Supplementary Note

Additional Results

Perturb-CITE-Seq identifies modules and programs of immune evasion

For example, *SOX4*[1,2], *RB1*[3], *SP100*[3] *and IRF3*[4] all comprise one co-functional module (module 1), despite their different known roles in EMT (*SOX4*), cell cycle regulation (*RB1*), response to IFNγ (*SP100*) and transcriptional regulation of type I interferon response (*IRF3*) (**Fig. 4d**). Interestingly, both SOX4 inactivation and RB1 inactivation (genetically or through hyperactivation of CDK4/6 phosphorylation) alter responses to immunotherapies in melanoma models[2] and breast cancer[3], respectively. Notably, *SOX4*, *IRF3* and *SP100* were repressed in the context of defects in the IFNγ -JAK/STAT pathway in our screen (**Fig. 4b**). The co-regulated program affected by their perturbation in our system (**Extended Data Table 6**) includes genes involved in cell cycle (*CDC80, CDKN3, CENPM*), metabolism (*COX20, DCTPP1, FH, IDO1, UCHL5, ZDHHC4, PAPSS1, PTGES*), mRNA maturation (*MRTO4, NSRP1, SNHG15*), DNA repair (*RAD51C*) and inflammation (*IFI16* and *IL1B*) or checkpoints (*IDO1*).

Loss or downregulation of *CD58* confers immune evasion

CD58 KO had a significantly enriched resistance in the viability screen at a TIL:tumor cell ratio of 2:1 (and borderline significance at 4:1), and our validation experiments demonstrated a survival advantage at all tested ratios, including at 4:1 (**Fig. 5f**). This suggests that limited power (likely due to number of cells in the screen) rather than biological differences (e.g. discordance between tested sgRNAs) underlie the weaker signal for *CD58* KO enrichment in the pooled viability screen.

Because CD2, the CD58 binding partner, is also expressed on Natural Killer (NK) cells, we hypothesized that *CD58* KO would also increase immune evasion from these cells. To test this, we co-cultured melanoma cells with *CD58, B2M* or *PD-L1* KO as well as unperturbed control cells, with human NK cells (NK-92, **Methods**) and found that *CD58* KO indeed conferred resistance to NK cell killing, while *B2M* KO and *PD-L1* KO were lysed at similar rates as parental control (**Fig. 5g)**. Of two additional patient models (MaMel134 and MaMel80), *CD58* KO in MaMel134 also showed increased (and dose dependent) survival to NK cell lysis compared to otherwise *B2M* KO and parental cells (**Extended Data Fig. 7f**).

METHODS

Rapid T cell expansion from tumor infiltrating lymphocytes

Previously established TILs[5] were grown in 24 well plates in RPMI 1640 supplemented with heat-inactivated 10% human AB serum (Corning, Corning, NY), GlutaMax, 10 mM HEPES, 55 μM 2-Mercaptoethanol (all ThermoFisher Scientific, Waltham, MA), 50 Units/mL penicillin, and 50 μg/mL streptomycin at 37°C and 5% CO_2 in a humidified incubator. TIL media was supplemented with 300-3000 IU recombinant human IL2 (Chiron Therapeutics, Emeryville, CA) and media with IL2 was replenished every two days by half media exchange. For rapid expansion, 0.5-1x10^6 proliferating TILs were seeded with 100x10^6 normal donor peripheral blood mononuclear cells irradiated with 5,000 rad in a 1:1 mixture of TIL media and AIM-2 media (ThermoFisher), supplemented with 30 ng/mL OKT3 (Miltenyi Biotec, Bergisch Gladbach, Germany) in a G-Rex-10 bottle (Wilson Wolf, St. Paul, MN). On day 2, 6,000 IU/mL IL2 were added followed by a half media exchange on day 5 and addition of another 6,000 IU/mL IL2. On day 7, cells were counted and a minimum of 6x10^7 cells were transferred to a G-Rex-100 bottle in a total volume of 600 mL

AIM-5 medium supplemented with 6,000 IU/ml IL2. On day 10, 400 mL AIM-5 with 6,000 IU/ml IL2 were added. On day 12, 3,000 IU/ml IL2 were added. Expansion was completed on day 14 and cells were frozen in aliquots of 10-100x10^6 in Bambanker freezing medium (Nippon Genetics, Tokyo, Japan)

