Supplementary Note 1: Heterozygous/Homozygous integration prediction model.

An estimation of the percentage of cells with bi-allelic insertions at a single autosomal genomic locus (two potential alleles) can be made from only fluorescent phenotypes if two HDR templates integrating different fluorescent proteins into that same site are introduced into the cell (by electroporation). A simple probability model requires only two assumptions.

Assumption 1: There are no off-target integrations at other sites besides the target locus nor concatemers or multiple integration events at the target locus that contribute to fluorescent phenotypes.

Assumption 2: Integration of a specific second fluorescent protein (i.e. RFP) does not depend on which fluorescent protein was integrated at the cell's other allele (i.e. GFP or RFP integrations at the first allele are equally likely to have an RFP integration at the second).

Following the labelling in **Extended Data Fig. 3**, the percentages of four different phenotypic populations are known:

- % GFP^-RFP^-
- % GFP^+RFP^-
- % GFP^-RFP^+
- % GFP^+RFP^+

From these, immediately two genotypes are known:

1) Genotype A = NA/NA = % GFP^-RFP^-
2) Genotype E = GFP/RFP = % GFP^+RFP^+

The four remaining genotypes sum to the two remaining single fluor positive phenotypes:

3) Genotype B + Genotype D = GFP/NA + GFP/GFP = % GFP^+RFP^-
4) Genotype C + Genotype F = RFP/NA + RFP/RFP = % GFP^-RFP^+

The probabilities that a RFP^+ cell will also be GFP^+, and vice versa, are also known from the phenotypes:

5) Probability of being GFP^+ given being RFP^+ = P(GFP|RFP) = (% GFP^+RFP^+) / (% RFP^+ + %GFP^+RFP^+)
6) Probability of being RFP^+ given being GFP^+ = P(RFP|GFP) = (% GFP^+RFP^+) / (% GFP^+ + %GFP^+RFP^+)

Following from assumption 2, if the probability that a cell receives a GFP integration at its second allele is independent of whether the first integration was a GFP or RFP, then a relationship between the single positive genotypes can be determined (fig S13):

7) D = P(GFP|RFP) * B
8) F = P(RFP|GFP) * C

Inserting the equations 7 and 8 into equations 3 and 4 respectively and simplifying solves for the remaining genotypes in terms of the known phenotypes:

9) B = % GFP^+RFP^- / (1 + (% GFP^+RFP^+) / (% RFP^+ + %GFP^+RFP^+))
10) C = % GFP^-RFP^+ / (1 + (% GFP^+RFP^+) / (% GFP^+ + %GFP^+RFP^+))

11) D = % GFP^+RFP^-- B
12) F = % GFP^-RFP^+ - C

From the known genotypes, the observed % of cells that are have mono-allelic or bi-allelic insertions, as well as other statistics, can be calculated readily:

- Observed % Cells Heterozygous = B + C
- Observed % Cells Homozygous = D + E + F
- Observed % Cells with at least 1 insertion = B + C + D + E + F = 1 - A = 1 - % GFP-RFP-

- Observed % Alleles that have a GFP = (B + E + 2D) / 2
- Observed % Alleles that have a RNP = (C + E + 2F) / 2
- Observed % Alleles with an insertion = % $Alleles_{GFP}$ + % $Alleles_{RFP}$

An expected % of cells homozygous if the HDR alleles were distributed randomly (in essence at Hardy-Weinberg Equilibrium) can be calculated from the observed % of cells with at least one insertion (HDR):

- p = HDR allele (GFP or RFP)

- q = non-HDR allele (NA)
- X = % of cells observed to have at least one HDR

13) $p + q = 1$

14) $p^2 + 2*p*q + q^2 = 1$

As any cell that has an HDR (GFP or RFP) allele will show the phenotype (in this case GFP+ or RFP+):

15) $X = p^2 + 2*p*q$

Substituting X into equation 14 and simplifying:

16) $q = (1 - X)^{½}$

17) $p = 1 - q$

18) $p = 1 - (1 - X)^{½}$

p^2 will give then give the expected % of cells homozygous for HDR integration if HDR template insertion was random among the target alleles:

19) $p^2 = 2 - 2(1 - X)^{½} - X$

As X is known, the expected % of homozygous cells can be calculated directly from the observed total % of cells with at least one HDR, and can then be compared the observed % of homozygous cells calculated by taking into account the information provided by integration of two separate fluorophores.

Supplementary Note 2: Analysis of off target effects.

