An Efficient Single-Cell RNA-Seq Approach to Identify Neoantigen-Specific T Cell Receptors

Yong-Chen Lu,[1] Zhili Zheng,[1] Paul F. Robbins,[1] Eric Tran,[1] Todd D. Prickett,[1] Jared J. Gartner,[1] Yong F. Li,[1] Satyajit Ray,[1] Zulmarie Franco,[1] Valery Bliskovsky,[2] Peter C. Fitzgerald,[3] and Steven A. Rosenberg[1]

[1]Surgery Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA; [2]Genetics Branch, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA; [3]Genome Analysis Unit, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA

The adoptive transfer of neoantigen-reactive tumor-infiltrating lymphocytes (TILs) can result in tumor regression in patients with metastatic cancer. To improve the efficacy of adoptive T cell therapy targeting these tumor-specific mutations, we have proposed a new therapeutic strategy, which involves the genetic modification of autologous T cells with neoantigen-specific T cell receptors (TCRs) and the transfer of these modified T cells back to cancer patients. However, the current techniques to isolate neoantigen-specific TCRs are labor intensive, time consuming, and technically challenging, not suitable for clinical applications. To facilitate this process, a new approach was developed, which included the co-culture of TILs with tandem minigene (TMG)-transfected or peptide-pulsed autologous antigen-presenting cells (APCs) and the single-cell RNA sequencing (RNA-seq) analysis of T cells to identify paired TCR sequences associated with cells expressing high levels of interferon-γ (IFN-γ) and interleukin-2 (IL-2). Following this new approach, multiple TCRs were identified, synthesized, cloned into a retroviral vector, and then transduced into donor T cells. These transduced T cells were shown to specifically recognize the neoantigens presented by autologous APCs. In conclusion, this approach provides an efficient procedure to isolate neoantigen-specific TCRs for clinical applications, as well as for basic and translational research.

INTRODUCTION

Adoptive cell therapy (ACT) using autologous tumor-infiltrating lymphocytes (TILs) can be an effective immunotherapy for patients with metastatic melanoma.[1,2] Recent clinical trials have extended the reach of this therapy to patients with additional types of metastatic cancer.[3,4] Post-treatment analyses of ACT, as well as immune checkpoint blockade therapies, have suggested that effective cancer immunotherapies are strongly associated with the activation of neoantigen-reactive T cells.[3,5–14] However, most patients with common epithelial cancers do not respond to current immunotherapy approaches, including ACT.[14] To improve the efficacy of ACT targeting tumor-specific mutations, we have proposed a new therapeutic strategy that involves the following steps: (1) neoantigen-specific T cell receptors (TCRs) are isolated from TILs grown from a cancer patient's resected tumors, (2) T cells obtained from the patient's own peripheral blood are genetically modified to express these neoanti-

gen-specific TCRs, and (3) autologous T cells with the new neoantigen specificities can then be transferred back to the cancer-bearing patient following host manipulations to enhance the activity of T cells.[8,14]

One of the major challenges for this proposed strategy is the efficient isolation of neoantigen-specific TCRs. In humans, a TCRα chain comprises a variable (V) gene segment, a joining (J) gene segment, and a constant (C) gene segment. A TCRβ chain contains an additional diversity (D) gene segment between V and J gene segments. Human TCR nucleotide sequences are highly diverse due to the recombination of V(D)J gene segments, the imprecise joining of nicked segments, the addition of non-germline nucleotides, and the pairing of TCRα chain and TCRβ chain.[15,16] The specificity of the TCR is predominantly determined by the peptide contact region complementarity-determining region 3 (CDR3), which encompasses the highly diverse V(D)J junction. Because of the high diversity of CDR3, the CDR3 nucleotide sequences can be used as unique signatures for each individual TCR.

To isolate paired TCRα/β sequences, the conventional approach involves T cell cloning by limiting dilution, identification of reactive T cell clones, and subsequent isolation and sequencing of TCR cDNA.[17] This approach is time consuming and technically challenging, with a high failure rate. These challenges can be summarized in the following four points. (1) T cell cloning by limiting dilution can take 2–4 weeks to grow T cell clones in 96-well plates. Because of T cell exhaustion, some antigen-specific T cell clones may fail to grow to a sufficient number of cells for testing reactivity and the subsequent molecular cloning. In addition, more than one T cell clone may grow from the same well, leading to unclear or incorrect TCR-sequencing results. (2) Molecular cloning can be challenging because of the low amount of cDNA isolated from
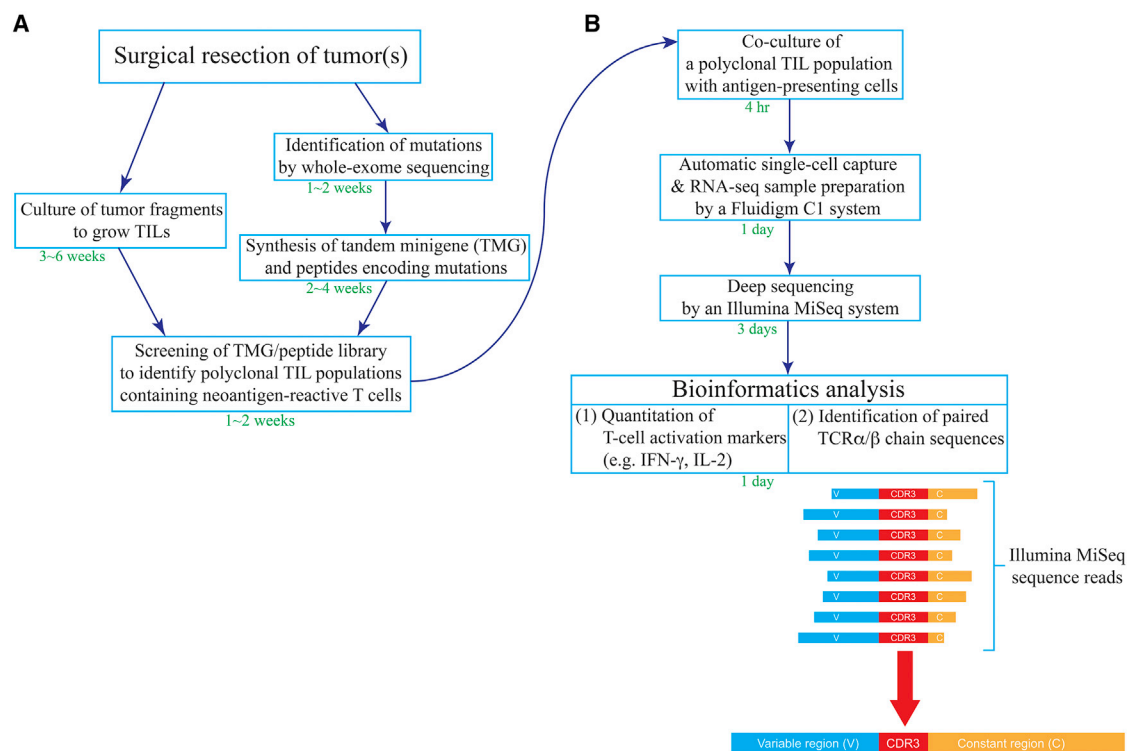
CrossMark

Figure 1. Schema for a New Approach to Identify Neoantigen-Specific TCRs

(A) Polyclonal tumor-infiltrating lymphocytes (TILs) were cultured from fresh tumor fragments and screened against tandem minigenes (TMGs) or a peptide library to identify neoantigen-reactive TIL populations. (B) Once a neoantigen-reactive TIL population was identified, this polyclonal TIL population was co-cultured with neoantigen-transfected/pulsed antigen-presenting cells (APCs) for 4 hr, and it was subjected to single-cell RNA-seq sample preparation and sequencing, followed by bioinformatics analysis.

T cell clones. Universal primers are used to identify V gene segments, and then specific V segment primers are used to amplify and clone the full-length TCRs. Due to the similarity between some of the V gene segments, wrong V gene segments might be identified because of the errors produced by the conventional Sanger sequencing. (3) TCRα and β chains must be paired correctly, otherwise T cells expressing an incorrectly paired TCR may lose specificity or gain unwanted specificities.[18] If more than one T cell clone grows from the same well after limiting dilution, it may lead to incorrect pairing. (4) Up to one-third of mature T cells may express two functional TCRα chains, and likely only one of the TCRα chains contributes the anticipated specificity.[19] The conventional Sanger sequencing results may become uninterpretable because of the mixed TCRα nucleotide sequences. Additional molecular cloning steps, which include cloning the PCR product into a vector and sequencing individual bacteria colonies, may take an additional week to identify the correct TCRα sequences.

In this study, we attempted to overcome these technical challenges and to establish an efficient procedure to identify paired TCRα/β specific to neoantigens. Utilizing this new approach, multiple TCRs were identified, and their specificities against mutations were tested.

