



JAN 19, 2024

OPEN  ACCESS**DOI:**

dx.doi.org/10.17504/protocols.io.261ged73yv47/v1

External link:

<https://www.biorxiv.org/content/10.1101/2022.11.01.514606v1.article-info>

Protocol Citation: Gavin R. Schnitzler, Helen Kang, Glen Munson, Philine Guckelberger, Drew Bergman, Brian Cleary, Eric S Lander, Jesse Engreitz, Rajat M. Gupta 2024. Systematic Perturb-seq to discover endothelial cell programs related to CAD. **protocols.io**
<https://dx.doi.org/10.17504/protocols.io.261ged73yv47/v1>

Systematic Perturb-seq to discover endothelial cell programs related to CAD

Gavin R. Schnitzler^{1,2,3}, Helen Kang^{4,5}, Glen Munson¹, Philine Guckelberger¹, Drew Bergman^{1,6}, Brian Cleary^{7,1}, Eric S Lander^{1,8,9}, Jesse Engreitz^{1,3,4,5,10}, Rajat M. Gupta^{1,3,2}

¹Broad Institute of Harvard & MIT, Cambridge, MA;

²Divisions of Genetics and Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston MA;

³The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute, Cambridge, MA;

⁴Department of Genetics, Stanford University School of Medicine, Stanford, CA;

⁵BASE Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, CA;

⁶Geisel School of Medicine at Dartmouth, Hanover, NH;

⁷Faculty of Computing and Data Sciences, Departments of Biology and Biomedical Engineering, Biological Design Center, and Program in Bioinformatics, Boston University, Boston, MA;

⁸Department of Biology, MIT, Cambridge, MA;

⁹Department of Systems Biology, Harvard Medical School, Boston, MA;

¹⁰Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA

Gavin R. Schnitzler: For Perturb-seq methods, dialout analysis, bulk RNAseq, TeloHAEC microscopic analyses, differential expression analysis, co-IP assays;

Helen Kang: For cNMF, G2P and V2G2P analyses, internal & external benchmarking, application to other traits & cell types.;

Jesse Engreitz: co-corresponding author;

Rajat M. Gupta: co-corresponding author;

RguptaLab

Engreitz Lab



Gavin R. Schnitzler

MANUSCRIPT CITATION:
Title: Convergence of coronary artery disease genes onto endothelial cell programs

Publication status: In press at **Nature**

Preprint available on BioRxiv:
<https://www.biorxiv.org/content/10.1101/2022.11.01.514606v1.article-info>

Authors & affiliations:

Gavin R. Schnitzler^{1,2,5*}, Helen Kang^{3,4,*}, Shi Fang^{1,5}, Ramcharan S. Angom⁶, Vivian S. Lee-Kim^{1,5}, X. Rosa Ma^{3,4}, Ronghao Zhou^{3,4}, Tony Zeng^{3,4}, Katherine Guo^{3,4}, Martin S. Taylor¹⁵, Shamsudheen K. Vellari^{1,5}, Aurelie E. Barry^{1,5}, Oscar Sias-Garcia^{1,5}, Alex Bloemendaal^{1,2}, Glen Munson¹, Philine Guckelberger¹, Tung H. Nguyen¹, Drew T. Bergman^{1,7}, Stephen Hinshaw¹⁶, Nathan Cheng¹, Brian Cleary^{1,8}, Krishna Aragam^{1,9}, Eric S. Lander^{1,10,11}, Hilary K. Finucane^{1,12,13,14}, Debabrata Mukhopadhyay⁶, Rajat M. Gupta^{1,2,5,†}, Jesse M. Engreitz^{1,2,3,4,17†}

1. Broad Institute of MIT and Harvard, Cambridge, MA
2. The Novo Nordisk Foundation Center for Genomic Mechanisms of Disease, Broad Institute, Cambridge, MA
3. Department of Genetics, Stanford University School of Medicine, Stanford, CA
4. BASE Initiative, Betty Irene Moore Children's Heart Center, Lucile Packard Children's Hospital, Stanford, CA
5. Divisions of Genetics and Cardiovascular Medicine, Department of Medicine, Brigham and Women's Hospital, Boston MA
6. Department of Biochemistry and Molecular Biology, Mayo Clinic College of Medicine and Science, Jacksonville, FL
7. Geisel School of Medicine at Dartmouth, Hanover, NH

ABSTRACT

Linking variants from genome-wide association studies (GWAS) to underlying mechanisms of disease remains a challenge^{1,4,6}. For some diseases, a successful strategy has been to look for cases where multiple GWAS loci contain genes that act in the same biological pathway^{1–6}. However, our knowledge of which genes act in which pathways is incomplete, particularly for cell-type specific pathways or understudied genes. Here we introduce a method to connect GWAS variants to functions, which links variants to genes using epigenomic data, links genes to pathways de novo using Perturb-seq, and integrates these data to identify convergence of GWAS loci onto pathways. We apply this approach to study the role of endothelial cells in genetic risk for coronary artery disease (CAD), and discover that 43 CAD GWAS signals converge on the cerebral cavernous malformations (CCM) signaling pathway. Two regulators of this pathway, CCM2 and TLNRD1, are each linked to a CAD risk variant, regulate other CAD risk genes, and affect atheroprotective processes in endothelial cells. These results suggest a model where CAD risk is driven in part by the convergence of causal genes onto a particular transcriptional pathway in endothelial cells, highlight shared genes between common and rare vascular diseases (CAD and CCM), and identify TLNRD1 as a new, previously uncharacterized member of the CCM signaling pathway. This approach will be widely useful for linking variants to functions for other common polygenic diseases.

Note: The list of authors for this protocol does not include all authors of the accompanying manuscript, only those who played a role in developing and executing the Perturb-seq method. For a complete list of manuscript authors, see the Manuscript Citation. Notes are provided to indicate which authors are best to contact for questions regarding specific methods.

IMAGE ATTRIBUTION

Helen Kang & Gavin Schnitzler

8. Faculty of Computing and Data Sciences, Departments of Biology and Biomedical Engineering, Biological Design Center, and Program in Bioinformatics, Boston University, Boston, MA
9. Cardiovascular Research Center, Massachusetts General Hospital, Boston, MA
10. Department of Biology, MIT, Cambridge, MA
11. Department of Systems Biology, Harvard Medical School, Boston, MA
12. Department of Medicine, Massachusetts General Hospital, Boston, MA
13. Analytic and Translational Genetics Unit, Massachusetts General Hospital, Boston, MA
14. Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, MA
15. Department of Pathology, Massachusetts General Hospital, and Harvard Medical School, Boston, MA
16. Department of Chemical and Systems Biology, ChEM-H, and Stanford Cancer Institute, Stanford University School of Medicine, Stanford, CA
17. Stanford Cardiovascular Institute, Stanford University, Stanford, CA, USA
*Equal contribution.
†Equal contribution.

License: This is an open access protocol distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited

Protocol status: Working

Created: Jan 12, 2024

Last Modified: Jan 19, 2024

PROTOCOL integer ID: 93454

GUIDELINES

INTRODUCTION TO THIS PROTOCOL

This protocol describes the steps performed for Perturb-seq in endothelial cells for the study: "**Convergence of coronary artery disease genes onto endothelial cell programs**", Schnitzler, Kang et al. (2024) Nature, in press, with a preprint available on BioRxiv: <https://www.biorxiv.org/content/10.1101/2022.11.01.514606v1.article-info>.

It is written to provide a complete account of the Perturb-seq experiments we performed: from establishing a CRISPRi cell line, to building a gDNA library, to creating and sequencing scRNASeq and dialout libraries (connecting guide sequences to cell barcodes). It is not meant to be a complete methods section for the manuscript. Where possible, we have also provided pointers for how researchers might apply this Perturb-seq approach to other cellular models and research questions.

Keywords: Perturb-seq, scRNA-seq, endothelial cells, coronary artery disease, GWAS, variant-to-function, consensus non-negative matrix factorization, CRISPRi

Funders Acknowledgement:

Broad Institute of Harvard & MIT
Grant ID: Variant-to-Function Initiative
NHLBI
Grant ID: R01HL159176
NHLBI
Grant ID: R01HL164811
NHGRI
Grant ID: UM1HG011972
NHGRI
Grant ID: R35HG011324
Stanford University
Grant ID: Gordon and Betty Moore and the BASE Research Initiative
NHGRI
Grant ID: K99HG009917
NHGRI
Grant ID: R00HG009917
Novo Nordisk Foundation
Grant ID: NNF21SA0072102
Harvard University
Grant ID: Harvard Society of Fellows
NHLBI
Grant ID: DP2HL152423
NHLBI
Grant ID: U01HL166060
Harvard Medical School
Grant ID: Khouri Innovation Award
Harvard Medical School
Grant ID: Braunwald Scholar Award
NHLBI
Grant ID: HL70567
Florida Department of Health
Grant ID: Cancer Research Chair's Fund Grant 3J-02
NIDDK
Grant ID: K08DK129824

Below, we provide a description of how the Perturb-seq approach described here was integrated into a Variant-to-Gene-to-Program method to link human Coronary artery disease variants to transcriptional programs in endothelial cells, followed by additional rationales for the study design (with a focus on Perturb-seq design).

INTRODUCTION TO THE V2G2P APPROACH FOR CAD

Genetic variants that influence complex traits are thought to regulate genes that work together in biological pathways. Identifying convergence on particular pathways can help in discovering genes and cellular functions that causally influence disease risk^{1–6}. However, it is often challenging to identify such convergence: complex traits involve contributions from multiple cell types; most risk variants are noncoding and can regulate multiple nearby genes; and it remains unclear which genes work together in which pathways in which cell types^{7–9}.

