 484 We loaded 15,300, 383,000 and 765,000 nuclei into single microfluidic channels of the Chromium device.
 485 Plotting all detected, ranked barcodes against the number of reads per barcode identifies a characteristic
 486 inflection point that separates nuclei from noise. The results indicate that nuclei pre-processed with scifi-RNA-
 487 seq are stable in the microfluidic device over a wide range of loading concentrations. **b)** The round1
 488 transcriptome index can deconvolute multiple nuclei per droplet into the respective single-cell transcriptomes. A
 489 1:1 mixture of 383,000 human (Jurkat) and mouse (3T3) cells and nuclei was processed and demultiplexed based
 490 on the microfluidic round2 barcode only (left plot), or based on the combination of round1 and round2 barcodes
 491 (right plot). **c)** Number of detected round1 barcodes per droplet for defined nuclei loading concentrations. **d)** Five
 492 human cell lines with variable cell size and transcript content (HEK293T, K562, Jurkat, NALM6, HL60) were
 493 processed with scifi-RNA-seq, using defined sets of round1 barcodes for each cell line. **e)** Performance plot
 494 showing unique molecular identifiers (UMIs) per nucleus as a function of the sequencing coverage. The
 495 efficiency (fraction of unique reads) is shown as a colour gradient. **f)** Fraction of unique reads plotted as a
 496 function of the sequencing coverage per nucleus. **g)** UMIs and the fraction of unique reads per nucleus are
 497 plotted against the number of nuclei contained in their respective droplet, indicating that there is no degradation
 498 in library complexity for high numbers of nuclei per droplet. **h)** Principal component analysis of the human cell
 499 line mixture experiment. The colours indicate the corresponding set of round1 indices for each cell line. **i)** Top 75
 500 specific genes per cell line.

501 **Supplementary figures**

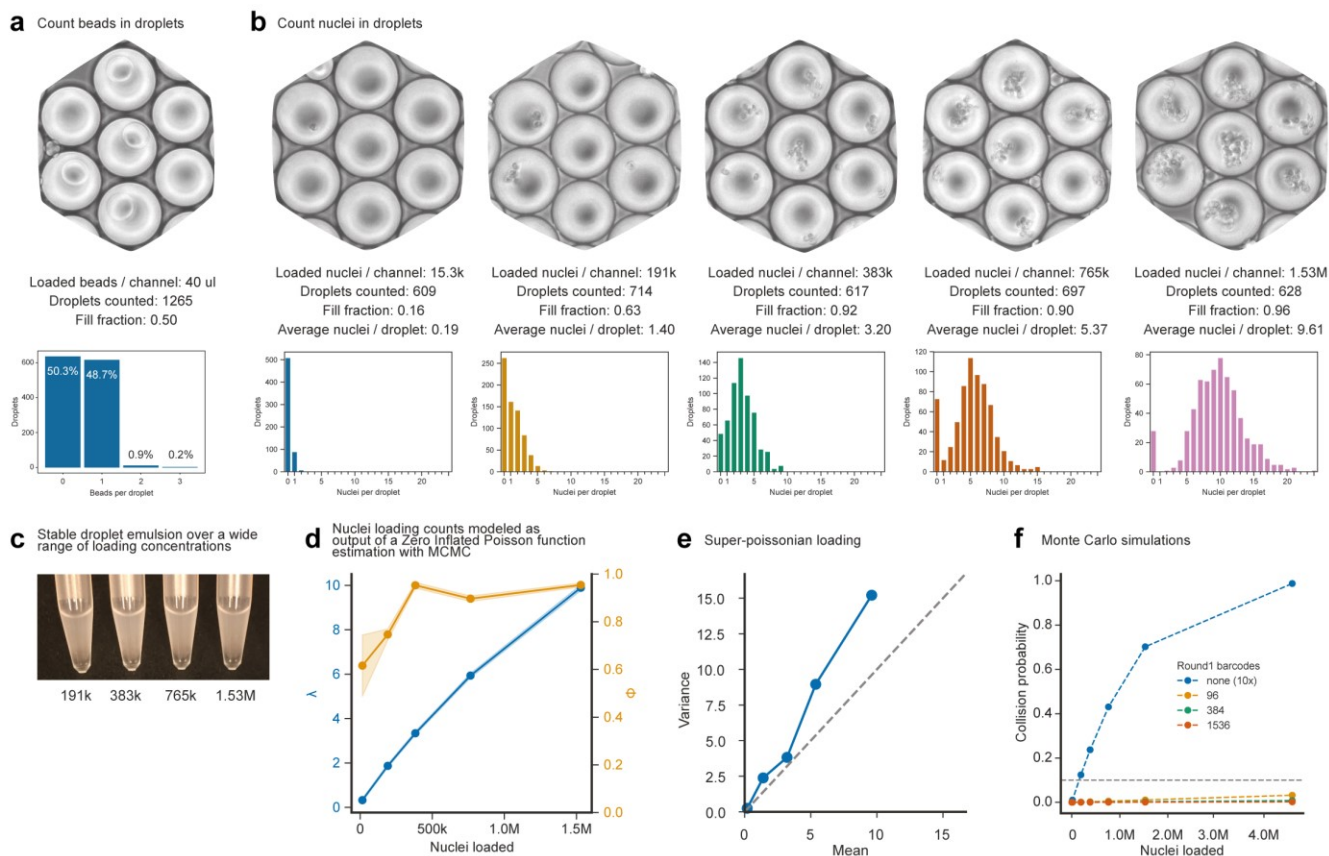
Design for 2-level single-cell combinatorial indexing with fluidic indexing RNA-seq (scifi-RNA-seq)
by Paul Datlinger, 2019-06-12



502
503 **Supplementary Figure 1: Detailed method design for scifi-RNA-seq**

504 Schematic with detailed oligonucleotide sequences, covering the entire scifi-RNA-seq protocol. The reverse
505 transcription is done on intact cells/nuclei in microwell plates, to introduce a well-specific round1 barcode into the
506 entire transcriptome. Pre-indexed cells/nuclei are pooled and encapsulated into emulsion droplets on a microfluidic
507 droplet generator. The round2 index is introduced through cycling ligation with a complementary bridge oligo and
508 thermostable ligase. Afterwards, the emulsion is broken, and a second defined end is introduced into the library
509 via template switching. cDNA is enriched and tagmented with a custom i7-only transposome. Finally, the library
510 is PCR-enriched, with the option to introduce an additional sample index.

511



Supplementary Figure 2: The Chromium droplet generator tolerates substantial overloading with cells/nuclei while maintaining a random loading distribution

a) By substituting the nuclei suspension with 1x Nuclei Buffer and omitting Reducing Agent B, intact gel beads could be visualized inside the emulsion droplets. Bead fill rates based on 1,265 evaluated droplet images are shown.

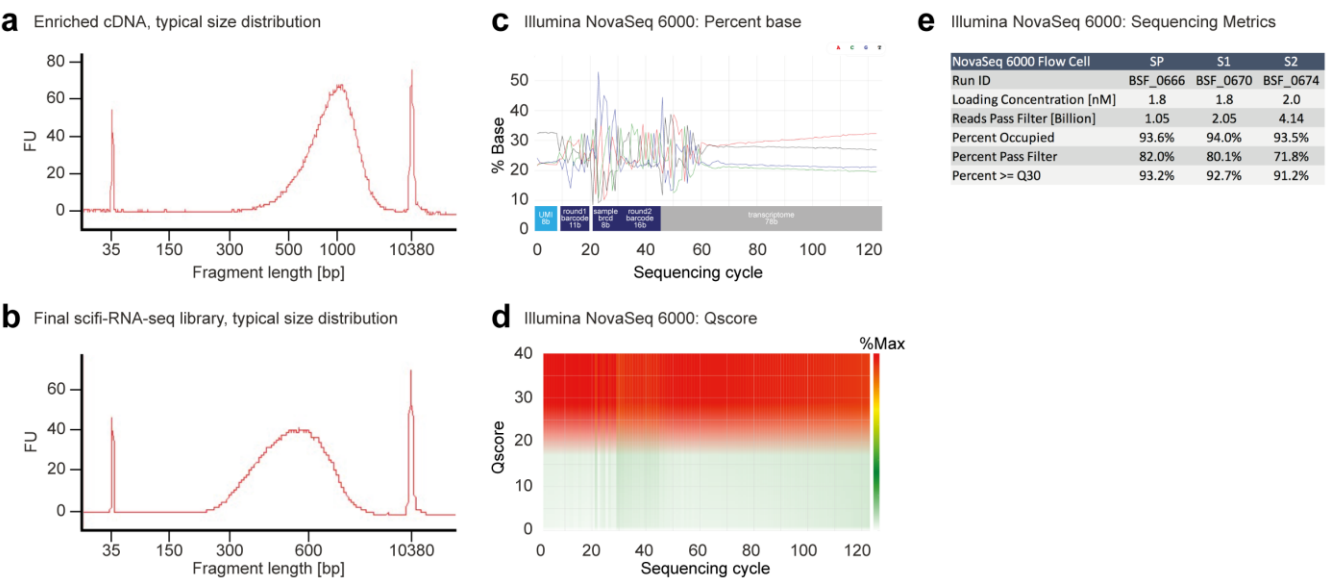
b) By omitting the lysis reagents, intact nuclei could be imaged using a standard microscope. This allows the exact counting of nuclei per droplet. Results for loading concentrations of 15,300, 191,000, 383,000, 765,000, and 1,530,000 cells/nuclei per channel are summarized as histograms.

c) Despite substantial overloading of the microfluidic device, we obtained stable droplet emulsions for all tested conditions.

d) Computational modeling of cells/nuclei loading as a zero-inflated Poisson function.

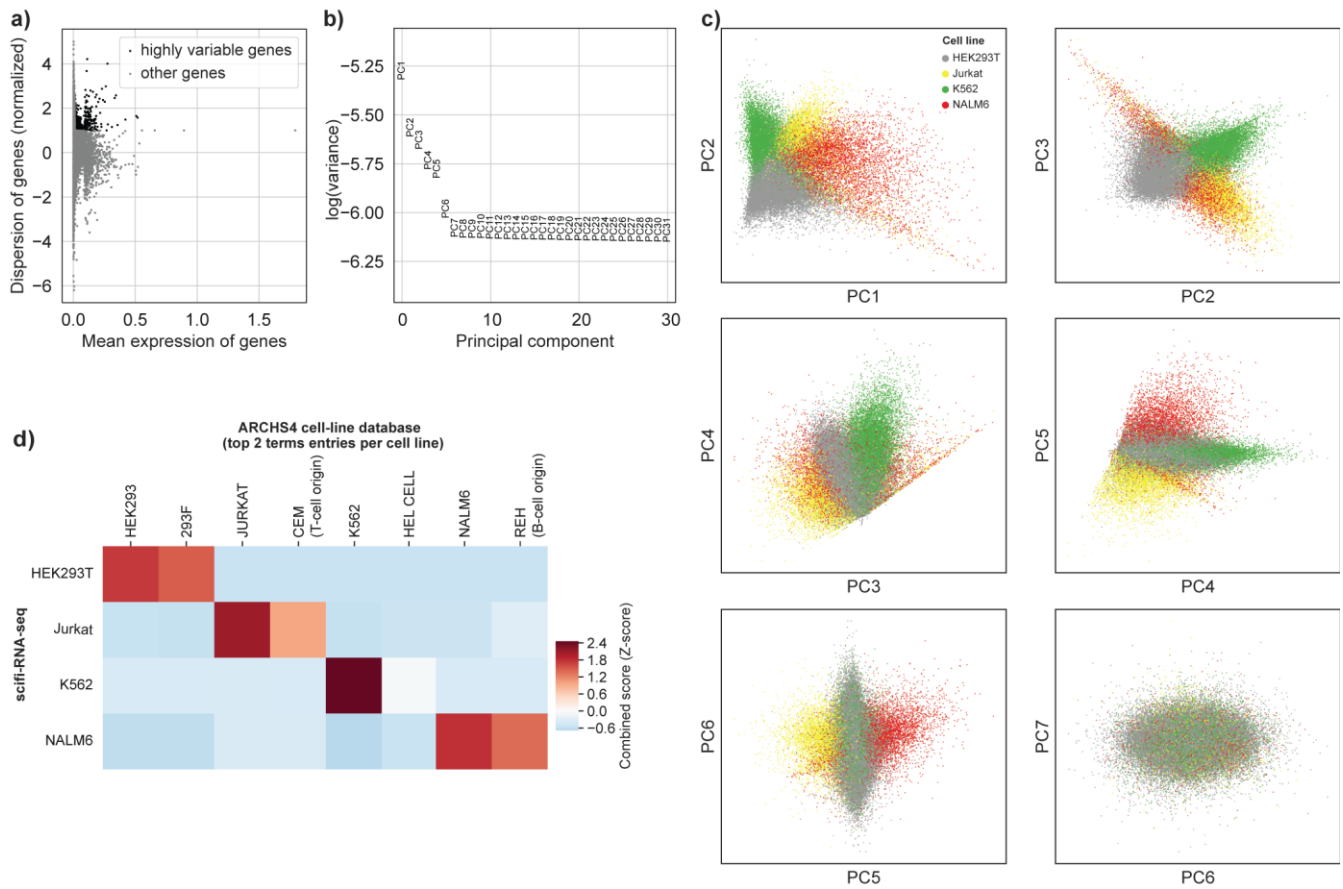
e) Nuclei loading displays super-poissonian properties.

f) Independent estimation of the cell doublet rates through Monte Carlo simulations of the scifi process.



Supplementary Figure 3: scifi-RNA-seq library preparation and next-generation sequencing

a) Typical size distribution of enriched cDNA. **b)** Typical size distribution of a final scifi-RNA-seq library. **c-e)** Key metrics from next-generation sequencing runs on the Illumina NovaSeq 6000 platform. scifi-RNA-seq was tested with SP, S1 and S2 reagents.



Supplementary Figure 4: Human cell line mixture experiment

a) Selection of highly variable genes as input for principal component analysis. **b)** Variance explained by principal components. **c)** PCR projections of PC1 to PC7. Cell line assignments based on round1 indices are coloured. **d)** Transcriptome signatures measured by scifi-RNA-seq correctly identify each cell line in an RNA-seq database. Of note, the CEM cell line was derived from T-cells and REH cells have a B-cell origin, which might explain their similarity to Jurkat and NALM6, respectively.

537 **References**

538 1. Macosko, E.Z. et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets.
539 *Cell* **161**, 1202-1214 (2015).
540 2. Klein, A.M. et al. Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* **161**, 1187-1201
541 (2015).
542 3. Zheng, G.X. et al. Massively parallel digital transcriptional profiling of single cells. *Nat Commun* **8**, 14049 (2017).
543 4. Cao, J. et al. Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661-667 (2017).
544 5. Rosenberg, A.B. et al. Single-cell profiling of the developing mouse brain and spinal cord with split-pool barcoding.
545 *Science* **360**, 176-182 (2018).
546 6. Stoeckius, M. et al. Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell
547 genomics. *Genome Biol* **19**, 224 (2018).
548 7. McGinnis, C.S. et al. MULTI-seq: sample multiplexing for single-cell RNA sequencing using lipid-tagged indices. *Nat*
549 *Methods* **16**, 619-626 (2019).
550 8. Satpathy, A.T. et al. Massively parallel single-cell chromatin landscapes of human immune cell development and
551 intratumoral T cell exhaustion. *bioRxiv* (2019).
552 9. Fan, H.C., Fu, G.K. & Fodor, S.P. Expression profiling. Combinatorial labeling of single cells for gene expression
553 cytometry. *Science* **347**, 628-637 (2015).
554 10. Gierahn, T.M. et al. Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* **14**,
555 395-398 (2017).
556 11. Han, X. et al. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* **172**, 1091-1107 e1017 (2018).
557 12. Cao, J. et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
558 13. Cusanovich, D.A. et al. Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing.
559 *Science* **348**, 910-914 (2015).
560 14. Vitak, S.A. et al. Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat Methods* **14**, 302-308
561 (2017).
562 15. Mulqueen, R.M. et al. Highly scalable generation of DNA methylation profiles in single cells. *Nat Biotechnol* **36**, 428-
563 431 (2018).
564 16. Ramani, V. et al. Massively multiplex single-cell Hi-C. *Nat Methods* **14**, 263-266 (2017).
565 17. Rozenblatt-Rosen, O., Stubbington, M.J.T., Regev, A. & Teichmann, S.A. The Human Cell Atlas: from vision to reality.
566 *Nature* **550**, 451-453 (2017).
567 18. Treutlein, B. et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* **534**,
568 391-395 (2016).
569 19. Datlinger, P. et al. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* **14**, 297-301 (2017).
570 20. Adamson, B. et al. A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the
571 Unfolded Protein Response. *Cell* **167**, 1867-1882 e1821 (2016).
572 21. Dixit, A. et al. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic
573 Screens. *Cell* **167**, 1853-1866 e1817 (2016).
574 22. Jaitin, D.A. et al. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* **167**,
575 1883-1896 e1815 (2016).
576 23. Vera Rykalina, A.S., Hans Lehrach, Tatiana Borodina qPCR-based characterization of DNA fragmentation efficiency of
577 Tn5 transposomes. *Biology Methods and Protocols* **2**, 1-8 (2017).
578 24. Salvatier J, W.T., Fonnesbeck C Probabilistic programming in Python using PyMC3. *PeerJ Computer Science*, e55
579 (2016).
580

3. MATERIALS AND METHODS

The following section contains detailed step-by-step protocols for CROP-seq (page 72) and scifi-RNA-seq (page 93). Both methods are provided as open source protocols and can be used for academic research without limitation. All oligonucleotide sequences, reagents and buffers required to reproduce our results are included in supplementary tables. The analysis code is available via public Github repositories, and raw data is available from the Gene Expression Omnibus (links and accession numbers can be found in the respective manuscript). To help interested labs to establish the protocol, we included detailed figures with exemplary results and data.

As both methods are actively used and may evolve over time, we will make future updates of the protocols available via dedicated supplementary websites:

<http://crop-seq.computational-epigenetics.org>

<http://scifi-rna-seq.computational-epigenetics.org>

3.1. Step-by-step protocol for CROP-seq (CRISPR droplet sequencing)

Authors: Paul Datlinger (pdatlinger@cemm.oeaw.ac.at), André F Rendeiro*, Christian Schmidl*, Thomas Krausgruber, Peter Traxler, Johanna Klughammer, Linda C Schuster, Amelie Kuchler, Donat Alpar and Christoph Bock (cbock@cemm.oeaw.ac.at).

Future updates of the protocol will be shared via <http://crop-seq.computational-epigenetics.org>.

3.1.1. Pooled, amplification-free cloning of gRNA libraries

Validation of the CROPseq-Guide-Puro plasmid preparation

Order the CROPseq-Guide-Puro plasmid from Addgene ([86708](#)), perform plasmid DNA extraction, and validate by restriction digestion and Sanger sequencing.

Plasmid validation by restriction digestion			
Restriction enzyme	Buffer	Temperature [°C]	Expected fragments [bp]
SapI	CutSmart	37 °C	1171, 3394, 5649
SphI	NEB Buffer 2.1	65 °C	1110, 3787, 5317

Plasmid validation by Sanger sequencing		
Primer ID	Primer sequence [5' > 3']	Validates element(s)
86708_p1	GAGGGCCTATTCCCATGATTCC	hU6, BsmBI site
86708_p2	CAAATGCATGCTCTTCAACCTC	BsmBI site, gRNA backbone, 3' LTR
86708_p3	GAGACAAATGGCAGTATTCATCCAC	EF-1a
86708_p4	TAAGTGCAGTAGTCGCCGTG	EF-1a, Puro
86708_p5	AGTCTTGTAATGCGGGCCA	Puro
86708_p6	CACACTGAGTGGGTGGAGAC	Puro
86708_p7	AGCAACAGATGGAAGGCCTC	WPRES
86708_p8.1	TTGGGCACTGACAATTCCTG	3' LTR, filler or gRNA
86708_p10	GACGAACCACTGAATTGCCG	Psi, Gag
86708_p11	GGGAGCTAGAACGATTCGCA	RRE
86708_p12	TTCCTTGGGTCTTGGGAGC	cPPT
86708_p13	CGAGTCAGTGAGCGAGGAAG	5' LTR

Preparation of CROPseq-Guide-Puro vector backbone

Plasmid digestion

Digest CROPseq-Guide-Puro plasmid and gel-purify the 8,333 bp fragment as indicated. Once prepared, the backbone fragment can be frozen at -20 °C and re-used for cloning further gRNA libraries.

Vector digestion mix							
Component	Per reaction		Master mix for # of digestions				
	Volume	Amount	4	6	8	10	12
CROPseq-Guide-Puro		5 µg					
NEB Buffer 3.1	2.5 µl	1x	10.0	15.0	20.0	25.0	30.0
BsmBI	2 µl	20 U	8.0	12.0	16.0	20.0	24.0
Nuclease-free water		to 25 µl					

Distribute 25 µl of vector digestion mix per well into a tube strip, incubate: 1 hour at 55 °C (digestion), 20 min at 80 °C (inactivation of the enzyme), hold at 10 °C.

UV-free gel extraction

<Note> Based on SNAP UV-Free Gel Purification Kit (Invitrogen cat. no. 45-0105)

Prepare required number of 0.8% agarose gels with TAE buffer containing 1.8 µg/ml Crystal Violet:

Example for one gel: 0.4 g agarose in 50 ml TAE, plus 40 µl of 2 mg/ml Crystal Violet Solution

Important: use a thin comb for more intense bands, TBE is not compatible with this kit

Load 8 µg of DNA ladder in 25 µl of water, mixed with 5 µl of 6x Crystal Violet Loading Dye

Load 25 µl of each plasmid digestion, mixed with 5 µl of 6x Crystal Violet Loading Dye

Run at 80 V and observe the separation of bands with the naked eye

Once the bands are sufficiently resolved, stop and isolate 8,333 bp fragment with gel-cutting tips

Transfer gel slice to 1.5 ml tube, weigh to estimate the agarose volume (assuming that 1 mg ~ 1 µl)

Add 2.5 volumes of Sodium Iodide Solution

Incubate at 42 °C with occasional vortexing, until the agarose is completely dissolved

Add 1.5 volumes of Binding Buffer, mix and load onto SNAP Purification Column

Centrifuge at 3,000 rcf for 30 seconds

Pour the flow through back onto the column and repeat the centrifugation, for a total of three times

After the last centrifugation step, discard the flow through

Wash two times with 400 µl of 1x Final Wash Buffer, discard flow through after second centrifugation

Centrifuge column for 1 min at > 10,000 rcf to remove residual wash buffer

Transfer the column to a fresh 1.5 ml tube and elute with 40 µl of EB buffer (10,000 rcf for 1 min)

Important: incubate for at least 1 min to allow absorption into the column

Transfer the eluate back to the column, and elute a second time as described above

Measure 1 µl of the eluate in a Qubit HS assay to determine the concentration of dsDNA

<Safe Stopping Point> The backbone fragment can be stored at -20 °C and re-used

<Troubleshooting> If required, perform ethanol precipitation to further concentrate the DNA: pool multiple gel-purification reactions, add 1/10 volume of 3M sodium acetate, pH 5.5 and 2.5 volumes of 100% ice cold ethanol. Precipitate DNA for 30 min at -80 °C, then spin for 20 min at 20,000 rcf at 4 °C to pellet. Wash pellet two times with room temperature 70 % ethanol, leave to dry (take care not to over-dry the pellet) and elute in a smaller volume of EB buffer.