RESULTS

The procedure for isolating a neoantigen-specific TCR is summarized in Figure 1. First, tumor specimens were resected from a cancer patient, and TIL fragment cultures were grown individually to generate multiple TIL populations, each containing a sufficient number of TILs ($>5 \times 10^6$ cells).[20,21] Nonsynonymous mutations were identified by whole-exome sequencing of tumor specimens and normal tissues, such as the patient' peripheral blood lymphocytes (PBLs). As described in detail previously, tandem minigenes (TMGs) and/or peptides encoding mutated amino acids, flanked on both sides by 12 additional amino acids present in the normal proteins, were synthesized and pulsed onto autologous antigen-presenting cells (APCs) to identify the potential polyclonal TIL populations containing the neoantigen-reactive T cells (Figure 1A).[7]

Once the polyclonal TIL populations containing neoantigen-reactive T cells were identified, the TMG-transfected or peptide-pulsed APCs were then co-cultured with the identified TILs for 4 hr. The stimulated T cells were subjected to a Fluidigm C1 system to prepare single-cell RNA sequencing (RNA-seq) samples. Each Fluidigm integrated fluidic circuit (IFC) plate could capture approximately 70 individual single T cells in 96 individual reaction chambers. To simplify the process, all 96 samples were barcoded, pooled, and
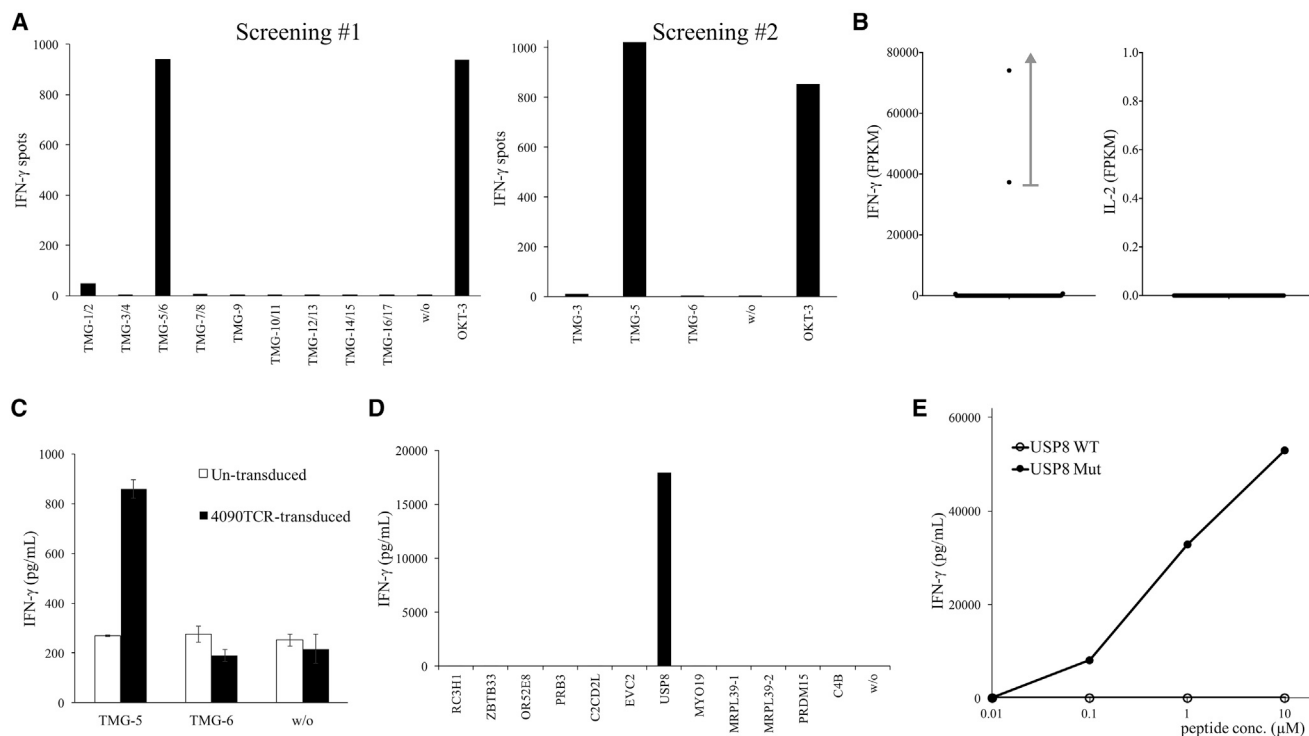
Figure 2. Isolation of a Mutated USP8-Specific TCR

(A) TIL4090F7 T cells were screened against a TMG library. The reactivity of T cells against TMG was measured by IFN-γ ELISPOT assay. (B) TIL4090F7 T cells were co-cultured with TMG-5-transfected autologous DCs for 4 hr, and then they were subjected to single-cell RNA-seq analysis. The expression of IFN-γ and IL-2 mRNA in each single cell was obtained by bioinformatics analysis. FPKM, fragments per kilobase of transcript per million mapped reads. (C) 4090TCR was transduced into donor T cells, and then transduced T cells were co-cultured with TMG-transfected autologous DCs. Error bars represent SD. (D) Mutated 25-mer peptides corresponding to each minigene within TMG-5 were pulsed on autologous DCs for 24 hr, and peptide-pulsed DCs were co-cultured with transduced T cells. (E) Purified 25-mer WT or mutated USP8 peptide (WAKFLDPITGTFHYYHSPTNTVHMY, R > H) was pulsed on autologous DCs for 24 hr, and peptide-pulsed DCs were co-cultured with transduced T cells. The secretion of IFN-γ from T cells was determined by ELISA.

deep-sequenced by Illumina MiSeq, regardless of whether each individual reaction chamber contained a single cell or not. Single-cell RNA-seq samples with high expressions of T cell activation markers, such as interferon (IFN)-γ and interleukin-2 (IL-2), were selected, and paired TCRα/β chain sequences from these samples were identified (Figure 1B).

To test this new approach, four polyclonal TIL populations isolated from four cancer patients were used in this study. TIL4090 cultures were grown from a metastatic lung lesion resected from a patient with colorectal cancer. Seventeen TMGs encoding 201 mutated minigenes were synthesized, and TMG mRNAs were made by *in vitro* transcription (Table S1). TIL4090 cultures were co-cultured with autologous dendritic cells (DCs) transfected with the TMG library, and one of the cultures, TIL4090F7, was found to be strongly reactive to TMG-5, determined by an IFN-γ enzyme-linked immunospot (ELISPOT) assay (Figure 2A).

To isolate the potential neoantigen-specific TCR, TIL4090F7 cells were co-cultured with TMG-5-transfected autologous DCs for 4 hr, and then they were subjected to single-cell RNA-seq analysis. Among

all 96 samples, two samples contained high levels of IFN-γ mRNA (74,105 and 37,316 fragments per kilobase of transcript per million mapped reads [FPKM]), while the remaining samples contained low levels of IFN-γ mRNA (0–716 FPKM). None of the samples contained any detectable IL-2 mRNA using this approach (Figure 2B). The data suggested that these two single T cells specifically reacted to neoantigens presented by DCs. In the next step, the TCRα/β variable regions and CDR3 sequences were identified from the single-cell RNA-seq data of these two samples. TCRα/β chain sequences from both samples were identical. The unique CDR3 sequences of 4090TCR are shown in Table 1.

To test the reactivity of this 4090TCR, the full-length TCRα and TCRβ sequences with modified mouse constant regions, linked by a furinSGSGP2A linker, were synthesized and cloned into a murine stem cell virus-based splice-gag vector (MSGV) retroviral expression vector. Peripheral blood T cells were transduced with 4090TCR and co-cultured with TMG-5-transfected autologous DCs overnight. Based on the secretion of IFN-γ by T cells, 4090TCR-transduced T cells recognized TMG-5-transfected DCs, but not DCs transfected with irrelevant TMG (Figure 2C). TMG-5 contained 12 minigenes

Table 1. The CDR3 Sequences of Four TCRs

TCR No.	TCR Variable Region	CDR3 (Nucleotide Sequence; 5′–3′)	CDR3 (Amino Acid Sequence)
4090	AV3	TGTGCTGTGAGAGACCATAGCAACTATCAGTTAATCTGG	CAVRDHSNYQLIW
	BV14	TGTGCCAGCAGCCAATCCGGTGGGGGCGGGTTCTCCTACAATGAGCAGTTCTTC	CASSQSGGGGFSYNEQFF
4095	AV4	TGCCTCGTGGGTGACATGGACCAGGCAGGAACTGCTCTGATCTTT	CLVGDMDQAGTALIF
	BV5-6	TGTGCCAGCAGCTTGGGGAGGGCAAGCAATCAGCCCCAGCATTTT	CASSLGRASNQPQHF
4112	AV38-1	TGTGCTTTCATGTGGGGATTAGGTCAGAATTTTGTCTTT	CAFMWGLGQNFVF
	BV28	TGTGCCAGCAGTGTGGAGCGGGAGAACACCGGGGAGCTGTTTTTT	CASSVERENTGELFF
4171	DV3	TGTGCCTTTACCCCAACTGGAGCCAATAGTAAGCTGACATTT	CAFTPTGANSKLTF
	BV7-8	TGTGCCAGCAGCGGACGGTCAGGGGGTGAGCAGTTCTTC	CASSGRSGGEQFF

encoding 12 nonsynonymous mutations. Each 25-mer peptide, corresponding to each minigene, contained the nonsynonymous mutation flanked on both sides by 12 normal amino acids. These 25-mer peptides were individually pulsed on autologous DCs for 24 hr, and peptide-pulsed DCs were then co-cultured with 4090TCR-transduced T cells overnight. 4090TCR-transduced T cells reacted only to mutated ubiquitin-specific peptidase 8 (USP8, WAKFLDPITGT FHYYHSPTNTVHMY [R > H])-pulsed DCs, suggesting that 4090TCR recognized mutated USP8 (Figure 2D). Lastly, high-performance liquid chromatography (HPLC)-purified mutated USP8 long peptide and the wild-type (WT) counterpart were pulsed on autologous DCs for 24 hr, and then peptide-pulsed DCs were co-cultured with 4090TCR-transduced T cells overnight. 4090TCR-transduced T cells reacted to mutated USP8 peptide at a minimum of 0.1 μM, but no significant recognition of WT USP8 peptide was observed (Figure 2E).