Tumor infiltrating lymphocyte characterization

Expanded TILs were thawed and rested in TIL media supplemented with 3,000 U/mL rhIL2. On day three, 1x10^5 TILs were seeded per well of a round bottom 96 well plate and stimulated with phorbol-12-myristate-13-acetate/Ionomycin in the presence of Brefeldin A (Biolegend) for 4 hours as per manufacturer's instruction. Cells were washed once in cold PBS and stained for 15 minutes with Zombi NIR (Biolegend) diluted 1:500 in PBS to label dead cells. Cells were then washed once with full staining buffer (3% FBS in PBS with 2 mM EDTA) and surface antigens were stained for 30 min on ice using the following antibodies: gdTCR-BV421 (B1, Biolegend, 331218), CD45-Pacific Blue (HI30, Biolegend, 304022), CD3-AF532 (eBiosciences, 58/0038-42). Next, cells were washed twice in ice cold staining Buffer and fixed and permeabilized using the eBioscience™ FoxP3/Transcription Factor Staining Buffer Set (ThermoFisher) per manufacturer's recommendation. Intracellular antigens and fixation-stable surface antigens that might be internalized during activation were stained for 45 minutes on ice using the following antibodies: CD8a-BV570 (RPA-T8, Biolegend, 301038), Granzyme B-PE-CF594 (GB11, BD, 562462), FoxP3-APC (236A/E7, eBiosciences, 17-4777-42), CD4-AF700 (RPA-T4, Biolegend, 4345826), IFNγ-AF488 (4S.B3, Biolegend, 502517), IL-17A-PE-Cy7 (BL168, Biolegend, 512315), and TNF-a-APC-Cy7 (Mab11, Biolegend, 502943). Fluorescent-minus-one (FMO) staining controls were prepared for gating of IFN-g, TNF-a, IL-17A, and Granzyme B. After

staining, cells were washed twice with permeabilization buffer, resuspended in staining buffer, and analyzed on a Sony SP6800 spectral analyzer (Sony, Tokyo, Japan).

Lentivirus production for fluorescent protein expression

Lentivirus encoding nucleoplasmin nuclear localization sequence (NLS) tagged fluorescent proteins dsRed (NLS-dsRed) and blue fluorescent protein (NLS-BFP) were generated using pHAGE vectors (Gift from W. Marasco). To generate lentivirus, HEK293T cells (ATCC) were seeded in 6-well plates in DMEM supplemented with 10% FBS (D10). The next day, cells had reached 80% confluence and were transfected with plasmids for lentivirus production using TransIT-LT1 transfection reagent (Mirus Bio, Madison, WI). 3 µl transfection reagent was added to a total volume of 15 µL Opti-MEM (ThermoFisher) and incubated for 5 minutes at room temperature. In the meantime, the plasmid mix was prepared by mixing 500 ng packaging plasmid psPAX2 (addgene #12260, gift from Didier Trono) and 250 ng envelope plasmid pMD2.G (addgene #12259, gift from Didier Trono) with 500 ng pHage expression vector in 37.5 µL Opti-MEM. The transfection reagent mix was added to the plasmid mix, incubated for 30 minutes at room temperature and then added dropwise to the cells. All volumes reported are for the transfection of one well and were scaled for transfection of multiple wells and plates. Cells were incubated for 18 hours, and then the transfection media was removed and DMEM supplemented with 20% FBS (D20) was added. Supernatant containing lentivirus was collected after 24 hours and fresh D20 was added. After another 24 hours, media was again collected and the lentivirus containing media was pooled, filtered through a 0.45 µM syringe filter (ThermoFisher) to remove debris, aliquoted, frozen and stored at -80°C.

Transduction of NLS-fluorescent proteins

Human melanoma cells were seeded in 12 well plates with increasing volumes of lentiviral supernatant in a final volume of 2 mL with 4 µg/mL polybrene (Millipore, Burlington, MA). Plates were centrifuged for 2 hours at 1,000x g at 30°C and cells were incubated for 16 hours at 37°C and 5% CO_2 in a humidified incubator. The next day, cells were detached and seeded in 6-well plates. Cell lines with a transduction efficiency of 30% were sorted on a FACS Aria cell sorter, expanded and stored frozen in Bambanker (Nippon Genetics) until further use.