A common concern for use of genetically modified T cells therapeutically is the potential for off-target effects. With any targeted editing strategy, there are at least three potential types of undesirable outcomes. First, the nuclease (such as a Cas9 RNP) can generate non-specific mutations via-nonhomologous repair mechanisms at the cleavage site or at off-target sites. Second, the DNA HDR template (either viral or non-viral) could integrate off-target, for example either at off-target Cas9 RNP cut sites or at naturally occurring double-stranded breaks in host DNA[18] (**Extended Data Fig. 4c, d**). Even integrase-deficient AAVs templates have been shown to stably integrate at off-target sites[4,17]. Third, the DNA HDR template could integrate at the desired target site but integrate incorrectly, for example through homology-independent mechanisms[19]. Off-target cutting and integrations should be minimized in cells destined for therapeutic use to ensure that integrated sequences remain under proper endogenous regulation and that critical off-target sites are not disrupted.

We looked for unintended non-homologous integrations with the non-viral system using an N-terminal GFP-RAB11A fusion construct that contained the endogenous *RAB11A* promoter sequence within its 5' homology arm. This construct can express GFP at off-target integration sites, which allowed us to assay for off-target events using flow cytometry (**Extended Data Fig. 4c**). Inclusion of a gRNA designed to cut a genome region that is not the homologous region to the targeting sequence can be used to approximate integration at an off-target cut site. While efficient GFP expression depended on pairing the HDR template with the correct gRNA targeting that site, rare GFP+ cells were observed when dsDNA HDR templates were delivered either alone or with an "off target" Cas9 RNPs (**Extended Data Fig. 4e-i**),

We tested whether nonviral genome targeting is compatible with a D10A Cas9 nickase variant engineered to reduce the potential for off-target double strand breaks[22,23]. This variant requires that two gRNAs bind and cleave in close proximity to each other to produce a double strand break, thus reducing the number of off-target double stranded DNA breaks because both gRNAs would need to have an off-target cut site in close proximity to each other. We tested a series of gRNA combinations at the *RAB11A* locus for the integration of GFP, and found that guides in a "PAM-Out" orientation led to the efficient introduction of GFP when the D10A nickase was used (**Extended Data Fig. 5e, f**). The D10A nickase also potentially reduced integrations at off-target nick sites, which we modelled using a gRNA that does not cut at a site

homologous to the template. With this individual "off-target" Cas9 nickase RNP, GFP integrations occurred only at background levels (**Extended Data Fig. 5g**).

While the D10A nickase may be useful to reduce off-target integration, a small, but significant number of GFP+ cells were generated even without targeting Cas9 with a gRNA to the site of homology. GFP+ cells were found at a similar rate when the donor along was used without any Cas9 nuclease, perhaps resulting from integration at naturally occurring double stranded DNA breaks (**Extended Data Fig. 4h**). We reasoned that remaining off-target integrations could be reduced by replacing the dsDNA HDR templates with similar ssDNA HDR templates, which should not integrate non-specifically at double strand breaks[20,21]. To test this hypothesis, we generated ssDNA HDR templates using two methods we recently developed to produce the large amounts of long ssDNA required for electroporation[21] (**Extended Data Fig. 5a, b**). ssDNA HDR templates reduced the number of functional off-target integrations approximately 100-fold, while maintaining efficient on-target integration (**Extended Data Fig. 5c, d**).

We next used a deep sequencing approach to assess off-target integrations and incorrect/non-homologous integration events that occurred with dsDNA and ssDNA HDR templates. We performed targeted locus amplification (TLA) sequencing in two donors with both dsDNA and ssDNA HDR templates (**Extended Data Fig. 4a**). Targeted locus amplifications were used due to the large amount of HDR template DNA that was retained in a non-integrated state in the cells. With both dsDNA and ssDNA HDR templates, no off-target integration sites were found above the limit of detection (approximately 1% of alleles) (**Extended Data Fig. 4a**). However, sequencing of the on-target locus revealed that some incorrect on-target integration events were detected with a dsDNA HDR template, potentially including concatemerization. One incorrect homology directed repair sequence was detected at a frequency of approximately 1%, resulting from 9 bps of overlap between a portion of GFP's sequence and a region within the 3' homology arm, which of course is also present at the target genomic locus. The ssDNA HDR template showed 10-fold fewer of this incorrect integration (**Extended Data Fig. 4b**).

Supplementary Note 3: Clinical history, genetic testing, and clinical phenotyping of family with compound heterozygous *IL2RA* mutations.