Assembly of gRNA-encoding ssDNA oligonucleotides into vector backbone by Gibson Assembly

Design gRNAs

Recommended for manual design: <https://benchling.com/>, for automated design: <https://github.com/boutroslab/cld/>. Order as separate ssDNA oligos, with 18 and 35 bases of homology to the hU6 promoter and gRNA backbone for Gibson Assembly. The 5' prefix also adds a G as a constant first base for more

efficient transcription from the hU6 promoter. Examples for positive controls (targeting essential genes) and negative controls (non-targeting) can be found in the original publication of CROP-seq (Supplementary Table 2).

<Example> TGGAAAGGACGAAACACCG (5' prefix, hU6 promoter and G as first base) + gRNA + GTTTTAGAGCTAGAAATAGCAAGTTAAAATAAGGC (3' postfix, gRNA backbone)

Gibson Assembly

Dilute gRNA-encoding ssDNA oligos to 100 µM with EB buffer. Pool all oligos contained in the library (including positive and negative controls) in equal amounts. Dilute the oligo pool 1:1000 in EB buffer, resulting in a molarity of 100 nM. Set up Gibson assembly reactions for each library and the included NEBuilder HiFi positive control, as described below.

Assembly of gRNA-encoding ssDNA oligos into the vector backbone by Gibson Assembly		
Per reaction		
Component	Volume	Amount
CROPseq-Guide-Puro backbone fragment (8,333 bp)		11 fmoles
Library ssDNA oligos, pooled (100 nM)	2.0 µl	200 fmoles
NEBuilder HiFi DNA Assembly Master Mix	10 µl	1x
Nuclease-free water		to 20 µl

<Note> We use the following online tool to calculate the volume of backbone fragment required:
<http://nebiocalculator.neb.com/#!/dsdnaends>

Incubate at 50 °C for 1 hour, then dialyze the reaction against water:

Fill a 6 cm cell culture dish with HPLC-quality water, place membrane filter on top

Pipette 20 µl of the Gibson Assembly reaction and let float on top of the filter for 30 min

<Note> This removes salts and makes the electroporation more efficient

About 10 µl of the Gibson Assembly reaction can be recovered after the filtration

Electroporation of Lucigen Endura cells

Preparations

Prepare mixture of ice and water

Thaw one vial of Lucigen Endura electrocompetent cells on ice (takes about 20 min, sufficient for cloning one gRNA library with two electroporations)

Pre-chill two 1 mm electroporation cuvettes on ice

Place two empty 2 ml tubes on ice

Pre-warm Lucigen recovery medium at 37 °C

Pre-warm two 25 x 25 cm LB-agar plates containing 100 µg/ml carbenicillin at room temperature

Electroporation

Aliquot 25 µl of bacteria into pre-chilled 2 ml tubes, and keep on ice

Add 10 µl of de-salted Gibson Assembly to the bacteria and gently tap the tube to mix

Transfer to electroporation cuvette, tap to bring contents to the bottom, check for air bubbles

Prepare P1000 pipette filled with 1 ml of Lucigen recovery medium

Electroporate with the following settings: 25 µF, 200 Ω, 1.5 kV

Immediately after the pulse, add the prepared Lucigen recovery medium

Gently pipet up and down two times, then transfer to round-bottom tubes

Take a note of the time constant for the electroporation (should be ~4.5 msec)

Incubate for 1 hour at 37 °C for recovery

Prepare a 1:100 dilution (10 µl bacteria plus 990 µl SOC medium), and plate 100 µl on a 10 cm LB-agar dish

Plate the remaining bacteria onto the 25 x 25 cm LB-agar plates

Incubate for 20 hours at 32 °C

Plasmid preparation

The next day, count colonies on the 1:1000 dilution plate and calculate the library coverage as follows:
(number of colonies x 1000) / (number of gRNAs in library)

Retrieve bacteria from plates as follows:

Add 20 ml of pre-warmed LB medium to the plate, scrape off bacteria, and transfer to 50 ml tube

Add another 30 ml of pre-warmed LB medium, tilt plate and transfer to 50 ml tube

Spin at 4,700 rcf for 15 min to collect a bacterial pellet and discard supernatant

Proceed with plasmid extraction using the Endo-Free Plasmid Mega kit (Qiagen #12381) according to the manufacturer's instructions

3.1.2. Next-generation sequencing of gRNA sequences for library quality control or pooled screens

gRNA sequences are amplified from a CROPseq-Guide-Puro plasmid library or from the gDNA of transduced cells (pooled screen) in a single PCR reaction with specially designed primers. This step adds all sequences required for compatibility with Illumina sequencers. In addition, i7 barcodes are provided for sample multiplexing, and stagger sequences can increase library complexity for Illumina sequencing.

Primers for amplification of gRNAs for next-generation sequencing

Primer ID	Primer sequence [5' > 3']
CROPseq_libQC_i5_s1	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATCTTCTTGTGGAAAGGACGAAACACCG
CROPseq_libQC_i5_s2	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATCTATCTTGTGGAAAGGACGAAACACCG
CROPseq_libQC_i5_s3	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATCTGATCTTGTGGAAAGGACGAAACACCG
CROPseq_libQC_i5_s4	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATCTCGATCTTGTGGAAAGGACGAAACACCG
CROPseq_libQC_i5_s5	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATCTTCGATCTTGTGGAAAGGACGAAACACCG

CROPseq_libQC_i5_s6	AATGATACGGCGACCACCGAGATCTACACACACTCTTTCCCTACACGACGCTCTTCCGATCTATCGATCTTGTGGAAAGGACGAAACACCG
CROPseq_libQC_i7:1	CAAGCAGAAGACGGCATACGAGATCGAGTAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTCTCAAGATCTAGTTACGCCAAGC
CROPseq_libQC_i7:2	CAAGCAGAAGACGGCATACGAGATTCTCCGGAGTACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTCTCAAGATCTAGTTACGCCAAGC
CROPseq_libQC_i7:3	CAAGCAGAAGACGGCATACGAGATAATGAGCGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTCTCAAGATCTAGTTACGCCAAGC
CROPseq_libQC_i7:4	CAAGCAGAAGACGGCATACGAGATGGAATCTCTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTCTCAAGATCTAGTTACGCCAAGC
CROPseq_libQC_i7:5	CAAGCAGAAGACGGCATACGAGATTCTGAATGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTCTCAAGATCTAGTTACGCCAAGC
CROPseq_libQC_i7:6	CAAGCAGAAGACGGCATACGAGATACGAATTCGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGTCTCAAGATCTAGTTACGCCAAGC

Functional elements

5' part of i5 or i7 Illumina adapter
 i7 index for sample multiplexing
 3' part of i5 or i7 Illumina adapter
 stagger sequence to increase complexity for Illumina sequencing
 primer binding site on CROPseq-Guide-Puro

To retain a good library representation and reduce bias, multiple PCR reactions should be performed for each gRNA plasmid library or gDNA preparation. The number ultimately depends on the size of the gRNA library, but we mostly use four parallel PCR reactions. These reactions can use the same i7 barcode, so that they are demultiplexed into the same file during data processing.

qPCR amplification of gRNAs from the plasmid library or gDNA of transduced cells		
Per reaction		
Component	Volume	Amount
CROPseq-Guide-Puro plasmid library or gDNA from a pooled screen		10 ng or 100 ng
Q5 Hot Start High-Fidelity 2x Master Mix	25 µl	1x
CROPseq_libQC_i5 primer (10 µM)	2.5 µl	500 nM
CROPseq_libQC_i7 primer (10 µM)	2.5 µl	500 nM
Optional: SYBR Green, 100x	0.5 µl	1x
Nuclease-free water		to 50 µl

By adding SYBR Green, the reactions can be run in a qPCR machine and stopped just before reaching the plateau phase. This ensures that the optimal number of enrichment cycles is used. When running multiple samples, the machine can be paused during the denaturation phase. Pierce the foil to remove finished reactions, and resume the amplification of the remaining samples.

Incubate as follows:

98 °C for 30 sec (initial denaturation)

40 cycles { 98 °C for 10 sec
72 °C for 45 sec }

stop qPCR reactions just before they reach the plateau phase

Clean PCR products with a 2.0x AMPure XP bead clean-up

Quality control on an Agilent Bioanalyzer High Sensitivity DNA chip should show a single band at 282 bp (an example is shown on the following page, in the *Expected results* section)

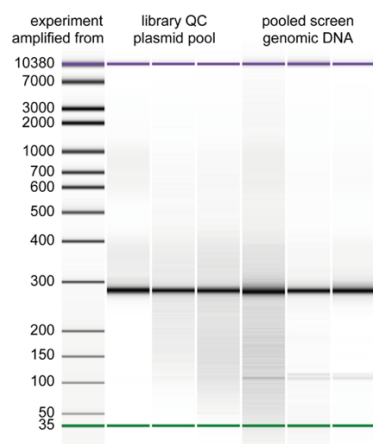
Measure the dsDNA concentration in a Qubit HS assay

Dilute library to 4 nM with EB buffer + 0.1% Tween based on size and concentration measurements

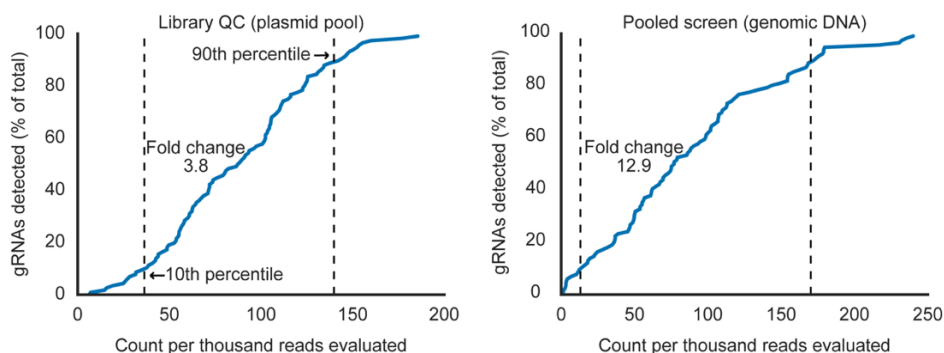
Sequence on an Illumina MiSeq machine, using a 150-cycle v3 flowcell with a read configuration of 167 bases (read 1) and 8 bases (i7 index). While this is the cheapest option, NextSeq and HiSeq instruments can be used instead

3.1.3. Expected results for the pooled, amplification-free cloning of gRNA libraries

gRNAs amplified from the plasmid pool during library quality control, or from genomic DNA in a classical pooled screen. The 282 bp PCR product is compatible with Illumina sequencing without further processing.



Typical representation of the gRNA library in the plasmid pool (left) and when amplified from genomic DNA in a pooled screen (right, day 10 post transduction).



3.1.4. Lentivirus production and cell culture for CROP-seq screens

Production of lentiviral particles

Prerequisites:

Plasmid pool of gRNA library cloned as described in part A, and validated by NGS

Plasmid preparations of the 3rd generation lentiviral packaging plasmids pMDLg/pRRE (Addgene [12251](#)), pRSV-Rev (Addgene [12253](#)), and pMD2.G (Addgene [12259](#))

Cell line(s) stably expressing Cas9 after transduction with lentiCas9-Blast (Addgene [52962](#)). The lentivirus production and transduction protocols below can be used to produce this cell line

<Note> The required scale of the lentivirus production depends on the size of the gRNA library and the planned experiments. For most applications **10 cm dishes**, or even **6-well plates** should suffice, due to the relatively high titers achieved with the CROPseq-Guide-Puro construct. Numbers for both formats are provided below (color-coded).

<Important> Dispose of lentivirus-containing materials according to your institute's biosafety guidelines. Persons performing experiments with lentivirus must be properly trained.

Prepare lentivirus packaging medium: Opti-MEM I, 5% FCS, 200 µM sodium pyruvate, no antibiotics

Seed HEK293T cells at **7 million cells per 10 cm dish in 12 ml of lentivirus packaging medium**, or at **1.2 million cells per well on a 6-well plate in 2 ml of lentivirus packaging medium**

The next morning proceed with the transfection

Prepare transfection mixtures A and B as indicated below

Transfection of HEK293T cells for lentivirus production					
Transfection mix	Component	10 cm dish		6-well plate	
		Volume	Amount	Volume	Amount
A	CROPseq-Guide-Puro gRNA library, plasmid pool		10.2 µg		1.7 µg
	pMDLg/pRRE		5.4 µg		0.9 µg
	pRSV-Rev		5.4 µg		0.9 µg
	pMD2.G		5.4 µg		0.9 µg
	Opti-MEM I	1.5 ml		250 µl	
	P3000 enhancer reagent	36 µl		6 µl	
B	Opti-MEM I	1.5 ml		250 µl	
	Lipofectamine 3000 reagent	42 µl		7 µl	

Prepare lipid-DNA complexes by combining transfection mixes A and B, and incubating for 20 min at room temperature

In the meantime, remove **6 ml**, or **1 ml** of medium from the HEK293T cells

Add **3 ml** or **500 µl** of lipid-DNA complexes to the HEK293T cells, and gently agitate

After 6 h of incubation at 37 °C and 5% CO₂ exchange medium for **12 ml** or **2 ml** of fresh lentivirus packaging medium

Collect lentivirus-containing medium after 24 h, store at 4 °C, add **12 ml** or **2 ml** of fresh lentivirus packaging medium

Collect lentivirus-containing medium after 48 h, pool with supernatant obtained the previous day and filter through a 0.45 µm filter to remove cells

Aliquot lentivirus preparation into cryotubes and freeze at -80 °C

Proceed with titration of the lentivirus preparation and pooled screening

Titration of lentivirus preparations (optional)

<Note> To save time, transductions for pooled screens can be performed with variable volumes of the virus preparation. All transduction conditions are then selected for 2 days, and the condition with the desired MOI can be chosen for the screen.

Seed HEK293T cells onto 24-well plates at 50,000 cells per well in 500 µl of complete culture medium (DMEM, 10% FCS, penicillin-streptomycin)

Grow overnight at 37 °C and 5% CO₂ so that cells reach 30% to 50% confluence

Exchange culture medium for 450 µl complete culture medium, supplemented with 8 µg/ml polybrene

Thaw one vial of the lentivirus preparation from storage at -80 °C

In a 96-well plate, prepare a 1:5 dilution series of the lentivirus preparation in complete culture medium with polybrene, ranging over 10 wells (dilution factors from 1:5 to 1:9,765,625)

Test the original stock, and each dilution in duplicate, by adding 50 µl per well to the 24-well plate with HEK293T cells (this requires 2 x (1 + 10) = 22 wells).

The remaining two wells serve as untransduced controls. Add 50 µl complete culture medium containing polybrene to test for toxicity of polybrene to cells (possible for some cell types).

24 h post transduction exchange the medium for 500 µl complete culture medium without polybrene

48 h post transduction start selection with puromycin and renew selective medium every 2-3 days.

The table below provides common antibiotic concentrations we use for puromycin selection, but ultimately the concentration will depend on the cell line used.

Examples of selection conditions used in our lab		
Cell line	c(Puromycin) [µg/ml] to select for gRNA	c(Blasticidin) [µg/ml] to select for Cas9
K562	2	30
Jurkat	2	25
MV4-11	1	20
KBM7	0.5	20
HEK293T	2.25	22.5
NALM-6	0.75	30

As soon as cells in the untransduced controls are dead, remove medium and wash once with 1x PBS

Stain colonies for 15 min in a solution of 1% (w/v) crystal violet in 10% ethanol

Wash several more times with 1x PBS, count colonies for two dilution factors, and two replicates each

Calculate the virus titer (transducing units/ml) as:

(average number of resistant colonies) x (dilution factor of well) x (1,000 µl/50 µl)

Example for a typical gRNA library:

56 and 66 colonies were counted for the well with dilution factor 15,625

$((56 + 66) / 2) \times (15,625) \times (1,000 \mu\text{l} / 50 \mu\text{l}) = 1.9 \times 10^7 \text{ TU/ml}$

Transduction of suspension cells for pooled CROP-seq screens

Seed suspension cells in 6-well plates at 5 million cells per well in 2 ml of complete culture medium (depends on cell line), supplemented with 8 $\mu\text{g/ml}$ polybrene and blasticidin (we recommend to keep up selection for lentiCas9-Blast throughout the screen, as it is easily silenced; please refer to the table above for suggestions on blasticidin concentrations for various cell lines).

Option 1 (virus titer was measured): Add calculated volume of lentivirus preparation for the desired MOI to one well, use a second well as untransduced control.

Option 2 (virus titer unknown): Transduce five wells with different volumes of the lentivirus preparation (e.g. 250 μl or 50 μl of the original stock, or 50 μl of 1:3, 1:6 and 1:12 dilutions). The sixth well is left untransduced.

Immediately after addition of lentiviral particles, centrifuge the plate for 45 min at 1,200 rcf and 37 °C (spinfection), followed by over night incubation at 37 °C and 5% CO₂.

24 h post transduction, pellet cells in 15 ml tubes, remove supernatant. Transfer cells to T75 cell culture flasks in 30 ml of fresh complete culture medium supplemented with blasticidin (still only selecting for lentiCas9-Blast). 48 h post transduction, start selection for the gRNA library by adding puromycin directly to the flask. Renew selective medium containing blasticidin and puromycin every 2-3 days for a total of 7-10 days to allow for efficient genome editing. If option 2 was chosen, select the flask with about 30% surviving cells after 2 days of puromycin selection.

Transduction of adherent cells for pooled CROP-seq screens

Seed adherent cells in 15 cm dishes at 5 million cells per plate in 20 ml of complete culture medium (depends on cell line) supplemented with 8 $\mu\text{g/ml}$ polybrene and blasticidin (we recommend to keep up selection for lentiCas9-Blast throughout the screen, as it is easily silenced; please refer to the table above for suggestions on blasticidin concentrations for various cell lines).

Option 1 (virus titer was measured): Add calculated volume of lentivirus preparation for the desired MOI to one plate, use a second plate as untransduced control.

Option 2 (virus titer unknown): Transduce five plates with different volumes of the lentivirus preparation (e.g. 250 μl or 50 μl of the original stock, or 50 μl of 1:3, 1:6 and 1:12 dilutions). A sixth plate is used as the untransduced control.

After addition of lentiviral particles, swirl the plate gently to mix and incubate over night at 37 °C and 5% CO₂. 24 hours post transduction, exchange medium for fresh complete culture medium supplemented with blasticidin (selecting for lentiCas9-Blast). At 48 h post transduction, start selection for the gRNA library by adding puromycin directly to the plate. Renew selective medium every 2-3 days for a total duration of 7-10 days to allow for efficient editing. If option 2 was chosen, select the plate with about 30% surviving cells after 2 days of puromycin selection.

3.1.5. Single-cell RNA-seq based on Drop-seq

Modified from: Macosko et al. 2015 and additional protocols (<http://mccarrolllab.com/dropseq/>)

Droplet microfluidics

Prepare cells

Collect cells by centrifugation at 300 rcf for 5 min

<Note> Centrifugation speed and time might need to be adjusted for other cell types

Wash once with 1 ml of PBS-0.01% BSA (make fresh on the day of the run)

Remove supernatant and resuspend in 1 ml of PBS-0.01% BSA

Filter through a 40 µm cell strainer to obtain a single-cell suspension

Measure cell concentration twice using a cell counter

<Note> We use a CASY device, but any alternative method can be used

Dilute to 220 cells per µl using PBS-0.01% BSA in a tube that allows easy uptake by the syringes

Check concentration of diluted cells twice

Prepare beads

Mix 1.6 ml of Drop-seq lysis buffer (stored at 4 °C) with 80 µl of DTT (stored at room temperature)

Dilute Drop-seq beads as follows

Resuspend Drop-seq beads (stored at 4 °C) by very gentle shaking

Immediately transfer the equivalent of 225,000 beads to a 1.5 ml tube

Centrifuge for 1 min at 1,000 rcf

Remove supernatant and take up beads in 1.5 ml of lysis buffer with DTT (150 beads per µl)

Transfer to a tube that allows easy uptake by the syringes

Drop-seq

<General comment> To load syringes, cut a 200 µl pipet tip without filter on both ends (to reduce shear force and to facilitate attachment to the syringe). After loading, invert the syringe, pull up whatever is left in the tip, discard tip, tap or push to remove air bubbles and attach a safe-lock needle and tubing using tweezers (be careful not to puncture your fingers!). Press syringe until tubing is free of air bubbles.

Load oil into a 10 ml syringe

Load cells into a 3 ml syringe

Place Drop-seq microfluidic device on microscope stage

Attach tubing for oil and cells and connect the outlet to a 50 ml tube (labelled “waste”)

Prepare a second 50 ml tube for collecting droplets once the run is stable

Load beads

<Critical step> Loading the beads is a critical step, as they can easily settle or clump. This can clog the needle or lead to an uneven bead concentration throughout the run.

Remove plunger from 3 ml syringe, then insert magnetic mixing disc and push plunger back
Set plunger of the empty syringe to 1.5 ml, attach to the pump
Set up the magnet to 1 jump per second and keep switched on
Detach empty syringe from pump
Use a P1000 to gently mix the bead solution (careful not to form air bubbles)
Load the syringe while gently agitating the bead solution; immediately invert the syringe to avoid clogging the needle (hold syringe with needle facing upwards)
Gently push the bead solution through the tubing, then connect tubing for bead solution to the microfluidic device
Finally, attach the syringe to the pump with the needle facing downwards (the magnet is still running and starts mixing immediately)
Set flow rate to 1.6 ml/h for cells and beads, 8 ml/h for the oil, and make sure the pump is set to the specific syringe type
Start the pumps in the order cells > beads > oil

<Note> Cells are started first to avoid a backflow of lysis buffer contained in the bead solution

<Troubleshooting> If you want to stop the run for whatever reason, the order is beads > cells > oil

Check droplets from time to time under the microscope by adding a few drops from the outlet tubing onto a glass slide.

<Note> About 5% of droplets should contain a bead. Cells are lysed within seconds and are therefore not visible.

Start collecting droplets in a 50 ml tube if the droplet size is homogenous

As the run progresses, the mixing disk will stop jumping because it gets blocked by the plunger (approximately when the plunger reaches the 0.4 ml marking). A typical run lasts for 35-40 min, filling the 50 ml tube to 7.5 – 10 ml. Make sure to retrieve the mixing disc from the bead syringe at the end.

Proceed with droplet breakage on the same day. When performing multiple runs per day, 50 ml tubes containing droplets can be stored at 4 °C and further processed together.

Droplet breakage

Remove as much oil below the droplet layer as possible

Add 30 ml of 6x SSC buffer

Add 1 ml of perfluorooctanol

Shake forcefully by hand 6 times to break the droplets

Bead capture

Label the top side of a 0.22 µm filter unit, remove plunger from a 20 ml syringe, attach syringe to 0.22 µm filter unit, place on a fresh 50 ml tube and pour the bead-containing sample into the syringe
Then place plunger back and filter as usual

Discard 50 ml tube with flowthrough and the syringe, but keep the filter containing the beads

<Note> The beads are retained in the 0.22 µm filter unit, due to their relatively large diameter

Washing beads

Place filter unit (top side up) on a fresh 50 ml tube, pour 6x SSC buffer into a reservoir, take up 20 ml from the reservoir with a fresh syringe, attach to filter unit on 50 ml tube and wash the beads

Repeat once for a total of two washes

Keep the wash fraction and filter, and discard the syringe

<Note> Beads are still retained in the filter unit, and only washed in the process

Eluting beads

Invert the filter unit and place it on a fresh 50 ml tube

Cut a 200 µl tip without filter on both ends, it will serve as an adapter to connect the bottom side of the 0.22 µm filter to a 10 ml syringe

Then elute beads three times with 10 ml of 6x SSC buffer

<Note> Beads have now been collected in the 50 ml tube in a total volume of 30 ml

Concentrating beads

Spin 50 ml tubes for wash and elution fractions for 2 min at 1,250 rcf with brake speed set to 50% (or use 'soft' setting if available)

<Note> Reducing the brake speed is critical, to prevent kicking beads back into solution

Compare pellets between wash and elution fraction, discard wash fraction unless there is a pronounced pellet as well

Remove supernatant from the elution fraction until only about 1 ml remains in the tube

Wash bead pellet once with 10 ml 6x SSC buffer, centrifuging with the above settings

Remove the supernatant carefully with a P1000 pipette

Take up the pellet in 200 µl of 5x reverse transcription (RT) buffer and transfer to a 1.5 ml tube

Spin 1.5 ml tube at 1,250 for 2 min with brake speed set to 50% (or 'soft' setting)

Remove as much of the RT buffer as possible without disturbing the beads

Immediately proceed with the reverse transcription

Preparation of cDNA library

Reverse transcription

Prepare reverse transcription master mix as indicated in the table below. The volumes provided already compensate for 10% dead volume.