In a similar context, TIL4095 cultures were also grown from metastatic lung lesions resected from the second patient with colorectal cancer.[22] Screening of 5 TMGs encoding 61 mutated minigenes showed that one of the cultures, TIL4095F5, recognized TMG-1 (Table S2). To isolate the potential neoantigen-specific TCR, TIL4095F5 cells were co-cultured with TMG-1-transfected autologous DCs for 4 hr, and they were subjected to single-cell RNA-seq analysis. All seven samples with high levels of IFN-γ mRNA (4,550–22,522 FPKM) contained identical TCRα/β CDR3 sequences (Figure 3A). Only two samples contained detectable IL-2 mRNA (15.62 and 233.5 FPKM), and these two samples co-expressed IFN-γ at high levels (6,755 and 6,469 FPKM, respectively) (Figures 3A and 3B).

To test the function of this TCR, the full-length TCRα and TCRβ chains were synthesized and cloned into an MSGV retroviral expression vector, and then they were transduced into donor T cells. In a previous study, we isolated an HLA-C*0802-restricted, mutated KRAS(G12D)-specific TCR from a patient with colorectal cancer.[4] Because patient 4095 was found to be positive for HLA-C*0802 and mutated KRAS(G12D) was encoded in TMG-1, we tested whether this 4095TCR could also recognize HLA-C*0802-restricted KRAS (G12D). As shown in Figure 3C, 4095TCR-transduced T cells were co-cultured with full-length KRAS(WT) or (G12D) mRNA-trans-

fected autologous DCs overnight. 4095TCR-transduced T cells recognized KRAS(G12D)-transfected DCs, but not DCs transfected with KRAS(WT). Lastly, autologous DCs were pulsed with the minimal epitope of HLA-C*0802-restricted KRAS(G12D) (GADGVGKSA) for 2 hr. 4095TCR-transduced T cells recognized KRAS(G12D) epitope at a minimum of 0.01 μM, but not the WT counterpart (Figure 3D).[22]

TIL4112 cultures were also grown from a metastatic liver lesion resected from a patient with cholangiocarcinoma. Twenty TMGs encoding 263 mutated minigenes were synthesized (Table S3). One of the cultures, TIL4112F5, recognized TMG-9, based on the results of TMG library screening. To identify the potential neoantigen-specific TCR, TIL4112F5 cells were co-cultured with TMG-9-transfected autologous DCs for 4 hr, and they were subjected to single-cell RNA-seq analysis. Twenty-two samples contained high levels of IFN-γ mRNA (10,857–47,741 FPKM) (Figure 4A), and all these samples contained the identical TCRβ CDR3 sequence (Table 1). Nine samples contained the identical TCRα CDR3 sequence (Table 1), but the other 13 samples did not contain any detectable TCRα CDR3 sequence. On the other hand, eight samples contained detectable IL-2 mRNA (9.526–619.3 FPKM) (Figure 4A). Among them, six samples had the same TCRα and TCRβ CDR3 sequences (Table 1). One additional sample had the same TCRβ CDR3 sequence, but this sample did not contain any detectable TCRα sequence. The other sample did not have any detectable TCRα or β sequence.

To test the reactivity of the TCR isolated from TIL4112F5, the full-length TCRα and TCRβ sequences with modified mouse constant regions were synthesized and then transduced into donor T cells. 4112TCR-transduced T cells recognized TMG-9-transfected autologous DCs, but not DCs transfected with irrelevant TMG (Figure 4C). Next, individual 25-mer peptides, corresponding to each minigene in TMG-9, were synthesized and tested, but none of the 25-mer peptides presented by DCs was significantly recognized by 4112TCR. Additionally, autologous DCs were no longer available. Therefore, autologous Epstein-Barr virus (EBV)-transformed B cells were generated and used in the subsequent experiments. Next, the mutated amino acid sequence of TMG-9 was submitted to the Immune Epitope Database (IEDB) and NetMHC websites to predict
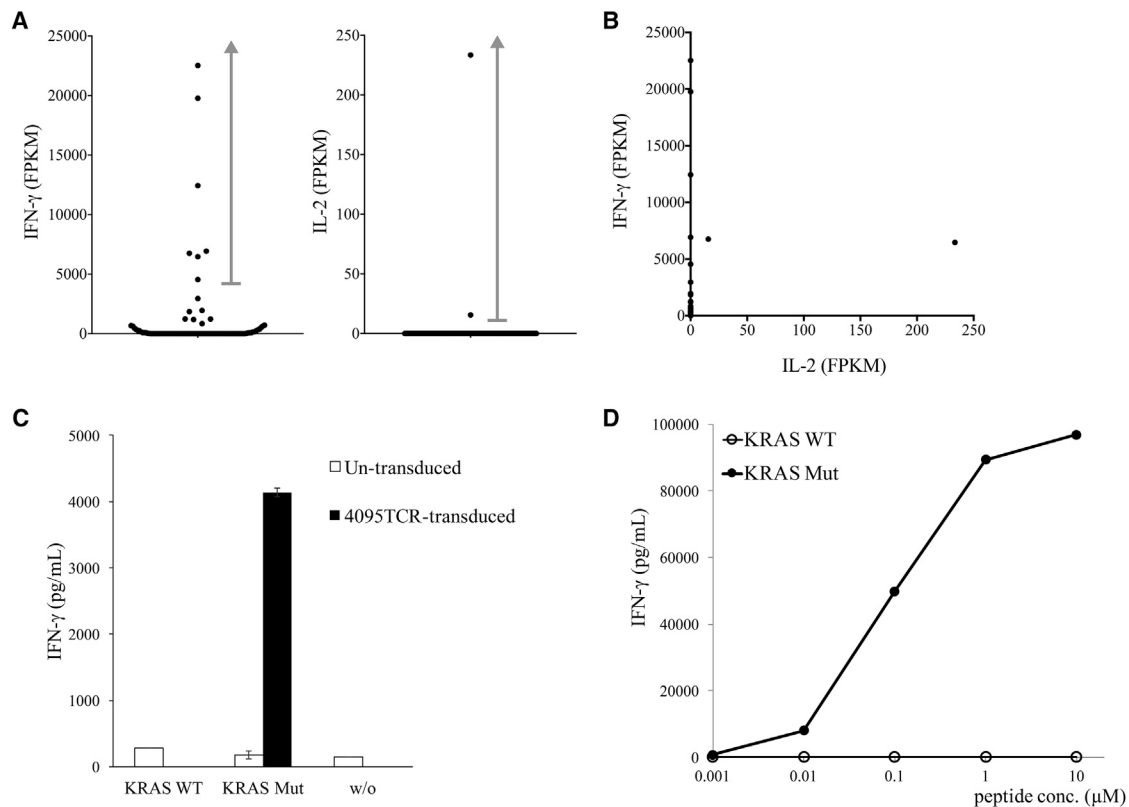
Figure 3. Identification of a Mutated KRAS-Specific TCR

(A and B) TIL4095F5 T cells were co-cultured with TMG-1-transfected autologous DCs for 4 hr, and then they were subjected to single-cell RNA-seq analysis. The expression of IFN-γ and IL-2 mRNA of each single cell is shown in the dot plots (A) and the scatter plot (B). (C) 4095TCR was transduced into donor T cells, and then transduced T cells were co-cultured with full-length WT or mutated KRAS mRNA-transfected autologous DCs. Error bars represent SD. (D) Purified 9-mer WT or mutated KRAS peptide (GA_D_GVGKSA, G > D) was pulsed on autologous DCs for 2 hr, and then peptide-pulsed DCs were co-cultured with transduced T cells. The secretion of IFN-γ from T cells was determined by ELISA.

potential peptides with high affinity to the six major histocompatibility complex class I (MHC class I) molecules identified from patient 4112. Totally, 67 predicted high-affinity peptides from IEDB (rank < 1%) and NetMHC (rank < 2%) were synthesized and combined into 10 pools (Table S4). 4112TCR-transduced T cells recognized short-peptide pool (SPP)-9 pulsed on autologous EBV-transformed B cells (Figure 4D). In the subsequent experiment, mutated neuroblastoma amplified sequence (NBAS) peptide WSYD_S_TLLAY (C > S) was identified as the minimum epitope recognized by 4112TCR-transduced T cells (Figure 4E). The 4112TCR-transduced T cells recognized mutated NBAS peptide, but not the WT counterpart (Figure 4F).