Cas9 lentivirus production

Cas9 lentivirus was generated as outlined for other vectors above using 6 well plates and the expression vector pLX311-Cas9 (addgene #118018, gift from William Hahn and David Root),

Lentiviral Cas9 transduction of human melanoma cell lines

Human melanoma cells were seeded in 12 well plates with increasing volumes of lentiviral supernatant in a final volume of 2 mL with 4 µg/mL polybrene (Millipore, Burlington, MA). Plates were centrifuged for 2 hours at 1,000x g at 30°C and cells were incubated for 16 hours at 37°C and 5% CO_2 in a humidified incubator. The next day, cells were detached and seeded in 6-well plates. In parallel, $1x10^5$ cells of infected and non-infected cells were seeded in 12 well plates to be used for transduction efficiency assessment. On day 2 after infection, selection was initiated with 2 µg/mL Blasticidin S (ThermoFisher). When all cells in the non-infected but selected control were dead, transduction efficiency was calculated as a ratio of cells surviving selection divided by cells in the non-selected control. Cells with approximately 30% transduction efficiency were chosen for further experiments.

Cas9 activity assay

To assess Cas9 editing activity, Cas9 expressing human melanoma cell line 2686 was transduced with lentiviral particles produced as described above using expression plasmid pXPR_011 (Addgene #59702, gift from John Doench and David Root), encoding for Enhanced Green Fluorescent Protein (EGFP) and a guide RNA targeting EGFP[6]. Cells with no Cas9 expression were transduced as non-editing, EGFP positive controls. Transduced cells were selected in media supplemented with 0.75 μg/mL puromycin. After selection was completed, EGFP expression was assessed by flow cytometry on a Sony SP6800 spectral analyzer (Sony, Tokyo, Japan). Control cells with and without EGFP expression were used as gating controls and Cas9 activity was defined as %EGFP negative cells. For the pooled screens (Perturb-CITE-seq and viability screens), we used melanoma model 2686, because in this cell line we achieved exceptionally high Cas9 enzymatic cutting efficiency.

Generation of pooled Perturb-CITE-seq gDNA library

To generate a Perturb-CITE-Seq vector compatible with FACS selection, we modified the lentivirus CROPseq-Guide-Puro vector[7] by Gibson Assembly to insert a gene expressing the fluorescent protein mKate2 (CROPseq-mKate2). After SPRI-purifying CROPseq-Guide-Puro which had been digested with BsiWI, we performed a Gibson Assembly to insert the Kate2 expressing oligo ordered as a gene string from Invitrogen (**Extended Data Table 7**).

For the pooled Perturb-CITE-seq CRISPR library, we designed a 744 single-guide (sg)RNA gDNA library targeting 248 genes selected from a previously defined signature of immune

checkpoint inhibitor resistance in melanoma[1], with three gRNAs targeting each gene with sequences picked from the Broad Institute Genetic Perturbation Platform Web sgRNA Designer (https://portals.broadinstitute.org/gpp/public/analysis-tools/sgrna-design)[8]. We also incorporated two types of control gRNAs, non-targeting and non-gene targeting intergenic controls, in the library, each at 5% of the total guide count (n=74 control guides for Perturb-CITE-Seq experiment). In addition to this library for Perturb-CITE-Seq, we generated a separate library (extended control library), including the same targeting sgRNAs, but a larger number of control guide RNAs (n=758,) for the large-scale CRISPR viability screen to increase the power for enrichment/depletion detection. All guide sequences can be found in **Extended Data Table 1**. The pool of guide sequences with appropriate overhangs for Golden Gate cloning was synthesized by Twist Biosciences. The gRNA oligo pool was amplified by PCR and inserted into a BsmBI-linearized CROPseq-mKate2 vector backbone via a Golden Gate ligation reaction. The resulting product was purified by isopropanol precipitation prior to transformation into electrocompetent Lucigen Endura cells (10 μF, 600 Ohms, 1800 V). Transformed libraries were grown overnight in shaking liquid culture at 37°C, and a library representation of >200 colonies per guide was determined prior to CRISPR library pool plasmid purification by Endofree Maxi (Qiagen). Coverage of the 818 guide library for Perturb-CITE-Seq as well as the extended control library was confirmed by PCR amplification of the guide sequence from the final library pool and deep sequencing on an Illumina MiSeq.

Perturb-CITE-seq ICR gDNA library lentivirus production

Lentivirus of the Perturb-CITE-seq sgDNA library was produced in multiple 6-well plates as described for other vectors above. To generate the extended control ICR library, the ICR library was mixed at equimolar ratios with the extended control library prior to transfection of the HEK cells.

