

TITLE:

Frequent Aneuploidy in Primary Human T Cells Following CRISPR-Cas9 cleavage

AUTHORS:

A.D. Nahmad^{1,2*}, E. Reuveni^{3*}, E. Goldschmidt^{4*}, T. Tenne⁵, M. Liberman⁵, M. Horovitz-Fried^{1,2}, R. Khosravi⁶, H. Kobo⁷, E. Reinstein^{5,8}, A. Madi^{4#}, U. Ben-David^{3#}, A. Barzel^{1,2#}

AFFILIATIONS:

¹The School of Neurobiology, Biochemistry and Biophysics, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

²Dotan Center for Advanced Therapies, Tel Aviv Sourasky Medical Center (TASMC) and Tel Aviv University, Tel Aviv

³Department of Human Molecular Genetics & Biochemistry, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

⁴Department of Pathology, Faculty of Medicine, Tel Aviv University, Israel

⁵Medical Genetics Institute, Meir Medical Center, Kfar-Saba, Israel

⁶Single-Cell Genomics Core, Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

⁷Genomics Research Unit, Faculty of Life Sciences, Tel Aviv University, Tel Aviv, Israel

⁸Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel

* These authors contributed equally

These authors contributed equally

Corresponding author email: adibarzel@gmail.com, ubendavid@tauex.tau.ac.il, asafmadi@tauex.tau.ac.il

SUMMARY:

Multiple ongoing clinical trials use site-specific nucleases to disrupt T cell receptor (TCR) genes in order to allow for allogeneic T cell therapy¹⁻⁵. In particular, the first U.S. clinical trial using CRISPR-Cas9 entailed the targeted disruption of the TCR chains and programmed cell death protein 1 (PDCD1) in T cells of refractory cancer patients⁶. Here, we used the same guide RNA sequences and applied single-cell RNA sequencing (scRNASeq) to more than 7000 primary human T cells, transfected with CRISPR-Cas9. Four days post-transfection, we found a loss of chromosome 14, harboring the TCR α locus, in up to 9% of the cells, and a chromosome 14 gain in up to 1.4% of the cells. We further identified truncations of chromosome 7, harboring the TCR β locus, in 9.9% of the cells. Loss of heterozygosity (LOH) was further validated using fluorescence *in situ* hybridization (FISH) and the temporal dynamics of cleavage and incomplete repair were monitored using digital droplet PCR (ddPCR). Aneuploidy was found among all T cell subsets and was associated with transcriptional signatures of reduced proliferation and metabolism as well as with induced p53 activation and cell death. We conclude that aneuploidy and chromosomal truncations are frequent outcomes of CRISPR-Cas9 cleavage in clinical protocols. Monitoring and minimizing these aberrant products is crucial for future applications of genome editing in T cell engineering and beyond.

MAIN:

Engineered T cells are revolutionizing cancer immunotherapy. Notably, four treatments entailing chimeric antigen receptor (CAR) T cells have received regulatory approval in different B cell malignancies⁷. At the same time, hundreds of ongoing clinical trials, for multiple additional indications, are applying different variations of CARs or engineered TCRs in an attempt to improve efficacy, safety and scalability¹. Importantly, the approved CAR T therapies all entail the cumbersome, expensive and time-consuming manipulation of autologous T cells. Many current efforts are therefore directed at facilitating the engineering and banking of allogeneic CAR/TCR-T cell products. Importantly, the endogenous TCR of engineered allogeneic T cells has to be disrupted in order to prevent graft vs. host disease (GVHD) and this is most often achieved using site-specific nucleases such as meganucleases², TALENs^{3,4}, megaTALs and CRISPR-Cas9⁵. Cleaving the TCR loci can further serve for the directed integration of CAR and engineered TCR genes, allowing uniform expression, enhancing T cell potency and delaying exhaustion⁸⁻¹⁰. Finally, additional genes, beyond the TCR chains, may be disrupted in order to improve T cell function¹¹, confer drug resistance¹², or avoid checkpoint inhibition¹³.

Site-specific nucleases can be highly potent¹⁴, but they are associated with a variety of undesired outcomes. Cas9 can be immunogenic *in vivo*, eliciting both humoral and cellular responses¹⁵. We recently demonstrated that Cas9 activates the p53 pathway and selects for p53-inactivating mutations¹⁶. Off-target cleavage can in turn be

reduced, but not eliminated¹⁷, and small on-target insertions and deletions (InDels) are common, even when providing a donor template for gene correction or insertion. Importantly, CRISPR-Cas9 cleavage can also lead to gross chromosomal aberrations. In particular, CRISPR-Cas9 cleavage leads to large deletions in early mouse¹⁸ and human embryos¹⁹ as well as in embryonic stem cells and induced pluripotent stem cells²⁰. Moreover, chromosomal truncations were reported following CRISPR-Cas9 cleavage in mouse and human cell lines^{21,22}, and entire chromosome loss has resulted from CRISPR-Cas9 cleavage in human embryos²³. Recently, chromothripsis, defined as multiplicity of gross genomic rearrangements in a one-off cellular crisis, was shown to occur in mobilized human CD34⁺ cells following CRISPR-Cas9 cleavage of the β -globin gene in a clinically relevant setting^{24,25}, with possible ramifications for the development of gene therapies for hemoglobinopathies. Finally, in the first U.S. clinical trial involving CRISPR-Cas9, Stadtmauer & Fraietta et al. aimed at disrupting the TCR and PDCD1 loci for allogeneic T cell therapy and reported the detection of chromosomal translocations whose frequency decreased over time after infusion into patients⁶.

Here, we use the same guide RNA sequences, as did Stadtmauer & Fraietta et al.⁶, to target the TCR and PDCD1 loci in primary human T cells with CRISPR-Cas9. Using a novel unbiased approach, based on scRNAseq, corroborated by FISH and ddPCR analyses, we detect frequent aneuploidy and truncations of the chromosomes harboring the targeted loci.

We hypothesized that double-strand DNA breaks, induced by site-specific nucleases, may sometime fail to be repaired and result instead in chromosomal truncations and aneuploidy. We further conjectured that such adverse outcomes could be detected and monitored, in a high throughput manner, by following concerted changes in gene expression along chromosomes. In order to analyze chromosomal truncations and aneuploidy in a clinically relevant setting, we first performed CRISPR-Cas9 ribonucleoprotein (RNP) electroporation into primary human T cells using a single guide RNA (gRNA) targeting the TCR α locus (Fig. 1A). Importantly, we used the same gRNA sequence as used clinically in T cells by Stadtmauer & Fraietta et al.⁶. As a control, we used an irrelevant gRNA¹⁴, with no matching target in the human genome. The TCR α locus was successfully targeted in more than 50% of the cells (Fig. 1B-C, Extended Data Fig. 1). 4 days post-transfection, samples were subjected to scRNAseq (Methods). The transcriptional downregulation of TCR α was confirmed (Fig. 1D), and the transcriptional landscapes of the cells were used to infer copy number alterations using a well-established method (inferCNV)^{26,27}. Strikingly, 5.3% of the cells in the TCR α targeted sample had expression patterns indicating a chromosome 14 loss (Fig. 1E-F). In addition, among these cells, chromosome 14 was an extreme outlier in the mean number of genes with undetected expression (p value < 0.0001, Extended Data Fig. 2). Notably, a chromosome 14 q-arm truncation at the TCR α locus is functionally equivalent to a whole-chromosome aneuploidy, because almost all protein-coding genes on chromosome 14 are coded on the q-arm, distal to the TCR α locus (Fig. 1A). Interestingly, a smaller fraction of the cell population, entailing 0.5-0.9% of the cells, had an apparent gain of chromosome 14 (Fig 1E-F). This functional chromosome 14 gain is assumed to result from mis-segregation of an acentromeric q-arm²⁴. In particular, 100 chromosome 14 genes were underexpressed and 107 chromosome 14 genes were overexpressed in the cells categorized as harboring a chromosome 14 loss or gain, respectively (Extended Data Fig. 2, Supplementary Table 1). To set a clear statistical threshold for gains and losses, only cells with inferCNV scores >2 standard deviations from the mean score of the population were defined as having a chromosomal aberration (Methods).

In order to corroborate the scRNAseq results, we next performed FISH assays employing red and green probes proximal and distal to the TCR α locus on chromosome 14, respectively. Expectedly, most examined nuclei had two foci of co-localized red and green signals (often co-detected as a yellow signal). Importantly, in two independent FISH iterations, signal patterns indicative of aneuploidy were observed significantly more frequently in TCR α -targeted cells than in control cells (Fig. 2A, Extended Data Fig. 3). These cells had either a single focus of co-localized signals or one focus of co-localized signals and another red focus, interpreted as corresponding to a truncation distal to the targeted gene. In particular, the 4% excess rate of these FISH patterns in TCR α -targeted cells, compared to control cells, is in agreement with the scRNAseq analysis.

The rate of chromosome 14 breakage and its incomplete repair can further be monitored using digital droplet PCR (ddPCR)²⁸ (Fig. 2B). DNA was collected at different time points after CRISPR-Cas9 RNP electroporation, in order to assess the dynamics of dislinkage between the two sides of the break. As early as 3 hrs after electroporation, the rate of dislinkage has peaked at more than 45% (Fig. 2C-D). Dislinkage rates were greatly reduced by 24 hrs, but remained significantly higher in the TCR α -targeted sample, compared to the control sample, for at least 4 days, indicating incomplete repair. While this ddPCR analysis alone cannot discriminate between transient breaks, translocations and deletions of various sizes, its results are consistent with the results from the

scRNAseq and FISH analyses. Moreover, the ddPCR assay provides a scalable and cost-effective means to track the dynamics of breakage, repair and aberration.

We next used t-distributed stochastic neighbor embedding (t-SNE) analysis in order to cluster T cells according to their transcriptional signatures (Fig. 3A, Extended Data Fig. 4). We found a similar distribution, among the clusters, of T cells transfected with CRISPR-Cas9 RNPs entailing either the TCR α -targeting gRNA or an irrelevant gRNA with no human target (Fig. 3B). Moreover, among cells receiving the TCR α -targeting gRNA, neither cells with chromosome 14 loss nor cells with chromosome 14 gain were found to be considerably enriched in any T cell subset (Fig. 3C). This implies unbiased incidence of aneuploidy, as well as similar selection forces acting on aneuploid cells of various T cell states. Importantly, however, a gene set enrichment analysis (GSEA) detected significant differences in global gene expression patterns between the groups, beyond the expected reduced expression of the genes encoded on chromosome 14 (Fig. Extended Data Fig. 5). Specifically, the transcriptional signatures enriched in T cells with a chromosome 14 loss or gain reflected reduced expression of genes associated with the cell cycle and with various metabolic pathways, and increased expression of genes associated with p53 pathway activation and with apoptosis, implying reduced fitness of the aneuploid cells (Fig. 3D-F and Supplementary Table 2).

Targeting different loci concomitantly with several gRNAs may aggravate the risk for chromosomal aberrations. Indeed, Stadtmauer & Fraietta et al. identified multiple translocations between cleaved chromosomes when co-delivering 3 gRNAs targeting both TCR loci as well as the PDCD1 gene⁶. We next used the same gRNA combination for CRISPR-Cas9 RNP electroporation of primary human T cells (Fig. 4A). We verified cleavage efficiency (Extended Data Fig. 6) and performed scRNAseq in order to analyze gene expression patterns and infer copy number changes along the targeted chromosomes. The transcriptional downregulation of TCR α , TCR β and PDCD1 was confirmed (Fig. 4B). Importantly, in this experiment, we found expression patterns indicative of a chromosome 14 loss and gain in 9.0% and 1.4% of the cells, respectively (Fig. 4C-D, Extended Data Fig. 7). The higher chromosome 14 aberration rate, compared to the previous experiment, may be explained by the 3-fold higher Cas9 concentrations, used in order to preserve the Cas9-per-gRNA ratio in the combined targeting setting. Strikingly, we further found as many as 9.9% of treated cells to be characterized with chromosome 7 truncations, entailing the TCR β locus (Fig. 4C-D). This result suggests that the repair of nuclease-induced breaks in primary human T cells may be inefficient. Importantly, the extent of the detected aberrations is in good correspondence with the chromosomal position of the cleavage sites. Cleaving the TCR α locus, near the chromosome 14 centromere, leads to loss of the entire arm and to functional whole-chromosome aneuploidy; cleaving the TCR β locus, in the middle of the chromosome 7 q-arm, leads to the expected truncations and cleaving the PDCD1 gene, which resides near the chromosome 2 q-arm telomere, expectedly has little effect on copy number and gene expression (Extended Data Fig. 8). Finally, a gene set enrichment analysis confirmed the transcriptional differences observed in our first experiment in cells that have lost or gained chromosome 14, and identified very similar pathway enrichments in the cells with a chromosome 7 truncation (Fig. 4E-G, Extended Data Fig. 9 and Supplementary Table 3).