Reverse transcription master mix									
Component	Stored at	Per reaction		Master mix for # of Drop-seq runs					
		Volume	Amount	1	2	3	4	5	6
Nuclease-free water	RT	75 µl	to 200 µl	82.5	165	247.5	330	412.5	495
Maxima 5x RT buffer	-20 °C	40 µl	1x	44	88	132	176	220	264
20% Ficoll PM-400	4 °C	40 µl	4%	44	88	132	176	220	264
10 mM dNTPs	-20 °C	20 µl	1 mM	22	44	66	88	110	132
Recombinant RNase inhibitor	-20 °C	5 µl	200 U	5.5	11	16.5	22	27.5	33

50 µM template switching oligonucleotide	-80 °C	10 µl	2.5 µM	11	22	33	44	55	66
Maxima H-RTase	-20 °C	10 µl	2000 U	11	22	33	44	55	66

Resuspend bead pellet(s) from the previous step in 200 µl of reverse transcription master mix

Transfer to a 96-well plate and incubate in a thermocycler with heated lid

30 min at 25 °C

90 min at 42 °C

Storage at 4 °C

<Safe stopping point> The finished reverse transcription reaction can be stored at 4 °C over night or left at 4 °C in the thermocycler

Exonuclease I treatment

Wash 1 (centrifuge 96-well plate for 2 min at 1,250 rcf with brake speed set to 50% throughout)

Wash beads once with 200 µl of TE-SDS

Wash beads twice with 200 µl of TE-TW

Wash beads once with 200 µl of 10 mM Tris, pH 8.0

Prepare exonuclease I master mix as indicated in the table below. The indicated volumes already account for 10% dead volume

Exonuclease I master mix									
Component	Stored at	/reaction		Master mix for # of Drop-seq runs					
		Volume	Amount	1	2	3	4	5	6
Nuclease-free water	RT	170 µl	to 200 µl	187	374	561	748	935	1122
10x Exonuclease I buffer	-20 °C	20 µl	1x	22	44	66	88	110	132
Exonuclease I	-20 °C	10 µl	200 U	11	22	33	44	55	66

Remove supernatant carefully, and resuspend the bead pellet in 200 µl of exonuclease I master mix

Incubate for 45 min at 37 °C

Wash 2 (centrifuge 96-well plate for 2 min at 1,250 rcf with brake speed set to 50% throughout)

Wash beads once with 200 µl of TE-SDS

Wash beads twice with 200 µl of TE-TW

<Safe stopping point> Sample can be stored in TE-WE at 4 °C

SMART PCR to enrich cDNA library

Pellet beads by centrifuging for 2 min at 1,250 rcf with the brake speed set to 50%

Remove supernatant and resuspend bead pellet in 400 µl of nuclease-free water

Estimate bead concentration

<Note> Drop-seq beads have a very high sedimentation rate, make sure to mix thoroughly immediately before pipetting bead-containing solution

Prepare PCR tube strip, containing 7.5 µl of 6x gel loading dye per well (one well per bead sample)

<Note> The gel loading dye helps to evenly distribute beads in the counting chamber later on

Mix bead solution by pipetting up and down with a P200

Quickly remove 37.5 µl of bead suspension, mix with the 6x gel loading dye in the tube strip and load 20 µl into each of the two chambers of a disposable Incyto C-Chip with Fuchs-Rosenthal grid
Count >4 squares for both chambers and calculate the average number of beads per square

Calculate total amount of beads per tube as

$[(\text{Average number of beads per square}) \times 5,000 \times 1.2] \times 0.3625 = x$ beads in original tube

where 5,000 is the chamber volume factor, 1.2 the dilution factor (gel loading dye) and multiplication by 0.3625 converts to the total amount of beads left in the original tube (now containing 362.5 µl)

Dilute bead solution to 220 beads per µl as follows

Pellet beads by centrifuging for 2 min at 1,250 rcf with the brake speed set to 50%

Remove supernatant and resuspend in $(x / 220) = y$ µl of nuclease-free water, where x is the total number of beads in the pellet, as calculated above

Calculate number of PCR reactions as

$y / (20+1) = n$ PCR reactions

where y is the volume used for diluting beads, 20 is the volume for 4,400 beads used in a single PCR reaction, and +1 accounts for 5% dead volume during pipetting

Prepare SMART PCR master mix as indicated in the table below

SMART PCR master mix			
Component	Stored at	/reaction Volume	Amount
Nuclease-free water	RT	4.6 µl	to 50 µl
100 µM SMART PCR primer	-20 °C	0.4 µl	0.8 µM
2x Kapa HiFi HotStart ReadyMix	-20 °C	25 µl	1x

Aliquot 20 µl of beads into 96-well plates

<Note> This corresponds to 4,400 beads or approximately 220 cells per PCR

Add 30 µl of SMART PCR master mix per well

Incubate PCR as follows

95 °C for 3 min

4x { 98 °C for 20 sec,
65 °C for 45 sec,
72 °C for 3 min }

10x { 98 °C for 20 sec,
67 °C for 20 sec,
72 °C for 3 min }

72 °C for 5 min

Storage at 4 °C

<Note> Depending on the cell type, the number of PCR cycles might need optimization. These cycle numbers have worked well on cell lines.

Purification, quality control and pooling of SMART-PCR enriched cDNA

Purification of SMART PCR products

<Note> Perform with multichannel pipettes in 96-well format

Add 30 µl of room temperature AMPure XP beads to each PCR reaction (0.6x bead to sample ratio)

Incubate for 10 min at room temperature

Place on magnet for 5 min

Wash twice with 80% ethanol (5 sec per wash)

Leave to dry for 5 min

Elute in 10 µl nuclease-free water

Run 1 µl of randomly selected wells on a Bioanalyzer High Sensitivity Chip according to the manufacturer's instructions. This should confirm high quality of the cDNA (peak maximum >1000 bp without bands that would indicate degradation), and low or undetectable amounts of primer dimer formation (peak at 150 bp).

<Troubleshooting> If a pronounced primer dimer peak is present at 150 bp, repeat the 0.6x AMPure XP bead clean-up one more time. Usually, this removes the peak entirely.

Measure the dsDNA concentration for all wells using a Qubit HS DNA assay

Based on these measurements, pool all SMART PCRs for a particular Drop-seq run in equal amounts

Measure the dsDNA concentration of each pool using a Qubit HS DNA assay, this concentration is needed for the Nextera XT library preparation.

<Safe stopping point> Enriched cDNA pools can be kept at 4 °C, or stored long-term at -20 °C

Nextera XT library preparation

Tagmentation

In a PCR tube strip, prepare 1 ng of pooled cDNA per Drop-seq run in a total volume of 5 µl of nuclease-free water

Preheat a thermocycler to 55 °C

Transfer Nextera XT Neutralization Buffer from storage at 4 °C to room temperature

Prepare Nextera XT tagmentation master mix as indicated in the table, the volumes are already accounting for 10% dead volume

Nextera XT tagmentation master mix			Master mix for # of Drop-seq runs					
Component	Stored at	Volume	1	2	3	4	5	6
Nextera TD buffer	-20 °C	10 µl	11	22	33	44	55	55
Amplicon tagment enzyme	-20 °C	5 µl	5.5	11	16.5	22	27.5	27.5

Add 15 µl of tagmentation master mix per well, and pipette up and down to mix

Incubate at 55 °C for 5 min

Immediately add 5 µl of Neutralization Buffer to stop the reaction

Incubate at room temperature for 5 min

Nextera XT PCR enrichment

Prepare Nextera XT PCR master mix as indicated in the table below. The volumes provided already account for 10% dead volume lost during pipetting.

Nextera XT PCR master mix								
Component	/ reaction		Master mix for # of Drop-seq runs					
	Stored at	Volume	1	2	3	4	5	6
Nuclease-free water	RT	8 µl	8.8	17.6	26.4	35.2	44	44
Nextera PCR mix	-20 °C	15 µl	16.5	33	49.5	66	82.5	82.5
10 µM New-P5-SMART PCR hybrid oligo	-20 °C	1 µl	1.1	2.2	3.3	4.4	5.5	5.5
10 µM Nextera N70X oligo	-20 °C	1 µl	1.1	2.2	3.3	4.4	5.5	5.5

Add 25 µl of Nextera XT PCR master mix to the tagmentation reaction, bringing the volume to 50 µl

Incubate PCR

95 °C for 30 sec

10x { 95 °C for 10 sec,
55 °C for 30 sec,
72 °C for 30 sec }

72 °C for 5 min

Storage at 4 °C

Purification of the final library

Clean-up 1 (perform with multichannel pipettes in the PCR tube strip)

Add 30 µl of room temperature AMPure XP beads to each PCR reaction (0.6x bead to sample ratio)

Incubate for 10 min at room temperature

Place on magnet for 5 min

Wash twice with 80% ethanol (5 sec per wash)

Leave to dry for 5 min

Elute in 32 µl of nuclease-free water, then transfer 30 µl to a fresh PCR tube strip

Clean-up 2

<Note> Both clean-up steps are required to remove PCR primers that might interfere with sequencing

Add 30 µl of room temperature AMPure XP beads to each PCR reaction (1.0x bead to sample ratio)

Incubate for 10 min at room temperature

Place on magnet for 5 min

Wash twice with 80% ethanol (5 sec per wash)

Leave to dry for 5 min

Elute in 10 µl of nuclease-free water, then transfer to 1.5 ml tubes for long-term storage at -20 °C

Library quality control and next-generation sequencing

Quality control and dilution of the CROP-seq library

Run 1 µl of the final library on a Bioanalyzer High Sensitivity DNA chip according to the manufacturer's instructions.

<Note> To increase the likelihood that fragments contain the gRNA sequence, CROP-seq uses a slightly larger library fragment size. This is achieved by adding more cDNA as input for the Nextera XT tagmentation reaction. Sometimes, the final library might appear a bit skewed towards larger fragments, which does not have a negative impact on the sequencing results.

Estimate the average size of the library from the Bioanalyzer profile.

<Note> For symmetrical libraries the peak maximum can be used; for skewed libraries the average fragment size is estimated based on the fragment distribution. We usually use 575 bp.

Measure the concentration of dsDNA in a Qubit DNA HS assay, according to the standard protocol

Calculate the molarity of your library as

$c(\text{dsDNA})[\text{ng}/\mu\text{l}] / (\text{average size} [\text{bp}] \times 660 \text{ g/mol}) \times 1,000,000 = \text{molarity of library} [\text{nM}]$

where 660 g/mol is the molecular weight of a DNA basepair and 1,000,000 is a conversion factor

Dilute your library with EB-0.01% Tween to reach a molarity of 4 nM

Next-generation sequencing on Illumina HiSeq 3000/4000 instruments

Drop-seq Custom Read1 Primer is spiked into the HP10 primer solution, located in column 11 of the cBot Reagent Plate at 1 µM final concentration. High sequence complexity needed for optimal base calling performance is achieved by adding 20-30% PhiX as spike-in. Cluster generation and Read 1 primer hybridization should be completed using the Illumina cBot protocol:

'HiSeq_3000_4000_HD_Exclusion_Amp_v1.0'.

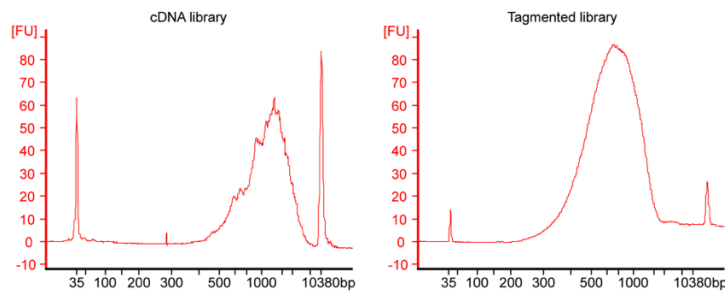
Next-generation sequencing on Illumina NextSeq 500 instruments

Libraries are sequenced on an Illumina NextSeq 500 instrument using the 75 cycle High Output v2 kit. Load 1.8 pM library without PhiX spike-in and provide Drop-seq Custom Read1 Primer at 0.3 µM in position 7 of the reagent cartridge. Set the read configuration to 20 bases (Custom Read1), 8 bases (Index1) and 64 bases (Read2).

<Note> When setting up the run in the Illumina Basespace, the read length of only 20 bases for Custom Read1 results in a warning. We ignore this warning and proceed with the run, because any additional bases would just read from the A-tail, thus reducing sequencing quality. However, when sequencing with this setup, the data has to be demultiplexed manually.

3.1.6. Expected results for single-cell RNA-seq by Drop-seq

Typical Bioanalyzer profiles for SMART-PCR enriched cDNA and final Nextera XT libraries



Measured cell doublet rate for the described Drop-seq setup, measured in a species mixing experiment on HEK293T (human) and 3T3 (mouse) cell lines. The mixing ratio was 1:1.

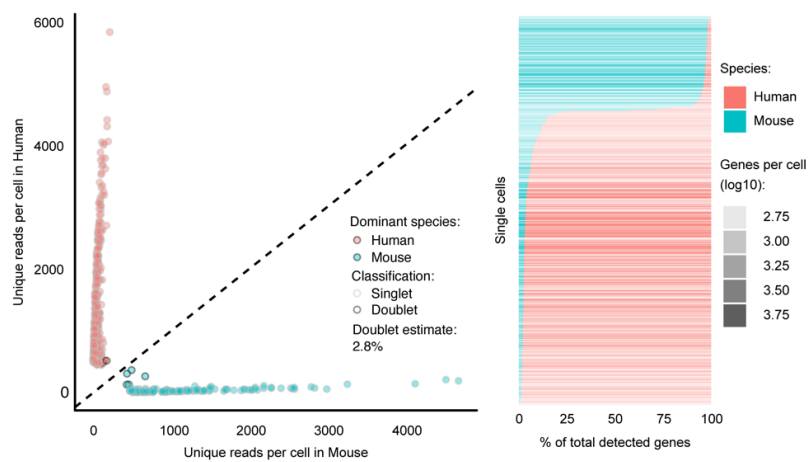


Table summarizing sequencing results for a typical NextSeq 500 run

LANE	READ	TILES	DENSITY (K / MM ²)	CLUSTER PF (%)	PHAS/PREPHAS (%)	READS	READS PF	%≥Q30
1	1	216	213 ±12	88.77 ±0.71	0.049 / 0.075	138,319,728	122,812,304	94.70
	2 (I)	216	213 ±12	88.77 ±0.71	0.000 / 0.000	138,319,728	122,812,304	81.42
	3	216	213 ±12	88.77 ±0.71	0.321 / 0.312	138,319,728	122,812,304	58.91
2	1	216	213 ±5	89.03 ±1.68	0.044 / 0.076	137,876,464	122,789,656	94.86
	2 (I)	216	213 ±5	89.03 ±1.68	0.000 / 0.000	137,876,464	122,789,656	81.99
	3	216	213 ±5	89.03 ±1.68	0.317 / 0.323	137,876,464	122,789,656	59.42
3	1	216	221 ±6	89.46 ±0.96	0.063 / 0.076	143,051,776	127,979,232	94.89
	2 (I)	216	221 ±6	89.46 ±0.96	0.000 / 0.000	143,051,776	127,979,232	83.66
	3	216	221 ±6	89.46 ±0.96	0.272 / 0.249	143,051,776	127,979,232	64.95
4	1	216	215 ±5	88.86 ±1.49	0.045 / 0.082	139,777,456	124,227,784	94.77
	2 (I)	216	215 ±5	88.86 ±1.49	0.000 / 0.000	139,777,456	124,227,784	81.89
	3	216	215 ±5	88.86 ±1.49	0.302 / 0.319	139,777,456	124,227,784	59.87

3.1.7. Materials

Reagents required for pooled, amplification-free cloning of gRNA libraries

<i>Reagent (alphabetical order)</i>	<i>Supplier</i>	<i>Cat. No.</i>	<i>Comment on use</i>
Absolute ethanol	Merck	64-17-5	
Agar, powder	Sigma	A5054-250G	
Agarose	Sigma	A9539-100G	
Agilent Bioanalyzer High Sensitivity DNA chip	Agilent	5067-4626	
AMPure XP beads	Beckman Coulter	A63880	
Bioassay dish, 245 x 245 x 25 mm	Sigma	D4803-1CS	Used for growing bacteria electroporated with gRNA libraries
BsmBI restriction enzyme	NEB	R0580S	
Carbenicillin, 5g	Roth	6344.2	
CROPseq-Guide-Puro	Addgene	86708	
EB buffer	Qiagen	1014609	
Electroporation cuvettes, 1 mm	BioRad	1652089	
Endo-Free Plasmid Mega kit	Qiagen	12381	For preparing plasmid containing entire gRNA libraries
Endura electrocompetent cells	Lucigen	60242-2	
Gel-cutting tips	Biostep	99-90-306	
GeneRuler 1 kb DNA Ladder	Thermo Fisher Scientific	SM0311	
LB (Miller), powder	Sigma	L3522-1KG	
Membrane filter	Merck	VMWP04700	For desalting Gibson assembly reactions
NEBuilder HiFi DNA assembly master mix	NEB	E2621S	
Plasmid Plus Midi kit, 100 preparations	Qiagen	12945	For CROPseq-Guide-Puro and lentivirus packaging plasmids
Q5 Hot Start High-Fidelity 2x Master Mix	NEB	M0494L	
Qubit dsDNA HS Assay kit	Thermo Fisher Scientific	Q32854	
Round-bottom tubes	Falcon	352051	
SapI restriction enzyme	NEB	R0569S	
SNAP UV-Free Gel Purification kit	Invitrogen	45-0105	
SOC medium	NEB	B9020S	
Sodium acetate, 3M, pH 5.5	Ambion	AM9740	
SphI restriction enzyme	NEB	R0182S	
SYBR Green; 10,000x concentrate in DMSO	Thermo Fisher Scientific	S7563	
TAE buffer, 50x	Invitrogen	24710-030	
Tween	Sigma	P7949-500ML	

Reagents required for lentivirus production and cell culture for CROP-seq screens

<i>Reagent (alphabetical order)</i>	<i>Supplier</i>	<i>Cat. No.</i>	<i>Comment on use</i>
Blasticidin	Invivogen	ant-bl-5	
Crystal violet	Sigma	61135-25G	To stain resistant colonies for titer estimation
DMEM	Gibco	10569010	
DPBS, 1x	Gibco	14190-094	
FCS	Sigma	N/A	
HEK293T cell line			
Lipofectamine 3000 transfection reagents	Invitrogen	L3000015	
Opti-MEM I	Gibco	51985-034	
Penicillin-streptomycin; 5,000 U/ml	Thermo Fisher Scientific	15070063	
pMD2.G	Addgene	12259	3rd generation lentiviral packaging system
pMDLg/pRRE	Addgene	12251	3rd generation lentiviral packaging system
Polybrene	Sigma	H9268-5G	To enhance contact between virus and cells
pRSV-Rev	Addgene	12253	3rd generation lentiviral packaging system
Puromycin	Thermo Fisher Scientific	A1113803	
Sodium pyruvate	Gibco	11360-070	
Syringe filter, 0.45 µm	VWR	514-8021	For removing cells from lentivirus preparations

Reagents required for single-cell RNA-seq based on Drop-seq

Modified from: Macosko et al. 2015 and additional protocols (<http://mccarrolllab.com/dropseq/>)

Reagent (alphabetical order)	Supplier	Cat. No.	Comment on use
Barcoded Bead SeqB	ChemGenes	MACOSKO-2011-10	Drop-seq beads
Bovine serum albumin (BSA)	Sigma	A7030-50G	
C-Chip Disposable Hemacytometers, Incyto, Pkg. of 50	Fisher Scientific	22-600-102	For estimating bead number
Cell strainer, 100 µm (for beads)	VWR	21008-950	
Cell strainer, 40 µm (for cells)	VWR	21008-949	
dNTP solution, 10 mM	ThermoFisher Scientific	R0192	
Droplet Generation Oil for EvaGreen, QX200	BioRad	1864006	Drop-seq oil
DTT	Sigma	646563-10x.5ML	Add on day of run
EDTA, 0.5 M, pH 8.0	ThermoFisher Scientific	AM92606	
Exonuclease I	NEB	M0293L	
Ficoll PM-400	Sigma	F5415-50ML	
KAPA HiFi HotStart ReadyMix	KAPA	KK2602	
Maxima H Minus Reverse Transcriptase, 4x 50 ul	ThermoFisher Scientific	EP0753	
Microfluidic devices	FlowJem	see order sheet for details	
N-Lauroylsarcosine sodium salt solution, 20%	Sigma	L7414	
Needle BD Microlance 3 (brown)	BD	300300	
Needle Neoject 25G 0.5 x 16 mm	Hartenstein	EU23	
Nextera XT kit for 24 samples	Illumina	FC-131-1024	
Perfluorooctanol (1H,1H,2H,2H-Perfluoro-1-octanol)	Sigma	370533-25G	For droplet breakage
Polyethylene tubing, ID 0.38 mm	Scicominc	BB31695-PE/2	
Recombinant Rnase inhibitor	Takara	2313A	
SSC buffer, 6x	Promega	V4261	For droplet breakage
Syringe BD Discardit II 10 ml	BD	309110	
Syringe BD Discardit II 20 ml	BD	300296	
Syringe BD Luer-Lok 1 ml	BD	309628	
Syringe BD Luer-Lok 3 ml	BD	309657	
Syringe filter unit, Millex-GV, 0.22 µm, PVDF	Merck Millipore	SLGV033RS	
Tris, 2M, pH 7.5	Sigma	T2944-100ML	
Tween-20	Sigma	P7949-500ML	

Oligonucleotide (order of use)	Supplier	Sequence	Comment
Template switching oligonucleotide (TSO)	Exiqon	AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG	r ... indicates RNA bases, store at -80 °C after resuspending
SMART PCR primer	Sigma	AAGCAGTGGTATCAACGCAGAGT	
New-P5-SMART PCR hybrid oligo	Sigma	AATGATACGCGACCCGAGATCTACACGCTGTCC GCGGAAGCAGTGGTATCAACGCAGAGT*A*C	
Custom Read1 primer	Sigma	GCCTGTCCGCGAAGCAGTGGTATCAACGCAGAGTAC	Custom sequencing primer for Read1 on Illumina machines

Solution (order of use)	Recipe
Drop-seq lysis buffer	For 40 ml (it's possible to store large stocks without DTT at 4 °C): 20 ml Nuclease-free water 12 ml 20% Ficoll PM-400 400 µl 20% N-Lauroylsarcosine sodium salt solution (Sarcosyl), <Caution> toxic 1.6 ml 0.5 M EDTA 4 ml 2 M Tris pH 7.5 50 µl per ml Lysis buffer 1 M DTT → add this on the day of the experiment
TE-SDS	10 mM Tris, pH 8.0 1 mM EDTA 0.5% SDS
TE-TW	10 mM Tris, pH 8.0 1mM EDTA 0.01% Tween