ICR-library titration

$2.5x10^5$ Cas9-expressing melanoma target cells (2686) were seeded in 1 ml of media in each well of a 12-well plate in the presence of 4 µg/ml polybrene (Millipore). ICR-library virus was added at increasing dilution with one well receiving no virus and serving as non-transduction control. The plates were centrifuged at 1000 x g for 2 h at 30°C and incubated overnight at 37°C and 5% CO_2 in a humidified incubator. After 16h, viral supernatants were removed, and fresh media was added. 8 hours later the cells were split in two plates and selection with 0.75 µg/ml puromycin was initiated in one of the plates. When all cells in the non-infected but selected control were dead, transduction efficiency was calculated as a ratio of cells surviving selection divided by cells in the non-selected condition. At the same time transduction efficiency in the non-selected cells was assessed by detecting the fluorescent marker mKate2 by flow cytometry using a Sony SP6800. Multiplicity of infection (MOI) was calculated as MOI = LN(1-x/100) with x=%$mKate^+$ or x=%survivors respectively. The percentage of cells receiving exactly one viral particle was calculated as Single-Infection%=((1-x/100)*LN(1-x/100))/(x/100)*100 as previously described[2].

Introducing ICR-libraries in melanoma target cells

Cas9-expressing melanoma target cells (2686) were transduced with the ICR-library (818 sgRNAs, for scRNA-seq) or the ICR-extended library (818 sgRNA from the ICR library + 818 sgRNA controls, for enrichment screening) aiming for a transduction rate <10% (>90% single infection rate) and at least 1000x coverage per sgRNA. To this end, $2.5x10^6$ cells per well of a 12 well plate were seeded in the presence of 4 μg/ml polybrene and titrated ICR-library virus was added. Cells with virus were centrifuged at 1000 x g for 2 h at 30°C and incubated overnight at 37°C and 5% CO_2 in a humidified incubator. After 16h, viral supernatants were removed, cells were detached and reseeded in T175 cell culture flasks, and Puromycin selection was started 24h after transduction. A 6-well plate with and without selection plated with transduced and non-transduced cells served as transduction control. Selection was completed after day 6 and transduced cells were further propagated in the presence of puromycin. On day 7 and day 14, aliquots of $1x10^6$ cells were saved for gDNA read-outs.

Perturb-CITE-Seq library generation and sequencing

After loading of cells onto the 10X Chromium system (Chromium Single Cell 3' Library and Gel Bead kit v3) single cell gene expression libraries were generated following the manufacturer's instructions with one change during cDNA amplification: to increase abundance of CITE-seq oligos, 2 pmol of the ADT additive was spiked-in into the 10X cDNA amplification. Following cDNA amplification, a 0.6X SPRI was performed to separate large gene expression library cDNA from smaller CITE-seq oligos. The supernatant of this 0.6X SPRI was set aside to generate the CITE-seq expression library. The 0.6X cDNA amplification SPRI was then completed and gene expression libraries were generated following the manufacturer's instructions. Returning to the supernatant of the cDNA amplification 0.6X SPRI, 1.4X SPRI was added to the initial 0.6X SPRI

supernatant. This double sided SPRI was completed following two 200 μL 80% ethanol washes. The product was eluted in 50 μL TE buffer, and a second 2.0X SPRI was performed on this CITE-seq template eluting in 50 μL TE buffer. To generate the CITE-seq sequencing library, 5 μL of the SPRI-cleaned CITE-seq template was mixed with 25 μL of 2X NEBNext HiFi Master Mix, 11.5 μL of diH2O, and 1 μL of a 1:1 ADT primer mixture (10 μM stock concentration of each primer). To enable multiplexing, one of six unique P7 indices were included in each reaction. The CITE-seq sequencing library reaction was cycled with the following conditions using a thermocycler: 98°C for 10 minutes, (98°C for 2 sec, 72°C for 15sec)x18 then 72°C for 1 minute.

To amplify fragments containing the guide barcodes, a 10 ng fraction of the WTA was mixed with 15 μL Phusion High-Fidelity PCR Master Mix, 11.5 μL H2O, 1.25 μL CROPDialOut_R1 (25 μM), and 1.25 μL CROPDialOut_U6_F (25 μM). The reaction cycled through the following conditions 98°C for 30 sec, (98°C for 15 sec, 69°C for 15 sec, 72°C for 20 sec) x 9 cycles, 72°C for 2 mins. After cleaning the product with 1x SPRI and eluting in 15 μL of H2O, it was subsequently used as a template in a second PCR reaction to attach Nextera adapters. 12.5 μL of the SPRI-cleaned PCR product was mixed with 15 μL Phusion High-Fidelity PCR Master Mix, 1.25 μL CROPDialOut_P5_R1 (25 μM), and 1.25 μL of a unique P7 index primer. The reaction was amplified through an additional 9 cycles following the conditions of the first PCR, while adjusting the melting temperature to 57°C. A 0.7x SPRI clean was performed twice to purify the product, eluting in 50 μL H_2O and 35 μL H_2O respectively. All primer sequences are listed in **Extended Data Table 8**. scRNA-Seq and CITE-seq libraries were sequenced on Illumina HiSeq to 25,000 reads/cell and 10,000 reads/cell respectively.