Aneuploidy is associated with most types of cancers²⁹. Of note, chromosome 14 loss is frequently found in early-onset colon cancer³⁰, and is among the most commonly affected chromosomal regions in ovarian cancer³¹. Chromosome 14 monosomy, trisomy and tetrasomy are also frequently found in meningioma³². Similarly, high grade glioma is often characterized by chromosome 7 trisomy³³. Due to the role of aneuploidy in tumorigenesis, aneuploid cells are considered a tumorigenic risk and are strictly avoided in cell transplantations³⁴. Our observations raise the possibility that Cas9-induced aneuploidy and chromosomal truncations, observed in our study, will too be associated with increased risk of tumorigenesis. Importantly, the methods we described were optimized for the detection of aneuploidy and large truncations, therefore underestimating the overall occurrence of genetic alterations. While detected in some cases (Extended Data Fig. 3), translocations and focal deletions may better be assessed using complementary technologies⁶.

Aberration rates can potentially be reduced by using base editors instead of nucleases³⁵. In the liver, gene insertion by homologous recombination can be achieved without nucleases³⁶⁻³⁸. However, in T cells, nucleases may still be required for site-specific gene insertion of CAR and engineered TCR genes into TCR loci in order to enhance T cell potency and delay exhaustion⁸⁻¹⁰. Further studies are required in order to determine whether sorting for expression of the targeted receptor cassette will help reduce the rate of cells with nuclease-induced aberrations. Alternatively, endogenous genes that are differentially expressed in aberrant T cells may be used for cell sorting prior to adoptive transfer. As a first step towards this aim, we compiled a list of genes whose expression

significantly differed between the cells that have lost chromosome 14 and those that have not (Supplementary Table 1), and manually curated the list to identify cell surface proteins. Interestingly, CD70, which is expressed at low levels on naïve T cells³⁹, is upregulated in cells with a chromosome 14 loss (Extended Data Fig. 10). Similarly, L-Selectin (SELL/CD62L) is highly expressed in naïve T cells⁴⁰ and is downregulated in chromosome 7 truncated cells (Extended Data Fig. 10). Finally, our gene set enrichment analysis suggests that aneuploidy provides fitness disadvantage to the T cells *in vitro* (Fig. 3D-F, Supplementary Table 2-3), in line with multiple studies that showed the negative effect of aneuploidy induction on cell survival and proliferation^{29,41,42}. Additional propagation of the cells in culture prior to transplantation may reduce the prevalence of aneuploid cells, and extended culturing should thus be considered, while also weighing the effects on cell activation, differentiation and exhaustion.

In conclusion, inferring copy number variations from scRNASeq data is an unbiased high-throughput method to identify nuclease-induced chromosomal aberrations. Our discovery of frequent aneuploidy and chromosomal truncations in human T cells targeted using CRISPR-Cas9 and clinical gRNA sequences, highlights potential oncogenic risks and underscores the need for mitigation strategies in order to allow the safe application of nucleases in adoptive T cell transfer and beyond.

REFERENCES:

1. Weber, E. W., Maus, M. V. & Mackall, C. L. The Emerging Landscape of Immune Cell Therapies. *Cell* **181**, 46–62 (2020).
2. MacLeod, D. T. *et al.* Integration of a CD19 CAR into the TCR Alpha Chain Locus Streamlines Production of Allogeneic Gene-Edited CAR T Cells. *Mol. Ther.* **25**, 949–961 (2017).
3. Philip, L. P. B. *et al.* Multiplex genome-edited T-cell manufacturing platform for ‘off-the-shelf’ adoptive T-cell immunotherapies. *Cancer Res.* **75**, 3853–3864 (2015).
4. Qasim, W. *et al.* Molecular remission of infant B-ALL after infusion of universal TALEN gene-edited CAR T cells. *Sci. Transl. Med.* **9**, 1–10 (2017).
5. Osborn, M. J. *et al.* Evaluation of TCR Gene Editing Achieved by TALENs, CRISPR / Cas9 , and megaTAL Nucleases. *Mol. Ther.* **24**, 570–581 (2016).
6. Stadtmauer, E. A. *et al.* CRISPR-engineered T cells in patients with refractory cancer. *Science (80-.)* **367**, (2020).
7. Braendstrup, P., Levine, B. L. & Ruella, M. The long road to the first FDA-approved gene therapy: chimeric antigen receptor T cells targeting CD19. *Cytotherapy* **22**, 57–69 (2020).
8. Eyquem, J. *et al.* Targeting a CAR to the TRAC locus with CRISPR/Cas9 enhances tumour rejection. *Nature* **543**, 113–117 (2017).
9. Roth, T. L. *et al.* Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature* **559**, 405–409 (2018).
10. Schober, K. *et al.* Orthotopic replacement of T-cell receptor α - and β -chains with preservation of near-physiological T-cell function. *Nat. Biomed. Eng.* (2019). doi:10.1038/s41551-019-0409-0
11. Shifrut, E. *et al.* Genome-wide CRISPR Screens in Primary Human T Cells Reveal Key Regulators of Immune Function. *Cell* **175**, 1958-1971.e15 (2018).
12. Valton, J. *et al.* A Multidrug-resistant Engineered CAR T Cell for Allogeneic Combination Immunotherapy. *Mol. Ther.* **23**, 1507–1518 (2015).
13. Rupp, L. J. *et al.* CRISPR / Cas9-mediated PD-1 disruption enhances anti-tumor efficacy of human chimeric antigen receptor T cells. *Sci. Rep.* **737**, 1–10 (2017).
14. Nahmad, A. D. *et al.* Engineered B cells expressing an anti-HIV antibody enable memory retention, isotype switching and clonal expansion. *Nat. Commun.* 1–10 (2020). doi:10.1038/s41467-020-19649-1
15. Charlesworth, C. T. *et al.* Identification of preexisting adaptive immunity to Cas9 proteins in humans. *Nat. Med.* **25**, (2019).
16. Enache, O. M. *et al.* Cas9 activates the p53 pathway and selects for p53-inactivating mutations. *Nat. Genet.* **52**, 662–668 (2020).
17. Lazzarotto, C. R. *et al.* CHANGE-seq reveals genetic and epigenetic effects on CRISPR–Cas9 genome-wide activity. *Nat. Biotechnol.* (2020). doi:10.1038/s41587-020-0555-7
18. Adikusuma, F. *et al.* Large deletions induced by Cas9 cleavage. *Nature* **560**, E8–E9 (2018).
19. Alanis-lobato, G., Zohren, J., McCarthy, A., Fogarty, N. M. E. & Kubikova, N. Frequent loss-of-heterozygosity in CRISPR-Cas9 – edited early human embryos. *Proc. Natl. Acad. Sci. USA* (2021).

doi:10.1073/pnas.2004832117

20. Weisheit, I. *et al.* Detection of Deleterious On-Target Effects after HDR-Mediated CRISPR Editing. *Cell Rep.* **31**, 107689 (2020).
21. Przewrocka, J., Rowan, A., Rosenthal, R., Kanu, N. & Swanton, C. Unintended on-target chromosomal instability following CRISPR/Cas9 single gene targeting. *Ann. Oncol.* **31**, 1270–1273 (2020).
22. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* **36**, (2018).
23. Zuccaro, M. V. *et al.* Allele-Specific Chromosome Removal after Cas9 Cleavage in Human Embryos. *Cell* **183**, 1650–1664.e15 (2020).
24. Leibowitz, M. L. *et al.* Chromothripsis as an on-target consequence of CRISPR–Cas9 genome editing. *Nat. Genet.* **53**, 895–905 (2021).
25. Urnov, F. D. CRISPR–Cas9 can cause chromothripsis. *Nat. Genet.* **53**, 765–769 (2021).
26. Patel, A. P. *et al.* Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science (80-.).* **344**, 1396–1402 (2014).
27. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science (80-.).* **352**, 189–196 (2016).
28. Puig, M. *et al.* Determining the impact of uncharacterized inversions in the human genome by droplet digital PCR. *Genome Res.* **30**, 724–735 (2020).
29. Ben-David, U. & Amon, A. Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.* **21**, 44–62 (2020).
30. Mourra, N. *et al.* High Frequency of Chromosome 14 Deletion in Early-Onset Colon Cancer. *Dis. Colon Rectum* **50**, 1881–1886 (2007).
31. Bandera, C. A. *et al.* Deletion Mapping of Two Potential Chromosome 14 Tumor Suppressor Gene Loci in Ovarian Carcinoma. *Cancer Res.* **57**, 513–516 (1997).
32. Tabernero, M. D. *et al.* Characterization of Chromosome 14 Abnormalities by Interphase In Situ Hybridization and Comparative Genomic Hybridization in 124 Meningiomas. *Anat. Pathol.* 744–751 (2005). doi:10.1309/D7U997XD2PHBCQCN
33. Lopez-Gines, C. *et al.* Association of chromosome 7, chromosome 10 and EGFR gene amplification in glioblastoma multiforme. *Clin. Neuropathol.* **24**, 209–218 (2005).
34. Andrews, P. W. *et al.* Assessing the Safety of Human Pluripotent Stem Cells and Their Derivatives for Clinical Applications. *Stem Cell Reports* **9**, 1–4 (2017).
35. Webber, B. R. *et al.* Highly efficient multiplex human T cell engineering without double-strand breaks using Cas9 base editors. *Nat. Commun.* (2019). doi:10.1038/s41467-019-13007-6
36. Barzel, A. *et al.* Promoterless gene targeting without nucleases ameliorates haemophilia B in mice. *Nature* **517**, 360–364 (2015).
37. Porro, F. *et al.* Promoterless gene targeting without nucleases rescues lethality of a Crigler-Najjar syndrome mouse model. *EMBO Mol. Med.* **9**, 1346–1355 (2017).
38. Chandler, R. J. *et al.* Promoterless, Nuclease- Free Genome Editing Confers a Growth Advantage for Corrected Hepatocytes in Mice With Methylmalonic Acidemia. *Hepatology* **73**, 2223–2237 (2021).
39. Borst, J., Hendriks, J. & Xiao, Y. CD27 and CD70 in T cell and B cell activation. *Curr. Opin. Immunol.* **17**, 275–281 (2005).
40. Ivetic, A., Louise, H., Green, H. & Hart, S. J. L-selectin \square : A Major Regulator of Leukocyte Adhesion , Migration and Signaling. *Front. Immunol.* **10**, 1–22 (2019).
41. Rutledge, S. D. *et al.* Selective advantage of trisomic human cells cultured in non- standard conditions. *Sci. Rep.* **6**, (2016).
42. Sheltzer, J. M. *et al.* Single-chromosome Gains Commonly Function as Tumor Suppressors. *Cancer Cell* **31**, 240–255 (2017).
43. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data Resource Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
44. Liberzon, A. *et al.* The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* **1**, 417–425 (2015).
45. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15545–15550 (2005).
46. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).
47. Brinkman, E. K., Chen, T., Amendola, M. & Van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, 1–8 (2014).

FIGURES:

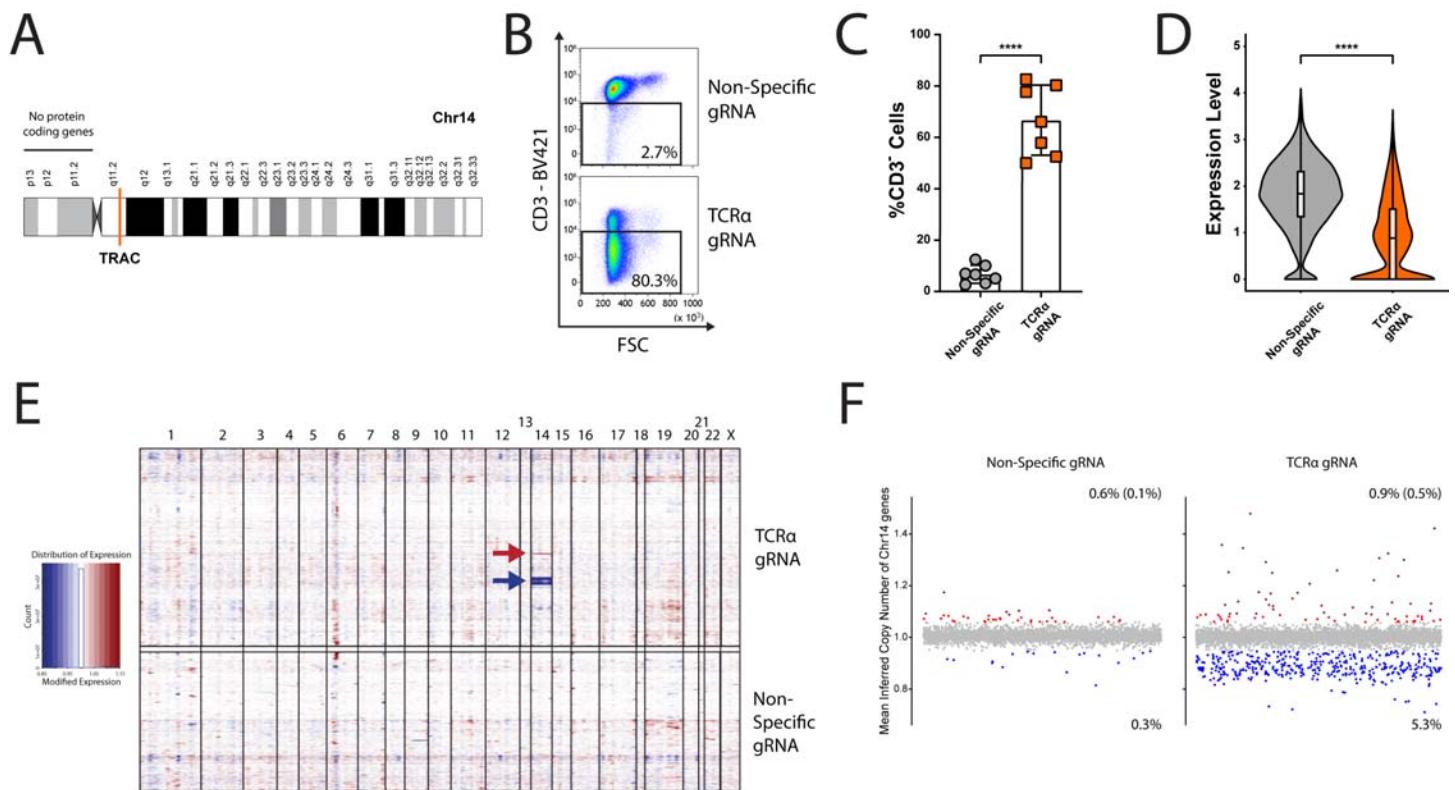
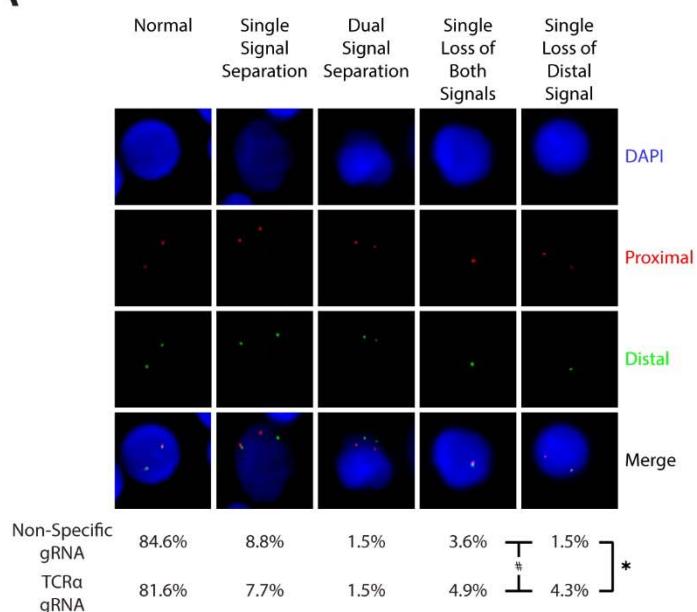
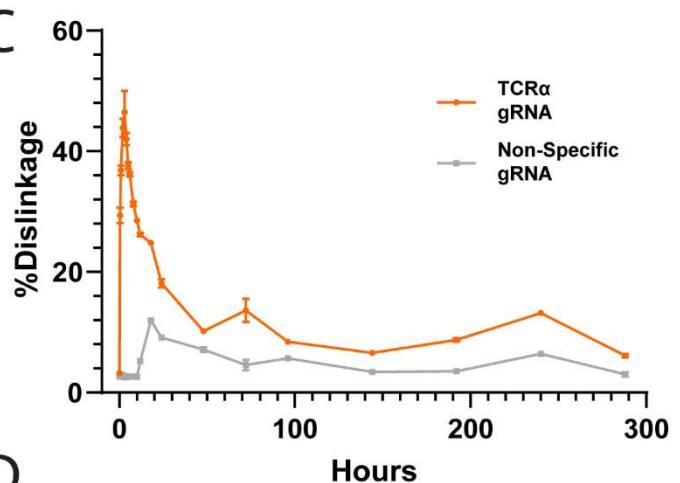


Figure 1: Targeting the TCR α locus using CRISPR-Cas9 leads to chromosome 14 aneuploidy. **A.** Schematic depiction of human chromosome 14 (Chr14). The target locus is indicated in orange. **B.** Flow cytometry example of TCR ablation in primary human T cells following CRISPR-Cas9 RNP electroporation. Cells were electroporated with Cas9 and either a non-specific gRNA or a TCR α -targeting gRNA. **C.** Quantification of B. Each dot represents an independent experiment. Mean value, standard deviation and individual experiments are indicated. $n=7$, ****, $p<0.0001$, two-sided unpaired Wilcoxon test. **D.** A reduction in TCR α expression is evident in the scRNAseq. The violin plots correspond to TCR α expression level in cells treated with either the non-specific gRNA or the TCR α -targeting gRNA. Upper and lower boundaries, median and quartiles are indicated. **** = $p<0.0001$, two-sided unpaired Wilcoxon test. **E.** A heat map depicting gene copy numbers inferred from scRNAseq analysis following treatment with either the non-specific or TCR α -targeting gRNAs. The Chromosomes are ordered in columns and the color coding indicates an increase (red) or decrease (blue) in copy number of genes along the chromosomes (x-axis) and each line represents an individual cell. Arrows indicate gain (red) or loss (blue) of chromosome 14. **F.** Each dot represents the mean inferred copy number of genes coded on chromosome 14 in each cell treated with a non-specific gRNA (left) or a TCR α -targeting gRNA (right). Cells are marked with red and blue dots, corresponding to a chromosome 14 gain or loss respectively, if their mean inferred gene copy number is >2 standard deviations (blue and red), or 3 standard deviations (dark red), from the population's mean. $n=6055$ and 7700 for Non-specific (left) or TCR α -targeting (right) gRNA treated cells, respectively.

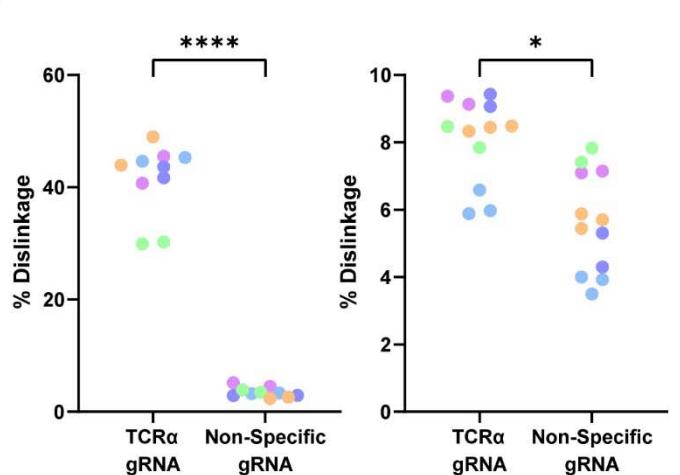
A



C



D



B

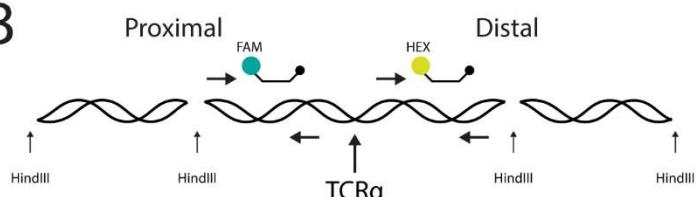


Figure 2: FISH and ddPCR analyses of Chromosome 14 aberrations. **A.** FISH analysis of cells treated with either a non-specific or a TCR α -targeting gRNA. In the pictures are examples of all signal patterns found among analyzed cells. The frequency of each signal pattern, among cells electroporated with the TCR α -targeting gRNA or a non-specific gRNA, is presented below the pictures. $n = 273$ and $n = 506$ for cells electroporated with the non-specific and TCR α -targeting gRNA, respectively. * $p=0.0129$ Fisher's exact test for loss of distal signal comparison. #, $p=0.0486$ Fisher's exact test for loss of either both signals or the distal signal. **B.** Schematic depiction of the ddPCR experiment. Genomic DNA is sheared using the HindIII restriction enzyme. The two sets of primers and probe are on the same restriction fragment are separated only upon CRISPR-Cas9 cleavage. **C.** Dislinkage, calculated based on the rate of single and dual color droplets (see methods) is analyzed at different time points following CRISPR-Cas9 electroporation with either a TCR α -targeting gRNA (orange) or a non-specific (gray) gRNA from a single experiment. Mean and standard deviation for technical replicates are indicated. **D.** Dislinkage at either 3 hours (left) or 4 days (right) following CRISPR-Cas9 electroporation. Each color represents a different human donor. Each dot represents a technical replication. $n=5$, ****, $p<0.0001$ and *, $p=0.0277$, two-sided unpaired t-test.

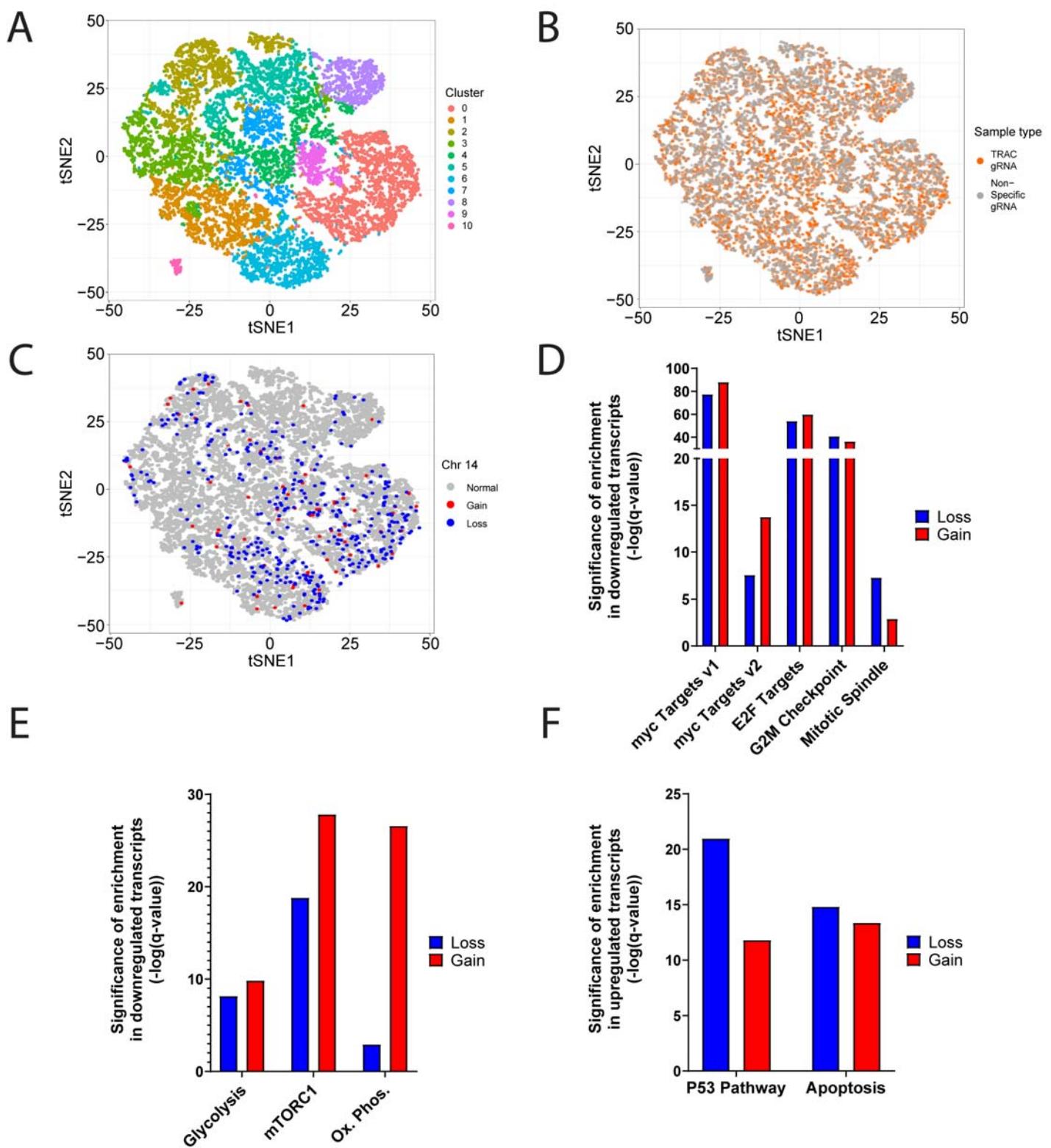


Figure 3: Global gene expression analysis following genome editing. **A.** scRNAseq data from primary human T cells showing 13,755 cells as individual dots. Data is displayed by tSNE with each cluster colored differently. **B.** Cells electroporated with a non-specific (gray) or TCR α targeting (orange) gRNAs are similarly distributed among the clusters. **C.** Cells with a chromosome 14 loss (blue) or gain (red) are similarly distributed among the clusters. **D-F.** Gene set enrichment analysis reveals that 'Hallmark' gene sets associated with cell cycle (D) and metabolism (E) are significantly downregulated; and the 'Hallmark' gene sets associated with p53 and apoptosis (F) are significantly upregulated, both in cells with chromosome 14 loss (blue) and in cells with chromosome 14 gain (red). Full enrichment matrix can be found in Extended Data Fig. 5 and Supplementary Table 2.