GWAS for coronary artery disease have discovered over 300 independent signals^{10–12}). 75% of these signals are not associated with circulating lipids, indicating the presence of undiscovered disease mechanisms that may function through cells in the coronary artery where atherosclerosis that causes CAD develops. Endothelial cells (ECs) are one of the most important of these arterial cells, controlling cholesterol uptake and efflux, smooth muscle cell responses, blood clotting and inflammatory immune cell recruitment^{13,14}, and are highly enriched for CAD heritability¹⁵. At a few individual CAD GWAS loci, noncoding risk variants have been shown to regulate the expression of key EC genes such as endothelial nitric oxide synthase (NOS3), endothelin 1 (EDN1), and others¹⁶. It remains unclear, however, which other genes in CAD GWAS loci might work together in which EC pathways to modulate disease risk.

To address these challenges, we have developed a new approach that systematically and unbiasedly links GWAS variants to genes and identifies their convergence onto specific disease-associated transcriptional programs. The 5 steps of this Variant-to-Gene-to-Program (**V2G2P**) approach, and their application to EC functions in CAD, are summarized below:

1. Identify a cell type and cellular model relevant to disease genetics, through enrichment of disease risk variants in enhancers in that cell type. Here, we focused on human arterial ECs, using telomerase-immortalized human aortic ECs (teloHAEC) as a model.
2. Build a map of variant-to-gene (V2G) links in that cell type, to link disease-associated variants to potential target genes. Here, we consider evidence from variants in EC enhancers, as well as coding regions and splice sites.
3. Build a map of gene-to-program (G2P) links in that cell type, by using Perturb-seq^{17–20} to systematically knock down all possible candidate disease genes and identify sets of genes that act together in biological pathways. Here, we knock down all expressed genes within ±500kb of 306 CAD GWAS signals, read out the effects of each

perturbation with single cell RNA-seq, and use unsupervised machine learning to define gene “programs,” unbiased by prior knowledge of gene sets or pathways.

4. Identify “disease-associated programs”, by developing a statistical test to determine whether the genes with links to risk variants are enriched in (that is, converge on) particular programs. Here, we find that many CAD GWAS loci converge on 5 gene programs identified de novo with Perturb-seq, which appear to correspond to branches of the cerebral cavernous malformations (CCM) signal transduction pathway.
5. Study the genes in disease-associated programs. Here, we nominate 41 genes likely to influence CAD risk through effects in ECs, and dissect two in detail: showing that knockdown of TLNRD1 or CCM2 mimics the effects of atheroprotective laminar blood flow, and that the poorly-characterized gene, TLNRD1, is a novel regulator in the CCM pathway.

In summary, the V2G2P approach defines cellular programs de novo using Perturb-seq, intersects these programs with enhancer-to-gene maps from the same cell type, and provides an interpretable, systematic, and unbiased framework for tracing the path from variant to gene to disease program simultaneously for all GWAS loci for a given disease and cell type.

GENERAL CONSIDERATIONS FOR APPLYING THE V2G2P APPROACH.

We aimed to create an approach to identify genes and programs relevant to disease risk that was cell-type specific, interpretable, unbiased with respect to prior information, and generally applicable to many cell types and complex traits. We and others have previously shown that combining both “top-down” information from gene programs and “bottom-up” approaches linking variants to genes can achieve higher specificity than either category of information alone^{3,31,32}. By combining GWAS, epigenomic, and Perturb-seq data, the variant-to-gene-to-program (V2G2P) approach expands upon these previous approaches by (i) generating variant-to-gene and gene-to-program maps in the same cell type; (ii) generating gene-to-program maps using Perturb-seq, providing a unbiased approach not dependent on previously known biological pathways or gene sets; and (iii) providing interpretable, testable hypotheses linking a specific variant to a gene to a program in a given cell type.

To implement this approach, we selected a cellular model enriched for heritability for the disease of interest. We constructed genome-wide enhancer-to-gene maps in endothelial cells by applying the Activity-by-Contact (ABC) model, which we recently showed performs well at linking noncoding variants to target genes in specific cell types^{9,22}. ABC outperforms other methods at predicting the effects of enhancers on target genes ^{9,22}, and requires minimal data inputs (e.g., ATAC-seq and H3K27ac ChIP-seq), allowing us, here, to apply the approach to link variants to candidate target genes in multiple endothelial cell states. We next created a catalog of gene programs and their regulators by applying Perturb-seq to systematically study all expressed genes in all GWAS loci for CAD.

Perturb-seq, which involves knocking down hundreds to thousands of genes in parallel and measuring their effects on gene expression using single-cell RNA-seq, has previously been shown to provide a high-content, unbiased view of cellular programs as represented in gene expression^{17–19}. Finally, we developed a simple statistical test to determine whether candidate disease genes might converge on particular gene programs by integrating gene-to-program information from Perturb-seq with variant-to-gene linking approaches.

Our approach to building a gene-to-program map using CRISPRi-Perturb-seq involved particular design and analysis considerations:

(i) We aimed to identify cellular programs and their related genes in an unbiased manner, such that we could look for enrichment of candidate CAD genes across a range of different endothelial cell pathways. This is in contrast to the approach of selecting a particular cellular phenotype (such as endothelial cell adhesion) that may or may not be important for the genetics of disease. Accordingly, we selected Perturb-seq due to its ability to perturb many hundreds or thousands of genes in parallel, and its ability to read out the effects on all genes in the genome, thereby providing a high-throughput and high-content readout of cell states.

(ii) Targeting all candidate genes near GWAS signals was important for the V2G2P approach. Specifically, we designed our Perturb-seq study to include all expressed genes within 500 Kb on either side of each CAD GWAS signal, as well as the two closest genes on either side if they were further than 500Kb, rather than selecting just a prioritized subset of genes. In practice, this resulted in us testing a median of 8 genes per CAD GWAS locus. This unbiased approach to selecting genes was essential for conducting the V2G2P enrichment test, which examines whether particular programs contain more genes with CAD variant-to-gene (V2G) links than expected by chance. This assessment would have been impossible if we had pre-selected only genes with V2G links to include in the Perturb-seq experiment. As such, the V2G2P enrichment test is applicable to experiments that perturb all expressed genes in all GWAS loci, or all genes in the genome.

(iii) The CRISPRi Perturb-seq approach was designed to read-out long term transcriptomic effects of gene knockdowns in the expected range of effect for common disease variants. We aimed to perturb genes in a way consistent with presumed effects of noncoding variants, which are thought to lead to quantitative changes in the expression of genes (rather than completely eliminating expression), and which might act over long periods of time to affect disease risk. Accordingly, we used CRISPRi to quantitatively knock down gene expression (average: 40% reduction). We then read out the effects after 5 days of doxycycline induction, to allow perturbations to propagate through the network and identify how perturbations affect stable gene expression programs.

(iv) We defined gene “programs” using an unsupervised machine learning approach (consensus non-negative matrix factorization), allowing us to identify sets of genes with similar properties (here, co-expression across single cells). This approach is independent

of and unbiased by prior knowledge about endothelial cell pathways — allowing us to avoid bias toward rediscovering or over-emphasizing known pathways, and identify new pathways if they exist. We did indeed discover gene programs that appeared to correspond to a wide range of biological pathways active in endothelial cells. Many of the 50 programs appeared to correspond to housekeeping pathways active in all cell types, because these genes/pathways (as well as EC-specific ones) are expressed and functional in endothelial cells.

MATERIALS

EC Media: Lifeline VEGF endothelial cell media (LL-0005) with all supplements added except for the Gentamycin antibiotic. In place of the provided antibiotic, instead supplement with 1x Penn/Strep (ThermoFisher #15070063 or equivalent, used at 1:100 dilution).

293T Media: DMEM (ThermoFisher #11965092 or equivalent) + 10% FBS (ThermoFisher #A5209501 or equivalent).

PBS: Any sterile phosphate buffered saline meant for tissue culture, such as Sigma-Aldrich #D8537-1L.

Trypsin: Any trypsin solution meant for tissue culture, such as ThermoFisher #R001100.

Tissue culture plates: For 10 cm plates use Corning tissue-culture treated culture dishes (Sigma Aldrich #CLS430599, or equivalent). Similarly for 15 cm dishes, or 6 or 12 well plates (be sure they are tissue-culture treated).

DMSO (any DMSO meant for cell culture, e.g. ThermoFisher #J66650.K2)

Opti-MEM (ThermoFisher #31985062 or equivalent)

Fugene6 transfection reagent (Promega #E2691)

Blasticidin 10mg/ml stock (Life Technologies #A1113903, or equivalent)

dCas9-KRAB-BFP (plasmid with CRISPRi machinery, which targets epigenetic repressors to efficiently silence enhancers or promoters^{53–55}, Addgene #85449)

rtTA (plasmid with the tetracycline activator, and with a hygromycin marker, Addgene #66810)

Hygromycin 50mg/ml stock (Life Technologies #10687010). Stored at 4°C in the dark.

Doxycycline (dox, a stable tetracycline analogue, Sigma #D5207 or equivalent).

Resuspended at 50mg/ml in water & kept frozen at -20°C. Working stock was 1mg/ml in PBS, also frozen at -20°C.