In the last example, TIL4171 cultures were grown from a metastatic lung lesion resected from a patient with colorectal cancer. 128 long peptides (25-mer) were synthesized, and each peptide contained a nonsynonymous mutation flanked on both sides by 12 normal amino acids (Table S5). TIL4171 cultures were screened against the peptide library, and one of the cultures, TIL4171F6, recognized peptide pool 3 (PP-3) (Figure 5A). TIL4171F6 cells were then co-cultured with PP-3-

pulsed autologous DCs for 4 hr, and they were subjected to single-cell RNA-seq analysis. Nine samples contained high levels of IFN-γ mRNA (2,209–24,845 FPKM). Among them, six samples had the same TCRβ CDR3 sequence (Table 1). Two samples did not contain any detectable TCRβ, and one sample contained two different TCRβ CDR3 sequences, which likely resulted from contamination by another T cell. However, none of these samples contained any detectable TCRα chain sequences. Similarly, four samples contained detectable IL-2 mRNA (331.2–1,497 FPKM). These samples all contained the identical TCRβ CDR3 sequence, but none of the samples had any detectable TCRα chain sequence.

In an attempt to discover the missing TCRα chain, we further investigated the single-cell RNA-seq data in this experiment, and we found that four IFN-γ[+] single cells and two IL-2[+] single cells expressed a unique TCR chain, which comprised a V gene segment DV3, a J gene segment AJ56, and a C gene segment AC. It has been known that several V gene segments are shared between TCRα and TCRδ chains, including AV14/DV4, AV23/DV6, AV29/DV5, AV36/DV7, and AV38-2/DV8.[23] These V gene segments have been found to be
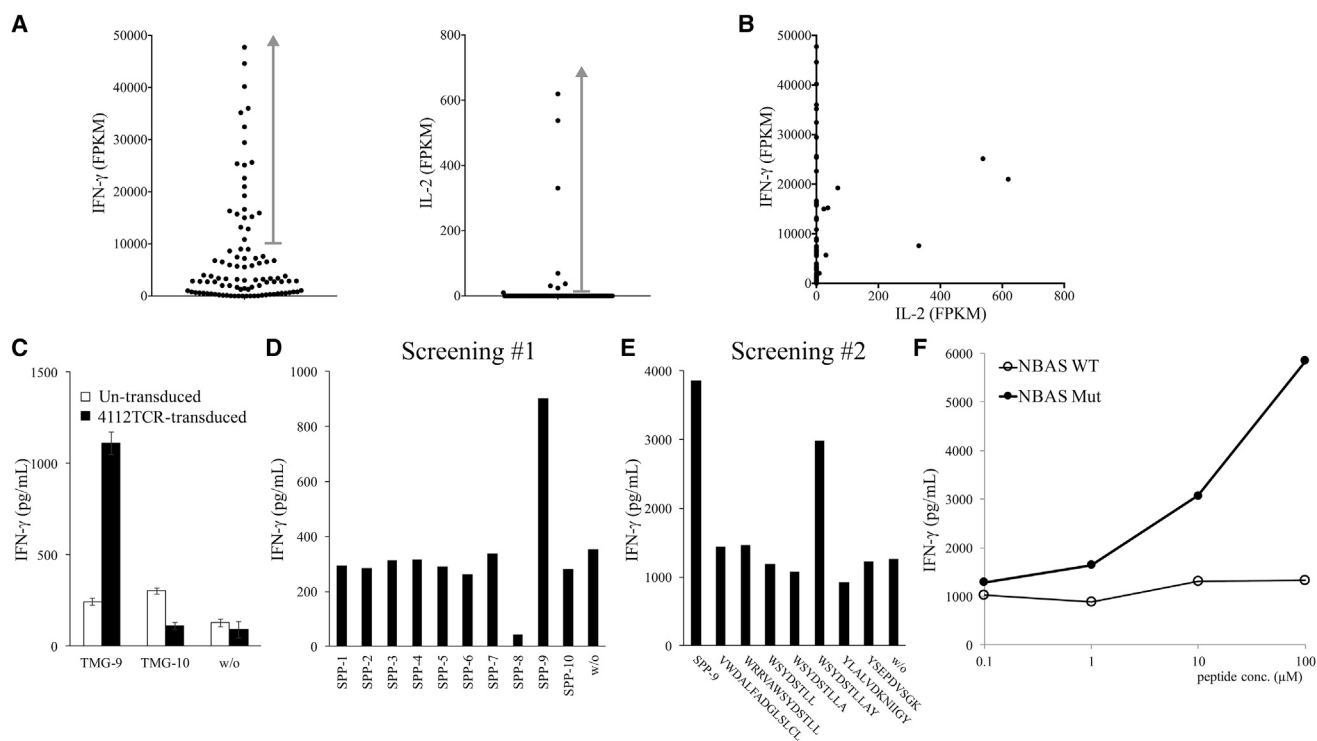
Figure 4. Isolation of a Mutated NBAS-Specific TCR

(A and B) TIL4112F5 T cells were co-cultured with TMG-9-transfected autologous DCs for 4 hr, and single-cell RNA-seq was performed. The expression of IFN-γ and IL-2 mRNA of individual single cells is shown in dot plots (A) and the scatter plot (B). (C) 4112TCR was transduced into donor T cells, and then transduced T cells were co-cultured with TMG-transfected autologous DCs overnight. Error bars represent SD. (D) 67 predicted short peptides were combined into 10 short-peptide pools (SPPs) and pulsed on autologous EBV-transformed B cells for 2 hr. Peptide-pulsed EBV-transformed B cells were then co-cultured with 4112TCR-transduced T cells overnight. (E) 7 predicted short peptides from SPP-9 were individually pulsed on autologous EBV-transformed B cells for 2 hr. Peptide-pulsed EBV-transformed B cells were then co-cultured with 4112TCR-transduced T cells overnight. (F) Purified mutated NBAS peptide (WSYDSTLLAY, C > S) or its WT counterpart was pulsed on autologous EBV-transformed B cells for 2 hr, and then peptide-pulsed B cells were co-cultured with 4112TCR-transduced T cells overnight. The secretion of IFN-γ from T cells was determined by ELISA.

rearranged to AJ-joining gene segments for TCRα and to be rearranged to DD diversity gene segments and DJ-joining gene segments for TCRδ. Notably, the orientation of DV3 transcription is inverted. So far, it has not been reported that a TCRα chain can utilize a DV3 gene segment.

To test the function of this unique TCR chain, this TCR chain was linked to the identified TCRβ chain and then cloned into a retroviral vector. 4171TCR-transduced T cells were strongly reactive to PP-3 (Figure 5D). This peptide pool PP-3 contained 14 mutated 25-mer peptides (Table S5). In the next step, autologous DCs were pulsed with individual peptides, and 4171TCR recognized mutated peptide SIN3 transcription regulator family member A (SIN3A)-pulsed DCs (Figure 5E). Lastly, 4171TCR-transduced T cells were shown to specifically recognize mutated SIN3A peptide (LGKFPELFNW FKIFLGYKESVHLET, N > I), but not the wild-type counterpart (Figure 5F). Therefore, this unique TCR was functional, and it could specifically recognize mutated SIN3A. Similar to other V gene segments, our data suggested that the DV3 gene segment could be shared between TCRα and TCRδ chains.