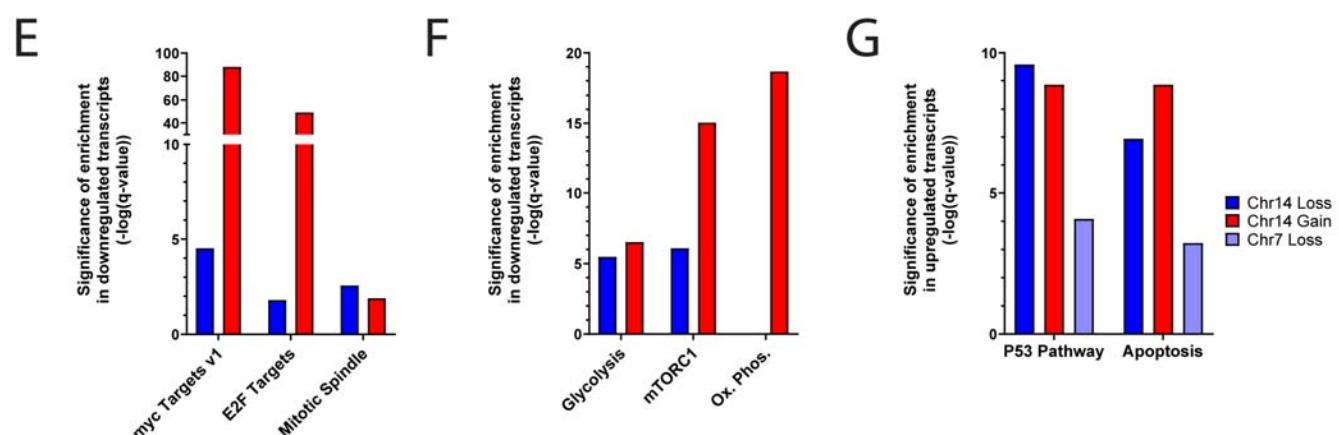
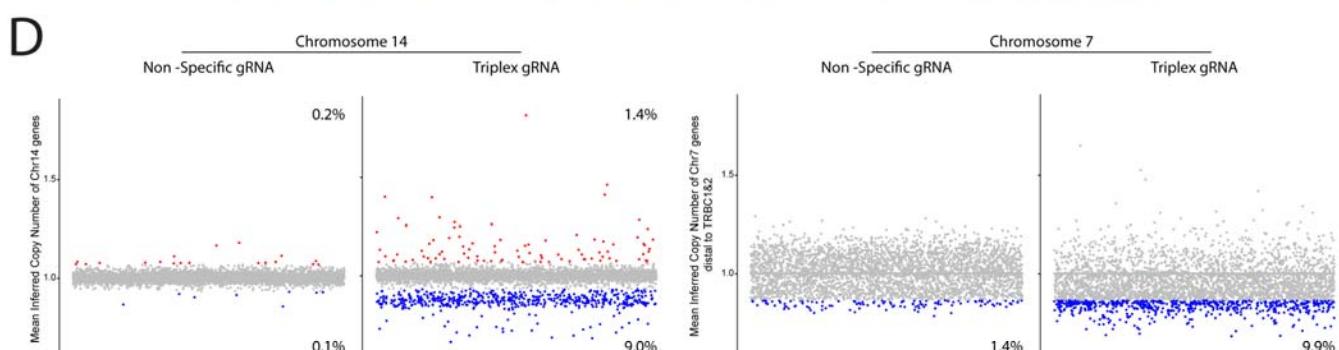
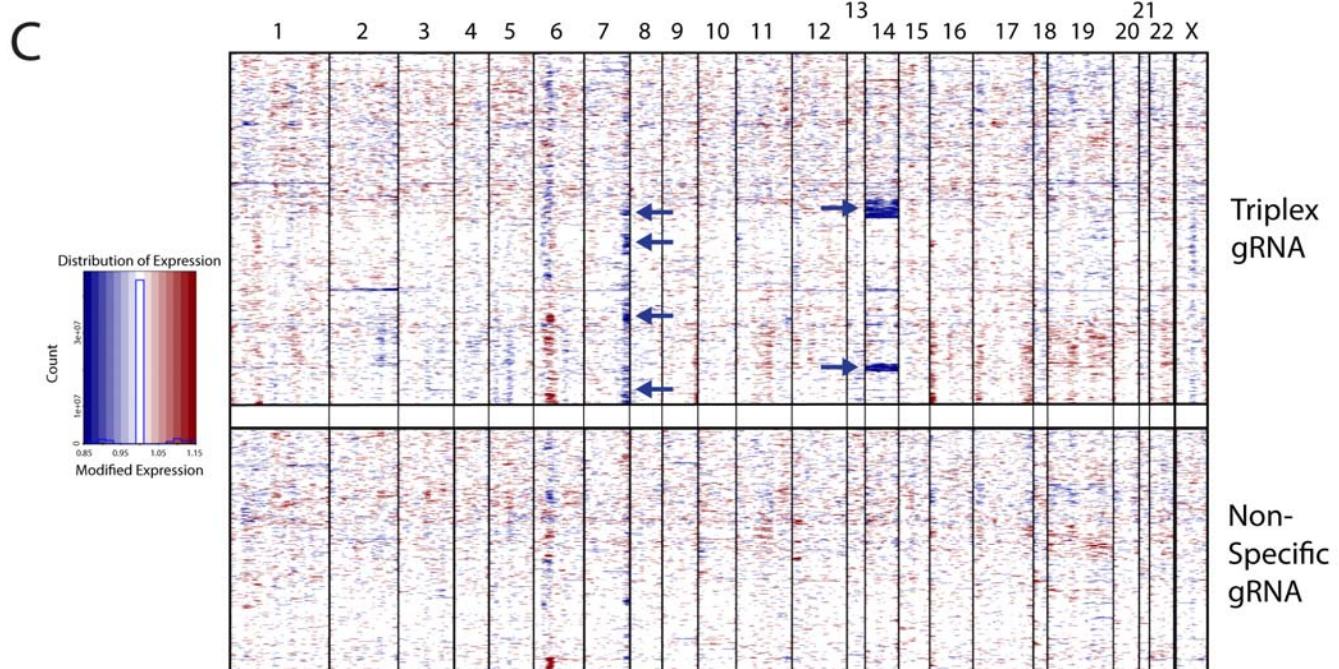
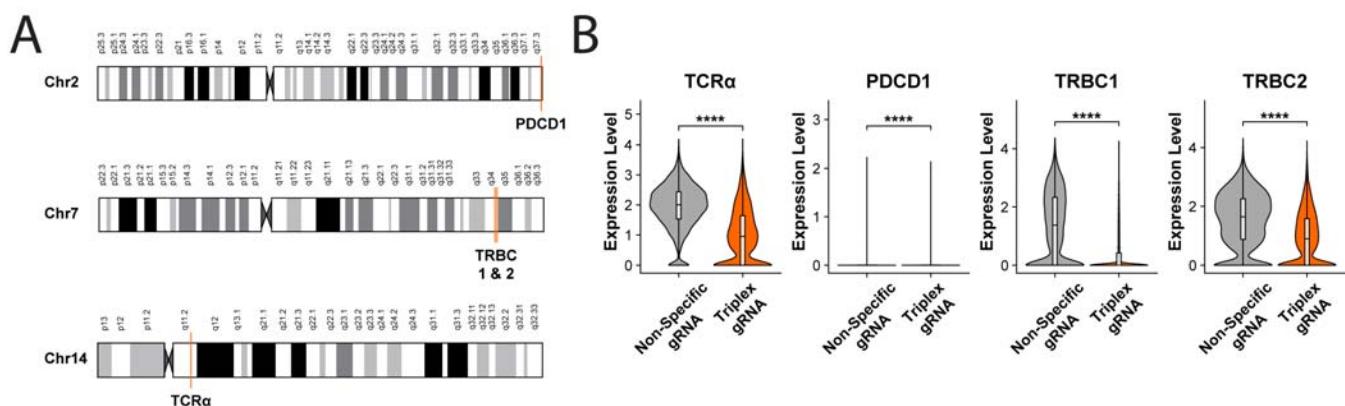


Figure 4: Targeting different loci concomitantly with several gRNAs aggravates chromosomal aberrations. **A.** Schematic depiction of the human chromosomes 2, 7 and 14. Target loci are shown in orange. A single gRNA targets both TRBC1 and TRBC2. **B.** RNA expression of TCR α , PDCD1, TRBC1 and TRBC2 in cells treated with either the non-specific gRNA or the combination of TCR α , TRBCs and PDCD1- targeting gRNAs. ****, p<0.0001, two-sided unpaired Wilcoxon test. **C.** A heat map depicting gene copy numbers inferred from scRNAseq analysis following treatment with either the non-specific or the combination of TCR α , TRBCs and PDCD1- targeting gRNAs. The chromosomes are ordered in columns and the color coding indicates an increase (red) or decrease (blue) in copy number. Blue arrows indicate loss of chromosome 14 or chromosome 7 segments. **D.** Each dot represents the mean inferred copy number of genes coded on the relevant chromosome in each cell treated with a non-specific gRNA (left) the combination of TCR α , TRBCs and PDCD1- targeting gRNAs (right). For chromosome 7, the mean inferred copy number of genes coded distal to the TRBC1&2 is analyzed. Cells are marked with red and blue dots, corresponding to a gain or a loss, respectively, if their mean inferred gene copy number is >2 standard deviations away from the population's mean. n=8619 and 6326 for cells treated with the non-specific gRNA (left) or the combination of TCR α , TRBCs and PDCD1- targeting gRNAs (right), respectively. **E-G.** Gene set enrichment analysis reveals that 'Hallmark' gene sets associated with cell cycle (E) and metabolism (F) are significantly downregulated; and the 'Hallmark' gene sets associated with p53 and apoptosis (G) are significantly upregulated, both in cells with chromosome 14 loss (blue) and in cells with chromosome 14 gain (red). Full enrichment matrix can be found in Extended Data Fig. 9 and Supplementary Table 3.

METHODS:

T cell editing

Whole blood was obtained with donor consent from the Israeli Blood Bank (Magen David Adom, Sheiba Medical Center) in accordance with Tel Aviv University Review Board. PBMCs were extracted using Lymphocyte Separation Medium (mpbio) and cryopreserved until subsequent use. Following thawing, cells were activated for 24-48hrs with 1ug/ml anti-human CD28 (biogems) and anti-human CD3 (biogems). Cells were cultured in MEM-Alpha (Biological Industries) supplemented with 10% Heat Inactivated FCS (Sigma), 50IU rhIL-2 (Peprotech) and P/S. For RNP electroporation, 18pmol of Alt-R spCas9 Nuclease V3 (IDT) and 66pmol of Alt-R CRISPR-Cas9 gRNA (IDT) per 1E6 cells per target were complexed in Buffer T (IDT). Experiments involving a triple targets (TRAC, TRBC, PD-1), quantities of Cas9:gRNA complexes were tripled for the non-specific control. gRNA sequences can be found in Supplementary Table 4. Cells were harvested, washed and electroporated at 1600v, 10ms, 3pulses in 10ul Buffer T and subsequently grown in culture media devoid of P/S.

Flow Cytometry

Cells were harvested, washed, and resuspended in Cell Staining Buffer (Biolegend) containing 1/100 diluted anti-human CD3 or anti-human TCR α / β (Biolegend), both targeting the TCR complex. Staining was performed for 15mins at room temperature in the dark. Finally, cells were washed and data acquisition was performed on an Attune NxT Flow Cytometer (Life Technologies).

Single-cell RNA sequencing

Edited T cells were harvested and live cells were purified using Lymphocyte Separation Medium (mpbio), washed and resuspended in PBS supplemented with 0.5% BSA to achieve optimal concentration of approximately 1000 cells per microliter. Cells were counted and viability assessed manually in Trypan Blue 0.4% (Biological Industries). 17,000 cells were loaded on Next GEM Chip (10x Genomics). Libraries were prepared at the Single-Cell Genomics Core, Faculty of Medicine, Tel Aviv University, using the 10x Genomics Chromium Controller in conjunction with the single-cell 3 μ v3.1 kit, protocol revision D. The cDNA synthesis, barcoding, and library preparation were then carried out according to the manufacturer's instructions. Briefly, cDNA amplification was performed for 11 cycles. Sample index PCR was performed for 13 cycles using Chromium i7 Sample Indices. Resulting libraries were quantified and analyzed by Qubit and Tapestation. Libraries were sequenced on the NextSeq 500 platform (Illumina), following the manufacturer's protocol, using a NextSeq 500/550 High Output Kit v2.5 (75 Cycles) kit (Illumina). Sequencing was performed at the Genomics Research Unit, at the Life Sciences Inter-Departmental Research Facility Unit, Tel-Aviv University.

scRNA-seq gene expression pre-processing

Raw BCL files for the DNA sequencing data were processed using Cellranger DNA (version 5.0.1). Data were aligned to the 10X Genomics GRCh38 genome. Results were visualized in the Loupe scDNA Browser (version 5.0.0). Raw gene expression data were extracted from the Seurat object as recommended in the "Using 10x data" section (inferCNV of the Trinity CTAT Project, <https://github.com/broadinstitute/inferCNV>).