BsmBI restriction enzyme (NEB #R0739S, or equivalent)

SPRIselect beads (e.g. Beckman-Coulter #B23318)

FastAP (ThermoFisher #EF0654)

T4 DNA Ligase (New England Biolabs #M0202S)

NEB 5-alpha competent cells (e.g. New England Biolabs #C2987I)

Carbenicillin (a more stable analog of ampicillin, Sigma, C3416_5g)

RNeasy kit (Qiagen LLC #74134)

Reverse transcriptase kit (e.g. Thermo Fisher #4368814)

NEBNEXT PCR Master Mix (#M0541S)

NEB Gibson Assembly 2X Master Mix (New England Biolabs #E2611S)

Endura Competent Cells (Lucigen #60242-2)

Qiaspin miniprep kit (Qiagen #27106)

EndoFree maxikit (Qiagen #12362)

Q5 Hot Start Master Mix (New England Biolabs #M0492S)

BSA (ThermoFisher #J65097.A1, or equivalent)

10X Genomics Chromium Controller (10X Genomics)
3' scRNA-seq V3 kit (10X Genomics)
Phusion HiFidelity Master Mix (New England Biolabs #M0531S)

Choice of cellular model for Perturb-seq

- 1 General considerations for selection of a cellular model for Perturb-seq.** The cellular model should be relevant to the GWAS trait of interest. Here, we chose an endothelial cell model as particularly relevant to the genetics of coronary artery disease, because: 1) endothelial cells play several key roles that are directly relevant to coronary artery atherosclerosis that leads to CAD, including: control of cholesterol influx from the blood, control of immune cell recruitment, and regulation of smooth muscle cell functions through release of vasoactive molecules, such as EDN1 and nitric oxide [13,14,47](#), 2) previous studies have demonstrated strong enrichment of CAD heritability in endothelial cells (e.g. [131](#)), and 3) detailed studies of individual CAD GWAS loci have identified likely causal genes that are clearly related to endothelial cell functions, including NOS3[132](#), EDN1[14](#), JCAD[133](#), ARHGEF26[134](#), PLPP3[135](#), and AIDA[136](#).

We chose telomerase immortalized human aortic endothelial cells (teloHAEC) for these studies, because, while immortalized, they maintain important in vitro EC functions such as tubing, lipid transport and response to inflammatory stimuli [21,137](#). We confirmed that teloHAEC enhancers were enriched for heritability for CAD using S-LDSC (see accompanying manuscript). We also compared their gene expression profiles to those of primary coronary artery endothelial cells *in vivo* [69](#) and found that genes near GWAS signals were similarly expressed (described below). We expect that similar analyses will be useful for future applications of the V2G2P approach.

Notably, although we conducted our Perturb-seq experiment in resting, unstimulated conditions, we identified several programs related to specific stimulus responses. These included non-cell type-specific programs for unfolded protein response (UPR), DNA damage, heat shock, and inflammation, as well as endothelial-specific programs such as flow response and the endothelial to mesenchymal transition (endMT). Thus, knocking down genes with Perturb-seq in resting cells can, nonetheless, reveal gene programs relevant to various stimuli that may be informative for understanding disease mechanisms. It remains possible, however, that prioritization of certain disease-associated programs will require specific atherogenic stimuli (e.g. inflammatory cytokines or oxidized LDL), and further studies will be required to test this possibility.

- 2 Comparison to the same cell type in human disease or animal models.**

To validate the choice of cellular model, it can be useful to compare transcriptomes against scRNaseq data for the same cell type from the relevant dissociated human tissue. In our study, to confirm the validity of teloHAEC as a relevant model for endothelial cells in human coronary artery (where atherosclerosis that leads to CAD develops), we compared single cell RNA-seq gene expression from control guide carrying

teloHAEC from our Perturb-seq screen to scRNAseq data from explanted human right coronary artery endothelial cells (RCAECs)⁶⁹. We compared the gene expression at two levels: for all perturbed genes (2,285 genes) and for the 41 CAD associated genes. Among the perturbed genes in teloHAEC, 2,107 genes are expressed at TPM > 1 in healthy or disease RCAECs. We observed high correlation of gene expression in transcripts per million (TPM) between teloHAECs and RCAECs (Pearson correlation = 0.66, p-value = 6.45 x 10⁻²⁸⁰).

Note that, for this analysis, scRNAseq datasets are collapsed to single values per gene (e.g. pseudobulk). Thus, while we compared scRNAseq data from human tissues to scRNAseq data from TeloHAEC Perturb-seq, it would have been equally effective to compare to TeloHAEC bulk RNAseq data.

General considerations for culturing of TeloHAEC

- 3** Telomerase-immortalized human aortic endothelial cells (TeloHAEC) were purchased from ATCC (CRL-4052), and grown in Lifeline VEGF endothelial cell media (LL-0005) with 1x Penn/Strep (ThermoFisher #15070063 or equivalent, used at 1:100 dilution) in place of the provided Gentamycin solution ("EC Media"). Cells were plated at a density of 0.5-1.0 x 10⁶ cells per 10 cm plate and split before reaching 4 x 10⁶/plate (3 to 4 days).

To study responses to CAD-associated cytokines, cells were untreated (control), or treated with 10 ng/ml recombinant human IL-1 β (Millipore IL038), 10 ng/ml recombinant TNF α (Millipore GF023), or with normal media lacking VEGF (for TeloHAEC) or supplemented with VEGF (1x concentration from LifeLine VEGF media, for Eahy926), for 24 hours.

Note: Be sure that cell lines being used are mycoplasma free and are the correct lines. For instance, in addition to authentication by the provider, we further authenticated each line by analysis of microscopic morphology (e.g. TeloHAEC displayed the characteristic EC cobblestone morphology and showed localization of VE-Cadherin to endothelial cell junctions), mapping of ATAC-seq, ChIP-seq and RNAseq reads to the human genome, and RNAseq profiles and responses (e.g. expression of EC-specific genes and observation of previously-observed responses to stimuli, such as IL-1 β).

Note: TeloHAEC are puromycin resistant, since that marker was used on the telomerase expression vector. Other selectable markers that work with TeloHAEC are blasticidin, hygromycin & G418.

Thawing TeloHAEC and derivatives

- 4** Take a cryovial of cells out of storage (liquid nitrogen for long term, or -80°C for short term), and thaw in a 37°C bead or water bath, with occasional swirling, until only a tiny bit of ice remains. Vials should contain ~0.5e6 to 1e6 cells to be plated to a single 10cm plate. Adjust the size and number of plates up or down if you have more or fewer cells than this.

- 5 Mix cells by gentle shaking, spray off the outside with 70% EtOH, flick down to bring all cells/media to the bottom of the tube & transfer to a tissue culture (TC) hood.
- 6 Remove cell solution to a 10 cm **tissue culture plate**. For 10 cm plates use Corning tissue-culture treated culture dishes (Sigma Aldrich #CLS430599, or equivalent). Similarly for 15 cm dishes, or 6 or 12 well plates (be sure they are tissue-culture treated). Add ~10x volume (e.g. 5ml for 0.5ml of cell solution) of 37°C prewarmed EC media dropwise, while swirling the plate to mix. After the first ~2ml, you can speed up the process and add 0.5ml at a time. Collect the diluted cell solution by pipette, move to a 15 ml conical tube & cap.
- 7 Centrifuge at 300xG for 5 minutes at room temperature, bring back to the TC hood and remove supernatant above the cell pellet by aspiration. Be careful not to disturb the pellet.
- 8 Add 8 ml EC media to a 10 cm plate (can be the same one you used to dilute the frozen cell solution). Resuspend the cells in 2 ml EC media (using a 2ml filter pipette), add to the plate and mix.
- 9 Incubate in a sterile tissue culture incubator at 37°C, with 5% CO₂, and kept humid by a pan of water in the bottom.

Check after several hours, or the next day, to confirm that most of the cells have adhered. Generally a high fraction of TeloHAEC will survive freezing, and one can assume that the number of cells that were frozen will be close to the number plated. If a large fraction of cells are rounded and floating after a day, however, you may need to allow more time for them to grow out before splitting.

Routine culturing of TeloHAEC and derivatives

- 10 Split cells when they reach ~4 million/10 cm plate. They will appear confluent but not dense (a cobblestone appearance). Do not grow to a density more than ~6 million/10 cm plate (above which cells will appear elongated and aligned, like a fingerprint pattern), as this will reduce their ability to be transduced by lentivirus, and may alter other phenotypes.

TeloHAEC have a doubling time in EC media of ~36 hours. Hence, if you plate 1M cells, you will have ~4M after 3 days. Or if you plate 0.5M cell you will have ~4M after a bit more than 4 days.

11 Splitting TeloHAEC:

In a TC hood, aspirate off the old media, and wash once with 10 ml room temperature **PBS** solution (any sterile phosphate buffered saline meant for tissue culture, such as Sigma-Aldrich #D8537-1L).

- 12 Aspirate off the PBS, add 1ml **Trypsin/EDTA** solution (any trypsin solution meant for tissue culture, such as ThermoFisher #R001100),, rock the plate to distribute well and return the plate to the incubator for ~3 to 5 mins.
- 13 When cells are evidently loose (visible as a cloud when tipping the plate or clearly loose and rounded up when examined at 4X magnification under a scope), return the plate to the hood, add 7 ml media, mix by pipetting and move to a 15 ml sterile conical tube.
- 14 Before cells settle down, use a sterile filter pipette tip to take a sample to count. Volumes will depend on your counting method. For instance, using a Countess cell counter, take 10 µl cells + 10 µl trypan blue solution and add to one chamber on the Countess slide.
- 15 Cap the conical tube and centrifuge 300xG for 5 mins at room temp. Meanwhile count your sample of cells. Count live cells, being sure to consider any dilution factor.
- 16 In the TC hood remove supernatant from above cell pellet by aspiration, and resuspend cells in 2 ml prewarmed EC media. Calculate the volume of cells to get 1 million (resuspended cell density will be 4x higher than your count, since you've gone from 8 ml to 2 ml after centrifugation), and plate this volume to a 10cm plate containing 10ml of prewarmed EC medium.