3.2. Step-by-step protocol for scifi-RNA-seq

Authors: Paul Datlinger (pdatlinger@cemm.oeaw.ac.at), Thorina Boenke and Christoph Bock (cbock@cemm.oeaw.ac.at)

Future updates of the protocol will be shared via <http://scifi-rna-seq.computational-epigenetics.org>

Preparations prior to the experiment

Prepare custom i7 transposome (Appendix A) and corresponding buffers (Appendix F)

For fixing cells, cool a 50 ml aliquot of 100% methanol at 4 °C

Cool a 500 ml bottle of 1x PBS at 4 °C

Thaw an aliquot of 10% BSA at 4 °C overnight in the fridge

Prepare 10% IGEPAL CA-630 solution (5 ml in 15 ml Falcon, vortex some minutes to dissolve, store in the dark at 4 °C)

Prepare plates with barcoded reverse transcription primers (1 µl of 25 µM per well, store at -20°)

For human/mouse mixing experiments, grow one T75 flask of Jurkat cells and four 15 cm dishes of 3T3 cells

Buffer preparations on the day of the experiment

For scifi-RNA-seq without fixation, use the tables in Appendix D to prepare:

PBS-BSA-SUPERase

Nuclei Preparation Buffer

Nuclei Wash Buffer

If following the nuclei fixation protocol, use the tables in Appendix E to prepare:

Nuclei Preparation Buffer without Digitonin and without Tween-20

Nuclei Wash Buffer without Tween-20

4% Formaldehyde in 1xPBS

PBS-BSA-SUPERase

3.2.1. Single-cell combinatorial fluidic indexing

Whole cell fixation for cell lines

<Note> Cool centrifuge to 4 °C and pellet cells at 300 rcf for 5 min throughout

Transfer 5 million cells in growth medium to a 15 ml tube and pellet

Wash once with 10 ml of ice-cold 1x PBS (resuspending carefully) and pellet

Fix cells in 5 ml ice-cold methanol at -20 °C for 10 min

Wash 2x with 5 ml ice-cold PBS-BSA-SUPERase

Resuspend cells in 200 µl of ice-cold PBS-BSA-SUPERase

Filter through cell strainer (40 µm or 70 µm depending on cell size)

Use 5 µl of the sample for cell counting (duplicates), dilute to 5000 cells/µl with ice-cold PBS-BSA-SUPERase

Proceed immediately to step 2 (reverse transcription)

Nuclei isolation for cell lines

Transfer 5 million cells in growth medium to a 15 ml tube and pellet (300 rcf, 5 min, room temperature)

Wash once with 10 ml of ice-cold 1x PBS and pellet (300 rcf, 5 min, 4 °C)

Resuspend in 500 µl ice-cold Nuclei Preparation Buffer

Incubate on ice for 5 min to lyse plasma membrane

Add 5 ml of ice-cold Nuclei Wash Buffer and pellet (500 rcf, 5 min, 4 °C)

Resuspend nuclei in 200 µl of ice-cold PBS-BSA-SUPERase

Filter through cell strainer (40 µm or 70 µm depending on cell size)

Use 5 µl of the sample for nuclei counting (duplicates), dilute to 5000 nuclei/µl with ice-cold PBS-BSA-SUPERase

Proceed immediately to step 2 (reverse transcription)

Nuclei isolation with formaldehyde fixation for primary cells

Transfer 5 million cells in growth medium to a 15 ml tube and pellet (300 rcf, 5 min, room temperature)

Wash once with 10 ml of ice-cold 1x PBS and pellet (300 rcf, 5 min, 4 °C)

Carefully resuspend in 500 µl ice-cold Nuclei Preparation Buffer without Digitonin and without Tween-20

Incubate on ice for 5 min to lyse plasma membrane

Add 5 ml of ice-cold Nuclei Wash Buffer without Tween-20 and pellet (500 rcf, 5 min, 4 °C)

Remove supernatant, resuspend in 5 ml of ice-cold 4% Formaldehyde in 1x PBS

Fix on ice for 15 min, then spin immediately (500 rcf, 5 min, 4 °C)

Remove supernatant (pellet is hard to see in this step)

Add 1.5 ml of ice-cold Nuclei Wash Buffer without Tween-20, mix and transfer to 1.5 ml tube
Spin (500 rcf, 5 min, 4 °C)

Wash one more time with 1.5 ml of ice-cold Nuclei Wash Buffer without Tween-20

Spin (500 rcf, 5 min, 4 °C)

Remove supernatant, carefully resuspend pellet in 200 µl of ice-cold Nuclei Wash Buffer without Tween-20

Snap-freeze in liquid nitrogen, then store at -80 °C <Safe Stopping Point>

Thaw frozen samples in a water bath at 37 °C for exactly 1 min, then immediately place on ice

Spin (500 rcf, 5 min, 4 °C)

Remove supernatant, and resuspend pellet in 250 µl of ice-cold Permeabilization Buffer, incubate for 5 min on ice

Add 250 µl of ice-cold Nuclei Wash Buffer without Tween-20 per sample

Spin (500 rcf, 5 min, 4 °C)

Wash with 250 µl of ice-cold Nuclei Wash Buffer without Tween-20 per sample

Spin (500 rcf, 5 min, 4 °C)

Take up in 100 µl of ice-cold PBS-BSA-SUPERase

Use 5 µl of the sample for nuclei counting (duplicates), dilute to 5000 nuclei/µl with ice-cold PBS-

BSA-SUPERase

Proceed immediately to step 2 (reverse transcription)

Reverse transcription (round 1 indexing)

<Note> Avoid harsh centrifugation, spin down plates for a maximum of 30 sec at 50 rcf

Set thermocycler to 55 °C (incubation mode, with heated lid set to 60 °C)

Thaw plate containing barcoded reverse transcription primers (1 µl of 25 µM per well) and place on ice

Add 10000 cells or nuclei (2 µl per well) to the pre-dispensed primer, record well assignments

Incubate for 5 min at 55 °C, then place on ice immediately (denaturation of RNA secondary structures)

Prepare Reverse Transcription Master Mix as described below and distribute into 8-tube strip on ice

Reverse Transcription Master Mix							
Component	Per reaction		Number of 96WP				384WP
	Volume (µl)	Amount	1	2	3	4	
Nuclease-free water	3	to 10 µl	330	660	990	1320	1416
RT buffer, 5x	2	1x	220	440	660	880	944
DTT, 100 mM	0.5	5 mM	55	110	165	220	236
dNTPs, 10 mM	0.5	0.5 mM	55	110	165	220	236
RNaseOUT RNase inhibitor (40 U/µl)	0.5	20 U	55	110	165	220	236
Maxima H Minus reverse transcriptase (200 U/µl)	0.5	100 U	55	110	165	220	236
Dispense into 12-tube strip for multichannel			62	123	185	246	250

Add 7 µl of Reverse Transcription Master Mix with a multichannel pipette and slowly pipette two times to mix

Seal and spin plate (maximum of 30 sec at 50 rcf)

Incubate (with heated lid set to 60 °C):

50 °C for 10 min

3 Cycles { 8 °C for 12 sec
 15 °C for 45 sec
 20 °C for 45 sec
 30 °C for 30 sec
 42 °C for 2 min
 50 °C for 3 min }

50 °C for 5 min

store at 4 °C

Pooling and cell counting

Cool two centrifuges suitable for 15 ml tubes and 1.5 ml tubes to 4 °C, thaw 10x Ampligase Buffer (Lucigen)

Remove 96-well plate(s) from thermocycler, place on ice

Add 40 µl of 1xPBS-1%BSA per well, then pool via a tube strip in a 15 ml tube on ice

Wash with an additional 50 µl of 1xPBS-1%BSA per well, and add to pool for better recovery
 Fill up to 15 ml with 1xPBS-1%BSA
 Pellet for 5 min at 500 rcf at 4 °C and carefully remove supernatant
 Slowly resuspend nuclei in 1.0 ml of 1x Ampligase Buffer (dilute Lucigen stock 1/10 with H2O)
 Filter through cell strainer (40 µm or 70 µm depending on cell size) into 1.5 ml tube
 Pellet for 5 min at 500 rcf at 4 °C and remove supernatant almost completely without disturbing the pellet

<Note> Aim for a volume smaller than 15 µl, which is the maximum input for the Thermocycling Ligation

Centrifuge for 30 sec at 500 rcf at 4 °C to bring the remaining liquid to the bottom of the tube
 Dilute sample 1:200 prior to cell counting (1 µl sample + 199 µl 1x Ampligase Buffer)
 Load 20 µl into a Fuchs Rosenthal counting chamber (Incyto DHC-F01)
 Count five squares per sample and calculate the cell concentration as follows

$$\text{Cells per } \mu\text{l} = [(\text{Cells in five squares} / 5) \times 200 (\text{dilution factor}) \times 5000 (\text{volume factor})] / 1000$$

Thermocycling ligation (round 2 indexing)

<Note> Some parts of this section were adapted from the 10x Genomics documentation: CG000168_ChromiumSingleCellATAC_ReagentKits_UserGuide_RevA.pdf

<Note> scifi-RNA-seq was tested for loading amounts between 383k and 2M nuclei per microfluidic channel

<Note> How many single-cell transcriptomes are recovered depends on sample quality and type

Transfer between 383k and 2M cells/nuclei to a tube-strip on ice
 Bring the volume to 15 µl with 1x Ampligase Buffer
 Thaw the following reagents from the 10x Chromium Single Cell ATAC kit:
 Gel Beads, Reducing Agent B
 Place Chromium Chip E in 10x Chip Holder and close
 For unused channels: add 75 µl (row 1), 40 µl (row 2) and 240 µl (row 3) of 50% glycerol solution
 Prepare Thermoligation Master Mix as described below, keep on ice

Thermoligation Master Mix									
Component	Per reaction Volume (µl)	1	2	3	4	5	6	7	8
Nuclease-free water	47.4	52.1	104.3	156.4	208.6	260.7	312.8	365.0	417.1
Ampligase Buffer, 10x	11.5	12.7	25.3	38.0	50.6	63.3	75.9	88.6	101.2
Ampligase Enzyme (100 U/µl)	2.3	2.5	5.1	7.6	10.1	12.7	15.2	17.7	20.2
Reducing Agent B	1.5	1.7	3.3	5.0	6.6	8.3	9.9	11.6	13.2
Bridge Oligo, 100 µM	2.3	2.5	5.1	7.6	10.1	12.7	15.2	17.7	20.2

Add 65 µl of Thermoligation Master Mix to each sample, bringing the volume to 80 µl
 Mix gently, then load 75 µl into Row 1 of the chip (bottom center, avoid air bubbles), wait > 30 sec to allow priming
 In the meantime, vortex 10x Gel Beads for 30 sec (use 10x Vortex Adapter) and bring contents to

bottom of the wells

Load 40 µl of Gel Beads into Row 2 of the chip (bottom center, avoid air bubbles)

Dispense 240 µl of Partitioning Oil into Row 3 of the chip (pipette along the wall of the inlet)

Attach gasket, aligning the notch with the top left-hand corner of the chip

Run chip in the 10x Chromium Controller (takes ~ 7 minutes, immediately proceed with next step)

Place 8-tube strip on ice

Remove chip from instrument, discard gasket and fold back chip holder lid to present wells at 45 degrees

Transfer 100 µl of the emulsion into the tube strip on ice (very slowly, should take 20 sec in both directions)

Incubate emulsion droplets in thermocycler (lid: 105 °C, volume: 100 µl) according to the program below:

12 Cycles { 95 °C for 30 sec
59 °C for 2 min }

store at 15 °C

During the incubation:

Thaw Cleanup Buffer for 10 min at 65 °C, check for precipitate, cool to RT

Equilibrate to room temperature: Reducing Agent B, Dynabeads MyOne Silane, SPRIselect beads

Proceed immediately with the cleanup steps

Silane bead cleanup

Add 125µl of Recovery Agent to each sample at room temperature

Invert ten times to mix, centrifuge briefly

<Note> This results in a pink oil phase at the bottom and a smaller aqueous phase containing the sample on top

Slowly remove 125 µl of the pink oil phase from the bottom of the tube, discard

Prepare Dynabead Cleanup Master Mix as described below, keep on ice

Dynabead Cleanup Mix									
Component	Per reaction Volume (µl)	Number of channels (+ 10% overhead)							
		1	2	3	4	5	6	7	8
Cleanup Buffer	182	182	400.4	600.6	800.8	1001.0	1201.2	1401.4	1601.6
Dynabeads MyOne SILANE	8	8	17.6	26.4	35.2	44.0	52.8	61.6	70.4
Reducing Agent B	5	5	11.0	16.5	22.0	27.5	33.0	38.5	44.0
Nuclease-free water	5	5	11.0	16.5	22.0	27.5	33.0	38.5	44.0

Add 200 µl of Dynabead Cleanup Mix per sample, pipet mix

Incubate for 10 min at room temperature

Prepare Elution Solution I as described below

Elution Solution I									
Component	Per reaction Volume (μl)	1	2	3	4	5	6	7	8
Buffer EB	49	49	107.8	161.7	215.6	269.5	323.4	377.3	431.2
10% Tween 20	0.5	0.5	1.1	1.65	2.2	2.75	3.3	3.85	4.4
Reducing Agent B	0.5	0.5	1.1	1.65	2.2	2.75	3.3	3.85	4.4

Place tube strip on magnetic separator (position: high) until supernatant clears

Remove and discard supernatant along with white precipitate at the bottom of the tube

Wash with 300 μl of 80% ethanol for 5 seconds, then wash with 200 μl of 80% ethanol for 5 seconds

Spin tube strip, place on magnetic separator (position: low) and remove residual ethanol

Remove from magnet and immediately add 40.5 μl Elution Solution I

<Note> Shear clumps with a thin pipette tip or needle

Incubate for 1 min at room temperature to elute sample from beads

Place on magnetic separator (position: low) and transfer 40 μl of each sample to a fresh tube strip

SPRI cleanup

Add 40 μl of SPRIselect beads (ratio 1.0x) per sample

Pipet mix and incubate for 5 min at room temperature to allow DNA binding to the beads

Place on magnetic separator (position: high) until supernatant clears

Wash twice with 200 μl of 80% ethanol

Spin tube strip briefly, then place on magnetic separator (position: low) and remove residual ethanol

Resuspend beads in 22 μl of EB Buffer, incubate 2 min at room temperature to elute

Place on magnetic separator (position: low) and transfer 20 μl of each sample to a fresh tube strip

Template switching

Prepare Template Switching Master Mix as described below

Template Switching Master Mix									
Component	Per reaction Volume (μl)	1	2	3	4	5	6	7	8
RT buffer, 5x	10	11	22	33	44	55	66	77	88
Ficoll PM-400, 20%	10	11	22	33	44	55	66	77	88
dNTPs, 10 mM	5	5.5	11	16.5	22	27.5	33	38.5	44
Recombinant Ribonuclease Inhibitor	1.25	1.38	2.75	4.13	5.5	6.88	8.25	9.63	11
Template Switching Oligo, 100 μM	1.25	1.38	2.75	4.13	5.5	6.88	8.25	9.63	11
Maxima H Minus reverse transcriptase (200 U/μl)	2.5	2.75	5.5	8.25	11	13.75	16.5	19.25	22

Add 30 μl of Template Switching Master Mix per sample, mix and incubate (heated lid: 60 °C):

25 °C for 30 min

42 °C for 90 min

store at 4 °C

<Safe stopping point> Samples can be left in the thermocycler overnight, and stored at 4 °C for several days

3.2.2. cDNA enrichment, library preparation and next-generation sequencing

SPRI cleanup

Add 50 µl of AMPure XP beads (ratio 1.0x) per sample
Pipet mix and incubate for 10 min at room temperature to allow DNA binding to the beads
Place on magnetic separator until supernatant clears
Wash three times with 100 µl of 80% ethanol
Remove residual ethanol by pipetting
Resuspend beads in 17 µl of EB Buffer, and incubate 2 min at room temperature to elute sample from beads
Place on magnetic separator and transfer 15 µl of each sample to a fresh tube strip

cDNA enrichment

Prepare cDNA Enrichment Master Mix as described below

cDNA Enrichment Master Mix									
Component	Per reaction Volume (µl)	Number of reactions (+ 10% overhead)							
		1	2	3	4	5	6	7	8
Nuclease-free water	33	36.3	72.6	108.9	145.2	181.5	217.8	254.1	290.4
NEBNext High Fidelity 2x Master Mix	50	55	110	165	220	275	330	385	440
Partial-P5 Primer, 100 µM	0.5	0.55	1.1	1.65	2.2	2.75	3.3	3.85	4.4
TSO Enrichment Primer, 100 µM	0.5	0.55	1.1	1.65	2.2	2.75	3.3	3.85	4.4
SYBR Green, 100x in DMSO	1	1.1	2.2	3.3	4.4	5.5	6.6	7.7	8.8

Add 85 µl of cDNA Enrichment Master Mix per sample, bringing the volume to 100 µl
Run reactions in qPCR cycler (heated lid: 105 °C)
<Note> When samples reach > 1000 RFU, pause run after the plate read and transfer samples to second thermocycler
98 °C for 30 sec
Cycle and stop when ready { 98 °C for 20 sec
65 °C for 30 sec
72 °C for 3 min, plate read }
72 °C for 5 min (in second thermocycler)
store at 4 °C

cDNA cleanup (0.8x, 0.6x) and quality control

Add 80 µl of AMPure XP beads (ratio 0.8x) per sample
Pipet mix and incubate for 10 min at room temperature to allow DNA binding to the beads
Place on magnetic separator until supernatant clears
Wash three times with 100 µl of 80% ethanol and remove residual ethanol by pipetting
Resuspend beads in 42 µl of EB Buffer, and incubate 2 min at room temperature to elute sample from beads
Place on magnetic separator and transfer 40 µl of each sample to a fresh tube strip

Add 60 µl of nuclease-free water to bring the sample volume to 100 µl
 Add 60 µl of AMPure XP beads (ratio 0.6x) per sample
 Pipet mix and incubate for 10 min at room temperature to allow DNA binding to the beads
 Place on magnetic separator until supernatant clears
 Wash three times with 100 µl of 80% ethanol and remove residual ethanol by pipetting
 Resuspend beads in 17 µl of EB Buffer, and incubate 2 min at room temperature to elute sample from beads
 Place on magnetic separator and transfer 15 µl of each sample to a fresh tube strip
 Use 1 µl of each sample to measure the concentration of double-stranded DNA with the Qubit HS assay
 Load 1.5 ng on the Bioanalyzer High Sensitivity DNA chip to check the size distribution of the cDNA
<Note> The size distribution of a typical cDNA library is shown in Appendix B, Figure 1a

Tagmentation with i7-only transposome

Dilute cDNA to 0.2 ng/µl with nuclease-free water
 Per tagmentation reaction, distribute 5 µl (1 ng) into a 96-well plate on ice
<Note> Tagment entire sample for maximum library complexity. Typically, 6-15 tagmentations of 1 ng are required
 Prepare Tagmentation Master Mix as described below

Tagmentation Master Mix			
Component	Per reaction [µl]	Number of reactions (+ 10% overhead)	x
Nuclease-free water	11.25		
Tn5 Reaction Buffer, 5x	5		
Dimethylformamide	2.5		
Freshly diluted i7-only transposome (1:4.5 in Tn5 Dilution Buffer)	1.25		

Add 20 µl of Tagmentation Master Mix per well and mix by pipetting
 Incubate for 10 min at 55 °C
 Place reactions on ice for 1 min
 Add 2.5 µl of 10x Stop Solution (Lucigen), mix by pipetting and incubate for 5 min at room temperature
 Add 22.5 µl of nuclease-free water per well
 Add 50 µl of AMPure XP beads (ratio 1.0x) per sample
 Pipet mix and incubate for 10 min at room temperature to allow DNA binding to the beads
 Place on magnetic separator until supernatant clears
 Wash three times with 100 µl of 80% ethanol
 Remove residual ethanol by pipetting
 Resuspend beads in 17 µl of EB Buffer, and incubate 2 min at room temperature to elute sample from beads
 Place on magnetic separator and transfer 15 µl of each sample to a fresh plate
 Add 5 µl of 10 µM barcoded P7 primer per well

<Note> If processing 1-2 samples that should be sequenced on the same flow cell, sequencing complexity of Index1 can be increased by using multiple P7 primers per sample

Prepare Enrichment PCR Master Mix as described below

Enrichment PCR Master Mix Component	Per reaction [μl]	Number of reactions (+ 10% overhead)	x
NEBNext High Fidelity 2x Master Mix	50		
Partial-P5 Primer, 100 μM	0.5		
SYBR Green, 100x in DMSO	1		
Nuclease-free water	28.5		

Primer name	Sequence	Index
Partial-P5	AATGATACGGCGACCACCGAGA	No index
P7-001	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG	TAAGGCGA
P7-002	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG	CGTACTAG
P7-003	CAAGCAGAAGACGGCATACGAGATTCTGCCTGTCTCGTGGGCTCGG	AGGCAGAA
P7-004	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG	TCCTGAGC
P7-005	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG	GGACTCCT
P7-006	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG	TAGGCATG
P7-007	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG	CTCTCTAC
P7-008	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG	CGAGGCTG

Add 80 μl of Enrichment PCR Master Mix per well and mix by pipetting

Run reactions in qPCR cycler (heated lid: 105 °C)

<Note> When samples reach > 4000 RFU, pause run after the plate read and transfer samples to second thermocycler

72 °C for 3 min (for end fill-in after tagmentation)

98 °C for 30 sec

Cycle and stop when ready { 98 °C for 10 sec
65 °C for 30 sec
72 °C for 30 sec, plate read }

72 °C for 5 min (in second thermocycler)

store at 4 °C

Library cleanup (0.7x, 0.8x) and quality control

Add 70 μl of AMPure XP beads to each PCR reaction (0.7x ratio)

Pipet mix, and incubate for 10 min at room temperature to allow DNA binding to the beads

Wash twice with 200 μl 80% ethanol, then remove residual ethanol

Resuspend bead pellets in 27 μl of EB Buffer, incubate for 2 min at room temperature to elute

Place on magnetic separator, transfer 25 μl of sample to a fresh plate

Pool all samples with the same P7 index in a 1.5 ml tube

Add AMPure XP beads at a ratio of 0.8x

Pipet mix, and incubate for 10 min at room temperature to allow DNA binding to the beads

Wash twice with 200 μl 80% ethanol, then remove residual ethanol

Resuspend bead pellets in 0.2 bead volumes, incubate for 2 min at room temperature to elute
 Place on magnetic separator, transfer sample to a fresh tube
 Measure final library concentration with the Qubit High-Sensitivity DNA assay
 Run 1.5 ng on a Bioanalyzer High Sensitivity DNA chip

<Note> The size distribution of a typical scifi-RNA-seq library is shown in Appendix B, Figure 1b

Next-generation sequencing on the Illumina NovaSeq 6000 platform

Dilute library to 2.0 nM with EB + 0.1% Tween-20 and submit for sequencing

The volume of 2.0 nM library required for loading differs between flow cells

<Note> Choose reagent kit depending on the number of cells in the library. The sequencing depth per cell will greatly influence the quality of single-cell transcriptomes. Aim for at least 10,000 reads per cell.