Pre-processing of single-cell data

Expression matrices, representing Unique Molecular Identifier (UMI) counts for both scRNA-seq and CITE-seq data were obtained using the Cumulus[9] version 0.14.0 implementation of the CellRanger v.3 workflow with genome reference GRCh38 v3.0.0 and default parameters. Cells with fewer than 200 detected genes or with >18% of detected genes labeled mitochondrial were removed from subsequent analysis. Genes detected in fewer than 200 cells were also removed from further analysis.

Differential expression testing

Differential expression for CITE antibodies across experimental conditions was performed using the Seurat implementation of logistic regression[10] with the "ADT" assay and default parameters. Differential expression for transcripts across experimental conditions was performed using MAST[11] with default parameters. Differential expression testing relating to Figure 3 was performed on cells with non-targeting sgRNAs only.

Identification of cellular programs using jackstraw PCA

Jackstraw PCA was implemented according to the algorithm described by Chung and Stoery[12]. Briefly, expression matrices from each condition were processed separately using MOI = 1 cells only. A subset of 150 features (out of all genes passing QC and all 20 CITE proteins) was randomly sampled, followed by PCA, for 150,000 iterations. This allowed a PC-wise calculation of the synthetic null distribution of PC loadings. PCA was then run on the full expression matrix, to calculate an empirical p value for each feature. A threshold ($p < 0.03$) was used to assign specific

features to each program (PC) to obtain programs of reasonable biological size. The Scikit-learn implementation of PCA was used in all cases[13].

Initial sgRNA assignment to cells

Dial-out PCR sequencing data was processed with custom code to identify cell barcode (CBC) and sgRNA barcode (GBC) pairs. A dictionary was built mapping each CBC-GBC pair to a list of UMIs, each with its associated number of reads. A GBC was assigned to a particular CBC if it exceeded *any one* of the following thresholds: (**1**) a UMI with more than 200 reads, (**2**) more than 20 total UMIs, or (**3**) a UMI with more than 50 reads that accounts for more than 20% of the total reads for the CBC-GBC pair. These criteria are intended to correct for chimeric products produced during any round of PCR[14]. Specifically, we use "reads" to refer to the number of reads for a given UMI, and we use UMI to refer to the sequence itself regardless of the number of reads. A UMI with a very large number of reads (*e.g.*, 200) is likely to be associated with a true cell barcode – sgRNA barcode pair, motivating threshold #1. A chimeric product can also be formed in an early round of PCR. However, a true match of a cell barcode with an sgRNA barcode is more likely to have several *UMIs* (regardless of the number of reads of each UMI), motivating threshold #2. Finally, a single UMI with many reads (*e.g.*, 50) accounting for a large proportion of the total reads for a given pair of cell barcode – sgRNA barcode (*e.g.*, 20%) is likely to represent a non-chimeric UMI indicating a true cell barcode – sgRNA barcode pair, motivating threshold #3. The output of the computational pipeline is robust to changes in these thresholds.

Scoring programs across cells and with perturbations

Programs were scored according to a previously described procedures[15,16]. Briefly, expression values are first sorted into 50 expression bins. The expression of a target list of genes is then scored as their average expression less the expression of a control set. The control set is a randomly sampled set of genes with size equal to the number of genes in the target list (unless the target list has fewer than 50 genes, in which case the number of control genes was 50). In the case of programs with both an up- and down-regulated components (*e.g.*, positive and negative loading for the jackstraw programs), the two components are scored separately and the composite score is calculated as the score of the up-regulated component minus the score of the down-regulated component.

The sgRNA to target dictionary (above) was used to assess the effect of perturbations on program scores. Cells were scored according to the procedure above. Cells with sgRNAs mapping to the same target were grouped. Each target had the program score of its cells compared to the score of cells with a non-targeting sgRNA. Welch's T test was used to test significance for enrichment or depletion of a given program.