DISCUSSION

In this report, we describe a new approach to isolate the sequences of neoantigen-specific TCRs. We further demonstrated that these TCRs could recognize neoantigens presented by autologous APCs in four examples. We found that the early T cell activation marker IFN-γ mRNA was a valuable marker to identify T cells that had been activated by a specific neoantigen. Although another early T cell activation marker, IL-2 mRNA, was an alternative indicator of neoantigen-specific TCRs, it appeared to be less useful because of its low expression levels. 4-1BB (CD137) cell surface protein was originally identified as a late T cell activation marker with the optimal protein expression at approximately 24–48 hr.[24,25] In our previous studies, 4-1BB protein was demonstrated as a good cell surface marker to isolate neoantigen-reactive T cells after 16-hr stimulation with neoantigens presented by autologous DCs.[26] Additionally, programmed cell death-1 (PD-1, CD279) protein was a valuable cell surface marker to enrich neoantigen-reactive T cells from peripheral blood or tumors, but notably PD-1 was also expressed on self-antigen-reactive T cells.[27] However, TILs in this study were only stimulated for a short period of time (4 hr). As a result, the expression of PD-1 and 4-1BB
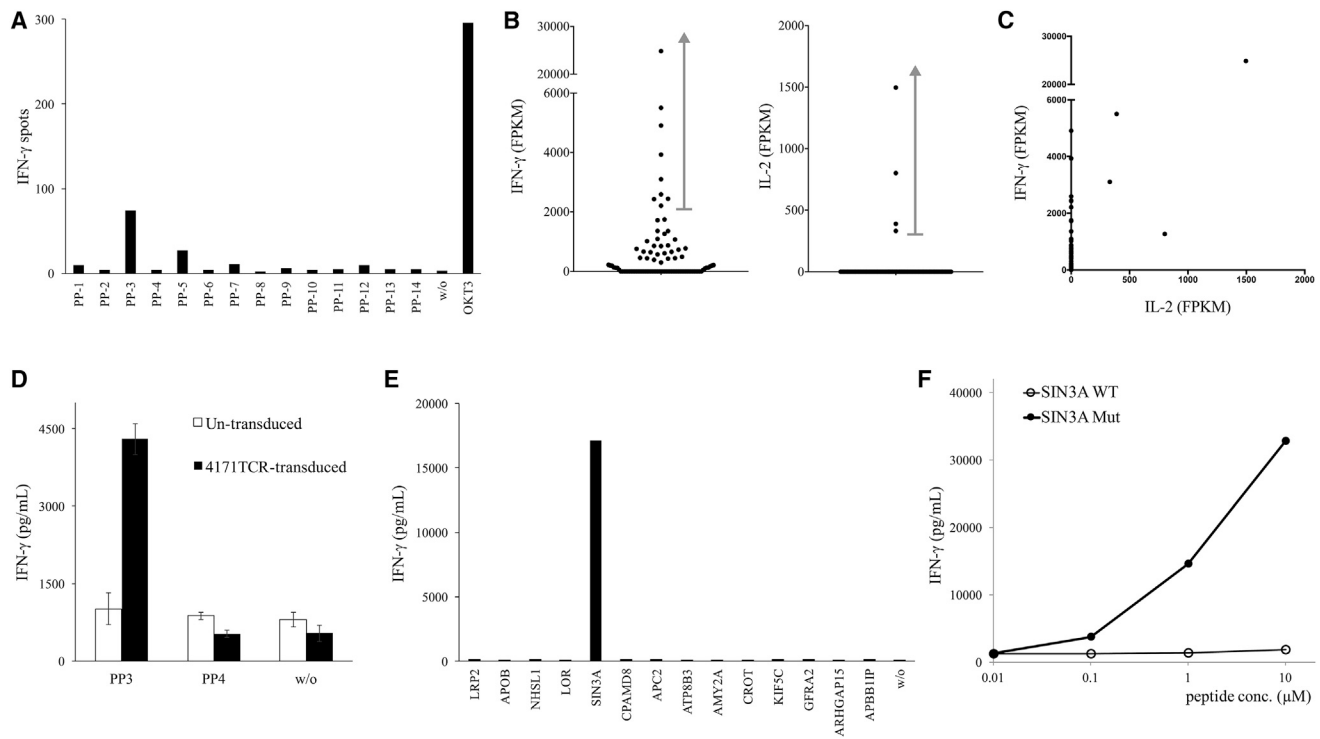
Figure 5. Isolation of a Mutated SIN3A-Specific TCR

(A) TIL4171F6 T cells were screened against a library of 25-mer long-peptide pools (PPs) encoding mutations. The reactivity of T cells against mutation was measured by IFN-γ ELISPOT assay. (B and C) TIL 4171 F6 T cells were co-cultured with PP-3-pulsed autologous DCs for 4 hr, and then they were subjected to single-cell RNA-seq analysis. The expression of IFN-γ and IL-2 mRNA of each single cell is shown in dot plots (B) and the scatter plot (C). (D) 4171TCR was transduced into donor T cells, and then transduced T cells were co-cultured with PP-pulsed DCs. Error bars represent SD. (E) Individual mutated 25-mer peptides corresponding to PP-3 were pulsed on autologous DCs for 24 hr, and peptide-pulsed DCs were co-cultured with 4171TCR-transduced T cells. (F) Purified 25-mer WT or mutated SIN3A peptide (LGKFPELFNWFKIFLGYK ESVHLET, N > I) was pulsed on autologous DCs for 24 hr, and peptide-pulsed DCs were co-cultured with transduced T cells. The secretion of IFN-γ from T cells was determined by ELISA.

mRNA was relatively low at this early time point, and PD-1 and 4-1BB were poor indicators for neoantigen-specific TCRs in this specific experimental setting.

In the past few years, several research groups have attempted to improve the process of TCR identification. The first approach involved sorting single T cells into individual wells of 96-well plates by fluorescence-activated cell sorting (FACS), followed by conventional PCR amplification and Sanger sequencing.[28] Although this approach removed the step of T cell cloning, it still required sub-cloning if a T cell expressed two TCRα chains. A subsequent study solved this problem by utilizing next-generation sequencing (NGS) techniques to analyze amplified TCR sequences.[29] This NGS technique could obtain TCRα chain sequences from a single T cell expressing either one or two TCRα chains. The second approach involved the deep sequencing of TCRα/β CDR3 from an oligoclonal population, and then TCRα and β chains were paired based on frequency-based matching.[30] However, this approach was not useful to T cells with more than one functional TCRα chain, as well as T cells in a highly diverse population. The latest approach, called pairSEQ, involved splitting a pool of T cells into a 96-well plate. TCR cDNA from individual wells was

then barcoded and deep-sequenced. The pairing of TCRα/β chains was predicted by finding the same paired TCRα/β CDR3 sequences found in several individual wells.[31] Although these new approaches could overcome some of the difficulties mentioned in the Introduction, these approaches still required significant labor and time.

Similar to our study, a recent study used a mouse model of *Salmonella* infection to study the expansion and phenotypes of clonal CD4+ T cells after infection.[32] At different time points after infection, CD4+ T cells from spleens were sorted by FACS, and single-cell RNA-seq data were obtained using Illumina HiSeq2500 (paired-end 100-bp reads). To obtain TCR sequences, RNA-seq data were mapped against all possible combinations of mouse V and J regions. In addition, ambiguous "N" nucleotide sequence characters were introduced into the junction between V and J regions to improve the alignments of reads. Lastly, single cells from the same TCRα/β were grouped to analyze the gene expression profile. In contrast to this published method, we took advantage of known TCR biology, allowing us to develop a simplified bioinformatics approach. As detailed in the Materials and Methods, the single-cell RNA-seq data were aligned by human V region sequences, and TCR sequences with the same

CDR3 nucleotide sequences were piled up and counted. In addition, longer sequences (paired-end 250-bp reads by Illumina MiSeq) enabled us to identify CDR3 sequences and assemble full-length TCRs more easily. Most importantly, we went further to test the specificity of these TCRs by expressing these newly identified TCRs in donor T cells, and we showed that these TCR-transduced T cells could recognize neoantigens presented by autologous APCs.

The technology for single-cell transcriptome analysis has evolved significantly in the past few years, and it's likely to continue to improve in the near future.[33] New single-cell technologies may help to overcome some of the current technical limitations, such as the number and percentage of single cells that can be captured and analyzed in each experiment. In addition, better data quality may help to identify TCRs expressed at low levels. In the future, these improvements may enable us to identify TCRs from more challenging specimens, such as isolating TCRs directly from uncultured, unexpanded TILs from tumor specimens.

Although this study significantly improved the technique for TCR isolation, it remains labor intensive and time consuming in some other parts of the process. As a result, it may take 3–5 months to prepare the good manufacturing practice (GMP)-grade cell products for this highly personalized TCR therapy targeting neoantigens. In comparison, it took 103 days (range of 89–160 days) to prepare the personalized RNA vaccines targeting neo-epitopes.[34] We are actively developing new approaches and utilizing new technologies in the attempts to simplify and optimize several steps of this process. For instance, currently it can take 3–6 weeks to expand TILs in order to obtain a sufficient number of cells for screening, and then an additional week to screen against a TMG/peptide library (Figure 1A). Despite the amount of time and labor in this current approach, we could identify approximately 1–6 TIL cultures containing neoantigen-reactive T cells among 24 TIL cultures generated from a tumor specimen, and we were able to identify neoantigen-reactive T cells from 42 of 54 patients with gastrointestinal cancer (M.R. Parkhurst, F.R. Robbins, E.T., R.P. Somerville, J.J.G., L. Jia, T.D.P., Y.F.L., S.R., L.T. Ngo, S.A.R., unpublished data).[4,14] We will continue to streamline the entire process, such as utilizing new technologies to reduce the number of cells required for screening, so it will take less time to expand TILs. Ultimately, we hope to reduce the time, labor, and cost to the minimum; thus, this type of T cell therapy can become feasible and affordable to the majority of cancer patients.

We plan to initiate this proposed neoantigen-specific TCR clinical trial in the near future. This trial will allow us to test the hypothesis that neoantigen-reactive T cells can induce tumor regressions. A potential limitation is that tumors may resist this therapy through the loss of antigens or the components in the antigen presentation pathway, such as β-2-microglobulin, as demonstrated in several recent studies.[22,34–37] Targeting multiple antigens or MHC class II-restricted antigens at the same time may overcome such resistance. Additionally, combining checkpoint blockade therapy with T cell therapy may help to enhance the clinical efficacy by preventing T cell exhaustion.[38–40] These hypotheses should be tested in the future clinical trials.

MATERIALS AND METHODS

Generation of TILs, DCs, and EBV-Transformed B Cells

All patient materials were obtained from a clinical trial approved by the National Cancer Institute Institutional Review Board (Clinical Trial registration ID: NCT01174121). The method to generate TILs has been described in detail previously.[20,21] Briefly, a tumor specimen was cut into 24 tumor fragments (2–3 mm), which were then cultured in RPMI 1640 media containing human serum (10%) and IL-2 (6,000 IU/mL) in 24-well plates. Half of the medium was changed on day 5 after the initiation of TIL culture and every 2–3 days thereafter. TILs were split to 2 wells when reaching confluence. It took approximately 3–6 weeks to obtain a sufficient number of TILs (>5 × 10^6 cells) for screening and TCR isolation.