NovaSeq 6000 Flow Cell Type	Loading Volume [μl]
SP	100
S1	100
S2	150

Illumina NovaSeq 6000 Run Parameters	
Reagent Kits	NovaSeq 6000 SP Reagent Kit (100 Cycles), Illumina #20027464 NovaSeq 6000 S1 Reagent Kit (100 Cycles), Illumina #20012865 NovaSeq 6000 S2 Reagent Kit (100 Cycles), Illumina #20012862
Read Structure	Read 1: 21 bases Index 1 (i7): 8 bases Index 2 (i5): 16 bases Read 2: 78 bases

Appendix A: Preparation of the custom i7 transposome and activity check

Oligonucleotides for custom i7 transposome assembly	
Oligo ID	Oligo sequence [5' > 3']
Tn5-top_ME	[5Phos]CTGTCTCTTATACACATCT
Tn5-bottom_Read2N	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG
Functional elements	
5' phosphorylation (required for transposition mechanism)	
mosaic end for association with Tn5 transposase	
i7 primer binding site for enrichment PCR	

Anneal oligonucleotides

Prepare Annealing Buffer as described in Appendix F

Dissolve lyophilized oligonucleotides at 100 μ M in EB buffer

Set up annealing reactions according to the table below

Anneal oligonucleotides for transposome assembly		
Component	Per reaction	
	Volume	Amount
Tn5-top_ME, 100 μ M	22.5 μ l	45 μ M
Tn5-bottom_Read2N, 100 μ M	22.5 μ l	45 μ M
Annealing Buffer, 10x	5 μ l	1x

Anneal in thermocycler: 95 °C for 3 min, 70 °C for 3 min, ramp to 25 °C at a rate of 2 °C per minute

Add 180 μ l of nuclease-free water to the annealed oligonucleotides and mix

<Safe stopping point> The diluted oligo cassette can be aliquoted and frozen for future transposome assemblies

Load Tn5 transposase

Mix 20 μ l of the diluted oligo cassette from the previous step with 20 μ l of 100% glycerol

<Note> Glycerol is extremely viscous, use slow reverse pipetting

Add 10 μ l of EZ-Tn5 Transposase (Lucigen #TNP92110) and mix well by pipetting

Incubate for 30 min at 25 °C

Store at -20 °C until ready to use

Optional activity test of custom transposome in a qPCR assay

<Note> This optional test allows you to compare the activity of your custom i7-only Tn5 with that of the ATAC Enzyme provided by 10x Genomics, to better estimate the amount needed. Generally, a 1:4.5 dilution of custom Tn5 works well, but in case of problems this test is recommended.

Set up PCR reactions as indicated in the table below

PCR product for tagmentation test		
Component	Per reaction	
	Volume per 6.6 reactions	Amount per reaction
pUC19 plasmid (NEB, 50 pg/μl)	33 μl	250 pg
Q5 HotStart High-Fidelity Master Mix, 2x	165 μl	1x
pUC19-FWD (AAGTGCCACCTGACGTCTAAG), 100 μM	1.65 μl	500 nM
pUC19-REV (CAACAATTAATAGACTGGATGGAGGCGG), 100 μM	1.65 μl	500 nM
Nuclease-free water	128.7 μl	to 50 μl

Distribute 50 μl into six wells of a tube strip

Run PCR in a thermocycler:

98 °C for 30 sec

31 cycles { 98 °C for 10 sec
 68 °C for 30 sec
 72 °C for 1 min }

72 °C for 2 min, storage at 12 °C

To each 50 μl PCR reaction, add 6.25 μl of CutSmart Buffer (10x) and 6.25 μl of DpnI

Incubate at 37 °C for 1 hour, 80 °C for 20 min, storage at 12 °C

<Note> This step digests the template plasmid

Pool the six PCR reactions in a 2 ml tube, add 5 volumes (1500 μl) of PB Buffer (QiaQuick PCR Purification Kit), mix

Continue with QiaQuick PCR purification using two columns, elute in 30 μl each and pool

Measure Qubit HS of a 1/10 dilution and dilute the PCR product to 25 ng/μl with EB buffer

<Note> This typically results in about 400 μl of 25 ng/μl solution

Run 100 ng (4 μl) of PCR product on a 1% agarose gel containing ethidium bromide, next to 500 ng (1 μl) of GeneRuler 1 kb DNA Ladder

<Note> The size of the desired band is 1961 bp

Set up tagmentation reactions as indicated below

Tagmentation Test Master Mix		
Component	Per 3.3 reactions	
	Volume	Amount per reaction
pUC19 PCR product from above (25 ng/μl)	6.6 μl	50 ng
ATAC Buffer from 10x Genomics Single Cell ATAC Kit	23.1 μl	N/A

Transfer 9 μl of Tagmentation Test Master Mix to three wells of a tube strip on ice

Add nuclease-free water and Tn5 enzymes as indicated below

Tagmentation Test Conditions			
Component	Tagmentation		
	Positive Control	Negative Control	Test
ATAC Enzyme from 10x Genomics Single Cell ATAC Kit	3	-	-
Custom i7 transposome	-	-	6
Nuclease-free water	3	6	-

Incubate at 37 °C for 60 min, storage at 4 °C

Add 1.75 µl of 1% SDS solution and incubate at 70 °C for 10 min to dissociate Tn5 from the DNA

Dilute reactions 1/100 with EB buffer

Run qPCR reactions in triplicates, as indicated below

qPCR reactions for transposome activity check		
Component	Volume	Per reaction Amount
Tagmentation reaction (1/100 diluted)	2 µl	
GoTaq qPCR Master Mix, 2x	10 µl	1x
FWD primer (AAGTGCCACCTGACGTCTAAG), 100 µM	0.1 µl	500 nM
REV primer (CAACAATTAATAGACTGGATGGAGGCGG), 100 µM	0.1 µl	500 nM
Nuclease-free water	7.8 µl	to 20 µl

Run qPCR reactions in thermocycler:

95 °C for 2 min

40 cycles { 95 °C for 30 sec
 68 °C for 30 sec
 72 °C for 2 min + plate read }

stop when plateau is reached

Calculate tagmentation efficiency for each condition as:

$$\text{Tagmentation efficiency [\%]} = 100 / [2 ^ {(\text{average Ct}^{\text{Sample}} - \text{average Ct}^{\text{Negative Control}})}]$$

Appendix B: Expected results

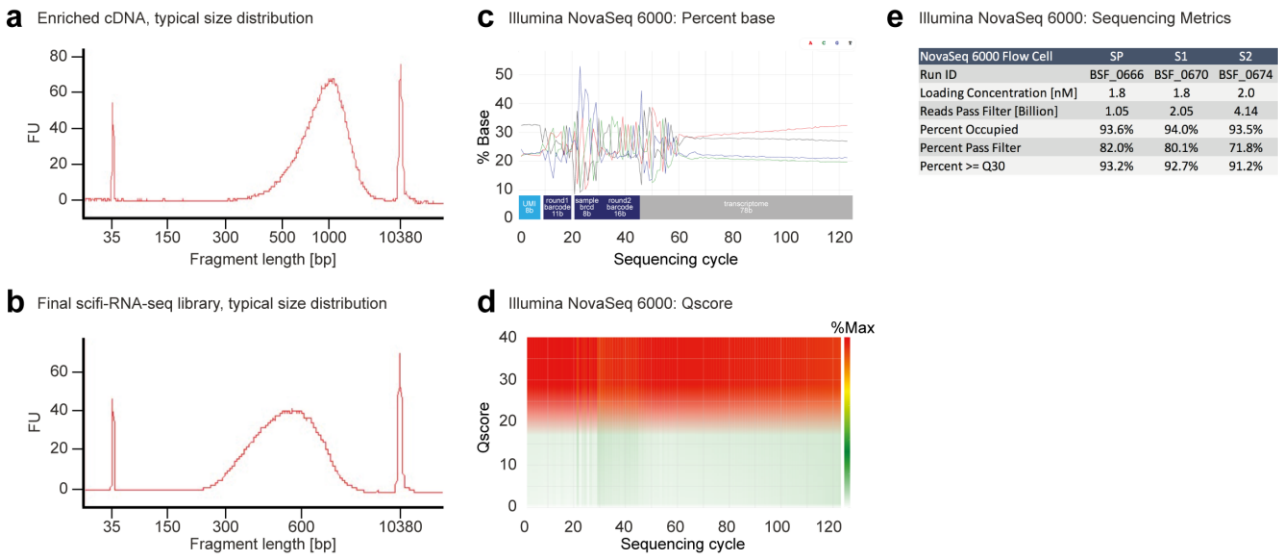


Figure 1 Expected results for scifi-RNA-seq. **a)** Typical size distribution of enriched cDNA measured on the Agilent Bioanalyzer. **b)** Final fragment distribution of a final scifi-RNA-seq library measured on the Agilent Bioanalyzer. **c-e)** Typical performance metrics on the Illumina NovaSeq 6000 platform.

Appendix C: Materials

Reagent (alphabetical order)	Supplier	Cat. No.	Comment on use
Agarose	Sigma	A9539-100G	custom i7-only Tn5 activity check
Agilent Bioanalyzer High Sensitivity DNA chip	Agilent	5067-4626	scifi-RNA-seq library prep
Ampligase Enzyme & Buffer, High Concentration (100 U/ul)	Lucigen	A0102K	scifi-RNA-seq library prep
AMPure XP beads	Beckman Coulter	A63880	scifi-RNA-seq library prep
BSA	Sigma	A8806-5	scifi-RNA-seq library prep
Cell strainer 40 um	A. Hartenstein	Z540	scifi-RNA-seq library prep
Cell strainer 70 um	A. Hartenstein	Z570	scifi-RNA-seq library prep
Chromium Single Cell ATAC Library & Gel Bead Kit, 16 rxns	10x Genomics	1000110	scifi-RNA-seq library prep
Chromium Single Cell E Chip Kit, 48 rxns	10x Genomics	1000082	scifi-RNA-seq library prep
Counting Chamber, Fuchs Rosenthal	Incyto	DHC-F01	scifi-RNA-seq library prep
Digitonin, 2%	Promega	G944A	scifi-RNA-seq library prep
Dimethylformamide	Sigma	D4551-250ML	scifi-RNA-seq library prep
dNTPs, 10 mM	Thermo Fisher Scientific	R0193	scifi-RNA-seq library prep
DpnI	NEB	R0176L	custom i7-only Tn5 activity check
DTT	Sigma	646563-10x.5ML	scifi-RNA-seq library prep
EB Buffer	Qiagen	1014609	scifi-RNA-seq library prep
EDTA, 0.5M	Invitrogen	AM9260G	custom i7-only Tn5 activity check
Ethanol, absolute	Merck	603-002-00-5	scifi-RNA-seq library prep
Ethidium bromide	AppliChem	A2273,0015	custom i7-only Tn5 activity check
EZ-Tn5 Transposase	Lucigen	TNP92110	scifi-RNA-seq library prep
Ficoll PM-400	Sigma	F5415-50ML	scifi-RNA-seq library prep
Formaldehyde, 16%	Thermo Fisher Scientific	28908	scifi-RNA-seq library prep
GeneRuler 1kb DNA Ladder	Thermo Fisher Scientific	SM0311	custom i7-only Tn5 activity check
Glycerol	Sigma	G5516-100ML	scifi-RNA-seq library prep
GoTaq qPCR Master Mix	Promega	A600A	custom i7-only Tn5 activity check
IGEPAL CA-630	Sigma	I8896-50ML	scifi-RNA-seq library prep
Maxima H Minus Reverse Transcriptase, 4x 10000 U	Thermo Fisher Scientific	EP0753	scifi-RNA-seq library prep
Methanol	Fisher Scientific	M/4000/17	scifi-RNA-seq library prep
MgCl ₂ , 1M	Ambion	AM9530G	scifi-RNA-seq library prep
NaCl, 5M	Sigma	S5150-1L	scifi-RNA-seq library prep
NaOH, 10M	Sigma	72068-100ML	scifi-RNA-seq library prep
NEBNext High Fidelity 2x Master Mix	NEB	M0541S	scifi-RNA-seq library prep
NovaSeq 6000 S1 Reagent Kit (100 Cycles)	Illumina	20012865	next-generation sequencing
NovaSeq 6000 S2 Reagent Kit (100 Cycles)	Illumina	20012862	next-generation sequencing
NovaSeq 6000 SP Reagent Kit (100 Cycles)	Illumina	20027464	next-generation sequencing
PBS, 1x	Gibco	14190-094	scifi-RNA-seq library prep
pUC19 plasmid	NEB	N3041A	custom i7-only Tn5 activity check
Puradisc 25 mm Polyethersulfone Syringe Filter, 0.2 um, sterile	GE Healthcare	6780-2502	scifi-RNA-seq library prep
Q5 HotStart High-Fidelity 2x Master Mix, 500 rxns	NEB	M0494L	custom i7-only Tn5 activity check
QiaQuick PCR Purification Kit	Qiagen	28106	custom i7-only Tn5 activity check
Qubit dsDNA HS Assay Kit	Thermo Fisher Scientific	Q32854	scifi-RNA-seq library prep
Recombinant RNase inhibitor	Takara	2313A	scifi-RNA-seq library prep
RNaseOUT RNase inhibitor	Invitrogen	10777019	scifi-RNA-seq library prep
SDS solution, 10% in water	Sigma	71736-100ML	custom i7-only Tn5 activity check
SPRIselect beads	Beckman Coulter	B23318	scifi-RNA-seq library prep
SUPERase In RNase Inhibitor	Thermo Fisher Scientific	AM2696	scifi-RNA-seq library prep
SYBR Green, 10000x in DMSO	Life Technologies	S7563	scifi-RNA-seq library prep, custom i7-only Tn5 activity check
TAPS	Sigma	T9659-100G	scifi-RNA-seq library prep
Tris-HCl, pH 7.5, 2M	Sigma	T2944-100ML	scifi-RNA-seq library prep
TritonX-100	Sigma	X100-100ML	scifi-RNA-seq library prep
Tween-20	Sigma	P7949-500ML	scifi-RNA-seq library prep

Appendix D: Buffers for isolation of nuclei or whole cells without fixation

PBS-BSA-SUPERase						
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei	Per 5 Mio Whole Cells
PBS, 1x					890	9890
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		100	1100
SUPERase In Rnase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		10	10

Nuclei Preparation Buffer					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
Nuclease-free water					470.25
Tris-HCl, pH 7.5, 2M	2 M		10 mM		2.75
NaCl, 5M	5 M		10 mM		1.1
MgCl ₂ , 1 M	1 M		3 mM		1.65
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		55
SUPERase In Rnase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		5.5
Tween-20	10 % (v/v)		0.1 % (v/v)		5.5
IGEPAL CA-630, 10% (v/v)	10 % (v/v)		0.1 % (v/v)		5.5
Digitonin	2 % (v/v)		0.01 % (v/v)		2.75

Nuclei Wash Buffer					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
Nuclease-free water					4785
Tris-HCl, pH 7.5, 2M	2 M		10 mM		27.5
NaCl, 5M	5 M		10 mM		11
MgCl ₂ , 1 M	1 M		3 mM		16.5
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		550
SUPERase In Rnase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		55
Tween-20	10 % (v/v)		0.1 % (v/v)		55

Appendix E: Buffers for isolation of nuclei with formaldehyde fixation and permeabilization

DAY1: Nuclei isolation, fixation and flash-freezing

Nuclei Preparation Buffer without Digitonin and without Tween-20					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
Nuclease-free water					4545.75
Tris-HCl, pH 7.5, 2M	2 M		10 mM		26.125
NaCl, 5M	5 M		10 mM		10.45
MgCl ₂ , 1 M	1 M		3 mM		15.675
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		522.5
SUPERase In RNase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		52.25
IGEPAL CA-630, 10% (v/v)	10 % (v/v)		0.1 % (v/v)		52.25

Nuclei Wash Buffer without Tween-20					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
Nuclease-free water					75407.2
Tris-HCl, pH 7.5, 2M	2 M		10 mM		428.45
NaCl, 5M	5 M		10 mM		171.38
MgCl ₂ , 1 M	1 M		3 mM		257.07
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		8569
SUPERase In RNase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		856.9

4% Formaldehyde in 1x PBS					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
PBS, 1x					39188
Formaldehyde, 16%	16 %		4 %		13063

PBS-BSA-SUPERase					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
PBS, 1x					890
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		100
SUPERase In RNase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		10

DAY2: Thawing, nuclei permeabilization

Nuclei Wash Buffer without Tween-20					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
Nuclease-free water					2992
Tris-HCl, pH 7.5, 2M	2 M		10 mM		17
NaCl, 5M	5 M		10 mM		6.8
MgCl ₂ , 1 M	1 M		3 mM		10.2
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		340
SUPERase In RNase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		34

Permeabilization Buffer					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
Nuclei Wash Buffer without Tween-20 (from above)					1083.5
Digitonin	2 %		0.01 %		5.5
Tween-20	10 %		0.1 %		11

PBS-BSA-SUPERase					
Component	Stock	Unit	Final	Unit	Per 5 Mio Nuclei
PBS, 1x					1958
BSA, 10% (w/v)	10 % (w/v)		1 % (w/v)		220
SUPERase In RNase Inhibitor (20 U/ul)	20 U/ul		0.2 U/ul		22

Appendix F: Buffers for transposome assembly and tagmentation

Annealing Buffer						
Component	Stock	Unit	Final	Unit	Per 1 ml	Unit
Nuclease-free water					875	μl
Tris-HCl, pH 7.5, 2M	2	M	10	mM	5	μl
NaCl, 5M	5	M	500	mM	100	μl
EDTA, 0.5M	0.5	M	10	mM	20	μl
Tn5 Reaction Buffer, 5x						
Component	Stock	Unit	Final	Unit	Per 100 ml	Unit
Nuclease-free water					87.5	ml
TAPS	powder		50	mM	1.216	g
MgCl ₂ , 1 M	1	M	25	mM	2.5	ml
adjust pH to 8.5 with 10 M NaOH, then fill up to 100 ml with nuclease-free water filter through 0.2 μm filter under a cell culture hood and aliquot						
Tn5 Dilution Buffer						
Component	Stock	Unit	Final	Unit	Per 10 ml	Unit
Tris-HCl, pH 7.5, 2M	2	M	50	mM	250	μl
NaCl, 5M	5	M	100	mM	200	μl
EDTA, 0.5M	0.5	M	0.1	mM	2	μl
Glycerol	100	%	50	%	5000	μl
Triton X-100	100	%	0.1	%	10	μl
Before using this buffer, add 1/100 volumes of fresh 0.1M DTT	0.1	M	1	mM		

Appendix G: Oligonucleotides in tubes

Oligonucleotide ID	Used for step	Oligonucleotide sequence [5' > 3']	Index sequence
Tn5-top_ME	Transposome Assembly	[5Phos] <u>CTGTCTCTTATACACATCT</u>	
Tn5-bottom_Read2N	Transposome Assembly	GTCTCGTGGGCTCGG <u>AGATGTGTATAAGAGACAG</u>	
pUC19-FWD	Transposome Activity Test	AAGTGCCACCTGACGTCTAAG	
pUC19-REV	Transposome Activity Test	CAACAATTAATAGACTGGATGGAGGCGG	
Bridge Oligo	Thermoligation Barcoding	CGTCGTGTAGGGAAGAGTGTGACGCTGCCGACGA[ddC]	
Template Switching Oligo	Template Switching	AAGCAGTGGTATCAACGCAGAGTGAATrGrGrG	
Partial-P5	cDNA Enrichment, Enrichment PCR	AATGATACGGCGACCAACCGAGA	
TSO Enrichment Primer	cDNA Enrichment	AAGCAGTGGTATCAACGCAGAGT	
P7-001	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCTCGTGGGCTCGG	TAAGGCGA
P7-002	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCTCGTGGGCTCGG	CGTACTAG
P7-003	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCTCGTGGGCTCGG	AGGCAGAA
P7-004	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCTCGTGGGCTCGG	TCCTGAGC
P7-005	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATAGGAGTCCGTCTCGTGGGCTCGG	GGACTCCT
P7-006	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATCATGCCTAGTCTCGTGGGCTCGG	TAGGCATG
P7-007	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATGTAGAGAGGTCTCGTGGGCTCGG	CTCTCTAC
P7-008	Enrichment PCR	CAAGCAGAAGACGGCATACGAGATCAGCCTCGGTCTCGTGGGCTCGG	CGAGGCTG

Modifications	
[5Phos]	5' phosphorylation
Underlined	Tn5 mosaic end
[ddC]	Dideoxycytosine to block 3'-extension
r	The following base is RNA