Use 2 million cells if plating to 15 cm dishes, 160,000 if plating to 6 well plates, 80,000 if plating to 12 well plates, etc.
- 17 Split cells again after 3 days. Or, if you plated 500k to 600k cells, split after 4 days.

Freezing TeloHAEC and derivatives

- 18 Follow the procedure for splitting up to the point where you have counted cells, spun them down and aspirated off the media.
- If you wish to continue to culture a portion of the cells while freezing the rest, you can aliquot to two 15 ml conical tubes, using one for splitting and the other for freezing.
- Alternatively, you may follow the splitting procedure, then re-spin any unused cells at 300xG for 5 mins and aspirate off the media over the cell pellet, before continuing with the freezing protocol.
- 19 Prepare a "Freeze mix" solution of 90% FBS 10% **DMSO** (any DMSO meant for cell culture, e.g. ThermoFisher #J66650.K2) to give ~400 µl per million cells you wish to freeze.
- 20 Resuspend the cell pellet in the correct volume of freeze mix and aliquot to labeled cryotubes (~1M cells per tube). Freeze tubes at -80°C overnight. After this you may move the tubes to liquid nitrogen storage.
TeloHAEC will survive storage at -80°C for a year or more, with liquid nitrogen recommended for longer term storage of banked stocks.

Culturing HEK 293 T cells

- 21 HEK293T cells were purchased from ATCC (CRL-2922).
- HEK293 T cells are thawed, split and maintained the same as for TeloHAEC, with the following exceptions:
- 1) Use DMEM (ThermoFisher #11965092 or equivalent) + 10% FBS (ThermoFisher #A5209501 or equivalent)+1x Penn/Strep (ThermoFisher #15070063 or equivalent) as a growth medium ("**293T medium**").
 - 2) Plate at ~1.5M cells/10 cm plate and split when they reach ~12M cells/plate (~ 3 days, should be visibly confluent or near-confluent but not overcrowded). Do not grow too dense, or this may reduce their transfectability.
 - 3) The recommendations above assume a doubling time of ~24 hours. Different lab stocks of 293T cells may have different doubling times, and it is worth calculating the doubling time for your stock, and adjusting splitting parameters accordingly.
 - 4) Freeze mix for 293T cells is DMEM + 30% FBS + 10% DMSO.

Preparation of lentivirus (3 plasmid system)

- 22 **Considerations for using lentivirus.**
Using the 3 plasmid system (separate plasmids encoding the viral genome, env and gag/pol proteins), it is extremely unlikely that you can produce virus that is capable of infecting cells and producing more virus.

Nonetheless, lentiviral genomes integrate into the DNA (potentially causing damaging mutations) and the proteins you may express in your viral vector can themselves be harmful (particularly oncogenes, etc.). Thus, it is recommended to use particular care when handling lentivirus. Many institutions have specific safety protocols (and potentially dedicated TC rooms) for lentiviral work. Check with your institutions health and safety regulations and lab protocols, and follow those guidelines.

23 Day 0:

Harvest and count a plate of 293T cells (grown as described under "Culturing HEK 293 T cells" above). Make a cell suspension at the concentration of 250,000 cells/ml in 293T medium, and plate to 6 well tissue culture treated sterile plates (2ml/well). Include 1 well per ~1.5 ml of lentivirus solution you wish to make.

Alternatively, you can plate the morning of transfection, using 500k cells/ml, and allowing at least 4 hours for cells to settle down and adhere before transfection.

24 Day 1:

A: Prepare DNA mix of plasmids:

In a sterile Eppendorf tube combine:

- 1) 25 µl **Opti-MEM** (ThermoFisher #31985062 or equivalent).
- 2) 1.2 µg transfer plasmid (expressing guide, dCAS9-KRAB-BFP, selectable marker, etc. between viral LTR sequences)
- 3) 1 µg psPAX2 (encoding gag & pol proteins, Addgene #12260)
- 4) 0.3 µg pMD2.G (encoding envelope protein, Addgene #12259)

DNA can be purified by standard miniprep or midiprep kits, but should have a good A260/A280 ratio and not be too dilute (<100ng/µl).

B: Dilute transfection reagent

To 167 µl of OptiMEM add 8 µl **Fugene6** (Promega #E2691) and mix by swirling pipette tip or flicking the tube (do not pipette up and down or vortex). Incubate at room temperature for 5 mins.

C: Mix DNA and diluted transfection reagent

Add 175 µl of transfection reagent to each DNA mix, and mix by swirling pipette tip or flicking the tube (do not pipette up and down or vortex). Incubate at room temperature for 15 mins.

D: Add to cells

Add the 200 µl of transfection reagent/DNA to cells dropwise, distributing drops around the well. Let sit ~5 mins at room temp, then mix gently by swirling. Return cells to incubator.

25 Day 2:

~18 hours after transfection, aspirate out the media in your transfection wells, and replace with DMEM + 20% FBS (and with no Penn/Strep).

26

Day 3:

Harvest virus by pulling up the media in each well with a 5 ml Luer-lock syringe. Secure a Luer-lock syringe type 0.45 µm filter (large enough for virus to pass through, but not cells) to the end of the syringe, and press viral stock solution through the filter into labeled cryotubes (2 per well).

Virus can be stored at 4°C for a week or -80°C for over a year. Do not store virus in liquid nitrogen (to avoid chances of viral contamination if a tube explodes), unless you have a specific approved protocol that minimizes this danger.

Lentiviral titer and small scale transductions of TeloHAEC cells

27

Introduction

This protocol is useful for generating cell lines with specific combinations of vectors. For our studies, we used it to 1) Engineer our CRISPRi TeloHAEC cell line, and 2) Create CRISPRi TeloHAEC derivatives expressing single guides to individual target promoters.

28

Day 0: Plate TeloHAEC

Plate TeloHAEC at 50,000 cells per well in 1 ml EC media to 12 well tissue culture dishes. Allow 3-5 wells per viral stock to be tested, and an additional 2 wells for controls (plus 1 extra well for each selectable marker to be used, if more than one). For 2 wells prepare control cell dilutions 1:10 (900 µl media + 100 µl cell suspension) and 1:100 (990 µl media +10 µl cell suspension).

Note: If you want to move straight from viral production to transduction, perform this step at the same time as the Day2 media change in the viral production protocol.

29

Day 1:

Add 1 µl of 10mg/ml polybrene transfection reagent to each well of cell (10µg/ml final) and mix gently.

Add different volumes of each virus to sets of 3 to 5 wells. For a rough titration try 2, 20, and 200 µl virus. For a finer titration, try 1,4,16, 60 and 200 µl virus. Mix gently.

Put the plate in a plate-adaptor fitted swinging bucket rotor and centrifuge at 2250 RPM for 30 minutes at 30°C.

After centrifugation, move the plate to the 37°C incubator.

30

Day 2:

Aspirate the media off each well and replace with EC medium + antibiotic to select for cells that received virus. For our TeloHAEC studies we generally used **Blasticidin** at 15µg/ml final (Life Technologies

#A1113903, or equivalent). If your vector expresses a G418 or Hygromycin resistance marker, you can also use those for selection (each at ~100 μ g/ml final concentration). Add antibiotic to 1 well of undiltued, uninfected cells.

31

Day 5:

Check your plates. By this point all cells in your uninfected + antibiotic well should be dead (for Blasticidin, it may take an additional 1 or 2 days for G418 or Hygromycin). Your + virus wells should have an increasing number of surviving cells based on the volume of virus.

Rough estimate of viral titer

This may be adequate if all you want to do is get a rough sense of how much virus to use so that a reasonable fraction of cells get virus but few cells get more than one virus (e.g. aiming for an ~20% transduction rate). Compare +virus+antibiotic wells to your 1:1, 1:10 and 1:100 no virus no antibiotic wells and use this to get a rough sense of titer. E.g. if the number of cells in your 2 μ l virus well is similar to the 1:10 novirus well, then 2 μ l virus gives a multiplicity of infection of ~0.1 (10% of cells get virus). If you did the 5 point titration you're likely to have one well with close to your aimed-for infection rate, which you could immediately grow out and use in your studies.

Measured calculation of viral titer

- A. Remove media & wash wells you wish to quantitate (including all no virus, no antibiotic 1:1, 1:10 and 1:100 wells) with 1 ml PBS,
- B. Remove PBS and add 150 μ l trypsin and incubate at 37°C for 3-5 mins.
- C. Resuspend cells with 850 μ l EC medium (1 ml total) and count a sample of cells from each well. Make sure to resuspend cells thoroughly, so the count will be accurate. One way to ensure this is to draw up the whole volume of cells in a 1ml pipette tip, place the tip flush against the bottom of the well, and dispense under pressure (~10 seconds for the whole volume).
- D. Multiply the counts for the 1:10 and 1:100 controls by 10 and 100. Consider the maximum value to be equivalent to 100% infection. Note, the 1:1 well will already be very dense and will likely have retarded growth, and, so, is likely not to be the best control. Typically, the 1:10 well is best.
- E. Plot cell counts for each volume of virus normalized the the control max to determine infection rate. You can use these curves, then, to extrapolate volumes needed for any particular MOI.