InferCNV

InferCNV was used to infer copy number changes from the gene expression profiles^{26,27}. The non-targeted T cell population was used as the reference, and the CRISPR/Cas9-targeted population was tested, with the following parameters: "denoise", default hidden markov model (HMM) settings, and a value of 0.1 for "cutoff".

Identification of aberrant cells

For each cell, the mean of inferCNV scores was calculated across genes and plotted. The PDCD1 gene resides near the chromosome 2 q-arm telomere and only 15 genes, expressed in T cells, reside between the TCR α gene and the chromosome 14 centromere. Therefore, for chromosome 14 and chromosome 2, all of the expressed genes on the respective chromosomes were used for the analysis. For chromosome 7, the 47 expressed genes that reside distal to TCR β were used. Cells with a mean lower or higher >2 standard deviations from the mean of the population were determined as cells with a loss or a gain, respectively.

Simulation analysis for downregulated genes

In order to determine whether cells categorized as harboring a chr14 loss of a chr7 distal loss, also had a significant increase in the fraction of zero expression calls (that is, whether these regions are enriched with genes not detected at all by scRNAseq), the ratio between the number of genes from each chromosome with expression = 0 and expression > 0 for each cell population (loss vs. non-loss) was calculated. The fold change of the proportion of zeros calls between the normal and aberrant cells was determined, and 10,000 simulations were then performed, selecting an equivalent number of random genes from other chromosomes. An empirical p-value was determined by comparing the empirical values to the simulated values.

Differential gene expression analysis

The “FindMarkers” package from the Seurat library⁴³ was used to detect the differentially expressed (DE) genes between two groups of cells. The function receives two identities of clusters in the data set and a value for the minimum percentage that is required for a feature to be detected in either of the two groups of cells. The minimum percentage value that we used is 0.25.

The first comparison was made to detect DE genes between the cells that have undergone loss in the TRAC gRNA and the cells that did not show loss or gain in the same treatment group. The second was to detect DE genes between the cell that have undergone gain in the TRAC gRNA and the cells that did not show gain or loss in the same treatment group.

Gene Set Enrichment Analysis

The lists of differentially expressed genes between each two conditions were determined using Seurat as described above. These lists were subjected to gene set enrichment analysis (GSEA) using the GSEA-MSigDB portal (<https://www.gsea-msigdb.org/gsea/msigdb/>). The analysis was run using the following curated gene sets: “Hallmark”, “KEGG”, “GO biological process” and “positional” gene sets from MSigDB⁴⁴⁻⁴⁶.

ddPCR

For digital droplet PCR, whole genomic DNA was extracted from cells using Gentra PureGene Tissue Kit (Qiagen). In order to remove sheared genomic fragments, resulting eluates were further purified using AmpureXP beads (Beckman Coulter) at a 0.5:1 ratio. DNA fragmentation by digestion was performed in reaction, using 66ng of purified genomic DNA and 10U HindIII-HF (NEB) in ddPCR Supermix for Probes (BioRad). Thermo-cycling reaction was performed as per manufacturer recommendation. Sequences for the primers and probes can be found in Supplementary Table 4. Reactions were performed using a QX200 Droplet Digital PCR System (Bio-Rad). To analyze for dislinkage, we used the following equation as per the resulting Quantasoft (BioRad) Linkage (*Linkage*) and Concentration (C_{HEX} , C_{FAM}) values:

$$\% \text{ Dislinkage} = 100 - \left(\frac{\text{Linkage}}{\frac{C_{HEX} + C_{FAM}}{2}} \times 100 \right)$$

For dislinkage followup, multiple electroporations of treated cells were pooled and then divided in separate wells for collection at each time point. Cells were seeded at 1E6 cells/ml in MEM-Alpha (Biological Industries) supplemented with 10% Heat Inactivated FCS (Sigma) and 50IU rhIL-2 (Peprotech).

Nucleic acid manipulations

For T7E1 assays, PCR amplification was performed on Gentra PureGene Tissue Kit (Qiagen) extracted genomic DNA. 200-500ng of genomic DNA was amplified using PrimeStar MAX (Takara) for 35 cycles. Primers for these reactions can be found in Supplementary Table 4. Resulting amplicons were denatured and reannealed in a thermocycler before nuclease reaction using T7 Endonuclease 1 (New England Biolabs) at 37C for 20min. Resulting fragments were analyzed by agarose gel electrophoresis and quantified using Biovision (Vilber Lourmat) using a rolling ball for background subtraction. Efficiency was calculated using the following equation:

$$\text{Cleavage efficiency} = 100 \times (1 - (1 - \text{fraction cleared})^{\frac{1}{2}})$$

For TIDE analysis, PCR amplicons were subjected to purification by AmpureXP beads (Beckman Coulter) at a 1:1 ratio. Sanger sequencing was performed at the DNA Sequencing Unit, Tel Aviv University. Samples were compared using TIDE (<https://tide.nki.nl/>)⁴⁷.

Fluorescence in-situ hybridization

Fluorescence in situ hybridization (FISH) analysis was performed following the manufacturer's instructions (Cytocell) on interphase human T cells from peripheral blood spreads' using the TRACD breakapart probe. Images were captured using GenASIs imaging system.

Statistics

For FISH, genomic DNA cleavage efficiency and flow cytometry knock out efficiency, statistical analyses were performed on Prism (GraphPad). For t-tests on dislinkage by ddPCR, for each donor, technical replicates were averaged and t-test was performed on the averaged values. Each figure legend denotes the statistic used, central tendency and error bars.

ACKNOWLEDGEMENTS:

We thank Natalie Zelikson (TAU), for helpful discussions and reviewing the manuscript. We thank the IDRNU, GRU and SICF units, Tel Aviv University for logistic support and council. This research was funded by the European Research Council (A.B. and U.B.-D.), the Gertner Institute Scholarship, the Yoran Institute Scholarship, the SAIA Foundation (A.D.N.), SCGC Tel Aviv University (A.B. and U.B.-D.), the Israel Cancer Research Fund Gesher Award (U.B.-D.), the Azrieli Faculty Fellowship (U.B.-D), the Alon Fellowship for Outstanding Young Scientists, Israel Council for Higher Education (A.M).

AUTHOR CONTRIBUTIONS:

A.D.N. designed, performed and analyzed the experiments; E.R. and E.G. analyzed the single-cell RNA sequencing data; T.T. and M.L. performed fluorescence in-situ hybridization; M.H-F. helped with sample culturing and processing; R.K. and H.K. performed single-cell RNA sequencing; E.R. supervised fluorescence in-situ hybridization; A.M. and U.B.-D. supervised single-cell RNA sequencing analysis; A.D.N. and A.B. drafted the manuscript, and revised it together with A.M. and U.B.-D; U.B.-D. and A.B. supervised the study; A.B. conceptualized the study.

COMPETING INTERESTS:

None

CODE AVAILABILITY

The software used for individual and integrated analyses are described and referenced in the individual sections in Methods.

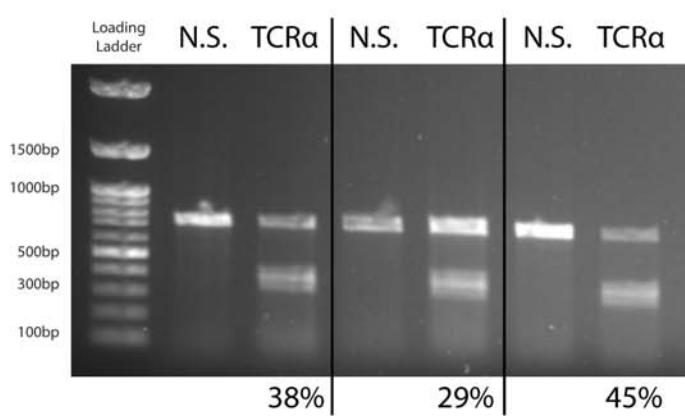
DATA AVAILABILITY

All datasets are available within the article and its Supplementary Information. 10X scRNAseq data have been deposited to SRA with BioProject accession number PRJNA745549.

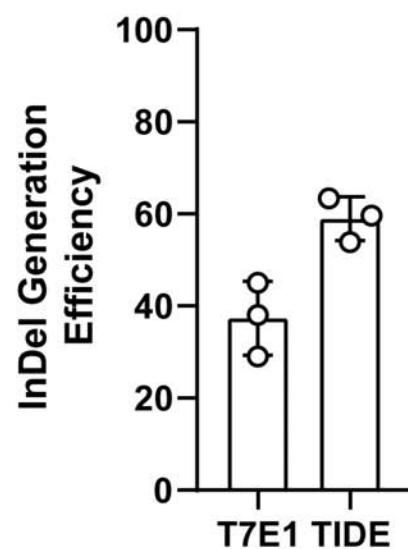
Supplementary Information and Extended Data Figures are available for this paper.

EXTENDED DATA FIGURES:

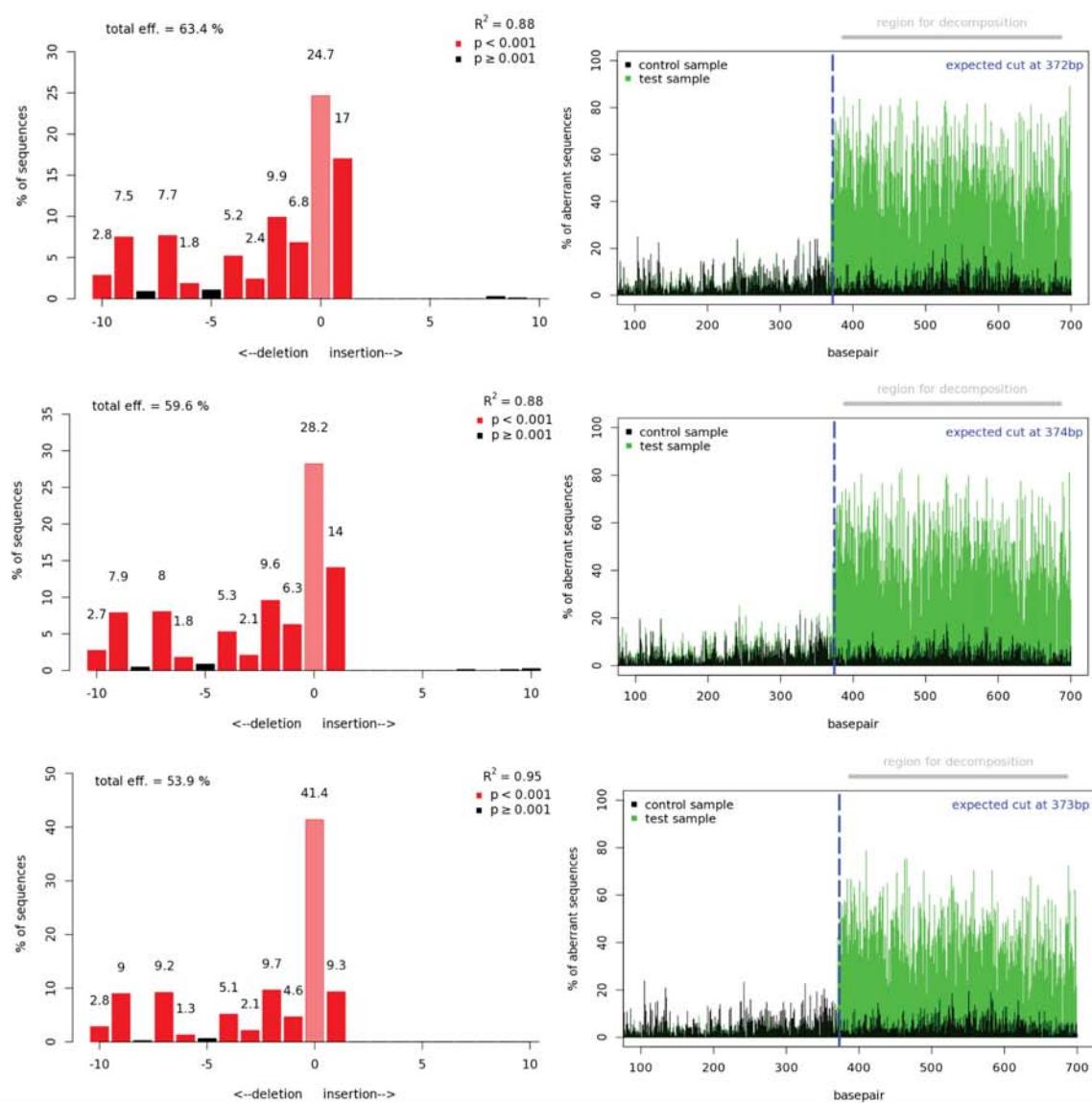
A



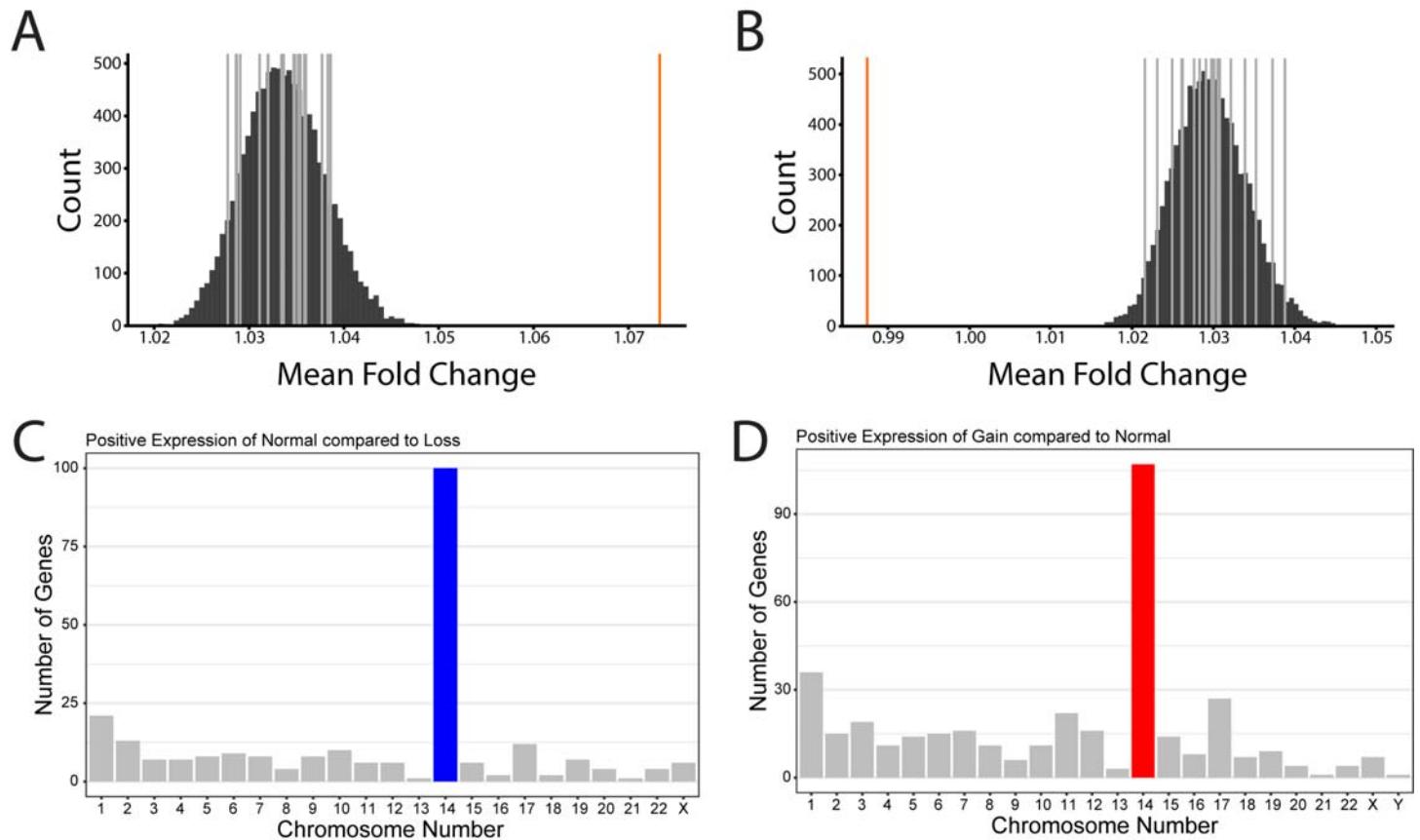
C



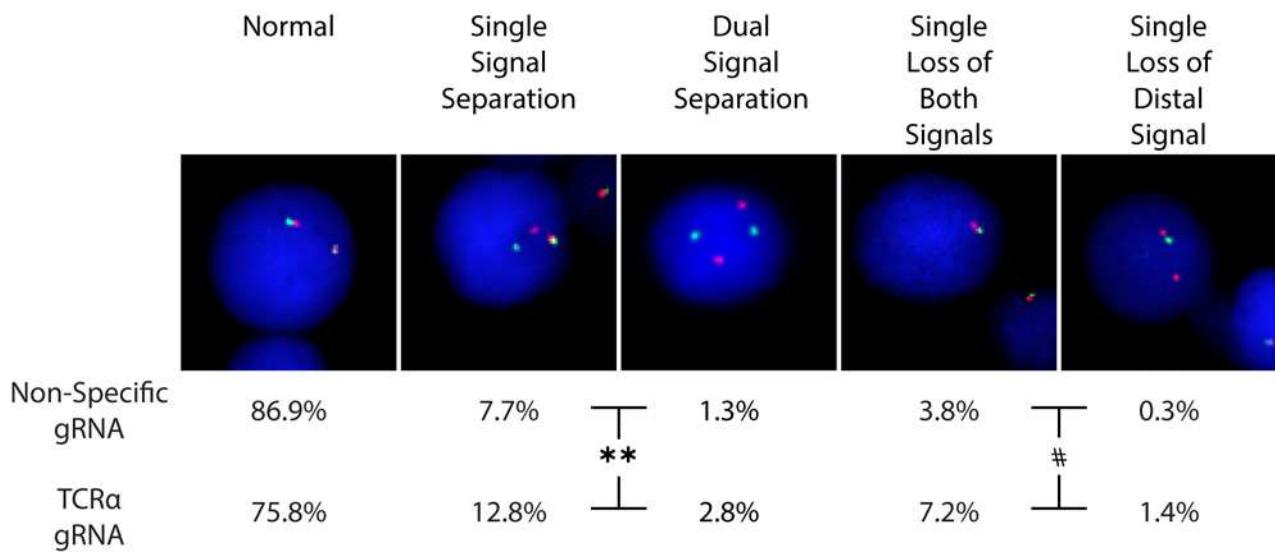
B



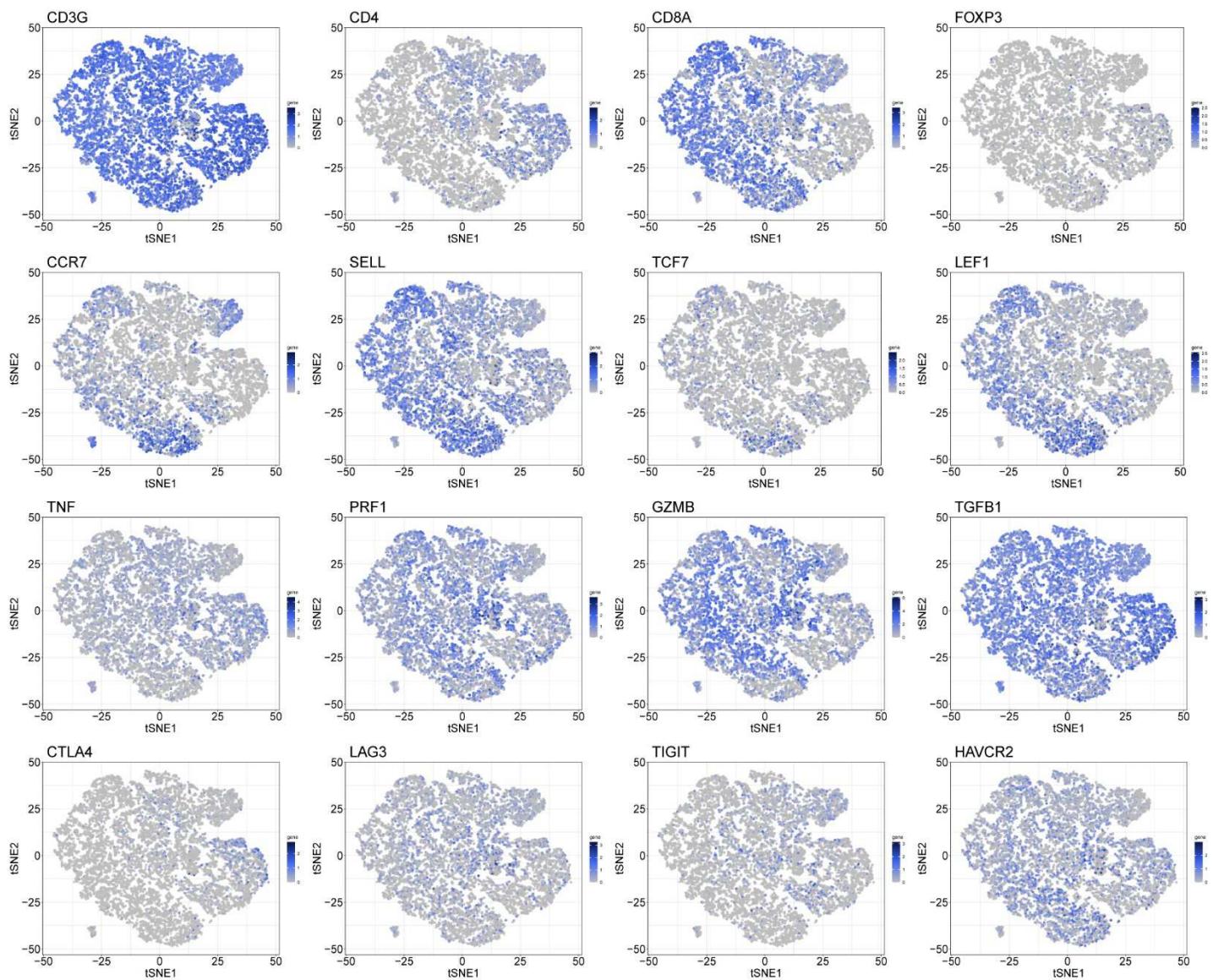
Extended Data Figure 1: Quantification of InDels produced by CRISPR-Cas9 activity at the TCR α locus indicates efficient cleavage. **A.** T7 Endonuclease 1 (T7E1) assay for three independent experiments. For each experiment, one lane for Non-Specific gRNA (N.S.) and one lane for TCR α -targeting gRNA treated cells are presented. Experiments are separated by a black bar. Percentages, presented below the lanes, refer to cleavage efficiency as inferred by densitometric analysis. Loading Ladder and relative sizes are indicated on the left. Unprocessed scan can be found in Supplementary Data 1. **B.** TIDE Analysis for the same experiments as in A. In the left panel, the height of each bar corresponds to the rate of sequences having the given number of nucleotides added or deleted. The right panel depicts the rate of sequence misalignments at each position of the PCR fragment amplified from the TCR α locus of cells treated with either the TCR α -targeting gRNA (green) or a Non-Specific gRNA (black) **C.** Quantification of A and B.