Appendix H: Set of 384 barcoded reverse transcription primers with fixed bases for error correction

plate_name	plate_well	oligo_ID	oligo_seq[5'>3']	index_ID	index_seq	fixed_base1_N	fixed_base2_V
SOFI_U384_plate1	A01	SOFI_U384_001	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAAGTGATAGCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_001	AGTGATTAGCA	A	A
SOFI_U384_plate1	A02	SOFI_U384_002	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAGAAATCCCCCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_002	GAATCCCCCTA	A	A
SOFI_U384_plate1	A03	SOFI_U384_003	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNACCTGGGAACTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_003	CTCGGGAACT	A	A
SOFI_U384_plate1	A04	SOFI_U384_004	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATACCTCCAGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_004	TACCTCCAGG	A	A
SOFI_U384_plate1	A05	SOFI_U384_005	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAATTTGTGGTATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_005	ATTTTGTGGTAT	A	A
SOFI_U384_plate1	A06	SOFI_U384_006	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNACCCGAGAGATCATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_006	CCGAGAGAGAT	A	A
SOFI_U384_plate1	A07	SOFI_U384_007	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAGAGTATAGGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_007	GAGTATAGGG	A	A
SOFI_U384_plate1	A08	SOFI_U384_008	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATCTTAATGAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_008	CTTAAATGAG	A	A
SOFI_U384_plate1	A09	SOFI_U384_009	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNACGTCTGTGCGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_009	ACGTCTGTGC	A	A
SOFI_U384_plate1	A10	SOFI_U384_010	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNACATATGGCGAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_010	CATAATGGCA	A	A
SOFI_U384_plate1	A11	SOFI_U384_011	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAGGCGGAACATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_011	GGCGGGAACAT	A	A
SOFI_U384_plate1	A12	SOFI_U384_012	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATCACTTCAAGAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_012	TCACCTCAAGA	A	A
SOFI_U384_plate1	B01	SOFI_U384_013	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAGGTGGCTCTCATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_013	GGCTGGCTCTC	A	A
SOFI_U384_plate1	B02	SOFI_U384_014	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATGACATTTGATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_014	CTGCACATTTG	A	A
SOFI_U384_plate1	B03	SOFI_U384_015	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAGTCTTCGGATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_015	GCTCTTCGAT	A	A
SOFI_U384_plate1	B04	SOFI_U384_016	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATTCGACCCGAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_016	TTTCGACCCGA	A	A
SOFI_U384_plate1	B05	SOFI_U384_017	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNACGTAAATACATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_017	CGGTAAATAC	A	A
SOFI_U384_plate1	B06	SOFI_U384_018	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATAGTCGATATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_018	GTAGTACGAT	A	A
SOFI_U384_plate1	B07	SOFI_U384_019	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATGAAGAATCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_019	TGTAAGAATCA	A	A
SOFI_U384_plate1	B08	SOFI_U384_020	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAGGCGCGCATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_020	AGGCGCGCGCT	A	A
SOFI_U384_plate1	B09	SOFI_U384_021	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNACCTCACGGAAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_021	CTCCACGGAAG	A	A
SOFI_U384_plate1	B10	SOFI_U384_022	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATGATTAACCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_022	GTTGTTGATTA	A	A
SOFI_U384_plate1	B11	SOFI_U384_023	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATAGAGCCGCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_023	TAGAGCCGCA	A	A
SOFI_U384_plate1	B12	SOFI_U384_024	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNAGAAATGGAGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_024	AGAAATGGAG	A	A
SOFI_U384_plate1	C01	SOFI_U384_025	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNATGAGCTGTGTTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_025	CAGGCTGTGTT	T	A
SOFI_U384_plate1	C02	SOFI_U384_026	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGTGACGAATAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_026	GTTGACGAATA	T	A
SOFI_U384_plate1	C03	SOFI_U384_027	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTGCATCAGGCAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_027	TGCATCAGGCG	T	A
SOFI_U384_plate1	C04	SOFI_U384_028	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAGCCATATGAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_028	AGCCATATGAG	T	A
SOFI_U384_plate1	C05	SOFI_U384_029	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCTGCCTCTGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_029	CTGCTGCTGTT	T	A
SOFI_U384_plate1	C06	SOFI_U384_030	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATAGCAATGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_030	GGTATAGCAAT	T	A
SOFI_U384_plate1	C07	SOFI_U384_031	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTAAAGTCATGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_031	TAAAGTCATG	T	A
SOFI_U384_plate1	C08	SOFI_U384_032	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAGAACTCAGAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_032	AGAACTCAGAA	T	A
SOFI_U384_plate1	C09	SOFI_U384_033	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCAGCTTAATCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_033	CAGCTTAATCG	T	C
SOFI_U384_plate1	C10	SOFI_U384_034	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTGGCTTAAGAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_034	GGCTTAAGAGA	T	C
SOFI_U384_plate1	C11	SOFI_U384_035	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCACCTGTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_035	TCACCTGTGCG	T	C
SOFI_U384_plate1	C12	SOFI_U384_036	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATAGTGATATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_036	ATAGTGATATC	T	C
SOFI_U384_plate1	D01	SOFI_U384_037	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCATCAGCAACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_037	CATCAGCAAC	T	C
SOFI_U384_plate1	D02	SOFI_U384_038	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCGCATGGGCACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_038	GGCGCATGGGA	T	C
SOFI_U384_plate1	D03	SOFI_U384_039	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTGTGAACCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_039	TGTTGAACCTC	T	C
SOFI_U384_plate1	D04	SOFI_U384_040	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATACGAGCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_040	ATATACGAGT	T	C
SOFI_U384_plate1	D05	SOFI_U384_041	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCACGACGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_041	CGCAGCAGTA	T	C
SOFI_U384_plate1	D06	SOFI_U384_042	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGGGACAGAAGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_042	GGGACAGAAGG	T	C
SOFI_U384_plate1	D07	SOFI_U384_043	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTAAACAAGTATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_043	TTAAACAAGTAT	T	C
SOFI_U384_plate1	D08	SOFI_U384_044	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTACTCTGGGACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_044	ACTCTGGGTA	T	C
SOFI_U384_plate1	D09	SOFI_U384_045	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCAGAAACCCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_045	CAGAAACCCCT	T	C
SOFI_U384_plate1	D10	SOFI_U384_046	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTACTACCCAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_046	CACACCCAGT	T	C
SOFI_U384_plate1	D11	SOFI_U384_047	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTATTATGCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_047	TATTATGCTCT	T	C
SOFI_U384_plate1	D12	SOFI_U384_048	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAGCAGTGTATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_048	AGCAGTGTAT	T	C
SOFI_U384_plate1	E01	SOFI_U384_049	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCTTGCAGAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_049	CTTTCGAGAG	C	C
SOFI_U384_plate1	E02	SOFI_U384_050	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGTAGATGCACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_050	GTTAGATGCA	C	C
SOFI_U384_plate1	E03	SOFI_U384_051	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAATACACATCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_051	TAAATACCAT	C	C
SOFI_U384_plate1	E04	SOFI_U384_052	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATTTTATGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_052	ATATTTTATG	C	C
SOFI_U384_plate1	E05	SOFI_U384_053	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCCCTAAAGCGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_053	CCCTAAAGCG	C	C
SOFI_U384_plate1	E06	SOFI_U384_054	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGATTTAGGACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_054	GATTTAGGAG	C	C
SOFI_U384_plate1	E07	SOFI_U384_055	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTACGAGTCTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_055	TACGAGTCTC	C	C
SOFI_U384_plate1	E08	SOFI_U384_056	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAGAACAAATGTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_056	AGAACAAATGT	C	C
SOFI_U384_plate1	E09	SOFI_U384_057	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGTGATCTCATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_057	CTTGATCTCAT	C	C
SOFI_U384_plate1	E10	SOFI_U384_058	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCCACGCGTACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_058	GCCACGCGTA	C	C
SOFI_U384_plate1	E11	SOFI_U384_059	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATAGATACGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_059	TATAGATACGG	C	C
SOFI_U384_plate1	E12	SOFI_U384_060	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATACGAGCACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_060	ATACCGAGCA	C	C
SOFI_U384_plate1	F01	SOFI_U384_061	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCCCGATGTACCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_061	CCCGATGTAC	C	C
SOFI_U384_plate1	F02	SOFI_U384_062	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGATCTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_062	GATCTGGCTT	C	C
SOFI_U384_plate1	F03	SOFI_U384_063	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAGCGCTTGTCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_063	TAGCGCTTGT	C	C
SOFI_U384_plate1	F04	SOFI_U384_064	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTAAATACAGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_064	CTAAATACAG	C	C
SOFI_U384_plate1	F05	SOFI_U384_065	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGGGTTTACAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_065	GGGTTTACCA	C	G
SOFI_U384_plate1	F06	SOFI_U384_066	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATCATTCATTTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_066	ATATCATTCG	C	G
SOFI_U384_plate1	F07	SOFI_U384_067	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCCCGAGAGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_067	CTCCGAGAGT	C	G
SOFI_U384_plate1	F08	SOFI_U384_068	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGAAATATCTCGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_068	GAATATCTCG	C	G
SOFI_U384_plate1	F09	SOFI_U384_069	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTACTACCCAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_069	TACTACCCAG	C	G
SOFI_U384_plate1	F10	SOFI_U384_070	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAAATGATACAAATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_070	AAATGATACAA	C	G
SOFI_U384_plate1	F11	SOFI_U384_071	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCCCATTTATTTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_071	CCCATTTATT	C	G
SOFI_U384_plate1	F12	SOFI_U384_072	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGAGTTCGAGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_072	GAGTTCGAGC	C	G
SOFI_U384_plate1	G01	SOFI_U384_073	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTGTGTGGCGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_073	TGTTGTGGCG	G	G
SOFI_U384_plate1	G02	SOFI_U384_074	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTCAAGGAAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_074	ATTCAGGAAGA	G	G
SOFI_U384_plate1	G03	SOFI_U384_075	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCAAGATCGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_075	CCAAAGATCTG	G	G
SOFI_U384_plate1	G04	SOFI_U384_076	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTATCTGTATACAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_076	ATATCTGTATCA	G	G
SOFI_U384_plate1	G05	SOFI_U384_077	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCTGTGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_077	CGATTTGCTG	G	G
SOFI_U384_plate1	G06	SOFI_U384_078	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCAACAGCTGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_078	TGCAACAGCTG	G	G
SOFI_U384_plate1	G07	SOFI_U384_079	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGGGCTGTCTATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_079	GGCTGTCTTAT	G	G
SOFI_U384_plate1	G08	SOFI_U384_080	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGACCTGGCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_080	ACCTGGCGTG	G	G
SOFI_U384_plate1	G09	SOFI_U384_081	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCGATCTATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_081	CGTACGCTTAT	G	G
SOFI_U384_plate1	G10	SOFI_U384_082	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTGCATTAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_082	TTGCACTTAAT	G	G
SOFI_U384_plate1	G11	SOFI_U384_083	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGGACAAAGAGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_083	GACAAAGAGAA	G	G
SOFI_U384_plate1	G12	SOFI_U384_084	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGACACAGCAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_084	ACACAGCAGC	G	G
SOFI_U384_plate1	H01	SOFI_U384_085	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTTATCTAGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_085	CTTATCTAGAG	G	G
SOFI_U384_plate1	H02	SOFI_U384_086	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAATGTGGTCAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_086	TAAATGTGGTCA	G	G
SOFI_U384_plate1	H03	SOFI_U384_087	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGGCGGAGCTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_087	GGCGGAGCTG	G	G
SOFI_U384_plate1	H04	SOFI_U384_088	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCATGTGTGTGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_088	TCCATGTGTG	G	G
SOFI_U384_plate1	H05	SOFI_U384_089	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGGAGGAATCAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_089	GGAGGAATCA	G	G
SOFI_U384_plate1	H06	SOFI_U384_090	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTCAAAATAGGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_090	ATCAAAATAGG	G	G
SOFI_U384_plate1	H07	SOFI_U384_091	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGACGAGCTGCAAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_091	CAAGAGTGCCA	G	G
SOFI_U384_plate1	H08	SOFI_U384_092	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTAGATGGACGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_092	TAGGATGGAC	G	G
SOFI_U384_plate1	H09	SOFI_U384_093	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGGACAGTATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_093	GATCGGACAT	G	G
SOFI_U384_plate1	H10	SOFI_U384_094	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGCAAGAAATGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_094	CCCAAGAAAT	G	G
SOFI_U384_plate1	H11	SOFI_U384_095	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTACCTAACGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_095	TACCTAACGCA	G	G
SOFI_U384_plate1	H12	SOFI_U384_096	[5'phos]JACACTCTTCCCTACACGAGCCTCTCCGATCTNNNNNNNTGAGTTGAGTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT	SASI_096	GTTAGTTGAG	G	G

113

114

115

4. DISCUSSION

Pooled CRISPR screens record the abundance of gRNAs, but cannot reveal more detailed cellular phenotypes. This limits them to simple readouts such as cell proliferation and survival, unless reliable screening reporters are developed to select relevant sub-populations. Due to their limited information content, pooled genetic screens suffer from high false-positive rates and require extensive follow-up experiments to validate potential hits. High-resolution phenotypes such as RNA-seq have previously been restricted to arrayed screens, which are severely limited by cost and labor and work best with cell lines that can be cultured indefinitely. Single-cell CRISPR screens constitute a new paradigm for functional genetics. They use individual cells as “micro-laboratories for genetic perturbations” (Wagner & Klein, 2017), and they allow phenotypes to be recorded as entire transcriptomes. This allows for reporter-free screening based on transcriptional signatures, which should be suitable for a wide range of phenotypes. Even for biological experiments without extensive prior knowledge, relevant screening signatures can be developed from non-targeted control cells included in the assay.

In the following section, I will compare CROP-seq to alternative technologies for single-cell CRISPR screening and discuss applications where single-cell CRISPR screening provides the most benefit over existing methods. Finally, I will discuss current limitations and how they could be addressed by future technology development.

4.1. Advantages of CROP-seq over competing technologies

CROP-seq was published as part of the first wave of methods combining pooled CRISPR screens with a single-cell RNA-seq readout. Alternative protocols include Perturb-seq (Adamson et al., 2016, Dixit et al., 2016) (jointly developed by the Jonathan Weissman and Aviv Regev labs), CRISP-seq (Jaitin et al., 2016) (Ido Amit lab) and Mosaic-seq (Xie et al., 2017) (Gary Hon lab). The distinguishing feature of CROP-seq is the direct detection of gRNA sequences from scRNA-seq data. All competing technologies

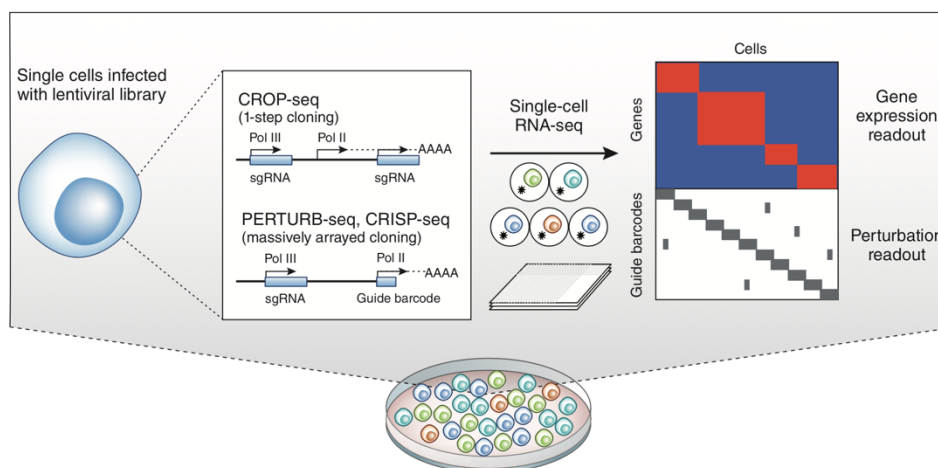


Figure 10 Comparison of CROP-seq with alternative methods. This figure was reprinted from Nature Methods with permission from the publisher (Wagner & Klein, 2017). CROP-seq has the unique feature of capturing the gRNA sequence in a polyadenylated transcript, such that it is directly detectable from single-cell RNA-seq data. In contrast, alternative methods rely on a co-expressed indexing transcript containing a guide barcode. Besides other advantages, this design decision makes CROP-seq compatible with standard one-step library cloning, while alternative technologies rely on arrayed cloning and lentivirus production to avoid recombination issues.

rely on the indirect detection of polyadenylated indexing transcripts containing guide barcodes (GBCs). This fundamental difference in the method design is illustrated in Figure 10.

One disadvantage of the indirect detection approach is that it necessitates a two-step, arrayed cloning procedure for the CRISPR library. GBCs and gRNA sequences are cloned in succession, and the link between a gRNA and its GBC has to be established for example by Sanger sequencing. While arrayed library cloning might still be feasible for single-cell CRISPR screens with limited cell numbers, the approach is inelegant, labor-intensive and reduces the library complexity. Inevitably, Perturb-seq, CRISP-seq and Mosaic-seq will run into a major bottleneck as screens get larger. In contrast, CROP-seq provides full compatibility with established one-step protocols for pooled library cloning, and it even allows researchers to re-use existing oligonucleotide pools from established libraries. In combination with new technologies for ultra-high throughput scRNA-seq, such as combinatorial indexing and scifi-RNA-seq, CROP-seq can easily scale to gene-family-wide or even genome-wide libraries without changes to the protocol.

An even more concerning problem with the Perturb-seq, CRISP-seq and Mosaic-seq designs is the pronounced recombination between the gRNA and GBC, with up to 50% re-shuffling of barcode to perturbation associations (Hill et al., 2018). The problem is rooted in lentivirus biology. Lentiviral particles are pseudo-diploid, meaning that one virion contains two separate copies of the single-stranded RNA genome, even though only one provirus ultimately integrates into the host genome.

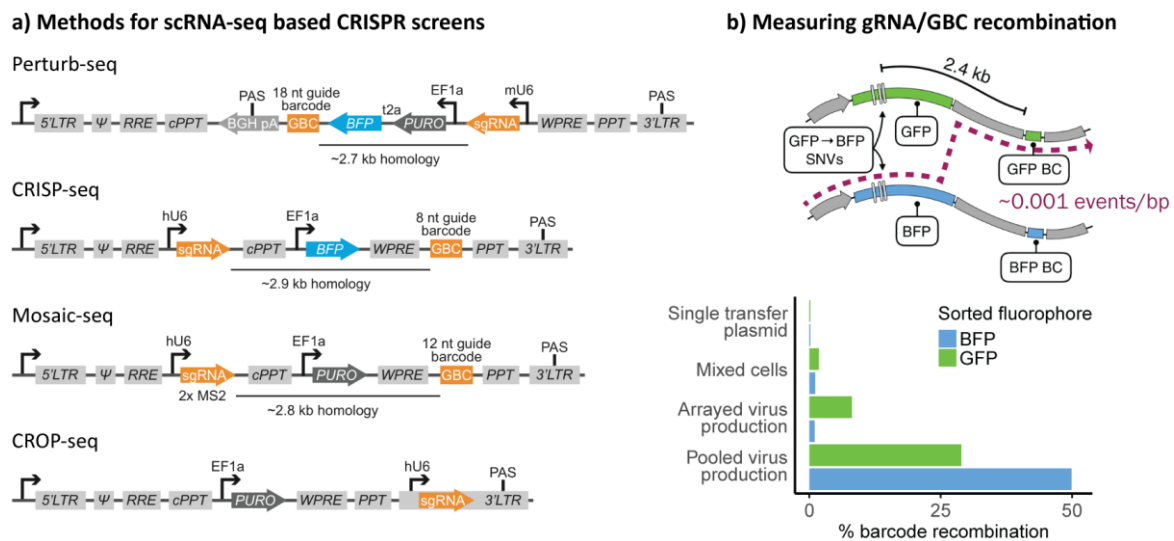


Figure 11 CROP-seq avoids lentiviral switching of gRNA to GBC associations. a) Perturb-seq, CRISP-seq and Mosaic-seq rely on the detection of indexing transcripts containing Guide Barcodes (GBCs), co-expressed with a pre-determined gRNA. CROP-seq can detect gRNAs directly from single-cell RNA-seq data. Panel a was reprinted under the CC-BY-NC-ND 4.0 International License (Adamson et al., 2018). b) Lentiviral particles are pseudo-diploid. When produced from a mixture of gRNA plasmids, they contain two heterotypic copies of the single-stranded RNA genome. Template switching during lentiviral reverse transcription leads to high rates of shuffling of gRNA to GBC associations. CROP-seq avoids this issue altogether, and is the only method fully compatible with a pooled workflow. Hill et al. measured the recombination rate using two lentiviral constructs in which a fluorophore was co-expressed with a pre-determined DNA barcode. Panel b was modified and data was re-plotted as percent barcode recombination for clarity (Hill et al., 2018).

When a library of gRNA-expressing transfer plasmids is supplied during the lentivirus production, virtually all of the resulting lentiviral particles will contain two heterotypic genome copies. When a lentiviral particle successfully transduces a cell, the contained reverse transcriptase associates with the RNA genome to synthesize a DNA copy that is later integrated into the host genome. During the reverse transcription of the minus strand DNA, template switching occurs at a rate of 1 event per kilobase. This can result in shuffling of the GBC to perturbation associations at a rate proportional to the distance between the pair of sequences.

Figure 11 provides detailed schematics of the four method designs. In Perturb-seq, CRISP-seq and Mosaic-seq, the distance between the gRNA and its associated barcode is between 2.7 and 2.9 kb, which is expected to result in the maximum possible swap rate of 50%. Indeed, this could be experimentally confirmed in an elegantly designed experiment (Hill et al., 2018): two separate lentiviral transfer plasmids expressing either GFP or BFP, separated from a pre-determined DNA barcode by a distance of 2.4 kb, were constructed and used for lentivirus production. Since the fluorescent proteins differ only by three base substitutions, the region had almost perfect homology. When lentivirus was produced in a pooled fashion from a mixture of both plasmids, up to 50% of fluorophore-positive cells carried the wrong barcode. However, producing the lentiviral particles in an arrayed format did not result in recombination. Consequently, the original papers describing Perturb-seq, CRISP-seq and Mosaic-seq avoided the issue by performing the lentivirus production, transduction and selection of target cells in an arrayed format and combining cells with different perturbations right before the scRNA-seq analysis. This process does not scale to larger CRISPR screens. In contrast, CROP-seq avoids the recombination problem altogether by reading the gRNA sequence directly and was applied in a pooled fashion throughout our experiments.

An updated version of Perturb-seq has been described that minimizes lentiviral recombination through decoy transfer plasmids without homology to the screening vector (Adamson et al., 2018). While quite effective, this method reduces the viral titer by about ten-fold, which severely limits its applicability to large screens or primary material. Another implementation of Perturb-seq, which was jointly developed with the biotechnology company 10x Genomics, abandons the original method design altogether. Instead, gRNAs are captured with specially designed reverse transcription primers that are used in addition to the standard poly-T primer targeting the transcriptome. Unfortunately, this locks the method into the 10x Genomics ecosystem and does not provide compatibility with any other single-cell technology platform. Furthermore, the protocol largely relies on artificial capture sequences, which are incorporated into the gRNA backbone and necessitate a compatible gRNA expression vector. Since capture sequences are placed at key locations of the gRNA backbone, the new gRNA enrichment method is not compatible with many interesting screening modes that rely on the recruitment of modifier proteins via RNA aptamers, such as CRISPR activation or epigenetic editing.

In summary, CROP-seq has the unique feature of direct gRNA detection from scRNA-seq data. This facilitates the cloning of large gRNA libraries, and maintains compatibility with already designed oligonucleotide pools that might be available in labs specializing in genetic screening. CROP-seq is the only method that provides full compatibility with pooled screening, without resorting to arrayed lentivirus production and transduction. Due to its unique design, CROP-seq avoids the gRNA/GBC shuffling issues associated with alternative methods. These practical and conceptual advantages have led to the widespread adoption of CROP-seq, as highlighted in Figure 12.

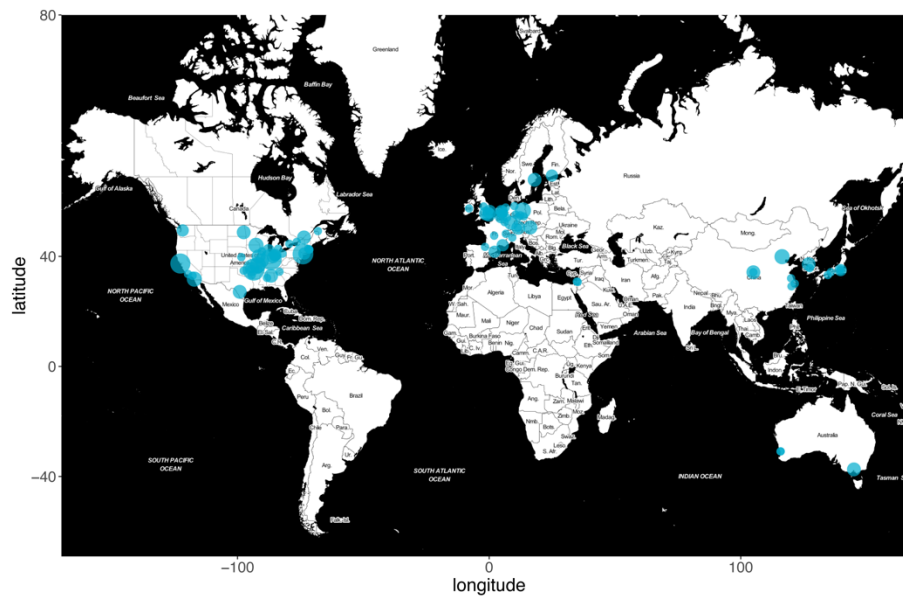


Figure 12 CROP-seq has been adopted by labs world-wide. The map highlights 195 labs that have ordered CROP-seq reagents from the plasmid distributor Addgene at the time of writing.

4.2. The value of single-cell resolution in CRISPR screens

Due to the high cost of single-cell analyses and next-generation sequencing, we cannot yet realize the full potential of single-cell CRISPR screens. Today, they are often used as a secondary line of screening, for instance to follow up on hits from a genome-wide CRISPR screen (Adamson et al., 2016). As the cost for a single-cell transcriptome drops even further, it seems realistic that pooled CRISPR screens with scRNA-seq readout will eventually become the primary method of CRISPR screening. However, single-cell CRISPR screens are not just a high-throughput alternative to arrayed platforms. Even with a limited number of perturbations, they can offer unique insights for biological questions where single-cell resolution matters.