Clinical History of Family with Autoimmunity/Immune Dysregulation

The proband is a Caucasian infant who presented at 15 weeks of age after vomiting, fussiness and tachypnea led to medical evaluation that revealed severe diabetic ketoacidosis and a serum glucose level of 920 mg/dL. One week after diagnosis, testing for GAD65, IA-2 and insulin autoantibodies was negative; however, autoimmune diabetes was confirmed when repeat antibody tests at 5-7 months of age in three different laboratories showed positive results for IA-2 and insulin autoantibodies, as well as very high levels of GAD65 antibodies in two of the laboratories [42.8 nmol/L (<0.02) at Mayo Laboratories and 896 IU/mL (0.0-5.0) at Barbara Davis Center]. Testing for thyroid dysfunction and celiac disease has been negative but mildly low IgA levels suggest partial IgA deficiency. C-peptide testing was repeatedly completely undetectable, including at 7 months of age when measured 90 minutes after a feed with a serum glucose level of 202 mg/dL, at which time proinsulin was also undetectable. After the initial DKA was treated with intravenous insulin, the patient was discharged on multiple daily injections of subcutaneous insulin (glargine and lispro) initially and later transitioned to an insulin pump with continuous glucose monitoring. He consistently required a high replacement dose of insulin in the range of 0.8-0.9 units/kg/day (48% basal at 7 months of age). He had been delivered by repeat c-section at 37 weeks gestation with a birth weight of 3.629 kg (75th percentile) without any complications and there have been no concerns about his developmental progress and his medical history has otherwise been unremarkable. His parents have disparate Caucasian ancestry and denied consanguinity.

Clinical information on family members is provided in **Supplementary Table 4**. More detailed information is as follows:

1. Mother (37):
    a. Pneumonia as a child – explained as viral
    b. Ear infections as a child treated with antibiotics
    c. Tooth problems (perhaps related to antibiotics)
    d. Her father developed insulin dependent diabetes in his 30's. He had a low WBC and also had nummular dermatitis of the scalp.

e. Her mother had lupus

2. Father (44)
   a. Moroccan descent
   b. No major medical problems
   c. Some possible concern this his response time to common viral infections may be prolonged.

3. Affected child (14)
   a. Immune thrombocytopenic purpura: (+ anti-platelet antibodies)
   b. Neutropenia (anti-neutrophil Ab)
   c. Autoimmune haemolytic anaemia (DAT+ i.e. direct Coombs+)
   d. Nummular dermatitis of the scalp
   e. Hypercellular bone marrow: inverted CD4+/CD8+ ratio (0.36).
   f. Mouth ulcers
   g. Ear infections treated with tubes
   h. Diarrhoea as a child
   i. 46XX – no known chromosomal abnormality
   j. Flow cytometry of peripheral blood: 82.7% of CD45+ cells are CD3+ and 5.9% are CD19+. CD19+CD5+ cells are the deficient B cells. 43.6% of CD45+ cells are CD8+ with an inverted CD4+/CD8+ ratio (0.6). There is a relative increase in TCR(alpha beta)+ CD3+ CD4- CD8- T lymphocytes (26% of TCR alpha beta+ CD3+ cells and 5% of CD45+ leukocytes).
   k. Has been treated with immunosuppression including prednisone (20 mg), IgG-pro-IgA, Flonase nasal spray and topical steroids and Symbicort. Also treated with Neupogen.

4. Affected child
   a. 3+ diabetes autoantibodies (anti-GAD, MIAA, ICA, negative ZnT8 and ICA512/IA-2) normal OGTT
   b. Ear infections treated with tubes at 1 year
   c. Eczema in the winter

5. Unaffected daughter (15)

a. Allergies, but otherwise healthy

6. Affected son (4)
   a. Eczema in winter
   b. Positive test for HSV
   c. Insulin dependent diabetes within the first year of life, C-peptide < 0.1 at presentation, anti-GAD ab+ (>30 (nl<1U/ml) 1 yr after dx but negative at dx, ICA512 Ab+ (1.3 (nl<1.0)) 1 yr after dx but negative at dx

7. Unaffected daughter (9)
   a. Asthma

Genetic Testing to Identify *IL2RA* Mutations

Initial genetic testing of the proband using an in-house targeted next-generation sequencing multi-gene panel of over 40 genes known to be involved in monogenic forms of diabetes was negative. Subsequent exome sequencing in the trio of proband and parents revealed the causative compound heterozygous mutations in the *IL2RA* gene. Two siblings carry only one mutation, but the other two with both mutations have evidence for autoimmunity: an older male sibling was found (at 4 or 5 years of age) to have positive diabetes autoantibodies in the absence of hyperglycemia and an older female sibling was diagnosed with autoimmune mediated pancytopenia at age 11 years. IL2RA expression was markedly reduced in the three compound heterozygous children.