Note: This approach does not control for potential viral toxicity (where cells are made sick or killed at very high volumes of virus). To properly account for this, you should count a no antibiotic control for every volume of virus used and calculate $(\text{counts}+\text{antibiotic})/(\text{counts}-\text{antibiotic})$ for each volume of virus. This is less important if you plan low MOI transductions (where toxicity should not be an issue), but may be critical if you want to titer the virus for high MOI experiments (e.g. 5 to 10 viruses per cell).

Considerations for viral storage and titrations

Virus can lose potency with freezing and (over the long term) with storage at -80C. If you wish to measure viral titer on one day and have it be accurate months later, freeze your virus in many aliquots. Thaw one

aliquot to perform your titration and do not refreeze any left over virus. In this way your titration should be accurate for any "thawed once" tube of virus.

Generation of an inducible CRISPRi cell line

32

Overview:

To create the TeloHAEC CRISPRi line, cells were transduced with lentiviral vectors containing 1) dox-inducible (tetracycline operator controlled) **dCas9-KRAB-BFP** (CRISPRi machinery, which targets epigenetic repressors to efficiently silence enhancers or promoters^{53–55}, Addgene #85449) and 2) **rtTA** (tetracycline activator) with a hygromycin marker (Addgene #66810).

Note: this same approach is likely to be applicable to other immortalized adherent culture cells to meet different research goals.

33

Preparation:

Generate lentivirus for each vector (dCas9-KRAB-BFP/CRISPRi & rtTA) and titer the virus according to the two protocols above.

34

Day 1:

Transduce TeloHAEC with both viruses so that each is at an MOI of 1 (on average 1 of each virus per cell), using the "Lentiviral titer and small scale transductions of TeloHAEC cells" protocol. Transducing multiple wells will speed up the process.

35

Day 2:

To select for cells that received the rtTA vector, perform **hygromycin** selection (Life Technologies #10687010 250 µg/ml) for 4 days.

36

Day 6:

To select for cells that also received the tet operator controlled dCas9-KRAB-BFP vector, split cells, and grow one plate of cells with normal EC medium and another with 1 µg/ml **doxycycline** (dox, a stable tetracycline analogue, Sigma #D5207 or equivalent) for 3 days.

37

Day 9:

Harvest both plates of cells, resuspend in EC medium, and do a FACS sort for BFP positive cells as follows.
i. Look at the no dox cells in a scatterplot with axes FSC-A and SSC-A, and establish a gate selecting for intact monomeric cells.

- ii. Plot BFP signal for gated cells (BP450-A channel), to establish a baseline for BFP negative cells.
- iii. Use the same FSC-A and SSC-A gate for the +Dox treated cells. You should see a second peak in the BP450-A channel for BFP positive cells (cells that received both vectors, and in which rtTA binds to the tet operators in the CRISPRi vector to turn on dCas9-KRAB-BFP expression).
- iv. Sort for the top 15% of BFP expressing cells.

38 Confirmation of stable CRISPRi expression:

Grow the sorted cells for at least 1 week in the absence of dox. Repeat the 3 day dox induction and examine by FACS. If more than 5% of dox-treated cells are now BFP negative (likely because of epigenetic silencing of the CRISPRi or rtTA vectors), repeat the FACS sort.

39

Growth and Storage:

Grow confirmed BFP+ sorted cells out in the absence of Dox, and freeze aliquots for later use.

40 Diagnostic test concurrent with Pertub-seq:

We performed diagnostic FACS with a sample of cells after expansion and dox induction for the "Perturb-seq library preparation" step, below, to confirm that these manipulations had not reduced the efficacy of doxycycline control of the CRISPRi machinery.

Creation of a blasticidin resistant CROPseq vector

41

Because TeloHAEC are puromycin resistant, we adapted the CROP-opti vector ([20](#), Addgene, #106280) for Blasticidin resistance ("CROP-opti Blast"), by digesting the vector with BsiWI and MluI, PCR-amplifying the Blasticidin resistance gene from lenti-dCas-VP64_Blast (Addgene, #61425) with added homology arms, and performing Gibson Assembly (Gibson Master mix, New England Biolabs). To create "CROP-opti-BC-Blast", we added HyPR-Seq barcodes between the WPRE element and the U6 promoter of CROP-opti-Blast, as described in [60](#).

These vectors are available on request. For most cell lines, that are not already puromycin resistant, however, the original CROP-opti vector is the best choice ([20](#), Addgene, #106280)

Preparation of single gene CRISPRi knockdown lentiviral vectors

42 Introduction and rationale.

Being able to knock down individual genes of interest can be useful at both early and later steps in the V2G2P process, including:

- 1) Validation of the efficacy of CRISPRi knockdown after BFP selection of your doxycycline-inducible CRISPRi line.

- 2) Validation of Perturb-seq transcriptional results, and comparison of Perturb-seq transcriptomic effects with those of related genes not tested in the original library.
 - 3) Downstream phenotypic assays to connect transcriptional effects of V2G2P genes with relevant cellular phenotypes.

We used this approach for all three of these purposes in our studies.

43 Selection or design of CRISPi guides

CRISPRi guides work best when they target the immediate promoter region of target genes (from -150 to +100 relative to the Transcription Start Site (TSS)). The efficacy of any given guide can not be accurately predicted, however, and even with the best design software some fraction of guides, perhaps ~50%, will be weak or non-effective.

One approach is to select guides from ones that have already been designed, one of the best sources for which are the Dolcetto libraries (Sanson, K.R., Hanna, R.E., Hegde, M. *et al.* Optimized libraries for CRISPR-Cas9 genetic screens with multiple modalities. *Nat Commun* **9**, 5416 (2018).

<https://doi.org/10.1038/s41467-018-07901-8>). A caveat to this approach is that, while this set of guides has been shown to be generally effective for genes that affect cell growth, every guide to each target gene has not been validated for efficacy.

Another approach is design your own guides. This allows you to target promoters that may not have been targeted in the Dolcetto libraries, and densely tile your guides (e.g. we designed 15 guides per target promoter). For our studies we used our established pipeline for CRISPRi guide design (9,22, available at <https://github.com/EngreitzLab/CRISPRDesigner>).

A final alternative is to choose guides validated by you or other labs. For instance, our Perturb-seq studies provided information about which of the 15 CRISPRi guides were effective, that we used to select guides to create single gene knockdown lines.

Well established control guide sets can be taken from published resources. We included 400 non-targeting guides (that do not have close matches to any region in the human genome) and 600 safe targeting guides (targeting non-genic regions lacking enhancer marks) 53

44 Generation of single-guide CRISPRi vectors and TeloHAEC derivatives

A. For each guide, order a pair of oligonucleotides (e.g. from IDT... standard non gel-purified guides work well). The oligo sequences will be:

Top: 5' CACCGNNNNNNNNNNNNNNNNNNNNNNNNNN N=guide

Bottom: 5'AAACnnnnnnnnnnnnnnnnnnnC n=reverse complement of guide

Note: Including a G as the first transcribed base (position 5 on the top strand, and it's reverse complement C as the 3' terminal base on the bottom strand) helps increase guide expression. If the genomic sequence targeted by the guide does not start with a G, add a G to the 5' end of the guide sequence.

Resuspend oligos at 100 µM in water and store frozen at -20C.

B. Prepare Backbone

1. Digest 5 µg of vector with 50 units of **BsmBI** restriction enzyme at 55°C for 4 hours (e.g. NEB #R0739S, using manufacturer-supplied buffers). Warning: BsmBI can lose activity soon after it's marked expiration date. If you are not using a new BsmBI stock, check a small sample of this digest on a gel to be sure you see cutting (clean ~7 kb vector and ~1.8kb filler bands before proceeding).
2. **SPRI bead purification:** Purify by adding 0.7 volumes of SPRIselect beads (**SPRI beads**, Beckman-Coulter #B23318), mixing and incubating for 5 minutes at room temperature.
3. Place tube in a magnetic stand until beads clear from the solution and adhere to the side.
4. Remove supernatant, and with the tube still on the magnetic stand wash 2x with 1.7 initial volumes of 80% ethanol (removing ethanol by pipette).
5. Briefly spin the tube, put back on the mag stand and remove any remaining ethanol.
6. Let air dry for 5 minutes, then resuspend in 22 µl of water.
7. Let sit for ~ 5 minutes, then put back on the mag stand until the beads clear, and collect 20 µl cut plasmid.
8. Redigest with 25 units BsmBI at 55°C for 2+ hours.
9. Dephosphorylate the 5' ends of the vector by adding 5µl **FastAP** (ThermoFisher #EF0654) and incubating at 37°C for 20 minutes. Note: this step is omitted when preparing the vector for Gibson cloning of complex guide libraries ("Preparation of lentiviral CRISPRi gRNA library" below).
10. Heat inactivate FastAP at 80°C for 15 minutes.
11. Purify the vector by separation on a 1% agarose gel. If adding loading dye directly to the BsmBI/FastAP reaction results in too great a volume to load, you can perform the SPRI select bead purification from steps 2 though 7 to reduce the volume. Note that gel purification is necessary to remove the ~7kb vector from the ~1.8kb filler sequence.
12. Measure cut, dephosphorylated vector concentration by nanodrop spectrophotometer. Make a working stock at 10ng/µl.