To generate autologous DCs and EBV-transformed B cells, we followed the protocols described in the *Current Protocols in Immunology* (Units 7.32 and 7.22), with minor modifications. Briefly, CD14+ monocytes were purified from patients' peripheral blood mononuclear cell (PBMC) samples by anti-human CD14 magnetic particles (BD Biosciences, Franklin Lakes, NJ). Purified monocytes (1 × 10^7 cells) were cultured in 10 mL RPMI 1640 media containing 10% fetal calf serum (FCS), 50 ng/mL granulocyte-macrophage colony-stimulating factor (GM-CSF), and 20 ng/mL IL-4 (R&D Systems, Minneapolis, MN) in a Petri dish. 5 mL fresh medium was added on day 3, and then the non-adherent DCs were harvested on day 6. To generate EBV-transformed B cells, patients' PBMCs (1 × 10^7 cells) were cultured in 4 mL complete RPMI 1640 medium (10% FCS) and 1 mL B95-8 culture supernatant containing EBV (ATCC, Manassas, VA) for approximately 3 weeks. The EBV-containing culture medium was then removed, and the EBV-transformed B cell lines were expanded and maintained in complete RPMI 1640 medium.

Screening of Neoantigen-Reactive TILs

The polyclonal TIL populations were screened to identify neoantigen-reactive TILs. As described in detail previously, nonsynonymous mutations in tumors were identified by whole-exome sequencing, and patients' PBL samples were used as the normal control.[7] The screening process started with synthesizing TMG and peptide libraries encoding mutated amino acids, flanked on both sides by 12 additional normal amino acids (GenScript, Piscataway, NJ). Each TMG mRNA was synthesized by *in vitro* transcription using an mMESSAGE mMACHINE T7 ultra transcription kit (Ambion, ThermoFisher Scientific, Waltham, MA). TMG mRNA was then transfected by electroporation using a Neon Transfection System (1,500 V, 30 ms, 1 pulse) by following the manufacturer's instructions (Invitrogen, ThermoFisher Scientific, Waltham, MA). TMG-transfected or peptide-pulsed DCs were then co-cultured with TILs for 16 hr. Upon stimulation, neoantigen-reactive TILs produced IFN-γ, which was detected by an ELISPOT assay.

Identification of Neoantigen-Specific TCR Sequences from Single-Cell RNA-Seq Data

The main purpose of this new approach was to reduce the total process time with minimum labor. To achieve this, several steps

were eliminated or simplified, but data with reasonable quality were obtained. After a polyclonal TIL population was identified by screening, 1×10^6 TILs were co-cultured with 1×10^6 TMG-transfected or peptide-pulsed DCs for 4 hr. After co-culture, T cells were re-suspended and washed extensively, and then they were loaded on a small-sized (5–10 μm) IFC plate. Lysis buffer, reverse-transcription reaction mix, and PCR reaction mix were also loaded on an IFC plate, according to the manufacturer's instruction (Fluidigm, South San Francisco, CA). Single cells were automatically captured, and single-cell RNA-seq samples were also prepared automatically within the Fluidigm C1 system. All 96 single-cell RNA-seq samples were barcoded by Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA), and then they were sequenced by Illumina MiSeq system using reagent kit V3 (2×250 bp).

The following bioinformatics pipelines were used for NGS data analysis. The FPKM values of single-cell RNA-seq samples were calculated by a Partek Flow pipeline using STAR 2.4.1d and Cufflinks 2.2.1 (Partek, St. Louis, MO). Individual single cells with high levels of IFN-γ or IL-2 mRNA were selected to identify the TCR chain sequences in the subsequent steps. Next, selected single-cell RNA-seq data were aligned by Burrows-Wheeler Aligner (BWA) using the TCRα/β V region sequence database from the international immunogenetics information system (IMGT).[41] Using an in-house bioinformatics pipeline, CDR3 region sequences were identified and analyzed based on the conservative amino acid residuals (Cys…Phe/Trp) near the C terminus of the V region.[42] TCR chains with non-productive (out-of-frame) sequences were removed from the analysis. Additionally, some samples might contain more than one T cell due to the imperfect capturing mechanism of the Fluidigm C1. To streamline the process, samples with more than one TCRβ CDR3 sequence were eliminated. Individual CDR3 sequences with less than four reads within a sample were considered as sequencing noise. To assemble full-length TCR chain sequences, the partial V gene segment sequences were assembled with the identified human full-length TCR V gene segment sequences obtained from the IMGT database. To enhance pairing and avoid mispairing of TCRα/β, the partial C gene segment sequences were replaced by modified mouse constant region sequences (Figure 1B).[43–45]

Functional Testing of Neoantigen-Specific TCRs
The detailed protocol has been described previously, with some minor modifications described here.[46] Full-length TCRα and TCRβ sequences with modified mouse constant regions, linked by a fur-inSGSGP2A linker (RAKRSGSGATNFSLLKQAGDVEENPGP), were synthesized and cloned into an MSGV retroviral expression vector.[47] MSGV-TCR plasmid (1.5 μg) and 0.75 μg vesicular stomatitis virus glycoprotein (VSV-G; RD114) plasmid were co-transfected into 1×10^6 293GP cells in each 6-well plate using Lipofectamine 2000 Transfection Reagent (Invitrogen, Thermo Fisher Scientific). After 48 hr, the supernatant was harvested and spun at 3,000 rpm for 10 min to remove debris. The retrovirus supernatant was loaded on RetroNectin- (Takara, Otsu, Japan) coated 6-well plates by centrifugation at $2,000 \times g$ for 2 hr.

Separately, 1×10^6/mL PBMCs from health donors were stimulated with 50 ng/mL anti-CD3 mAb OKT3 and 1,200 IU/mL IL-2 in AIM V medium containing 5% human serum. After 2 days, stimulated cells were harvested and re-suspended in the same medium without OKT3. Stimulated PBMCs were added to each retrovirus-loaded well at 2×10^6 cells/well and spun at $1,000 \times g$ for 10 min. Plates were incubated overnight at 37°C, and the next day the PBMCs were transferred to new retrovirus-loaded wells and the transduction procedure was repeated. TCR-transduced T cells were continuously cultured in AIM V medium with 1,200 IU/mL IL-2 and 5% human serum for 5 additional days before performing co-culture experiments.

To test the specificity of TCR-transduced T cells, autologous DCs were transfected with TMG mRNA by a Neon Transfection System. Alternatively, autologous DCs or EBV-transformed B cells were pulsed with peptides. 1×10^5 T cells were then co-cultured with 1×10^5 autologous DCs or EBV-transformed B cells overnight in 96-well U-bottom plates. The supernatant was harvested, and the secretion of IFN-γ from T cells was determined by an ELISA (Thermo Fisher Scientific).

SUPPLEMENTAL INFORMATION
Supplemental Information includes five tables and can be found with this article online at https://doi.org/10.1016/j.ymthe.2017.10.018.

AUTHOR CONTRIBUTIONS
Y.-C.L., Z.Z., P.F.R., E.T., T.D.P., Y.F.L., S.R., Z.F., and V.B. conducted the experiments. J.J.G. and P.C.F. developed the bioinformatics pipelines and analyzed the data. Y.-C.L. and S.A.R. designed the experiments and wrote the paper.

ACKNOWLEDGMENTS

REFERENCES
1. Rosenberg, S.A., Yang, J.C., Sherry, R.M., Kammula, U.S., Hughes, M.S., Phan, G.Q., Citrin, D.E., Restifo, N.P., Robbins, P.F., Wunderlich, J.R., et al. (2011). Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. Clin. Cancer Res. 17, 4550–4557.

2. Rosenberg, S.A., and Restifo, N.P. (2015). Adoptive cell transfer as personalized immunotherapy for human cancer. Science 348, 62–68.

3. Tran, E., Turcotte, S., Gros, A., Robbins, P.F., Lu, Y.C., Dudley, M.E., Wunderlich, J.R., Somerville, R.P., Hogan, K., Hinrichs, C.S., et al. (2014). Cancer immunotherapy based on mutation-specific CD4+ T cells in a patient with epithelial cancer. Science 344, 641–645.

4. Tran, E., Ahmadzadeh, M., Lu, Y.C., Gros, A., Turcotte, S., Robbins, P.F., Gartner, J.J., Zheng, Z., Li, Y.F., Ray, S., et al. (2015). Immunogenicity of somatic mutations in human gastrointestinal cancers. Science 350, 1387–1390.

5. Lu, Y.C., Yao, X., Li, Y.F., El-Gamil, M., Dudley, M.E., Yang, J.C., Almeida, J.R., Douek, D.C., Samuels, Y., Rosenberg, S.A., and Robbins, P.F. (2013). Mutated PPP1R3B is recognized by T cells used to treat a melanoma patient who experienced a durable complete tumor regression. J. Immunol. 190, 6034–6042.

6. Robbins, P.F., Lu, Y.C., El-Gamil, M., Li, Y.F., Gross, C., Gartner, J., Lin, J.C., Teer, J.K., Cliften, P., Tycksen, E., et al. (2013). Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. Nat. Med. 19, 747–752.