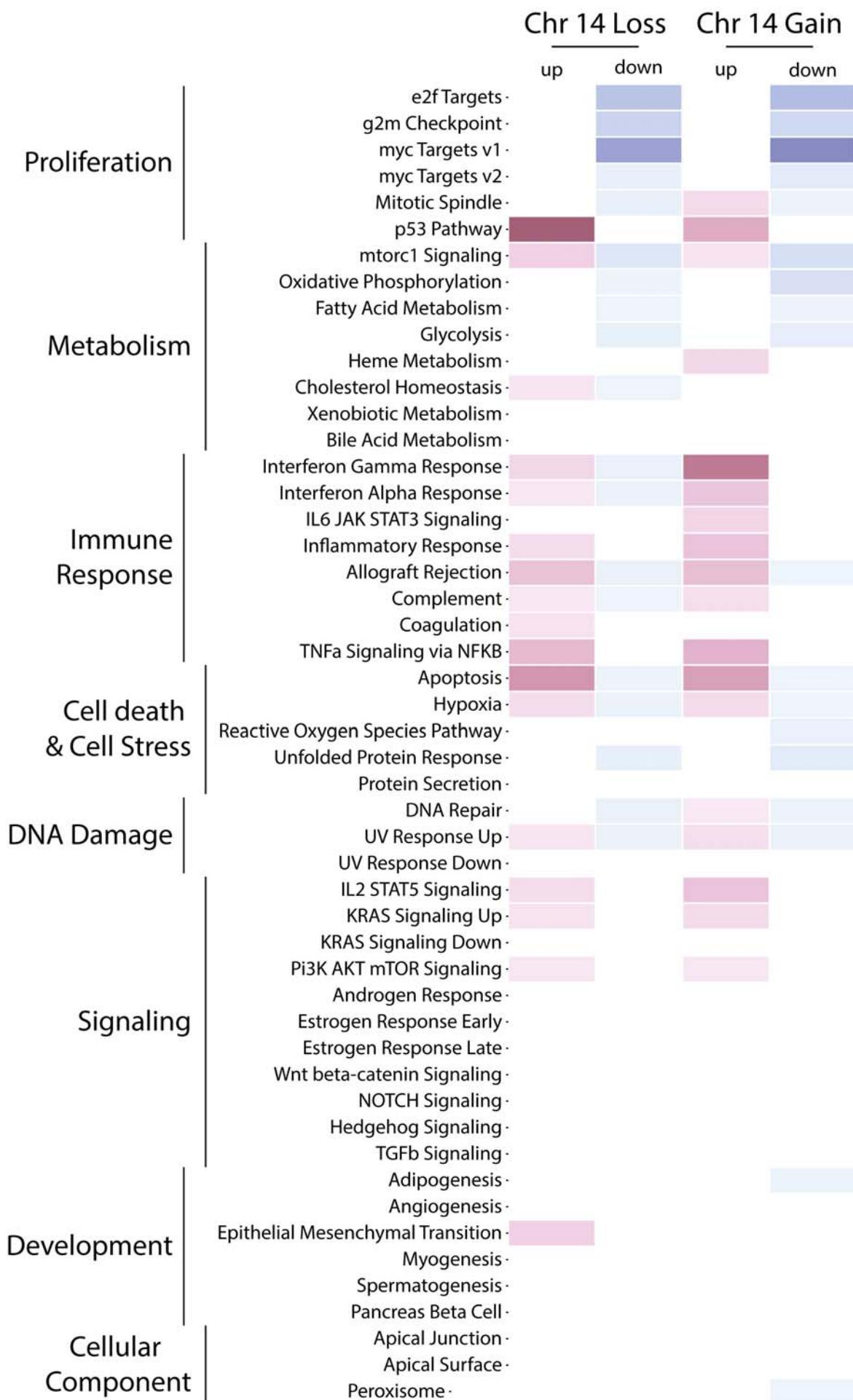
Extended Data Figure 2: A-B. Enrichment in the number of genes with no detected expression among cells identified as having a chromosome 14 loss (A) or gain (B) (Fig. 1F). The x-axis represents the fold-change in the number of genes with no detected expression between cells with or without a chromosome 14 loss, based on the InferCNV analysis (Fig. 1F). The dark gray lines represent the empirical values obtained for each chromosome, except for chromosome 14. The orange line is the empirical value for chromosome 14. The black bars are the results of 10,000 permutations. **C.** Number of differentially expressed genes in each chromosome, as compared between the chromosome 14 loss populations to chromosome 14 normal (see Fig. 1F). **D.** Number of differentially expressed genes in each chromosome, as compared between the chromosome 14 gain populations to chromosome 14 normal (see Fig. 1F).



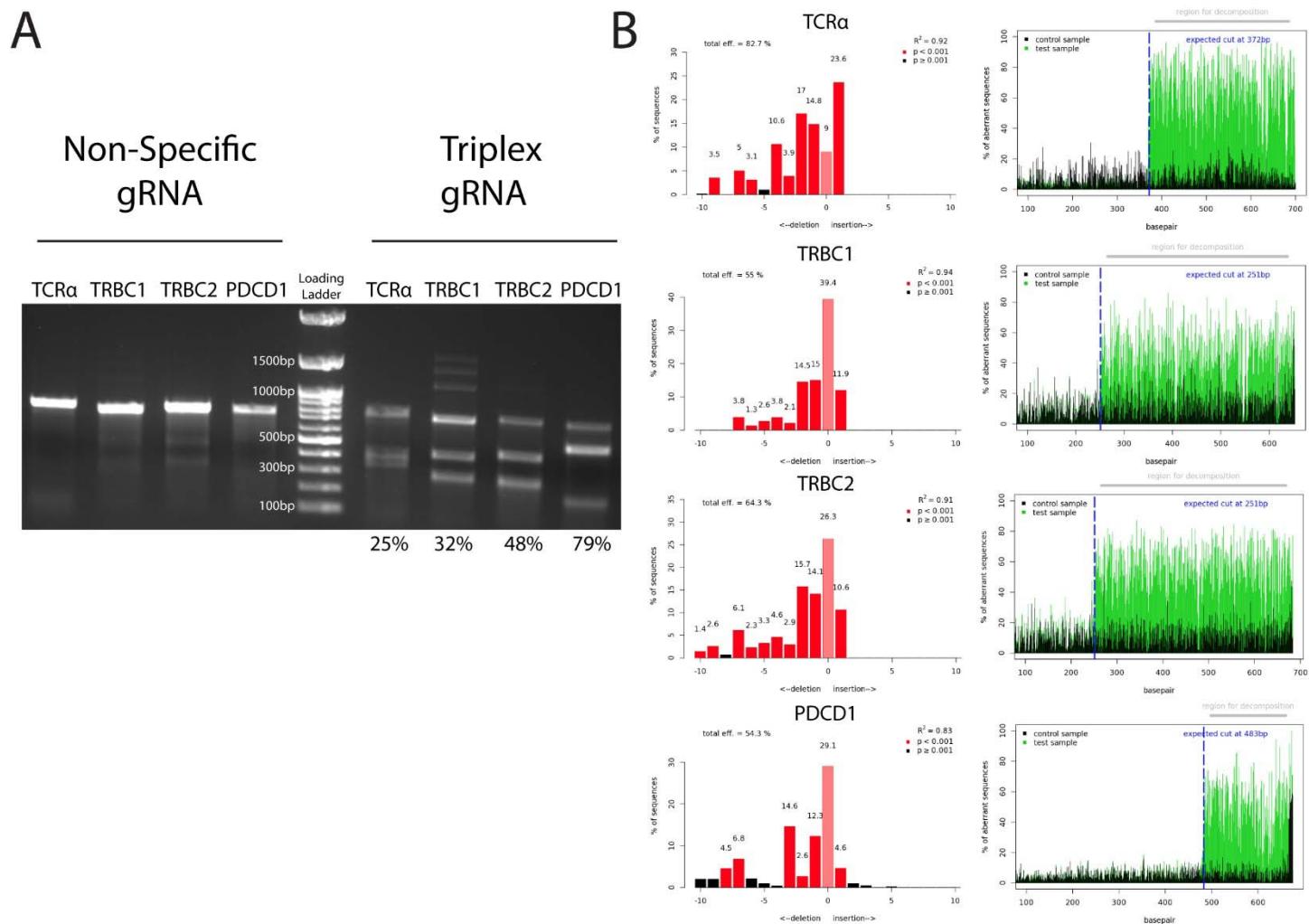
Extended Data Figure 3: FISH analysis of cells treated with either a non-specific or a TCR α -targeting gRNA. In the pictures are examples of all signal patterns found among analyzed cells. The frequency of each signal pattern, among cells electroporated with the TCR α -targeting gRNA or a non-specific gRNA, is presented below the pictures. n = 313 and n = 360 for non-specific and TCR α targeted cells, respectively. #, p= 0.0115 Fisher's exact test for loss of either both signals or the distal signal. ** p =0.0048 for either single or dual Signal Separation.



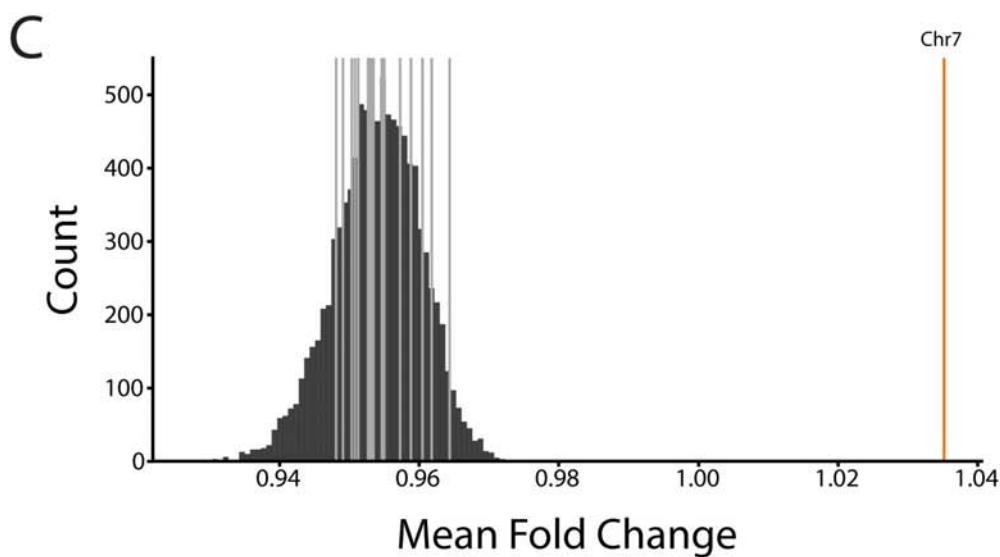
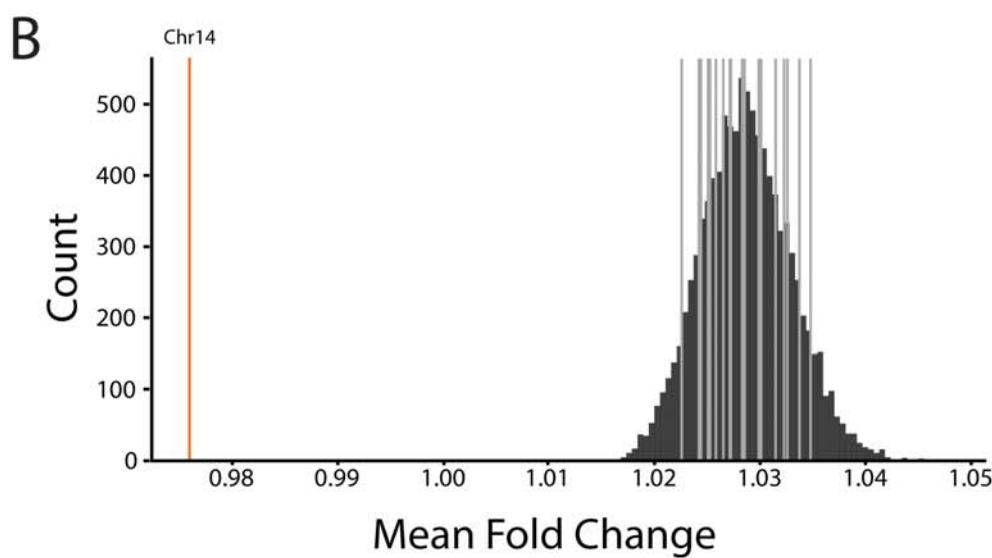
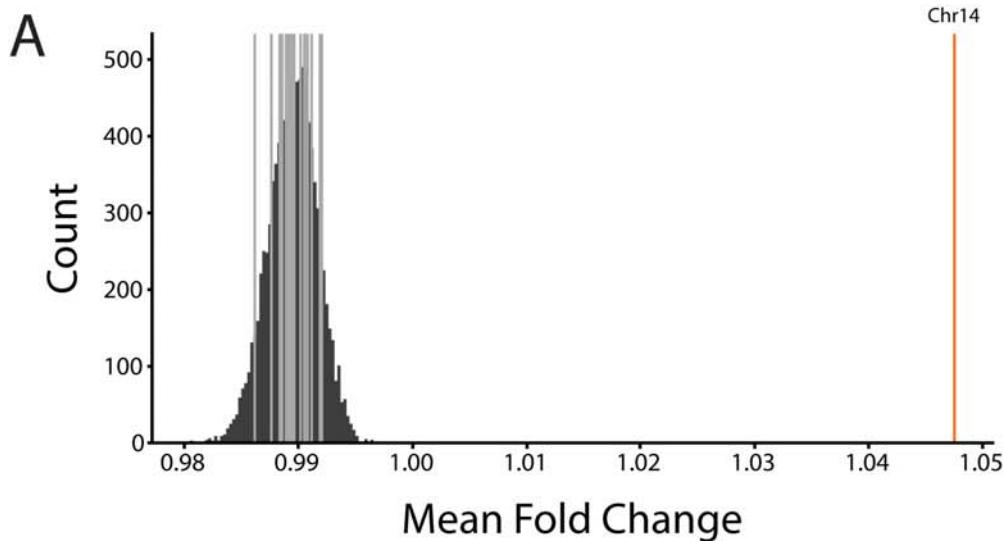
Extended Data Figure 4: Selected gene expression patterns across the cells. Each t-SNE plots represent all the cells in the experiment (Fig. 3A-C). Each plot represents the expression pattern of a different gene, indicated on the top left, among the clusters. A darker shade of blue corresponds to higher RNA expression.