C. Phosphorylate and anneal oligos

1. In a PCR tube strip, dilute top and bottom guideRNA oligos to 2.5 µM separately in 4 µl water. Add an extra well 4 µl water only to measure self-ligation of backbone.
2. Heat to 85°C for 2 minutes. Quickly transfer to ice.
3. Phosphorylate each oligo (top and bottom, separately) by adding 4 µl of master mix:

H₂O 1.6 µl
 10× NEB PNK Buffer 0.8 µl (contains DTT)
 10 mM ATP 1.2 µl
 NEB T4 PNK (add last) 0.4 µl (NEB T4 polynucleotide kinase, #M0201S)

Note: T4 PNK does not come with ATP, which must be added separately. Alternatively, use NEB T4 DNA Ligase buffer, which includes ATP and in which T4 PNK has 100% activity.

Mix gently but thoroughly, spin down, and incubate at 37°C for 30-60 minutes.

4. Mix Top and Bottom oligos together, denature and slowly anneal them in a PCR machine:
 2 min 98°C, 2 min 85°C, 5 min 75°C, 5 min 65°C, 10 min 22°C.

D. Ligation, Transformation and confirmation of guide inserts

1. Make the following master mix (in this order) and combine 5 µl mix with 5 µl annealed oligos from above.

H₂O 2.5 µl
 10× NEB T4 DNA Ligase Buffer 1 µl
 BsmBI-digested vector (10ng/µl) 1 µl
NEB T4 DNA Ligase (add last) 0.5 µl. (New England Biolabs Inc #M0202S)
(annealed oligo mix) (5 µl)
 10 µl total

2. Incubate at room temperature for 30 minutes.

3. Thaw competent **NEB 5-alpha cells** (New England Biolabs #C2987I, or, alternatively, NEB stable) on ice.

4. Chill ligation reaction on ice and add 1 µl ligation reaction to 5 µl cells. Mix gently.

5. Incubate on ice for 20 min.

6. Heat shock 42°C for 40 seconds, then return to ice for 2 minutes.

7. Add 50 µl SOC media (provided with competent cells) and incubate at 30°C for 1 hour with shaking.

8. Streak out onto **carbenicillin** (a more stable analog of ampicillin, Sigma, C3416_5g, or equivalent) plates (50 µg/ml) and incubate overnight at 30°C.

9. Pick colonies, and grow in 3 ml LB plus carbenicillin (50 µg/ml) at 30°C overnight.

10. Miniprep the DNA and send for Sanger sequencing using primers flanking the BsmBI insertion site.

Guide reference: Some guides from our studies that we have validated as effective for knock down of target gene expression in CRISPRi TeloHAEC (TargetGene_CloneIndex: ForwardSequence) are: CCM2_C2: GGCAAGAAGGTGAGCGTGCG, CCM2_F6: GAGCCGCTACATGCTCGACCC, CDH5_B8: GCCAGCTGGAAAACCTGAAG, CDH5_D5: GTTGGACTGCCTGTCCGTCCA, ITGB1BP1_C7: GAAGGCCGCGGCACTCCCACG, ITGB1BP1_G8: GAAGTCCGCAACCCGGGGAT, KLF2_C9:

GGACCCGGGGAGAAAGGACG, KLF2_G10: GCCGCCTATATAAGCCGGC, MAP2K5_B5: GTCTGCCCAACCGGAGACAC, MAP3K3_A4: GTTCCTGAGGTGGAGAACGG, MAP3K3_C3: GCCAATAACAAGAAGGAAGT, MEF2A_C10: GCGGCGCGAAGCGCTGGTGG, MEF2A_H10: GACTGAATTATCCTCTCGGT, NFAT5_D4: GGCCTCGCTTCCTGCCGGCG, NFAT5_D7: GGTCCCCGTCCCGCCGGGG, PDCD10_D11: GACCGAGCAGAAGAGGTCTA, PDCD10_G1: GCCGCTTACGCCACTCGCGT, TLNRD1_B3: GTGGCTGCCGCCGCCGCA, TLNRD1_D12: GCCTCCGGCAGCCCCCTGCGGG. If any of these genes is expressed well in your model cell line, these should work as effective positive controls to test your CRISPRi cell line.

For negative controls, we used nontargeting guides: Negative_control_B6: GCAACGGTGTACCGCGGATC, Negative_control_D2: GTGGTTCACAACCGGACCCA, Negative_control_D8: GGTGGTTCGGTTGCGTGGCC, Negative_control_F4: GCTGGCGGACGTTGGATA,

One guide, MAP2K5_A11: GCCGAGGCCGCGCGGACTGG, was not effective.

Confirmation of CRISPRi efficacy with sgRNA guides

45

- A. Prepare single guide virus for a gene well expressed in your cell line and a negative control guide virus, as per the "Preparation of Lentivirus (3 plasmid system)" method above.
- B. Transduce your dox-inducible CRISPRi line with a titration of virus using the "Lentiviral titer and small scale transductions of TeloHAEC cells" protocol.
- C. After drug selection, choose a well in which 10 to 40% of cells received virus, count, and split to give at least 2 6-well wells per transduction (seeded at 150,000 cells per well).
- D. Treat all but one well with 2 μ g/ml doxycycline for 3 days, to induce knockdown of the targeted gene.
- E. Harvest RNA using an **RNeasy** kit (Qiagen LLC #74134), following the manufacturer's instructions for adherent cultured cells. Alternatively, use any kit capable of purifying RNA for cDNA synthesis. Measure concentration on a nanodrop spectrophotometer.
- F. Convert 0.5 μ g of RNA to cDNA using any **reverse transcriptase** kit (e.g. Thermo Fisher #4368814).
- G. Perform qRT-PCR using primers to the targeted gene as well as a reference control (e.g. GAPDH). One straightforward approach to this is to use pre-optimized TaqMan probes, a protocol for which can be found here: https://tools.thermofisher.com/content/sfs/manuals/cms_041280.pdf.

Analyze your data and determine fold change in normalized target gene expression relative to controls. If your CRISPRi cell line is working, you should see at least 2-fold decrease in expression of the target gene

with the guide directed to it in the presence of Dox (but no effect with the control guide, or in the absence of Dox).

Selection of target genes for Perturb-seq

- 46** Design of your perturb-seq gDNA library will depend on your system and your experimental goals. Here we describe our study of CAD GWAS loci in endothelial cells, but the overall approach should be generally applicable.

We constructed a library of promoter-targeted CRISPRi guides to all potential causal CAD genes. First, we identified all coding genes within a 1 megabase window surrounding the lead SNPs from CAD loci identified in either or both of van der Harst et al.¹⁰ and Aragam et al.¹² that were expressed in TeloHAEC (1+ TPM, from bulk RNA-seq). If fewer than 2 expressed genes were found within 500kb up- or downstream of the lead SNP, the window was expanded to include the closest 2 genes to each side (for a total of 1661 genes). Non-coding genes were generally excluded, unless there was strong evidence for regulatory functions, particularly in ECs. Selected genes with TPM <1 were included, particularly if they were known to be important for CAD in tissues where they were more highly expressed (e.g. PCSK9), or were regulated by IL1-beta in bulk RNA-seq data in TeloHAEC (FDR<0.05, fold change >1.3).

As negative controls, we included guides targeting 48 coding genes expressed in other cell types but not detectably expressed in ECs, and the 132 expressed coding genes within 1 Mb of 16 randomly-selected lead SNPs associated with Inflammatory bowel disease, Crohn's disease or Ulcerative colitis ⁵⁹, and which did not overlap with CAD loci.

As positive controls, and to aid in connecting candidate CAD genes to known pathways in ECs, we targeted the promoters of an additional 284 genes with known roles in a wide range of CAD-relevant EC functions such as barrier formation, TGF-beta signaling and inflammation, as well as major classes of expressed transcription factors and common essential genes.

We also targeted an additional 160 promoters of expressed genes predicted to be regulated by EC enhancers containing fine-mapped variants associated with other disease phenotypes expected to be modulated by ECs (migraine, blood clotting in leg, systolic blood pressure, diastolic blood pressure & mean arterial blood pressure, from UKBB). This gave a total of 2285 genes, some of which were members of more than one category.

Preparation of lentiviral CRISPRi gRNA library

- 47** **A.** sgRNA guides were designed to target promoters of the chosen CAD and control genes (15 guides spanning from -150 to +100 relative to the Transcription Start Site (TSS)), using our established pipeline (^{9,22}, <https://github.com/EngreitzLab/CRISPRDesigner>). See "Selection or design of CRISPRi guides", above, for guide design options. We included 400 non-targeting guides (that do not have close matches to any

region in the human genome) and 600 safe targeting guides (targeting non-genic regions lacking enhancer marks) 53. All together, this library had 37,637 guides.

B. A pool of oligos encoding the guide sequences, plus extensions with homology to the U6 promoter and downstream scaffold (F: 5' TATCTTGTGGAAAGGACGAAACACCG & R: 5' GTTTAAGAGCTATGCTGGAAACAGCATAG) was synthesized by Agilent Technologies. Note, as described in "Preparation of single gene CRISPRi knockdown lentiviral vectors", the terminal G in the forward primer is to optimize gRNA expression, with the guide protospacer sequence beginning after that (unless the protospacer sequence happens to start with G).

C. Resuspended the oligo pool in water at 100 µM concentration. Dilute an aliquot to 1ng/µl.

D. Make a mix of:

30µl 2× **NEBNext PCR Master Mix (#M0541S)**
3µl 10 µM primer CI0111_CRISPRPoolAmp_F (GGCTTTATATATCTTGTGGAAAGGACGAAACACCG)
3µl 10 µM primer CI0112_CRISPRPoolAmp_R
(CTTATTAAACTTGCTATGCTGTTCCAGCATAGCTCTTAAAC)
2 µl 1ng/µl oligo pool
22 µl water

Be sure to include a no oligo insert control.