For example, single-cell technologies proved particularly useful for the analysis and deconvolution of heterogeneous cell populations. Likewise, technologies such as CROP-seq are promising approaches to pinpoint the effect of a particular perturbation to various cellular subsets in a heterogeneous mixture. Indeed, most biological problems cannot be faithfully modelled by cell lines. Our initial CROP-seq screens were done in Jurkat cells, an immortalized human cell line established from a patient with T cell leukemia in 1977 (Schneider et al., 1977). Using T cell receptor activation as our screening signature, we could identify most but not all of the expected gene interactions. It seems that already when screening for a relatively simple gene expression signature, the results can be confounded by pre-existing mutations and signaling deficiencies of the investigated cell line (Gioia et al., 2018). Going forward, we would like to study more complex behaviors of human T cells, such as T-cell mediated cytotoxicity of CD8⁺ T cells, or the profound specification of CD4⁺ T helper cell subsets. To ask these questions, we need to consider the complexity of the human T cell repertoire, with diverse naive, effector and memory subsets, which will be greatly facilitated by single-cell CRISPR screens.

Another research field revolutionized by single-cell methods is the study of development at the cellular, organ or even whole organism level. Differentiation is a continuous process, but not necessarily perfectly synchronized between cells. Therefore, a large population of cells sampled by scRNA-seq will

contain single cells at multiple stages of the differentiation process, and the proportion of cells corresponds to the duration of a particular stage. The single-cell data can then be projected into a lower-dimensional space, and developmental trajectories can be derived by computational methods (Qiu et al., 2017). In a screening setting, this allows the unbiased and continuous classification of measured single cell phenotypes along the differentiation path, instead of relying on binary selection based on a marker protein. Developmental single-cell CRISPR screens are particularly interesting for processes where the exact outcome is unclear and marker proteins fail to achieve a fully representative readout. This is often the case in cellular transdifferentiation and reprogramming, where single-cell studies have proven useful to uncover competing cell fates that are limiting the conversion efficiency (Treutlein et al., 2016). Technologies such as CROP-seq can introduce defined genetic perturbations into these systems, and can accurately describe the identity of cells based on single-cell transcriptomes. For instance, the induced differentiation process could be computationally reconstructed to assess the enrichment of gRNAs in certain developmental stages.

Single-cell CRISPR screening is especially useful when little is known about the biological system and the induced response. Cells with non-targeting control gRNAs can be used as an internal reference to derive a relevant signature for screening. For example, in our CROP-seq screen in Jurkat cells, we were able to derive a relevant T-cell activation signature purely from the scRNA-seq data, using a principle component analysis of unstimulated and stimulated control cells. This makes the assay self-calibrating, and it allows researchers to screen for multiple gene expression signatures defined *in silico* that might represent different aspects of the biological process (for instance T cell proliferation, activation, differentiation into subsets and exhaustion due to overstimulation with antigen). As our scientific knowledge on a topic grows, it is even possible to revisit old screens with an updated screening signature without repeating the experiment in the lab.

Another especially interesting application is the combination of genome-wide, positive selection screens with a single-cell transcriptome readout. Positive selection screening has been introduced in section 1.1.7 as an effective way to identify acquired resistance mechanisms to cancer drugs. Genome-wide CRISPR activation could be used to induce resistance in a small subset of the perturbed cells, and scRNA-seq could provide a full transcriptome signature for each of the screening hits. This allows researchers to filter out false positive hits and uncovers the exact molecular mechanism of resistance, such as activation of an alternative signaling pathway for survival.

Single-cell CRISPR screening is also particularly relevant for rare cells that cannot be cultured easily. The high capture efficiency of single-cell RNA-seq ensures that most of the material is effectively converted into a sequencing library, and the detailed whole-transcriptome readout can get away with a lower screening coverage when performing screens on scarce primary material.

Another interesting perspective that only single-cell technologies can provide is CRISPR screening with a multi-omics readout from the same cell. For instance, RNA and protein expression have been quantified from the same single cell by methods such as CITE-seq (Stoeckius et al., 2017), REAP-seq (Peterson et al., 2017) and Abseq (Shahi et al., 2017). More recently, a combined scRNA-seq and scATAC-seq readout has been achieved (Cao et al., 2018). A method called Perturb-ATAC combined CRISPR perturbations with single-cell ATAC-seq (Rubin et al., 2019), although at a very low throughput

using a small integrated fluidic circuit. Taken together, these new developments could open up new and exciting possibilities for screening at the single-cell level.

4.3. Current limitations of single-cell CRISPR screens

One limitation of CRISPR screens in general is that they are still largely applied to cell line models. The efficiency of single-cell RNA-seq protocols to process scarce material, along with the complexity of the transcriptome readout, promise screens in smaller cell populations such as those derived from primary material or patient samples. However, the delivery of the components required for CRISPR genome editing into these cells is challenging. One promising approach is to derive cells from a Cas9-expressing mouse model (Jaitin et al., 2016). Shortening the gRNA spacer sequence to 15 nucleotides results in *dead RNAs* (or *dRNAs*), which still bind to the target locus but lose their ability to cut the genome. This approach allows to suppress cutting in the SpCas9 mouse, and introduce other functions such as CRISPR activation (Liao et al., 2017). Screening in primary human cells is even more challenging, and will require the delivery of CRISPR components as protein (Shifrut et al., 2018) or mRNA to achieve efficient editing. Comparing the two approaches, mRNA seems the preferred choice because it can be produced cheaply and in large quantities with one standardized in vitro transcription protocol. By cloning the desired open-reading frames into a construct containing a T7 promoter, 3' and 5' untranslated regions for transcript stability and a hard-coded poly-A tail, this approach can be extended to a wide range of desired genome editing functions such as CRISPR activation, enhancer editing, random mutagenesis, base editing and many more.

Another major limitation is the current throughput and high cost of scRNA-seq protocols, which limits the number of perturbations that can be screened. Two scenarios seem particularly interesting. Targeting entire gene families (epigenetic modifiers, transcription factors, kinases) at very high single-cell coverage in the range of 1,000 cells per gene can reveal fine details in the associated transcriptome signatures to fully characterize the effect of each perturbation. Given that a typical gene family has between 500 and 1,000 members, this would require the profiling of 500,000 to 1,000,000 cells. On the upper end of the throughput requirements, we could think about rudimentary genome-wide single-cell screens at a limited coverage of 100 cells per gene. For human cells with their roughly 20,000 genes (Willyard, 2018), this would require a throughput of at least 2 million profiled cells. There is reason for optimism that disruptive technologies such as combinatorial indexing and scifi-RNA-seq will help reach these ambitious goals in the coming years.

One CROP-seq specific disadvantage is that the cloning capacity for combinatorial perturbations is limited. As already mentioned, CROP-seq places the gRNA expression cassette in the 3' LTR to achieve a direct readout of gRNAs. Large inserts, such as dual or triple gRNA expression cassettes, might render the LTR nonfunctional and could result in low viral titers. While CROP-seq can still achieve combinatorial perturbations by superinfection of cells with multiple lentiviral particles, the multiplicity of infection cannot be tightly controlled and leads to the sampling of unwanted single perturbations and higher order interactions.

Finally, since CROP-seq, Perturb-seq, CRISP-seq and Mosaic-seq are all based on scRNA-seq in microfluidic droplet generators, they require the dissociation of the material into a single-cell suspension.

Thus, they are limited to cell-autonomous processes. With pooled optical screens, introduced in section 1.2.1, there is an emerging complementary platform to assess important biological phenomena such as cell-cell interactions, cell morphology and other spatial phenotypes.

In summary, through our work on CROP-seq, we have demonstrated that CRISPR screening at the single-cell level is feasible, and that it provides unique advantages over existing methods. As the throughput of scRNA-seq continues to grow, so will the potential of single-cell CRISPR screens. With scifi-RNA-seq, we provide a promising method to scale CROP-seq screens to whole gene families and eventually to the genome-wide level. Technologies such as CROP-seq have the potential to revolutionize functional genetic screening, and will enable many exciting applications in the future.

5. REFERENCES

- 1000 Genomes Project Consortium (2015) A global reference for human genetic variation. *Nature* 526: 68-74
- Abudayyeh OO, Gootenberg JS, Essletzbichler P, Han S, Joung J, Belanto JJ, Verdine V, Cox DBT, Kellner MJ, Regev A, Lander ES, Voytas DF, Ting AY, Zhang F (2017) RNA targeting with CRISPR-Cas13. *Nature* 550: 280-284
- Adamson B, Norman TM, Jost M, Cho MY, Nunez JK, Chen Y, Villalta JE, Gilbert LA, Horlbeck MA, Hein MY, Pak RA, Gray AN, Gross CA, Dixit A, Parnas O, Regev A, Weissman JS (2016) A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell* 167: 1867-1882 e21
- Adamson B, Norman TM, Jost M, Weissman JS (2018) Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. *bioRxiv*
- Barrangou R, Fremaux C, Deveau H, Richards M, Boyaval P, Moineau S, Romero DA, Horvath P (2007) CRISPR provides acquired resistance against viruses in prokaryotes. *Science* 315: 1709-1712
- Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, Vitale L, Pelleri MC, Tassani S, Piva F, Perez-Amodio S, Strippoli P, Canaider S (2013) An estimation of the number of cells in the human body. *Ann Hum Biol* 40: 463-471
- Black JB, Adler AF, Wang HG, D'Ippolito AM, Hutchinson HA, Reddy TE, Pitt GS, Leong KW, Gersbach CA (2016) Targeted Epigenetic Remodeling of Endogenous Loci by CRISPR/Cas9-Based Transcriptional Activators Directly Converts Fibroblasts to Neuronal Cells. *Cell Stem Cell* 19: 406-414
- Blomen VA, Majek P, Jae LT, Bigenzahn JW, Nieuwenhuis J, Staring J, Sacco R, van Diemen FR, Olk N, Stukalov A, Marceau C, Janssen H, Carette JE, Bennett KL, Colinge J, Superti-Furga G, Brummelkamp TR (2015) Gene essentiality and synthetic lethality in haploid human cells. *Science* 350: 1092-1096
- Bock C, Tomazou EM, Brinkman AB, Muller F, Simmer F, Gu H, Jager N, Gnirke A, Stunnenberg HG, Meissner A (2010) Quantitative comparison of genome-wide DNA methylation mapping technologies. *Nat Biotechnol* 28: 1106-1114
- Bollag G, Tsai J, Zhang J, Zhang C, Ibrahim P, Nolop K, Hirth P (2012) Vemurafenib: the first drug approved for BRAF-mutant cancer. *Nat Rev Drug Discov* 11: 873-886
- Bolotin A, Quinquis B, Sorokin A, Ehrlich SD (2005) Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology* 151: 2551-2561
- Brinkman EK, Chen T, de Haas M, Holland HA, Akhtar W, van Steensel B (2018) Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol Cell* 70: 801-813 e6
- Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ (2013) Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* 10: 1213-1218
- Burr ML, Sparbier CE, Chan YC, Williamson JC, Woods K, Beavis PA, Lam EYN, Henderson MA, Bell CC, Stolzenburg S, Gilan O, Bloor S, Noori T, Morgens DW, Bassik MC, Neeson PJ, Behren A, Darcy PK, Dawson SJ, Voskoboinik I et al. (2017) CMTM6 maintains the expression of PD-L1 and regulates anti-tumour immunity. *Nature* 549: 101-105
- Cano-Rodriguez D, Gjaltema RA, Jilderda LJ, Jellema P, Dokter-Fokkens J, Ruiters MH, Rots MG (2016) Writing of H3K4Me3 overcomes epigenetic silencing in a sustained but context-dependent manner. *Nat Commun* 7: 12284

Cao J, Cusanovich DA, Ramani V, Aghamirzaie D, Pliner HA, Hill AJ, Daza RM, McFaline-Figueroa JL, Packer JS, Christiansen L, Steemers FJ, Adey AC, Trapnell C, Shendure J (2018) Joint profiling of chromatin accessibility and gene expression in thousands of single cells. *Science* 361: 1380-1385

Cao J, Packer JS, Ramani V, Cusanovich DA, Huynh C, Daza R, Qiu X, Lee C, Furlan SN, Steemers FJ, Adey A, Waterston RH, Trapnell C, Shendure J (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* 357: 661-667

Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, Zhang F, Mundlos S, Christiansen L, Steemers FJ, Trapnell C, Shendure J (2019) The single-cell transcriptional landscape of mammalian organogenesis. *Nature* 566: 496-502

Capecchi MR (1989) Altering the genome by homologous recombination. *Science* 244: 1288-1292

Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N, Kuehne AI, Kranzusch PJ, Griffin AM, Ruthel G, Dal Cin P, Dye JM, Whelan SP, Chandran K, Brummelkamp TR (2011) Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* 477: 340-343

Casini A, Olivieri M, Petris G, Montagna C, Reginato G, Maule G, Lorenzin F, Prandi D, Romanel A, Demichelis F, Inga A, Cereseto A (2018) A highly specific SpCas9 variant is identified by in vivo screening in yeast. *Nat Biotechnol* 36: 265-271

Cebrian-Serrano A, Davies B (2017) CRISPR-Cas orthologues and variants: optimizing the repertoire, specificity and delivery of genome engineering tools. *Mamm Genome* 28: 247-261

Chavez A, Scheiman J, Vora S, Pruitt BW, Tuttle M, E PRI, Lin S, Kiani S, Guzman CD, Wiegand DJ, Ter-Ovanesyan D, Braff JL, Davidsohn N, Housden BE, Perrimon N, Weiss R, Aach J, Collins JJ, Church GM (2015) Highly efficient Cas9-mediated transcriptional programming. *Nat Methods* 12: 326-328

Chen JS, Dagdas YS, Kleinstiver BP, Welch MM, Sousa AA, Harrington LB, Sternberg SH, Joung JK, Yildiz A, Doudna JA (2017) Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* 550: 407-410

Chen KH, Boettiger AN, Moffitt JR, Wang S, Zhuang X (2015) RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* 348: 412-427

Chu VT, Graf R, Wirtz T, Weber T, Favret J, Li X, Petsch K, Tran NT, Sieweke MH, Berek C, Kuhn R, Rajewsky K (2016) Efficient CRISPR-mediated mutagenesis in primary immune cells using CrispRGold and a C57BL/6 Cas9 transgenic mouse line. *Proc Natl Acad Sci U S A* 113: 12514-12519

Cloonan N, Forrest AR, Kolle G, Gardiner BB, Faulkner GJ, Brown MK, Taylor DF, Steptoe AL, Wani S, Bethel G, Robertson AJ, Perkins AC, Bruce SJ, Lee CC, Ranade SS, Peckham HE, Manning JM, McKernan KJ, Grimmond SM (2008) Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods* 5: 613-619

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science* 339: 819-823

Cusanovich DA, Daza R, Adey A, Pliner HA, Christiansen L, Gunderson KL, Steemers FJ, Trapnell C, Shendure J (2015) Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* 348: 910-914

Dahlman JE, Abudayyeh OO, Joung J, Gootenberg JS, Zhang F, Konermann S (2015) Orthogonal gene knockout and activation with a catalytically active Cas9 nuclease. *Nat Biotechnol* 33: 1159-1161

Datlinger P, Rendeiro AF, Schmidl C, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C (2017) Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods* 14: 297-301

Deltcheva E, Chylinski K, Sharma CM, Gonzales K, Chao Y, Pirzada ZA, Eckert MR, Vogel J, Charpentier E (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature* 471: 602-607

Dixit A, Parnas O, Li B, Chen J, Fulco CP, Jerby-Arnon L, Marjanovic ND, Dionne D, Burks T, Raychowdhury R, Adamson B, Norman TM, Lander ES, Weissman JS, Friedman N, Regev A (2016) Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell* 167: 1853-1866 e17

Doench JG, Hartenian E, Graham DB, Tothova Z, Hegde M, Smith I, Sullender M, Ebert BL, Xavier RJ, Root DE (2014) Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat Biotechnol* 32: 1262-1267

Eng CL, Lawson M, Zhu Q, Dries R, Kouloua N, Takei Y, Yun J, Cronin C, Karp C, Yuan GC, Cai L (2019) Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH. *Nature* 568: 235-239

Evers B, Jastrzebski K, Heijmans JP, Grenrum W, Beijersbergen RL, Bernards R (2016) CRISPR knockout screening outperforms shRNA and CRISPRi in identifying essential genes. *Nat Biotechnol* 34: 631-633

Fan HC, Fu GK, Fodor SP (2015) Expression profiling. Combinatorial labeling of single cells for gene expression cytometry. *Science* 347: 628-637

Fantom Consortium (2014) A promoter-level mammalian expression atlas. *Nature* 507: 462-470

Feldman D, Singh A, Schmid-Burgk JL, Mezger A, Garrity AJ, Carlson RJ, Zhang F, Blainey PC (2018) Pooled optical screens in human cells. *bioRxiv*

Findlay GM, Boyle EA, Hause RJ, Klein JC, Shendure J (2014) Saturation editing of genomic regions by multiplex homology-directed repair. *Nature* 513: 120-123

Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J (2018) Accurate classification of BRCA1 variants with saturation genome editing. *Nature* 562: 217-222

Galonska C, Charlton J, Mattei AL, Donaghey J, Clement K, Gu H, Mohammad AW, Stamenova EK, Cacchiarelli D, Klages S, Timmermann B, Cantz T, Scholer HR, Gnirke A, Ziller MJ, Meissner A (2018) Genome-wide tracking of dCas9-methyltransferase footprints. *Nat Commun* 9: 597

Gasiunas G, Barrangou R, Horvath P, Siksnys V (2012) Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. *Proc Natl Acad Sci U S A* 109: E2579-E2586

Gierahn TM, Wadsworth MH, 2nd, Hughes TK, Bryson BD, Butler A, Satija R, Fortune S, Love JC, Shalek AK (2017) Seq-Well: portable, low-cost RNA sequencing of single cells at high throughput. *Nat Methods* 14: 395-398

Gilbert LA, Larson MH, Morsut L, Liu Z, Brar GA, Torres SE, Stern-Ginossar N, Brandman O, Whitehead EH, Doudna JA, Lim WA, Weissman JS, Qi LS (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154: 442-451

Gioia L, Siddique A, Head SR, Salomon DR, Su AI (2018) A genome-wide survey of mutations in the Jurkat cell line. *BMC Genomics* 19: 334

Gootenberg JS, Abudayyeh OO, Lee JW, Essletzbichler P, Dy AJ, Joung J, Verdine V, Donghia N, Daringer NM, Freije CA, Myhrvold C, Bhattacharyya RP, Livny J, Regev A, Koonin EV, Hung DT, Sabeti PC, Collins JJ, Zhang F (2017) Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* 356: 438-442

GTEX Consortium (2015) Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* 348: 648-660

Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, Huang D, Xu Y, Huang W, Jiang M, Jiang X, Mao J, Chen Y, Lu C, Xie J, Fang Q et al. (2018) Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell* 172: 1091-1107 e17

Hart T, Chandrashekar M, Aregger M, Steinhart Z, Brown KR, MacLeod G, Mis M, Zimmermann M, Fradet-Turcotte A, Sun S, Mero P, Dirks P, Sidhu S, Roth FP, Rissland OS, Durocher D, Angers S, Moffat J (2015) High-Resolution CRISPR Screens Reveal Fitness Genes and Genotype-Specific Cancer Liabilities. *Cell* 163: 1515-1526

Hashimshony T, Senderovich N, Avital G, Klochendler A, de Leeuw Y, Anavy L, Gennert D, Li S, Livak KJ, Rozenblatt-Rosen O, Dor Y, Regev A, Yanai I (2016) CEL-Seq2: sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* 17: 77

Henriksson J, Chen X, Gomes T, Ullah U, Meyer KB, Miragaia R, Duddy G, Pramanik J, Yusa K, Lahesmaa R, Teichmann SA (2019) Genome-wide CRISPR Screens in T Helper Cells Reveal Pervasive Crosstalk between Activation and Differentiation. *Cell* 176: 882-896 e18

Hill AJ, McFaline-Figueroa JL, Starita LM, Gasperini MJ, Matreyek KA, Packer J, Jackson D, Shendure J, Trapnell C (2018) On the design of CRISPR-based single-cell molecular screens. *Nat Methods* 15: 271-274

Hilton IB, D'Ippolito AM, Vockley CM, Thakore PI, Crawford GE, Reddy TE, Gersbach CA (2015) Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. *Nat Biotechnol* 33: 510-517

Hu JH, Miller SM, Geurts MH, Tang W, Chen L, Sun N, Zeina CM, Gao X, Rees HA, Lin Z, Liu DR (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature* 556: 57-63

Huang YH, Su J, Lei Y, Brunetti L, Gundry MC, Zhang X, Jeong M, Li W, Goodell MA (2017) DNA epigenome editing using CRISPR-Cas SunTag-directed DNMT3A. *Genome Biol* 18: 176

Ishino Y, Shinagawa H, Makino K, Amemura M, Nakata A (1987) Nucleotide sequence of the *iap* gene, responsible for alkaline phosphatase isozyme conversion in *Escherichia coli*, and identification of the gene product. *J Bacteriol* 169: 5429-5433

Jaitin DA, Kenigsberg E, Keren-Shaul H, Elefant N, Paul F, Zaretsky I, Mildner A, Cohen N, Jung S, Tanay A, Amit I (2014) Massively parallel single-cell RNA-seq for marker-free decomposition of tissues into cell types. *Science* 343: 776-779

Jaitin DA, Weiner A, Yofe I, Lara-Astiaso D, Keren-Shaul H, David E, Salame TM, Tanay A, van Oudenaarden A, Amit I (2016) Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell* 167: 1883-1896 e15

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337: 816-821

Kearns NA, Pham H, Tabak B, Genga RM, Silverstein NJ, Garber M, Maehr R (2015) Functional annotation of native enhancers with a Cas9-histone demethylase fusion. *Nat Methods* 12: 401-403

Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA, Kirschner MW (2015) Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell* 161: 1187-1201

Kleinstiver BP, Pattanayak V, Prew MS, Tsai SQ, Nguyen NT, Zheng Z, Joung JK (2016) High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* 529: 490-495

Kohli RM, Zhang Y (2013) TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 502: 472-479

Konermann S, Brigham MD, Trevino AE, Joung J, Abudayyeh OO, Barcena C, Hsu PD, Habib N, Gootenberg JS, Nishimasu H, Nureki O, Zhang F (2015) Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. *Nature* 517: 583-588

Koonin EV, Makarova KS, Zhang F (2017) Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol* 37: 67-78

Kosuri S, Church GM (2014) Large-scale de novo DNA synthesis: technologies and applications. *Nat Methods* 11: 499-507

Kriaucionis S, Heintz N (2009) The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324: 929-930

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W, Funke R, Gage D, Harris K, Heaford A, Howland J, Kann L, Lehoczky J, LeVine R, McEwan P, McKernan K et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921

Li H, van der Leun AM, Yofe I, Lubling Y, Gelbard-Solodkin D, van Akkooi ACJ, van den Braber M, Rozeman EA, Haanen J, Blank CU, Horlings HM, David E, Baran Y, Bercovich A, Lifshitz A, Schumacher TN, Tanay A, Amit I (2019) Dysfunctional CD8 T Cells Form a Proliferative, Dynamically Regulated Compartment within Human Melanoma. *Cell* 176: 775-789 e18

Liao HK, Hatanaka F, Araoka T, Reddy P, Wu MZ, Sui Y, Yamauchi T, Sakurai M, O'Keefe DD, Nunez-Delicado E, Guillen P, Campistol JM, Wu CJ, Lu LF, Esteban CR, Izpisua Belmonte JC (2017) In Vivo Target Gene Activation via CRISPR/Cas9-Mediated Trans-epigenetic Modulation. *Cell* 171: 1495-1507 e15