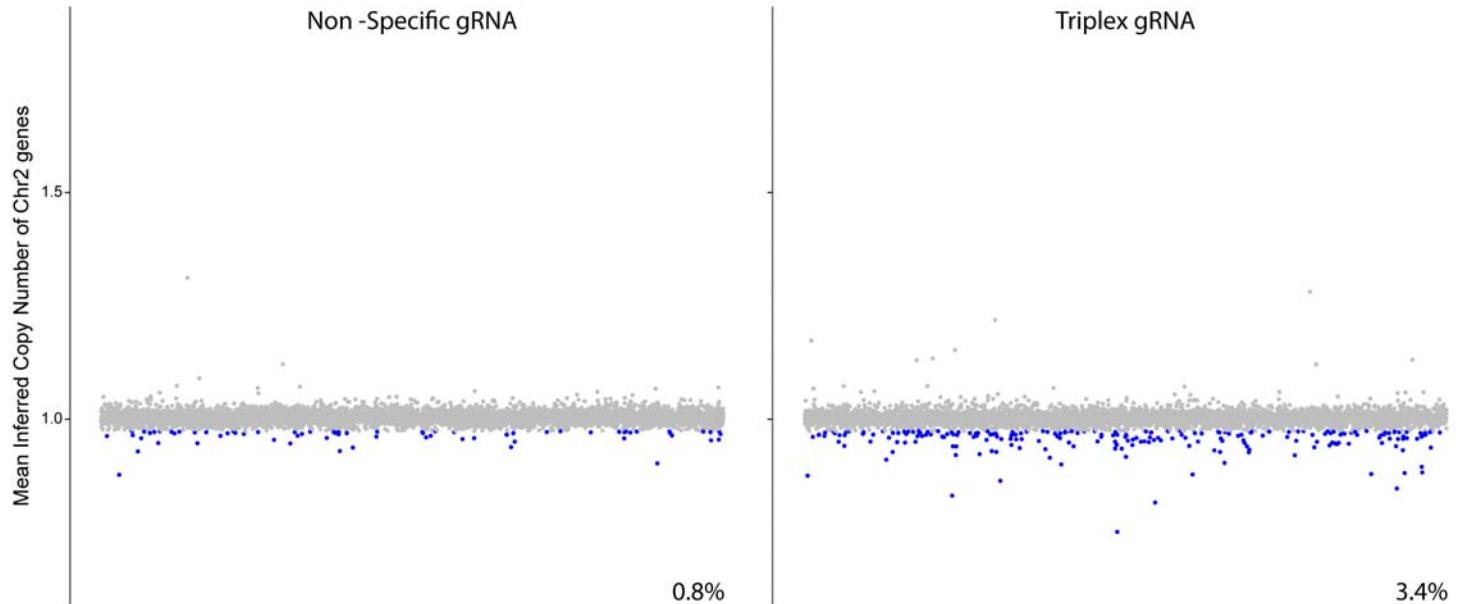
Extended Data Figure 5: A gene set enrichment analysis between cells that have lost a copy of chromosome 14 (Fig. 1F), to cells without these chromosomal aberrations. The 50 'Hallmark' MSigDB gene sets are shown. Gene sets enriched in up-regulated genes are depicted in red, and those enriched in down-regulated genes are depicted in blue. Values are scaled to $-\log(FDR)$ of the enrichments. The full enrichment scores are shown in Supplementary Table 2.



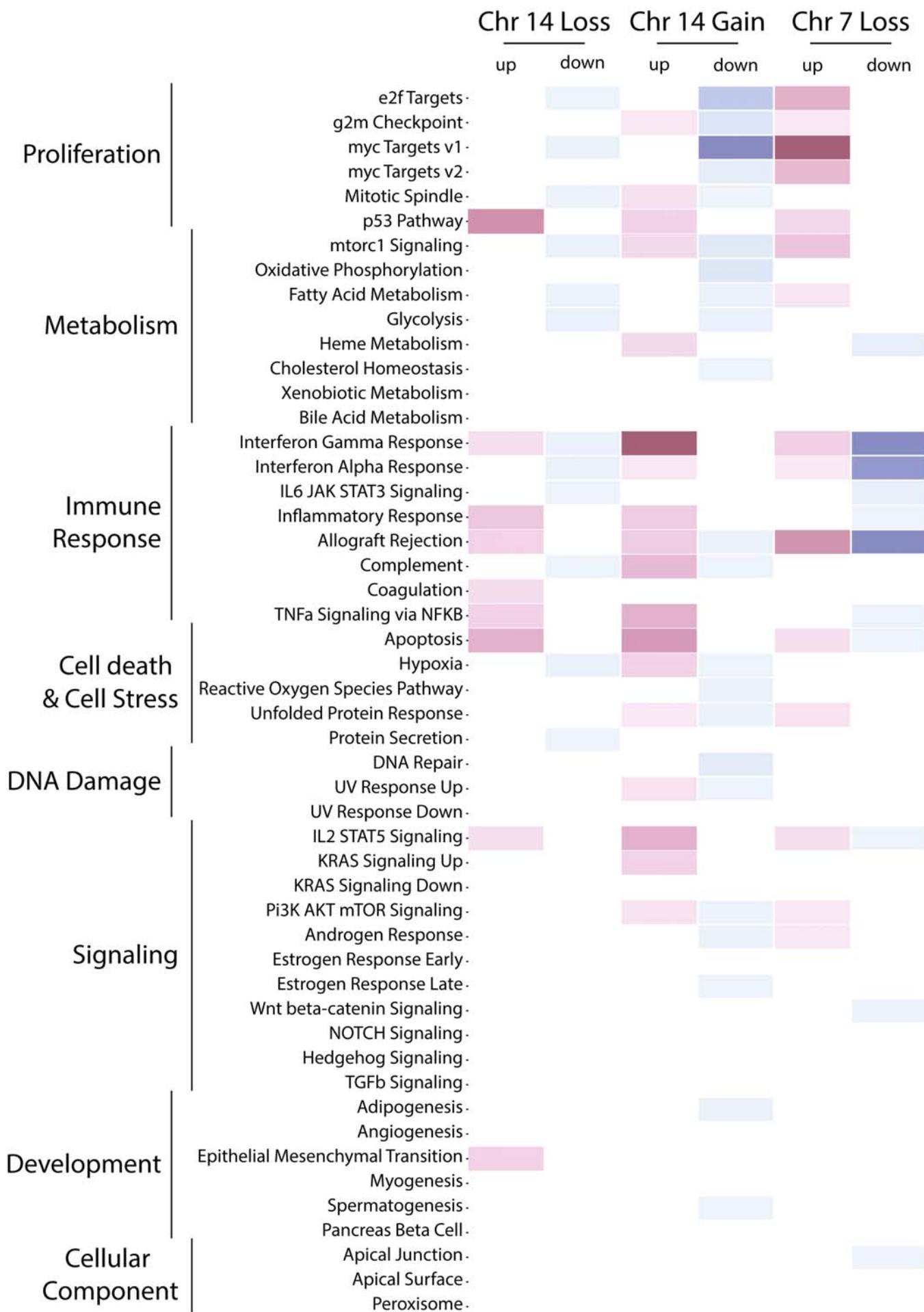
Extended Data Figure 6: Quantification of InDels produced by CRISPR-Cas9 activity at the TCR α , TCR β and PDCD1 loci indicates efficient cleavage. **A.** T7E1 assay for the 4 genomic target loci. Non-Specific gRNA treated cells (4 left lanes) are compared to cells treated with the combination of the TCR α , TCR β and PDCD1- targeting gRNAs (4 right lanes). In each lane, a different locus is analyzed. Calculated efficiency is indicated below. Loading Ladder and relative sizes are indicated in the middle. Unprocessed scan can be found in Supplementary Data 1. **B.** TIDE Analysis for the same experiments as in A. The rows represent (in this order, from top to bottom) the TCR α , TRBC1, TRBC2 and PDCD1 locus. In the left panels, the height of each bar corresponds to the rate of sequences having the given number of nucleotides added or deleted. The right panels depict the rates of sequence misalignments at each position of the PCR fragment amplified from the target locus of cells treated with either Non-Specific gRNA (black) or the combination of the TCR α , TCR β and PDCD1- targeting gRNAs (green).



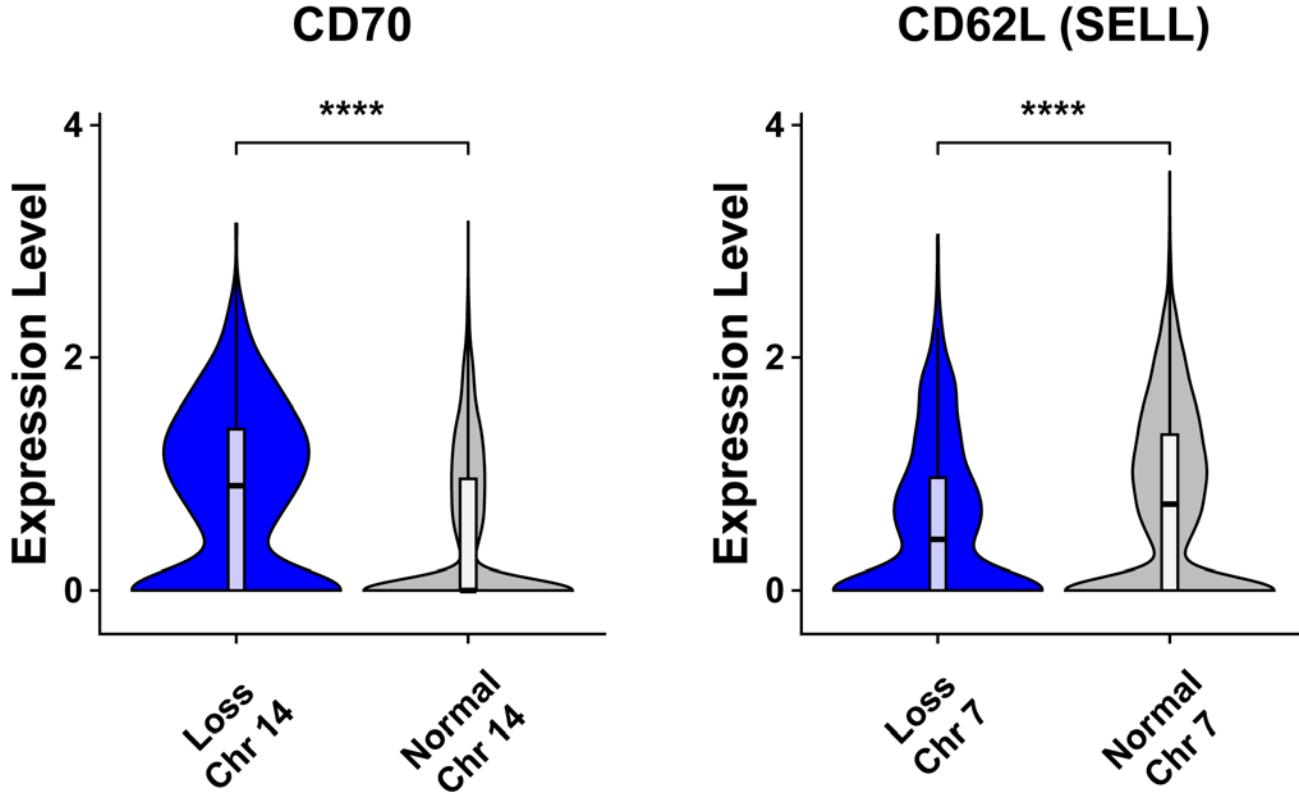
Extended Data Figure 7: A-C. Enrichment in the number of genes with no detected expression among cells identified as having a chromosome 14 loss (A), a chromosome 14 gain (B) or chromosome 7 loss (C) (Fig. 4D). The x-axis represents the fold-change in the number of genes with no detected expression between cells with or without a chromosome 14 loss or gain or 7 loss, based on the InferCNV analysis (Fig. 4D). The dark gray lines represent the empirical values obtained for each chromosome, except for chromosome 14 or 7. The orange line is the empirical value for chromosome 14 or chromosome 7. The black bars are the results of 10,000 permutations.



Extended Data Figure 8: Differential gene expression analysis for chromosome 2. Each dot represents the mean inferred copy number of genes coded on chromosome 2 in each cell treated with a non-specific gRNA (left) or the combination of the TCR α , TCR β and PDCD1- targeting gRNAs (right). Cells are marked with red and blue dots, corresponding to a chromosome 2 gain or loss respectively, if their mean inferred gene copy number is >2 standard deviations from the population's mean. n=8619 and 6326 for cells treated with the non-specific gRNA (left) or the combination of the TCR α , TCR β and PDCD1- targeting gRNAs (right), respectively.



Extended Data Figure 9: A gene set enrichment analysis between cells that have lost a copy of chromosome 14, or a segment of chromosome 7, to cells without these chromosomal aberrations. The 50 'Hallmark' MSigDB gene sets are shown. Gene sets enriched in up-regulated genes are depicted in red, and those enriched in down-regulated genes are depicted in blue. Values are scaled to $-\log(\text{FDR})$ of the enrichments. The full enrichment scores are shown in Supplementary Table 3.



Extended Data Figure 10: T cell surface markers CD70 and CD62 can potentially be used to sort out aberrant cells. The violin plots show gene expression of CD70 (left) and CD60L (SELL, right), for cells treated with the combination of the TCR α , TCR β and PDCD1- targeting gRNAs and have normal gene expression (grey) or a gene expression pattern indicating a chromosome 14 loss or a chromosome 7 truncation (blue). **** $p < 0.0001$, unpaired two-tailed Wilcoxon test.