E. PCR amplify in two stages:

98°C 30 secs
98°C 15 secs, 62°C 15 secs, 72°C 15 secs (4 cycles)
98°C 10 secs, 72°C 15 secs, 72°C 5 secs (10 cycles)
72°C 120 secs
4°C. hold

F. Run a small sample on a gel to verify the presence of an ~100 bp product. If the band is present but very faint, return the tube to the PCR machine and amplify for another 3 to 6 cycles with the program: 98°C 30 sec, [98°C 10 sec, 72°C 15 sec, 72°C 5 sec] (3 to 6 cycles), 72°C 120 sec, 4°C hold.

G. Clean up the PCR reaction by adding 90µl (1.5X volume) SPRI select beads and follow the SPRI bead purification method described under "Preparation of single gene CRISPRi knockdown lentiviral vectors" step B. Measure concentration on a nanadrop spectrophotometer.

H. Prepare your BsmBI vector backbone as described under the "Prepare backbone" step in "Preparation of single gene CRISPRi knockdown lentiviral vectors" above, **except** that you should omit the Fast AP dephosphorylation step. Measure concentration of the purified vector on a nanodrop spectrophotometer.

I. Assemble the Gibson Reaction as follows:

NEB Gibson Assembly 2X Master Mix 15 µL (New England Biolabs #E2611S)

BsmBI-digested Vector 500 ng

SPRICleaned PCR2 Product 70 ng

Water to 30 µl

Incubate for 1 hour at 50°C.

J. Purify Gibson-assembled plasmid with 21 µl (0.7x volume) AMPure XP SPRI beads (as per step G). Elute in 10 µl water.

K. Electroporate into **Endura Competent Cells** (Lucigen #60242-2), according to the Lucigen electroporation transformation protocol, steps 1 through 8

(https://biosearchtech.a.bigcontent.io/v1/static/manual_COMCEL-002_Endura-Competent-Cells), as modified below:

Step 0. Pre-chill 20µl and 200µl pipette tips at 4°C for 30 mins before beginning.

...

Altered step 5. Mix 8 µl of the SPRI bead-purified Gibson reaction product with 25µl cells (include a reaction for your no-insert control).

...

Altered step 8. Incubate for 1 hour at 30°C.

L. Take 10 µl (1% of the electroporation reaction) to 90 µl SOC medium, mix and serial dilute 5 times (10 µl of each prior mix into 90 µl SOC medium, for 1:1K, 1:10K, 1:100K, 1:1M, 1:10M, 1:100M dilutions from the electroporation). Streak out each of these dilutions onto an LB+ antibiotic plate (ampicillin or carbenicillin) and grow overnight at 30°C.

M. Inoculate the remaining cells into 100 mL LB+antibiotic (ampicillin or carbenicillin) grow at 30°C.

N. Estimate number of colonies from your serial dilution plates (e.g. if you have 32 colonies for your 1:10M dilution, you have ~ 32 million transformant cells). You should have at least 100, ideally 500, colonies per guide.

Colony counts from your no insert control should be <1% of your +insert Gibson reaction count. If not, consider cutting your vector for a 3rd time with BsmBI.

O. After 16-18 hours, or when shaking cells are at late log phase, take 4 mL and spin down for a miniprep (e.g. Qiagen **Qiaspin miniprep kit** #27106). Maxiprep the remaining cell with **Qiagen EndoFree maxikit** (Qiagen #12362). Measure DNA concentration from mini and maxipreps with a nanodrop spectrophotometer.

P. Amplify the library with sequencing primers:

1. Assemble the PCR reaction:

20µl 2× **Q5 Hot Start Master Mix** (New England Biolabs #M0492S)1.5 µl 25 µM Seq9971 primer (AATGATACGGCGACCACCGAGATCTACAC **ACGT**

CGATTCTTGGCTTATATATCTTGTG)

1.5 µl 25 µM Seq9981 primer (CAAGCAGAACGGCATACGAGAT **CCTGGTAG** ACAGTCGAGGCTGATCAGC)

1 µl of your miniprep library DNA diluted to 1ng/µl.

Water to 40 µl.

Note: the sequences in bold are index sequences. If you want to run multiple libraries on the same sequencing run, you can make additional primers with different indexes.

2. PCR:

98°C 30 sec

[98°C 15 sec, 64°C 15 sec, 72°C 15 sec] (4 cycles)

[98°C 15 sec, 72°C 20 sec] (16 cycles)

72°C 120 sec

4°C hold.

3. Check by running a small sample of your PCR reactions on a 2% gel. The amplicon should be ~206bp.

4. Clean PCR reactions by adding 1 volume (40µl) SPRI beads, and following the usual SPRI bead purification protocol, above. Elute in 40 µl of H2O.

5. Repeat SPRI Clean-up with 1.0 x beads (40 µl). Elute in 15 µL H2O. Measure concentration, ideally with Qubit dsDNA HS kit.

Q. Perform MiSeq Sequencing, spiking in these custom primers:

Into Port 18: Custom Read 1 Sequencing Primer: Seq999_hU6_R1

(CGATTCTTGGCTTATATCTTGTGGAAAGGACGAAACACCG)

Into Port 19: Custom Index Primer: Seq996_sgPuro_I

(AAGGCTAGTCGTTATCAACTGAAAAAGTGGCACCG)

R. Generate a text file with each of your guide IDs and sequences on separate lines. Use Bowtie2 to generate "genome" files, where the genome is all your guide sequences. Then run bowtie to identify exact matches to each of your guides, and count the number of unique matches to each guide. For bowtie2 methods see <https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>.**S.** Calculate "skew" as the difference in count frequency between the top and bottom 10th percentiles of guides. In our case, the library was sequenced and shown to include all 37,637 designed guides with relatively equal coverage of each (the difference in count frequency between the top and bottom 10th percentiles of guides was 2.8).

T. A lentiviral library was produced using a standard 3-plasmid protocol (above) at a scale to yield 10 ml of virus, stored in aliquots at -80°C, with each aliquot thawed only once. Viral titer was tested using the "Lentiviral transduction and titration at scale" protocol, below, on a frozen aliquot of virus.

Lentiviral transduction and titration at scale

48

Rationale:

For large, heterogenous guide libraries it is important to transduce enough cells that the complexity of the library is maintained (typically ~400-500 transduced cells per guide). It is difficult to achieve this using the small scale transduction protocol above. Instead cells are mixed with virus at high concentration and split out to larger plates for growth after transduction. Importantly, volumes of virus needed for a certain MOI will differ greatly between the small scale and large scale protocols, and it is important to do a viral titration under the conditions of the protocol you plan to use.

49

Preparation:

Grow 2 or more 10cm plates of TeloHAEC (or derivatives) to get the usual ~4M/plate concentration on the day you plan to do your viral titration.

50

Day 1:

- A.** Harvest and count cells per the usual maintenance protocol, and resuspend cells at 1e6/ml in EC media. Add 10 µg/ml polybrene (final) and mix.
- B.i.** For viral titration aliquot 1ml to sterile Eppendorf tubes and add virus (starting, perhaps with a titration of 4, 15, 50 and 200 µl virus). Mix gently by inversion. Plate the full volume of each to 1 well of a 24 well plate. Be sure to include 1ml cells & polybrene with no virus as a control.
- B.ii.** For full scale library transduction, after having titered your virus, plate 4 ml/well in a 6 well plate. Include a no virus control.
- C.** Centrifuge at 2000 rpm for 2 hrs at 30°C, and incubate at 37°C for 2 hrs.
- D.** After 2 hours add an additional equal volume of media without polybrene (1ml/24 well or 4ml/6 well).

51

Day 2:

- A.** Harvest the cells in each well (PBS wash, trypsinize, resuspend & spin)
- B.i.** For a titration, plate 1/3 of the cells from each 24-well well to 2ml EC medium in a 6 well well. Split the novirus well into 4 wells (1/3, 1/3, 1/9 and 1/30th volume), to make a novirus+antibiotic and 1:1, 1:3 and 1:10 dilution novirus-antibiotic control wells).
- B. ii.** For transduction at scale, resuspend cells from each 6-well well in 2.4 ml media, mix these all together, then aliquot 2.4 ml each to 20 ml media in 15cm plates (reserving at least 200 µl). For the novirus well, resuspend in 2.4 ml media, and prepare a 6-well control plate adding 200 µl to 2 wells, and 67 and 20 µl to single wells (your novirus+antibiotic and 1:1, 1:3 and 1:10 dilution novirus w/o antibiotic controls). Also add 200µl of your mixed infected cells to 1 well.

C. Add 15 μ g/ml final concentration blasticidin to each viral plate or well and to one of the 1:1 dilution novirus wells. Incubate at 37°C.

52

Day 6:

Determine % infected cells as follows:

Wash each 6-well well with 2 ml PBS, add 250 μ l trypsin, incubate at 37°C for ~5 mins, resuspend with 750 μ l EC media (1 ml total) and count a sample of cells from each well. Determine % infection as per the "Measured calculation of viral titer" in the "Lentiviral titer and small scale transductions of TeloHAEC cells" protocol.

Perturb-seq library preparation

53

A. Preparation: Titer lentivirus for large scale transduction.

We titered a 10 ml production stock of lentivirus prepared from our gRNA library of 37,637 guides, using the "Lentiviral transduction and titration at scale" protocol, above. We identified a volume of virus to use to get 15.7% transduction (15.7% blasticidin resistant cells).

54

B. Day 1: Transduce CRISPRi cells at low MOI, and select for cells that received the gDNA vector

1. For our Perturb-seq study, 127.5 million CRISPRi TeloHAEC were transduced and selected for blasticidin resistance (15 μ g/ml for 4 days).