Clinical Phenotyping of IL2RA-Deficient Patients

The IL2RA-deficient children have an almost complete loss of IL2RA cell surface expression on T cells and therefore virtually no detectable $CD3^+CD4^+IL2RA^{hi}CD127^{lo}$ Tregs in their blood, whereas family relatives carrying heterozygous *IL2RA* mutation display decreased IL2RA expression on their Tregs (**Extended Data Fig. 6a, b**). Frequencies of $CD3^+CD4^+CD127^{lo}FOXP3^+$ T cells in IL2RA-deficient subjects resemble those in HD and IL2RA+/- individuals, suggesting that Tregs may develop in the absence of IL2RA function (**Extended Data Fig. 6c**). Using a strategy to isolate Tregs without IL2RA expression, we found that $CD3^+CD4^+CD127^{lo}CD45RO^+TIGIT^+$ Treg-enriched cells from IL2RA-deficient subjects showed a defective ability to suppress the proliferation of responder T cells (Tresps) as compared to HD counterparts **(Extended**

Data Fig. 6e, f, h). In contrast, Tregs from relatives with a single heterozygous *IL2RA* mutation could inhibit Tresp proliferation, although with suboptimum capacity (**Extended Data Fig. 6g, h**). Hence, correcting functional IL2RA expression on the surface of $FOXP3^+$ T cells from these patients may represent a valuable approach for developing an *ex vivo* gene therapy.

Supplementary Note 4: Detailed description of endogenous TCR replacement constructs.

The genomic locus of the T cell receptor is complicated, with a large array of variable alleles (V and J alleles for the TCR-α chain and V, D and J alleles for the TCR-β chain) that undergo somatic gene rearrangement during T cell development in order to produce a functional T cell receptor. Important for the diversity of the TCR repertoire but challenging for targeted genomic editing at the TCR locus (whether knock-outs or knock-ins), these recombined sequences are variable across the polyclonal population of T cells. However, for both the TCR-α and TCR-β chains there is a constant domain at the C-terminus of the protein that is shared by all T cells, no matter what V-J or V-D-J segments have been rearranged (note: the *TCR-β* locus has two constant regions that can be utilized as shown in **e**). These constant sequences, termed T Receptor Alpha Constant (*TRAC*) Exons 1, Exon 2, etc. allows for one gRNA sequence to modify polyclonal T cells no matter what rearranged TCR they express (**Extended Data Fig. 9a**). A 2.1 kb HDR template was used to replace the endogenous TCR. Approximately 300 bp homology arms surround a ~1.5 kb inserted sequence encoding a self-cleaving peptide followed by the full-length sequence of the TCR-β chain of the desired antigen specific T cell (here the 1G4 NY-ESO-1 specific TCR). A second self-cleaving peptide follows the TCR-β chain, and separates it from the variable (recombined V and J) sequence of the desired antigen specific TCR-α chain. Only the variable sequence from the *TCR-α* chain and the sequence of the *TRAC* exon 1 prior to the gRNA cut site needs to be inserted as HDR will introduce it in frame with the remaining *TRAC* exons (**Extended Data Fig. 9b**).

Transcription in cells with successful HDR should yield a “polycistronic” mRNA encoding both TCR-β and TCR-α (**Extended Data Fig. 9c**). This targeting strategy is designed to yield three peptide chains: a remnant endogenous variable region peptide that does not possess a transmembrane pass (and thus should not be expressed on the cell surface), the full-length desired antigen specific TCR-β chain, and the full-length desired antigen specific TCR-α chain. The result is a T cell that expresses both chains of a desired antigen specific TCR under the control of the endogenous TCR-α promoter (**Extended Data Fig. 9d**). TCR replacement at TRAC was also possible in combination with electroporation of an RNP to knockout the TCR-β chain (**Extended Data Fig. 9h**).

TCR replacement can also be accomplished at the *TCR-β* locus with a similar strategy to targeting *TCR-α*, although the β locus is more complex as there are two

constant regions (*TRBC1* and *TRBC2*) that are highly homologous to each another. An HDR template inserts a new full length TCR-α and the VDJ regions of a new TCR-β at the 5' end of the first *TRBC1* exon using a gRNA targeting a sequence found in both *TRBC1* and *TRBC2*. Due to the sequence similarity between the *TRBC1* and *TRBC2* genomic regions, the 3' homology arm of this construct is almost perfectly homologous as well to the equivalent region in *TRBC2*, while the 5' homology arm has ~85% homology to the *TRBC2* genomic region in the 150 bps closest to the insertion site. Insertion thus likely predominates at *TRBC1*, but could also possible at *TRBC2* or with an intervening deletion between *TRBC1* and *TRBC2*. gRNAs that cut specifically at *TRBC1* or *TRBC2* could also used instead of the gRNA that targets both. Multiplexed replacement of the TCR, with a new TCR-α VJ domain targeted to *TRAC* and a new TRB-β VDJ domain targeted to *TRBC1* was also possible and could present a strategy to further reduce TCR mispairing (**Extended Data Fig. 9h, i**).