Enrichment of surface protein knockouts using fluorescence activated cell sorting

KO cell lines generated using nucleofection of Cas9 RNP were detached with Accutase (Stem Cell Technologies, Vancouver, Canada) and stained with APC conjugated antibodies targeting B2M (2M2), CD58 (TS2/9), and CD274 (29E.2A3, all Biolegend, San Diego, CA). Unperturbed cells stained with the same antibodies and isotype controls were used as gating controls. Cells negative

for B2M, CD58, or CD274 were sorted on an Influx Cell sorter (BD), grown out and banked until further use.

Prior to use in experiments, purity of KO cell lines was assessed using the same clonotypes as above using an Aurora Spectral Analyzer (Cytek Biosciences, Freemont, CA). In addition, loss of MHC class I expression in B2M knock-outs was assessed by staining with FITC-conjugated anti-HLA-A,B,C (W6/32, Biolegend). To maximize surface expression of IFNγ regulated proteins, cells were assessed with and without 16h pretreatment of 10 ng/ml recombinant human IFNγ (Peprotech, Rocky Hill, NJ).

NK co-culture assay

Melanoma target cells expressing NLS-dsRed were seeded in 100 µL full melanoma media (as described above) in a black walled 96-well plate (Corning) and left to adhere overnight. 2686 cells were seeded at a density of 5,000 cells per well, while MaMel80 and MaMel134 cells were seeded at a density of 10,000 cells per well. After 16 hours, media was replaced with 100 µL fresh pre-warmed melanoma media with 4 µM Caspase-3/7 activity dye (CellEvent, ThermoFisher). Plates were imaged using a Celigo Imaging Cytometer (Nexcelom) to obtain time point 0 cell counts. Prior to co-culture, NK-92 cells (ATCC) were maintained at a density of 2-4x10^5 cells per ml in upright cell culture flasks using MyeloCult Media (Stemcell technologies) supplemented with 100 UI/ml IL2 in a humidified incubator set to 37ºC and 5% CO2. On the day of the assay, NK-92 cells were centrifuged at 125 x g for 10 minutes and resuspended in melanoma media with 200 UI/mL IL2, counted and then added to melanoma target cells at effector:target ratios as indicated for different patient models in triplicates to a final volume of 200 µL per well (final Caspase activity dye = 2 µM, final IL2 = 100 UI/mL). Co-cultures were incubated at 37ºC and 5% CO2 in

a humidified incubator. Plates were reimaged 24 hours after NK-92 addition and analyses were performed as described in the miniaturized TIL co-culture assay.

References

1. Jerby-Arnon, L. *et al.* A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. *Cell* **175**, 984-997.e24 (2018).
2. Pan, D. *et al.* A major chromatin regulator determines resistance of tumor cells to T cell-mediated killing. *Science* **359**, 770–775 (2018).
3. Goel, S. *et al.* CDK4/6 inhibition triggers anti-tumour immunity. *Nature* **548**, 471–475 (2017).
4. Honda, K., Takaoka, A. & Taniguchi, T. Type I Inteferon Gene Induction by the Interferon Regulatory Factor Family of Transcription Factors. *Immunity* **25**, 349–360 (2006).
5. Peng, W. *et al.* Loss of PTEN Promotes Resistance to T Cell-Mediated Immunotherapy. *Cancer Discov* **6**, 202–216 (2016).
6. Doench, J. G. *et al.* Rational design of highly active sgRNAs for CRISPR-Cas9-mediated gene inactivation. *Nat. Biotechnol.* **32**, 1262–1267 (2014).
7. Datlinger, P. *et al.* Pooled CRISPR screening with single-cell transcriptome readout. *Nat. Methods* **14**, 297–301 (2017).
8. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nature Biotechnology* **34**, 184–191 (2016).
9. Li, B. *et al.* Cumulus provides cloud-based data analysis for large-scale single-cell and single-nucleus RNA-seq. *Nature Methods* **17**, 793–798 (2020).
10. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

11. Finak, G. *et al.* MAST: a flexible statistical framework for assessing transcriptional changes and characterizing heterogeneity in single-cell RNA sequencing data. *Genome Biology* **16**, 278 (2015).
12. Chung, N. C. & Storey, J. D. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics* **31**, 545–554 (2015).
13. Pedregosa, F. *et al.* Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON* 6.
14. Dixit, A. Correcting Chimeric Crosstalk in Single Cell RNA-seq Experiments. *bioRxiv* 093237 (2016) doi:10.1101/093237.
15. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* **344**, 1396–1401 (2014).
16. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).