2. Coverage (# of cells/guide) was estimated in 2 ways:

- a) By counting cells after blasticidin treatment and dividing by $2^{(\# \text{hours since transduction}/36.7\text{hr})}$, where 36.7 hrs is the doubling time for TeloHAEC observed in routine culture. By this measure we had transduced 360 cells/guide.
- b) By multiplying the number of cells transduced by the fraction of transduced cells calculated from our titration of virus. By this measure we had 461 cells per guide.

Neither of these measures is perfect (doubling time may differ in cells after transduction and in blasticidin relative to routine culture, and effective titer could differ from the predicted value). The combination of both measures, however, gives a reasonable sense of the range of coverage.

55

C. Day 6: Induce the CRISPRi machinery with doxycycline

After blasticidin selection, cells were treated with 2 μ g/ml dox for 5 days (plating 18e6 cells at each split, to maintain complexity of the library). We reasoned that, since atherosclerotic plaques develop slowly, the longer-term transcriptional effects of causal CAD gene disruption would provide the greatest insights into disease mechanisms. Thus, while we have found that knock down of guide-targeted genes is near maximal after 2 days of doxycycline treatment (inducing the CRISPRi machinery), we treated guide-containing cells with 2 μ g/ml doxycycline for 5 days, to measure the longer-term consequences of each perturbation.

As you grow out the cells during blasticidin selection and doxycycline treatment, monitor cell density and split when cells are ~8M/15 cm plate. When harvesting, pool all cells together, then be sure to maintain library complexity by splitting out enough cells to have ~400 per guide.

56

D. Day 10: Prepare scRNAseq libraries

The presence of guideRNAs in cells allows multiplets (droplets containing 2 or more cells) to be unambiguously identified, as droplets containing more than one guide. This allowed us to load ~10-fold more cells per 10X Genomics lane than the maximum number recommended in the manufacturer's protocol.

- 1) Harvest cells by trypsinization as normal.
- 2) Resuspend the cell pellet in ice cold PBS with 1% **BSA** (ThermoFisher #J65097.A1, or equivalent), return to ice, and count a sample.
- 3) Load 150,000 cells per lane on a **10X Genomics Chromium Controller** using a **3' scRNA-seq V3 kit** (10X Genomics, other 10X 3' scRNA seq kits will also work, but be sure to check the product literature for ways in which they differ from the V3 kit).
- 4) Follow the scRNAseq kit directions for preparation of your libraries. Be sure to retain some of the SPRI bead purified cDNA (step 2.3A in the 3' scRNA-seq V3 kit instructions) for dialout libraries. For final amplification, choose different indexes for each 10X controller lane.

If you have more lanes than fit into a single controller cassette (8 lanes max), you may split up your samples between multiple runs. Be very careful to treat the cells identically for each batch of lanes, to minimize batch effects. In our case, we ran 20 lanes, for a total of 3 million cells, with cells isolated in two batches, with 6 lanes for the first batch, and 14 lanes, across 2 cassettes, for the 2nd batch, 6 hours later. We found that even the 30 minutes difference between the 2nd batch of cells loaded on the 1st cassette and the 2nd cassette resulted in significant batch effects.

These batch effects did not prevent downstream analyses (as we were able to identify and rule out transcriptional programs that correlated with batch, and use batch as a covariate in other analyses). Nonetheless, it is preferable to minimize them. One way to reduce such batch effects might be, for each set of lanes run, to isolate and count cells from a subset of plates, preparing them as similarly as possible.

Dial out library preparation

57

From the initial amplified cDNA ("WTA" (whole transcriptome amplified) step 2.3a in the 10X 3' scRNAseq V3 instructions), we used a two stage PCR protocol to generate "dialout" libraries, for each lane. Because the CROP-seq vector expresses a Pol II polyadenylated transcript that ends just downstream of the guide sequence, the dialout libraries identify the guideRNA sequences associated with each droplet²⁰.

A) PCR1:

For each scRNAseq library Assemble a 30 μ l PCR reaction as follows:

| | | |
|--------------------------------------|------------|---|
| Phusion HiFidelity Master Mix | 15 μ l | (New England Biolabs #M0531S) |
| CropDialOut_R1 | 25 μ M | 1.25 μ l (CTACACGACGCTTCCGATCT) |
| CropDialOut_U6_F | 25 μ M | 1.25 μ l (GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTGTGGAAAGGACGAAACACC) |
| H2O | | 11.5 μ l |
| WTA cDNA (10 ng/uL) | | 1 μ l |

Run on a PCR machine with this program:

98°C 30 secs
7 cycles (98°C 15 secs, 69°C 15 secs, 72°C 20 secs)
72°C 120 secs
4 ∞

B) Clean with 1x SPRI-beads, as per the protocols above, and elute in 15 μ l warer.**C) PCR2:**

For each libray, assemble a 30 μ l PCR reaction as follows:

| | |
|--|--|
| Phusion HiFidelity Master Mix | 15 μ l |
| CropDialOut_P5_R1 | 25 μ M |
| (AATGATACGGCGACCACCGAGATCTACACTCTTCCCTACACGACGCTC) | 1.25 μ l |
| CropDialOut_P7_ | 25 μ M |
| GTGACTGGAGTTCA | 1.25 μ l (CAAGCAGAAGACGGCATACGAGAT-8bp index sequence- PCR1 product 12.5 μ l) |

Run on a PCR machine with this program:

98°C 30 secs
8 cycles* (98°C 15 secs, 57°C 15 secs, 72°C 20 secs)
72°C 120 secs
4 ∞

* The number of cycles can vary between libraries and cell lines. After 8 cycles, check by running a 1 μ l sample of a few of your PCR reactions on a 2% gel (looking for a clean amplicon band of 470 bp).

If the band is very faint, run 3 or 4 more cycles of PCR2 and check again. You can recover reasonable libraries with cycle numbers up to ~16, but much more than that may indicate a problem with reagents or primers that will prevent getting useful sequencing data.

- D) Clean with 0.7x volume (21 μ l) SPRI (Agencourt XP) beads, as per the protocols above. Elute in 50 μ l water.
- E) Repeat the SPRI clean up with 0.7 volumes beads (35 μ l), and elute in 12 μ l.
- F) Run a sample of selected libraries on a gel or Bioanalyzer to confirm band size and that primer is gone.
- G) Measure library concentrations with Qubit.

Identification of scRNAseq cells and assignment to guides

58 To get complete information about guide assignments, dialout libraries were sequenced to approximately 40-fold saturation. Guides were identified from read 1 sequences, using Bowtie2 (<https://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>) to align dialout reads to a “genome” composed of all 37,637 guide sequences, requiring no-mismatches. Aligning read 1 and read 2 sequences linked gRNA sequences with cell barcodes (CBCs, unique to each bead/droplet) and unique molecular identifiers (UMIs).

To avoid low-frequency PCR chimeras, we required that each CBC-UMI-guide combination be duplicated at least 4 times. We then identified the guides associated with each CBC, and the number of different UMIs for each CBC-guide combination.

In our study, we selected 4 UMIs for any single guide as the threshold to call a cell as containing a guide. We defined singlets (one cell & one guide per CBC) as having ≥ 4 UMIs for the most frequent guide and ≥ 4 x less than this for the 2nd most frequent guide (choosing these thresholds to give a good balance between power to detect transcriptional effects and accuracy in measuring the magnitude of these effects). Doublets and higher multimers, were cells with ≥ 4 UMIs for the top guide, and one or more additional guides with more than 1/4 this number of UMIs. For additional details on dialout sequencing and considerations for thresholds to call singlets, see Supplementary Fig. 2 in Schnitzler, Kang et al. (2024) Nature, in press.

Note: The CROPseq PolII transcript that reads through the guide cassette can be expressed at different levels in different cell lines. This expression was borderline OK in TeloHAEC, but may be too low to effectively assign guides in other cell lines. If so, one alternative is to use the "Feature barcoding" approach offered by 10X genomics.

scRNAseq library sequencing and assignment of reads to cells

59 scRNA-seq libraries were sequenced on two Illumina NovaSeq S4 flowcells, yielding 20,245,734,673 total reads, across all 20 libraries. The FASTQ files were processed on the 10X Cloud to run CellRanger count with

the hg38 reference genome. We used the “filtered” features (i.e., cell barcodes corresponding to droplets that contain a cell), and combined the outputs from all twenty 10X lanes into a single genes x cell matrix. This analysis identified 822,156 cell-containing droplets.

To measure the effects of individual guides on individual cells, we selected only those CBCs identified in the dialout analysis as corresponding to singlet cells. This identified 214,449 singlets (droplets containing one cell and one guide), defined as 4+ unique molecular identifiers (UMIs) for the top guide and ≥ 4 -fold fewer UMIs for any other guide. This gave an average of 5.7 cells per guide and 85.5 cells per target promoter. Average sequencing depth was 10,870 transcriptome-mapped UMIs per singlet cell, and 929,000 transcript UMIs, across all 15 guides, for each target promoter.

We found that this sequencing depth was sufficient to detect the effects of perturbations on transcriptional programs involving many coregulated genes (identified by consensus nonnegative matrix factorization, for details see Schnitzler, Kang et al. (2024) *Nature*, in press). It was generally not sufficient, however, to accurately measure the effects of perturbations of individual genes (except for genes expressed at an exceptionally high level, e.g. >200 TPM). If your experimental design requires measuring effects on individual genes, sequencing depth will need to be on the order of 5- to 10-fold higher.