7. Lu, Y.C., Yao, X., Crystal, J.S., Li, Y.F., El-Gamil, M., Gross, C., Davis, L., Dudley, M.E., Yang, J.C., Samuels, Y., et al. (2014). Efficient identification of mutated cancer antigens recognized by T cells associated with durable tumor regressions. Clin. Cancer Res. 20, 3401–3410.

8. Lu, Y.C., and Robbins, P.F. (2016). Cancer immunotherapy targeting neoantigens. Semin. Immunol. 28, 22–27.

9. Snyder, A., Makarov, V., Merghoub, T., Yuan, J., Zaretsky, J.M., Desrichard, A., Walsh, L.A., Postow, M.A., Wong, P., Ho, T.S., et al. (2014). Genetic basis for clinical response to CTLA-4 blockade in melanoma. N. Engl. J. Med. 371, 2189–2199.

10. Van Allen, E.M., Miao, D., Schilling, B., Shukla, S.A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Foppen, M.H.G., Goldinger, S.M., et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. Science 350, 207–211.

11. Le, D.T., Uram, J.N., Wang, H., Bartlett, B.R., Kemberling, H., Eyring, A.D., Skora, A.D., Luber, B.S., Azad, N.S., Laheru, D., et al. (2015). PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. N. Engl. J. Med. 372, 2509–2520.

12. Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. Science 348, 124–128.

13. Carreno, B.M., Magrini, V., Becker-Hapak, M., Kaabinejadian, S., Hundal, J., Petti, A.A., Ly, A., Lie, W.R., Hildebrand, W.H., Mardis, E.R., and Linette, G.P. (2015). Cancer immunotherapy. A dendritic cell vaccine increases the breadth and diversity of melanoma neoantigen-specific T cells. Science 348, 803–808.

14. Tran, E., Robbins, P.F., and Rosenberg, S.A. (2017). 'Final common pathway' of human cancer immunotherapy: targeting random somatic mutations. Nat. Immunol. 18, 255–262.

15. Nikolich-Zugich, J., Slifka, M.K., and Messaoudi, I. (2004). The many important facets of T-cell repertoire diversity. Nat. Rev. Immunol. 4, 123–132.

16. Davis, M.M., and Bjorkman, P.J. (1988). T-cell antigen receptor genes and T-cell recognition. Nature 334, 395–402.

17. Lefranc, M.-P. (2001). Using bioinformatics tools for the sequence analysis of immunoglobulins and T cell receptors. Curr. Protoc. Immunol. 71, A.1W.1–A.1W.15.

18. Bendle, G.M., Linnemann, C., Hooijkaas, A.I., Bies, L., de Witte, M.A., Jorritsma, A., Kaiser, A.D.M., Pouw, N., Debets, R., Kieback, E., et al. (2010). Lethal graft-versus-host disease in mouse models of T cell receptor gene therapy. Nat. Med. 16, 565–570.

19. Padovan, E., Casorati, G., Dellabona, P., Meyer, S., Brockhaus, M., and Lanzavecchia, A. (1993). Expression of two T cell receptor alpha chains: dual receptor T cells. Science 262, 422–424.

20. Dudley, M.E., Wunderlich, J.R., Shelton, T.E., Even, J., and Rosenberg, S.A. (2003). Generation of tumor-infiltrating lymphocyte cultures for use in adoptive transfer therapy for melanoma patients. J. Immunother. 26, 332–342.

21. Turcotte, S., Gros, A., Hogan, K., Tran, E., Hinrichs, C.S., Wunderlich, J.R., Dudley, M.E., and Rosenberg, S.A. (2013). Phenotype and function of T cells infiltrating visceral metastases from gastrointestinal cancers and melanoma: implications for adoptive cell transfer therapy. J. Immunol. 191, 2217–2225.

22. Tran, E., Robbins, P.F., Lu, Y.C., Prickett, T.D., Gartner, J.J., Jia, L., Pasetto, A., Zheng, Z., Ray, S., Groh, E.M., et al. (2016). T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer. N. Engl. J. Med. 375, 2255–2262.

23. Lefranc, M.-P. (2001). Nomenclature of the Human T Cell Receptor Genes. Curr. Protoc. Immunol. 40, A.10.1–A.10.23.

24. Wen, T., Bukczynski, J., and Watts, T.H. (2002). 4-1BB ligand-mediated costimulation of human T cells induces CD4 and CD8 T cell expansion, cytokine production, and the development of cytolytic effector function. J. Immunol. 168, 4897–4906.

25. Wolfl, M., Kuball, J., Ho, W.Y., Nguyen, H., Manley, T.J., Bleakley, M., and Greenberg, P.D. (2007). Activation-induced expression of CD137 permits detection, isolation, and expansion of the full repertoire of CD8+ T cells responding to antigen without requiring knowledge of epitope specificities. Blood 110, 201–210.

26. Parkhurst, M.R., Gros, A., Pasetto, A., Prickett, T.D., Crystal, J.S., Robbins, P.F., and Rosenberg, S.A. (2017). Isolation of T-cell receptors specifically reactive with mutated tumor-associated antigens from tumor-infiltrating lymphocytes based on CD137 expression. Clin. Cancer Res. 23, 2491–2505.

27. Gros, A., Parkhurst, M.R., Tran, E., Pasetto, A., Robbins, P.F., Ilyas, S., Prickett, T.D., Gartner, J.J., Crystal, J.S., Roberts, I.M., et al. (2016). Prospective identification of neoantigen-specific lymphocytes in the peripheral blood of melanoma patients. Nat. Med. 22, 433–438.

28. Kobayashi, E., Mizukoshi, E., Kishi, H., Ozawa, T., Hamana, H., Nagai, T., Nakagawa, H., Jin, A., Kaneko, S., and Muraguchi, A. (2013). A new cloning and expression system yields and validates TCRs from blood lymphocytes of patients with cancer within 10 days. Nat. Med. 19, 1542–1546.

29. Han, A., Glanville, J., Hansmann, L., and Davis, M.M. (2014). Linking T-cell receptor sequence to functional phenotype at the single-cell level. Nat. Biotechnol. 32, 684–692.

30. Linnemann, C., Heemskerk, B., Kvistborg, P., Kluin, R.J., Bolotin, D.A., Chen, X., Bresser, K., Nieuwland, M., Schotte, R., Michels, S., et al. (2013). High-throughput identification of antigen-specific TCRs by TCR gene capture. Nat. Med. 19, 1534–1541.

31. Howie, B., Sherwood, A.M., Berkebile, A.D., Berka, J., Emerson, R.O., Williamson, D.W., Kirsch, I., Vignali, M., Rieder, M.J., Carlson, C.S., and Robins, H.S. (2015). High-throughput pairing of T cell receptor α and β sequences. Sci. Transl. Med. 7, 301ra131.

32. Stubbington, M.J.T., Lönnberg, T., Proserpio, V., Clare, S., Speak, A.O., Dougan, G., and Teichmann, S.A. (2016). T cell fate and clonality inference from single-cell transcriptomes. Nat. Methods 13, 329–332.

33. Grün, D., and van Oudenaarden, A. (2015). Design and Analysis of Single-Cell Sequencing Experiments. Cell 163, 799–810.

34. Sahin, U., Derhovanessian, E., Miller, M., Kloke, B.P., Simon, P., Löwer, M., Bukur, V., Tadmor, A.D., Luxemburger, U., Schrörs, B., et al. (2017). Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. Nature 547, 222–226.

35. Zaretsky, J.M., Garcia-Diaz, A., Shin, D.S., Escuin-Ordinas, H., Hugo, W., Hu-Lieskovan, S., Torrejon, D.Y., Abril-Rodriguez, G., Sandoval, S., Barthly, L., et al. (2016). Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma. N. Engl. J. Med. 375, 819–829.

36. Ott, P.A., Hu, Z., Keskin, D.B., Shukla, S.A., Sun, J., Bozym, D.J., Zhang, W., Luoma, A., Giobbie-Hurder, A., Peter, L., et al. (2017). An immunogenic personal neoantigen vaccine for patients with melanoma. Nature 547, 217–221.

37. Riaz, N., Havel, J.J., Makarov, V., Desrichard, A., Urba, W.J., Sims, J.S., Hodi, F.S., Martín-Algarra, S., Mandal, R., Sharfman, W.H., et al. (2017). Tumor and Microenvironment Evolution during Immunotherapy with Nivolumab. Cell, S0092-8674(17)31122-4.

38. Abate-Daga, D., Hanada, K., Davis, J.L., Yang, J.C., Rosenberg, S.A., and Morgan, R.A. (2013). Expression profiling of TCR-engineered T cells demonstrates overexpression of multiple inhibitory receptors in persisting lymphocytes. Blood 122, 1399–1410.

39. Irving, M., Vuillefroy de Silly, R., Scholten, K., Dilek, N., and Coukos, G. (2017). Engineering Chimeric Antigen Receptor T-Cells for Racing in Solid Tumors: Don't Forget the Fuel. Front. Immunol. 8, 267.