Lubeck E, Coskun AF, Zhiyentayev T, Ahmad M, Cai L (2014) Single-cell in situ RNA profiling by sequential hybridization. *Nat Methods* 11: 360-361

Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, Tirosh I, Bialas AR, Kamitaki N, Martersteck EM, Trombetta JJ, Weitz DA, Sanes JR, Shalek AK, Regev A, McCarroll SA (2015) Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. *Cell* 161: 1202-1214

Maeder ML, Linder SJ, Cascio VM, Fu Y, Ho QH, Joung JK (2013a) CRISPR RNA-guided activation of endogenous human genes. *Nat Methods* 10: 977-979

Maeder ML, Linder SJ, Reyon D, Angstman JF, Fu Y, Sander JD, Joung JK (2013b) Robust, synergistic regulation of human gene expression using TALE activators. *Nat Methods* 10: 243-245

Marceau CD, Puschnik AS, Majzoub K, Ooi YS, Brewer SM, Fuchs G, Swaminathan K, Mata MA, Elias JE, Sarnow P, Carette JE (2016) Genetic dissection of Flaviviridae host factors through genome-scale CRISPR screens. *Nature* 535: 159-163

Marraffini LA, Sontheimer EJ (2008) CRISPR interference limits horizontal gene transfer in staphylococci by targeting DNA. *Science* 322: 1843-1845

Mezzadra R, Sun C, Jae LT, Gomez-Eerland R, de Vries E, Wu W, Logtenberg MEW, Slagter M, Rozeman EA, Hofland I, Broeks A, Horlings HM, Wessels LFA, Blank CU, Xiao Y, Heck AJR, Borst J, Brummelkamp TR, Schumacher TNM (2017) Identification of CMTM6 and CMTM4 as PD-L1 protein regulators. *Nature* 549: 106-110

Moffitt JR, Hao J, Bambah-Mukku D, Lu T, Dulac C, Zhuang X (2016a) High-performance multiplexed fluorescence in situ hybridization in culture and tissue with matrix imprinting and clearing. *Proc Natl Acad Sci U S A* 113: 14456-14461

Moffitt JR, Hao J, Wang G, Chen KH, Babcock HP, Zhuang X (2016b) High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc Natl Acad Sci U S A* 113: 11046-11051

Mojica FJ, Diez-Villasenor C, Garcia-Martinez J, Soria E (2005) Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60: 174-182

Mojica FJ, Juez G, Rodriguez-Valera F (1993) Transcription at different salinities of *Haloferax mediterranei* sequences adjacent to partially modified PstI sites. *Mol Microbiol* 9: 613-621

Morgan SL, Mariano NC, Bermudez A, Arruda NL, Wu F, Luo Y, Shankar G, Jia L, Chen H, Hu JF, Hoffman AR, Huang CC, Pitteri SJ, Wang KC (2017) Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat Commun* 8: 15993

Morgens DW, Deans RM, Li A, Bassik MC (2016) Systematic comparison of CRISPR/Cas9 and RNAi screens for essential genes. *Nat Biotechnol* 34: 634-636

Morita S, Noguchi H, Horii T, Nakabayashi K, Kimura M, Okamura K, Sakai A, Nakashima H, Hata K, Nakashima K, Hatada I (2016) Targeted DNA demethylation in vivo using dCas9-peptide repeat and scFv-TET1 catalytic domain fusions. *Nat Biotechnol* 34: 1060-1065

Mulqueen RM, Pokholok D, Norberg SJ, Torkenczy KA, Fields AJ, Sun D, Sinnamon JR, Shendure J, Trapnell C, O'Roak BJ, Xia Z, Steemers FJ, Adey AC (2018) Highly scalable generation of DNA methylation profiles in single cells. *Nat Biotechnol* 36: 428-431

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F, Nureki O (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* 156: 935-949

Nishimasu H, Shi X, Ishiguro S, Gao L, Hirano S, Okazaki S, Noda T, Abudayyeh OO, Gootenberg JS, Mori H, Oura S, Holmes B, Tanaka M, Seki M, Hirano H, Aburatani H, Ishitani R, Ikawa M, Yachie N, Zhang F et al. (2018) Engineered CRISPR-Cas9 nuclease with expanded targeting space. *Science* 361: 1259-1262

Park RJ, Wang T, Koundakjian D, Hultquist JF, Lamothe-Molina P, Monel B, Schumann K, Yu H, Krupczak KM, Garcia-Beltran W, Piechocka-Trocha A, Krogan NJ, Marson A, Sabatini DM, Lander ES, Hacohen N, Walker BD (2017) A genome-wide CRISPR screen identifies a restricted set of HIV host dependency factors. *Nat Genet* 49: 193-203

Perez-Pinera P, Kocak DD, Vockley CM, Adler AF, Kabadi AM, Polstein LR, Thakore PI, Glass KA, Ousterout DG, Leong KW, Guilak F, Crawford GE, Reddy TE, Gersbach CA (2013) RNA-guided gene activation by CRISPR-Cas9-based transcription factors. *Nat Methods* 10: 973-976

Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, Moore R, McClanahan TK, Sadekova S, Klappenbach JA (2017) Multiplexed quantification of proteins and transcripts in single cells. *Nat Biotechnol* 35: 936-939

Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R (2014) Full-length RNA-seq from single cells using Smart-seq2. *Nat Protoc* 9: 171-181

Pillay S, Meyer NL, Puschnik AS, Davulcu O, Diep J, Ishikawa Y, Jae LT, Wosen JE, Nagamine CM, Chapman MS, Carette JE (2016) An essential receptor for adeno-associated virus infection. *Nature* 530: 108-112

Platt RJ, Chen S, Zhou Y, Yim MJ, Swiech L, Kempton HR, Dahlman JE, Parnas O, Eisenhaure TM, Jovanovic M, Graham DB, Jhunjhunwala S, Heidenreich M, Xavier RJ, Langer R, Anderson DG, Hacohen N, Regev A, Feng G, Sharp PA et al. (2014) CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* 159: 440-455

Pourcel C, Salvignol G, Vergnaud G (2005) CRISPR elements in *Yersinia pestis* acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology* 151: 653-663

Qi LS, Larson MH, Gilbert LA, Doudna JA, Weissman JS, Arkin AP, Lim WA (2013) Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* 152: 1173-1183

Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, Trapnell C (2017) Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods* 14: 979-982

Ramani V, Deng X, Qiu R, Gunderson KL, Steemers FJ, Disteche CM, Noble WS, Duan Z, Shendure J (2017) Massively multiplex single-cell Hi-C. *Nat Methods* 14: 263-266

Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F (2013) Genome engineering using the CRISPR-Cas9 system. *Nat Protoc* 8: 2281-2308

Rao SS, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, Sanborn AL, Machol I, Omer AD, Lander ES, Aiden EL (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159: 1665-1680

Reyon D, Tsai SQ, Khayter C, Foden JA, Sander JD, Joung JK (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat Biotechnol* 30: 460-465

Roadmap Epigenomics Consortium (2015) Integrative analysis of 111 reference human epigenomes. *Nature* 518: 317-330

Rouet P, Smih F, Jasin M (1994) Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. *Mol Cell Biol* 14: 8096-8106

Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA (2017) The Human Cell Atlas: from vision to reality. *Nature* 550: 451-453

Rubin AJ, Parker KR, Satpathy AT, Qi Y, Wu B, Ong AJ, Mumbach MR, Ji AL, Kim DS, Cho SW, Zarnegar BJ, Greenleaf WJ, Chang HY, Khavari PA (2019) Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell* 176: 361-376 e17

Sack LM, Davoli T, Li MZ, Li Y, Xu Q, Naxerova K, Wooten EC, Bernardi RJ, Martin TD, Chen T, Leng Y, Liang AC, Scorsone KA, Westbrook TF, Wong KK, Elledge SJ (2018) Profound Tissue Specificity in Proliferation Control Underlies Cancer Drivers and Aneuploidy Patterns. *Cell* 173: 499-514 e23

Sanson KR, Hanna RE, Hegde M, Donovan KF, Strand C, Sullender ME, Vaimberg EW, Goodale A, Root DE, Piccioni F, Doench JG (2018) Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* 9: 5416

Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, de Rivera H, Bien E, Baum M, Bortolin L, Wang S, Goeva A, Nemesh J, Kamitaki N, Brumbaugh S, Kulp D, McCarroll SA (2018) Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell* 174: 1015-1030 e16

Savidis G, McDougall WM, Meraner P, Perreira JM, Portmann JM, Trincucci G, John SP, Aker AM, Renzette N, Robbins DR, Guo Z, Green S, Kowalik TF, Brass AL (2016) Identification of Zika Virus and Dengue Virus Dependency Factors using Functional Genomics. *Cell Rep* 16: 232-246

Schmidl C, Rendeiro AF, Sheffield NC, Bock C (2015) ChIPmentation: fast, robust, low-input ChIP-seq for histones and transcription factors. *Nat Methods* 12: 963-965

Schneider U, Schwenk HU, Bornkamm G (1977) Characterization of EBV-genome negative "null" and "T" cell lines derived from children with acute lymphoblastic leukemia and leukemic transformed non-Hodgkin lymphoma. *Int J Cancer* 19: 621-626

Sedlyarov V, Eichner R, Girardi E, Essletzbichler P, Goldmann U, Nunes-Hasler P, Srndic I, Moskovskich A, Heinz LX, Kartnig F, Bigenzahn JW, Rebsamen M, Kovarik P, Demareux N, Superti-

Furga G (2018) The Bicarbonate Transporter SLC4A7 Plays a Key Role in Macrophage Phagosome Acidification. *Cell Host Microbe* 23: 766-774 e5

Shah S, Lubeck E, Zhou W, Cai L (2016) In Situ Transcription Profiling of Single Cells Reveals Spatial Organization of Cells in the Mouse Hippocampus. *Neuron* 92: 342-357

Shah S, Takei Y, Zhou W, Lubeck E, Yun J, Eng CL, Koulina N, Cronin C, Karp C, Liaw EJ, Amin M, Cai L (2018) Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. *Cell* 174: 363-376 e16

Shahi P, Kim SC, Haliburton JR, Gartner ZJ, Abate AR (2017) Abseq: Ultrahigh-throughput single cell protein profiling with droplet microfluidic barcoding. *Sci Rep* 7: 44447

Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen T, Heckl D, Ebert BL, Root DE, Doench JG, Zhang F (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343: 84-87

Shapiro E, Biezuner T, Linnarsson S (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat Rev Genet* 14: 618-630

Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH (2017) DNA sequencing at 40: past, present and future. *Nature* 550: 345-353

Shi J, Wang E, Milazzo JP, Wang Z, Kinney JB, Vakoc CR (2015) Discovery of cancer drug targets by CRISPR-Cas9 screening of protein domains. *Nat Biotechnol* 33: 661-667

Shifrut E, Carnevale J, Tobin V, Roth TL, Woo JM, Bui CT, Li PJ, Diolaiti ME, Ashworth A, Marson A (2018) Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* 175: 1958-1971 e15

Slaymaker IM, Gao L, Zetsche B, Scott DA, Yan WX, Zhang F (2016) Rationally engineered Cas9 nucleases with improved specificity. *Science* 351: 84-88

Sondka Z, Bamford S, Cole CG, Ward SA, Dunham I, Forbes SA (2018) The COSMIC Cancer Gene Census: describing genetic dysfunction across all human cancers. *Nat Rev Cancer* 18: 696-705

Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, Satija R, Smibert P (2017) Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* 14: 865-868

Stunnenberg HG, International Human Epigenome C, Hirst M (2016) The International Human Epigenome Consortium: A Blueprint for Scientific Collaboration and Discovery. *Cell* 167: 1145-1149

Tahiliani M, Koh KP, Shen Y, Pastor WA, Bandukwala H, Brudno Y, Agarwal S, Iyer LM, Liu DR, Aravind L, Rao A (2009) Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324: 930-935

Tanenbaum ME, Gilbert LA, Qi LS, Weissman JS, Vale RD (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159: 635-646

Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA (2009) mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* 6: 377-382

Thakore PI, Black JB, Hilton IB, Gersbach CA (2016) Editing the epigenome: technologies for programmable transcription and epigenetic modulation. *Nat Methods* 13: 127-137

Treutlein B, Lee QY, Camp JG, Mall M, Koh W, Shariati SA, Sim S, Neff NF, Skotheim JM, Wernig M, Quake SR (2016) Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* 534: 391-395

Vasquez KM, Marburger K, Intody Z, Wilson JH (2001) Manipulating the mammalian genome by homologous recombination. *Proc Natl Acad Sci U S A* 98: 8403-8410

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M et al. (2001) The sequence of the human genome. *Science* 291: 1304-1351

Vitak SA, Torkenczy KA, Rosenkrantz JL, Fields AJ, Christiansen L, Wong MH, Carbone L, Steemers FJ, Adey A (2017) Sequencing thousands of single-cell genomes with combinatorial indexing. *Nat Methods* 14: 302-308

Wade N (2010) A Decade Later, Genetic Map Yields Few New Cures. *The New York Times*

Wagner DE, Klein AM (2017) Genetic screening enters the single-cell era. *Nat Methods* 14: 237-238

Wang G, Moffitt JR, Zhuang X (2018) Multiplexed imaging of high-density libraries of RNAs with MERFISH and expansion microscopy. *Sci Rep* 8: 4847

Wang T, Birsoy K, Hughes NW, Krupczak KM, Post Y, Wei JJ, Lander ES, Sabatini DM (2015) Identification and characterization of essential genes in the human genome. *Science* 350: 1096-1101

Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343: 80-84

Whitehurst AW, Bodemann BO, Cardenas J, Ferguson D, Girard L, Peyton M, Minna JD, Michnoff C, Hao W, Roth MG, Xie XJ, White MA (2007) Synthetic lethal screen identification of chemosensitizer loci in cancer cells. *Nature* 446: 815-819

Willyard C (2018) New human gene tally reignites debate. *Nature* 558: 354-355

Xie S, Duan J, Li B, Zhou P, Hon GC (2017) Multiplexed Engineering and Analysis of Combinatorial Enhancer Activity in Single Cells. *Mol Cell* 66: 285-299 e5

Yeo NC, Chavez A, Lance-Byrne A, Chan Y, Menn D, Milanova D, Kuo CC, Guo X, Sharma S, Tung A, Cecchi RJ, Tuttle M, Pradhan S, Lim ET, Davidsohn N, Ebrahimkhani MR, Collins JJ, Lewis NE, Kiani S, Church GM (2018) An enhanced CRISPR repressor for targeted mammalian gene regulation. *Nat Methods* 15: 611-616

Yin Y, Morgunova E, Jolma A, Kaasinen E, Sahu B, Khund-Sayeed S, Das PK, Kivioja T, Dave K, Zhong F, Nitta KR, Taipale M, Popov A, Ginno PA, Domcke S, Yan J, Schubeler D, Vinson C, Taipale J (2017) Impact of cytosine methylation on DNA binding specificities of human transcription factors. *Science* 356: eaaj2239

Zetsche B, Gootenberg JS, Abudayyeh OO, Slaymaker IM, Makarova KS, Essletzbichler P, Volz SE, Joung J, van der Oost J, Regev A, Koonin EV, Zhang F (2015) Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell* 163: 759-771

Zetsche B, Heidenreich M, Mohanraju P, Fedorova I, Kneppers J, DeGennaro EM, Winblad N, Choudhury SR, Abudayyeh OO, Gootenberg JS, Wu WY, Scott DA, Severinov K, van der Oost J, Zhang F (2017) Multiplex gene editing by CRISPR-Cpf1 using a single crRNA array. *Nat Biotechnol* 35: 31-34

Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, Ziraldo SB, Wheeler TD, McDermott GP, Zhu J, Gregory MT, Shuga J, Montesclaros L, Underwood JG, Masquelier DA, Nishimura SY, Schnall-Levin M, Wyatt PW, Hindson CM, Bharadwaj R et al. (2017) Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 8: 14049

Zhu Q, Shah S, Dries R, Cai L, Yuan GC (2018) Identification of spatially associated subpopulations by combining scRNAseq and sequential fluorescence in situ hybridization data. *Nat Biotechnol* 36: 1183-1190

6. CURRICULUM VITAE

► Personal Details

Paul Datlinger

Date of birth: 6 June 1988

Nationality: Austrian

Address: Lazarettgasse 14, AKH BT 25.3, 1090 Vienna, Austria

Email: pdatlinger@cemm.oeaw.ac.at

Phone: +43-1-40160-70032

Skype: paul_datlinger

► Education

PhD studies, Center for Molecular Medicine, Medical Epigenomics Lab, Vienna, Austria

October 2013 – present

Thesis: CRISPR screening at the single-cell level

Advised by: Christoph Bock

MSc studies in Molecular Biology at the University of Vienna

Jan 2011 – September 2013

Specialization: Molecular Medicine, Neuroscience

Thesis: Methods for DNA Methylation Mapping in Large Patient Cohorts at Single Base Pair Resolution

Advised by: Christoph Bock

BSc studies in Biology at the University of Vienna

Oct 2007 – Jan 2011

Majoring in: Molecular Biology

Thesis: The role of plectin in Epidermolysis bullosa

Advised by: Gerhard Wiche

Sep 1998 – June 2006

Secondary education: Gymnasium Mattersburg, with distinction

► Additional projects

June 2017:

‘A journalist tries CRISPR’, collaboration with the Austrian newspaper *Der Standard* in which I instructed a journalist to perform a genome engineering experiment himself.

<https://www.derstandard.at/story/2000059254039/ein-journalist-versucht-crispr-tage-im-labormantel>

Oct 2012 – Sep 2013:

Medical University of Vienna, AUT / CeMM / Christoph Bock Lab

DNA methylation dynamics in pediatric precursor B-cell ALL revealed by whole genome and reduced representation bisulfite sequencing.

Jan 2012 – Jun 2012:

University of Manchester, UK / Visiting Student specializing in cancer research.

Jul 2011 – Oct 2011:

Imperial College London, UK / Department of Molecular Medicine / Richard Harbottle Lab

Generation of a non-viral S/MAR vector for the gene therapy of alkaptonuria.

Mar 2010 – May 2010:

University of Vienna, AUT / Max F Perutz Laboratories / Gerhard Wiche Lab
The role of plectin in endothelial cell integrity in epidermolysis bullosa patients.

Jun 2009 – 2014:

Wiesen Music Festivals - film and post-production

Oct 2004 – June 2005:

Film adaptation of Franz Kafka's short story "The Judgement"
The film was awarded the youth culture award of the federal state of Burgenland and won a prize at the Viennese film festival "wienervideo&filmtage".

► Invited oral presentations

Labroots Genetics and Genomics Virtual Conference 2016 (invited by Diagenode)
Illumina User Group Meeting 2017, Cologne, Germany
Illumina User Day Vienna 2017, Vienna, Austria
Single Cell Sequencing Meeting 2017, Leuven, Belgium
10x Genomics European User Group Meeting 2017, Cambridge, United Kingdom
British Society for Immunology Congress 2017, Brighton, United Kingdom
ELRIG meeting 2017, Liverpool, United Kingdom
Horizon Genomics CRISPR Forum 2017, Dublin, Ireland
Annual conference of the Austrian Association of Molecular Life Sciences and Biotechnology, 2017, Innsbruck, Austria
Cold Spring Harbor conference on Genome Engineering, 2017, Cold Spring Harbor, USA
Red Stars Entrepreneur Meeting, Pörschach am Wörthersee, Austria
Single Cell Day, Universitätsklinik Hamburg-Eppendorf, Hamburg, Germany

► Awards

Life Science Research Award Austria 2017 by the Austrian Association of Molecular Life Sciences and Biotechnology (ÖGMBT) for the publication *Pooled CRISPR screening with single-cell transcriptome readout*

► Publications

Datlinger P, Rendeiro AF*, Boenke T*, Krausgruber T, Barreca D, Bock C **Ultra-high throughput single-cell RNA sequencing by combinatorial fluidic indexing. *In preparation.***

Ferreira da Silva J, Salic S, Wiedner M, Datlinger P, Essletzbichler P, Hanzl A, Superti-Furga G, Bock C, Winter G, Loizou JI **Genome-scale CRISPR screens are efficient in non-homologous end-joining deficient cells. *Submitted.***

Klughammer J*, Kiesel B*, Roetzer T, Fortelny N, Nemc A, Nenning KH, Furtner J, Sheffield NC, Datlinger P, Peter N, Nowosielski M, Augustin M, Mischkulnig M, Strobel T, Alpar D, Erguner B, Senekowitsch M, Moser P, Freyschlag CF, Kerschbaumer J, Thome C, Grams AE, Stockhammer G, Kitzwoegerer M, Oberndorfer S, Marhold F, Weis S, Trenkler J, Buchroithner J, Pichler J, Haybaeck J, Krassnig S, Mahdy Ali K, von Campe G, Payer F, Sherif C, Preiser J, Hauser T, Winkler PA, Kleindienst W, Wurtz F, Brandner-Kokalj T, Stultschnig M, Schweiger S, Dieckmann K, Preusser M, Langs G, Baumann B, Knosp E, Widhalm G, Marosi C, Hainfellner JA, Woehrer A, Bock C (2018) **The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space.** Nat Med 24: 1611-1624

Datlinger P, Rendeiro AF*, Schmidl C*, Krausgruber T, Traxler P, Klughammer J, Schuster LC, Kuchler A, Alpar D, Bock C (2017) **Pooled CRISPR screening with single-cell transcriptome readout.** Nat Methods 14: 297-301

Sheffield NC, Pierron G, Klughammer J, Datlinger P, Schonegger A, Schuster M, Hadler J, Surdez D, Guillemot D, Lapouble E, Freneaux P, Champigneulle J, Bouvier R, Walder D, Ambros IM, Hutter C, Sorz E, Amaral AT, de Alava E, Schallmoser K, Strunk D, Rinner B, Liegl-Atzwanger B, Huppertz B, Leithner A, de Pinieux G, Terrier P, Laurence V, Michon J, Ladenstein R, Holter W, Windhager R, Dirksen U, Ambros PF, Delattre O, Kovar H, Bock C, Tomazou EM (2017) **DNA methylation heterogeneity defines a disease spectrum in Ewing sarcoma**. Nat Med 23: 386-395

Blueprint consortium (2016) **Quantitative comparison of DNA methylation assays for biomarker development and clinical applications**. Nat Biotechnol 34: 726-737

Klughammer J, Datlinger P, Printz D, Sheffield NC, Farlik M, Hadler J, Fritsch G, Bock C (2015) **Differential DNA Methylation Analysis without a Reference Genome**. Cell Rep 13: 2621-2633

Farlik M*, Sheffield NC*, Nuzzo A, Datlinger P, Schonegger A, Klughammer J, Bock C (2015) **Single-cell DNA methylome sequencing and bioinformatic inference of epigenomic cell-state dynamics**. Cell Rep 10: 1386-1397

Tomazou EM*, Sheffield NC*, Schmidl C, Schuster M, Schonegger A, Datlinger P, Kubicek S, Bock C, Kovar H (2015) **Epigenome mapping reveals distinct modes of gene regulation and widespread enhancer reprogramming by the oncogenic fusion protein EWS-FLI1**. Cell Rep 10: 1082-1095

Kulinski TM, Casari MR, Guenzl PM, Wenzel D, Andergassen D, Hladik A, Datlinger P, Farlik M, Theussl HC, Penninger JM, Knapp S, Bock C, Barlow DP, Hudson QJ (2015) **Imprinted expression in cystic embryoid bodies shows an embryonic and not an extra-embryonic pattern**. Dev Biol 402: 291-305

Vienna, Austria, September 2019
Paul Datlinger