40. Lanitis, E., Dangaj, D., Irving, M., and Coukos, G. (2017). Mechanisms regulating T-cell infiltration and activity in solid tumors. Ann. Oncol., Published online September 21, 2017. https://doi.org/10.1093/annonc/mdx238.

41. Lefranc, M.P., Giudicelli, V., Duroux, P., Jabado-Michaloud, J., Folch, G., Aouinti, S., Carillon, E., Duvergey, H., Houles, A., Paysan-Lafosse, T., et al. (2015). IMGT®, the international ImMunoGeneTics information system® 25 years on. Nucleic Acids Res. 43, D413–D422.

42. Lefranc, M.P., Pommié, C., Ruiz, M., Giudicelli, V., Foulquier, E., Truong, L., Thouvenin-Contet, V., and Lefranc, G. (2003). IMGT unique numbering for immunoglobulin and T cell receptor variable domains and Ig superfamily V-like domains. Dev. Comp. Immunol. 27, 55–77.

43. Cohen, C.J., Zhao, Y., Zheng, Z., Rosenberg, S.A., and Morgan, R.A. (2006). Enhanced antitumor activity of murine-human hybrid T-cell receptor (TCR) in human

lymphocytes is associated with improved pairing and TCR/CD3 stability. Cancer Res. *66*, 8878–8886.

44. Cohen, C.J., Li, Y.F., El-Gamil, M., Robbins, P.F., Rosenberg, S.A., and Morgan, R.A. (2007). Enhanced antitumor activity of T cells engineered to express T-cell receptors with a second disulfide bond. Cancer Res. *67*, 3898–3903.

45. Haga-Friedman, A., Horovitz-Fried, M., and Cohen, C.J. (2012). Incorporation of transmembrane hydrophobic mutations in the TCR enhance its surface expression and T cell functional avidity. J. Immunol. *188*, 5538–5546.

46. Morgan, R.A., Dudley, M.E., Wunderlich, J.R., Hughes, M.S., Yang, J.C., Sherry, R.M., Royal, R.E., Topalian, S.L., Kammula, U.S., Restifo, N.P., et al. (2006). Cancer regression in patients after transfer of genetically engineered lymphocytes. Science *314*, 126–129.

47. Wargo, J.A., Robbins, P.F., Li, Y., Zhao, Y., El-Gamil, M., Caragacianu, D., Zheng, Z., Hong, J.A., Downey, S., Schrump, D.S., et al. (2009). Recognition of NY-ESO-1+ tumor cells by engineered lymphocytes is enhanced by improved vector design and epigenetic modulation of tumor antigen expression. Cancer Immunol. Immunother. *58*, 383–394.

Supplemental Information

An Efficient Single-Cell RNA-Seq Approach

to Identify Neoantigen-Specific T Cell Receptors

Yong-Chen Lu, Zhili Zheng, Paul F. Robbins, Eric Tran, Todd D. Prickett, Jared J. Gartner, Yong F. Li, Satyajit Ray, Zulmarie Franco, Valery Bliskovsky, Peter C. Fitzgerald, and Steven A. Rosenberg

Supplementary Table 1. 4090 tandem minigene (TMG) library

TMG ID	Minigene ID	Gene name	Position and change of mutated amino acid	Mutated amino acid sequence
1	1	KCNB2	S421Y	IALPIPIIVNNFYEFYKEQKRQEKA
1	2	ABR	P804L	AQVQVLLYYLQHLPISFAELKRNTL
1	3	MAG	V81A	YPPVVFKSRTQVAHESFQGRSRLLG
1	4	ZNF483	S536L	KCKDCGRPFSDSLSLIQHQRIHTGE
1	5	FSCB	Q659P	QPPPAEEAPAEVPPPPAEEAPAEVQ
1	6	MST1R	V188 frameshift	DCVASPLGTRVTG
1	7	MMP1	W13S	MHSFPPLLLLLFSGVVSHSFPATLE
1	8	PRR35	E488K	GPQALGEAWGRPKLGPVLTGGTPEP
1	9	KARS	S160 frameshift	RGEGVKLQVMANSEIINQKKNLFILITNCVGET
1	10	UGT2A3	P342 frameshift	PQKVLWRYKGKKTIHIRSQYSAV
1	11	SLC17A2	D391E	LLILIPGTSNLCESGFIINTLDIAP
1	12	TMX4	R102C	KVDVIQEPGLSGCFFVTTLPAFFHA
2	13	OLFML2A	R355W	DLSVHRETTWKTWLRRNSYGNCFLV
2	14	TLR5	R224K	VSVDWGKCMNPFKNMVLEILDVSGN
2	15	SETD1B	P1268L	PERAPEHDLEVELEPPMMLPLPLQP
2	16	PCDHA13	A642T	EISTTRPLDEVDTPHHRLLVLVKDH
2	17	HYOU1	R925W	KPRPRPKDKNGTWAEPPLNASASDQ
2	18	SRRM2	P1171L	LETAESKEKMALLPQEDATASPPRQ
2	19	CDRT15L2	G53R	VRRRTQVPQDSPRQALAGQATPEIP
2	20	PCNT	R2543H	THLQNQEKLQHLHTALTSAEARGSQ
2	21	RYR2	D1621Y	ISERQGWLVQCLYPLQFMSLHIPEE
2	22	PRSS12	R630W	VCGLRLLHRRQKWIIGGKNSLRGGW
2	23	TP53	V272M	SSGNLLGRNSFEMRVCACPGRDRRT
2	24	MAP2	A172V	EFHDQQELTPSTVEPSDQKEKESEK
3	25	KCND2	P358S	EKGSSASKFTSISAAFWYTIVTMTT
3	26	MED12	T1546M	KLRLNLVGGMFDMVQRSTQQTTEWA
3	27	ADAM28	I184T	DGVLWAHDLQQNTALPATKLVKLKD
3	28	ZNF843	A92V	APASPLGRSHSSVGVRQGFSGQLCC
3	29	GIMAP6	G62S	GSGKSATGNSILSRDVFESKLSTRP
3	30	SDK1	P772 frameshift	RLMLPEEPPSAPPKKYSGQWAD
3	31	PTCHD1	K619E	KNFTDMLRNSFLEAPQFSHFQEDII
3	32	KRAS	G12V	MTEYKLVVVGAVGVGKSALTIQLI
3	33	OR5H1	M136T	VAICKPLLYPAITTNGLCIRLLILS
3	34	KMT2E	I960 frameshift	FENISSPESSSPEIETHL
3	35	PKD1L1	F164I	ASCSSMMRHSLPRIFVAGLVGALML
3	36	MAOA	S2649R	PPGIMTQYGRVIHQPVGRIFFAGTE
4	37	HRNR	R424H	GSGSGRSSSSGQQGSGLGESSGFGH
4	38	FGD5	H2567Q	ETDEDYIVVPRVLLREDEPKDEGSV
4	39	ATF6	P72L	ELGYFTDTDELQMEAANETYENNFD
4	40	CDH7	F164S	FKRLADMYGTGQGSLYS
4	41	C5orf42	N19T	LRWSQLPVKENKGFSGAAKSHFECG
4	42	KIF5B	A406T	KDITLTNDKPATTIGVIGNFTDAER
4	43	ISL2	G269E	QQQQHSDKTSLQELTGTPLVAGSPI
4	44	R3HDM4	E12V	MVALENPECGPVAAEGTPGGRRLL
4	45	ATG101	R209L	TDALGTSVTTTMLRLIKDTLAL
4	46	PAPOLA	A414D	VGSLEKNEFITLDHVNPQSFPAPKE
4	47	MPRIP	L786V	DAESKHSMSMFTVRGRYEEEIRCVV
4	48	HIST1H1E	E42D	AAKRKASGPPVSDLITKAVAASKER
5	49	RC3H1	Y890C	RFGAISRTSKTICQGAGPMQAMAPQ
5	50	ZBTB33	R510Q	VCKRSYVCLTSLQRHFNIHSWEKKY
5	51	OR52E8	R178C	VFLLLRLPFCGHCIIPHTYCEHMGI
5	52	PRB3	K235Q	GKPEGPPSQGGNQPQGPPPHPGKPQ
5	53	C2CD2L	R212C	GEGLLISWAFTDCPDLSLTVLPKLQ
5	54	EVC2	294-294 deletion	DQMIDILSSEDPSMLQALEELEIAT
5	55	USP8	R667H	WAKFLDPITGTFHYYHSPTNTVHMY
5	56	MYO19	P550S	YHTAGLVEKNKDSIPPELTRLLQQS
5	57	MRPL39-1	L189V	VRAPEVPVLDSKVDEWMPTKENLRS
5	58	MRPL39-2	L189V	GAFCYDVVLDSKVDEWMPTKENLRS
5	59	PRDM15	E580K	CSATFLELQLLNKHLLGHLEQAKSL
5	60	C4B	S1242W	WGSVTGSQSNAVWPTPAPRNPSDPM
6	61	TMEM255B	C151Y	QTEVTCHSLDGKYQLKVRSNTCYCC
6	62	PCDHGA3	E523K	TGVLYALRSFDYKQFRDLKLLVTAS
6	63	PRAMEF11	D67V	PRRWKLQVLDLQVVCENFWMVWSEA
6	64	FIGNL2	Q474P	AAPGAAEGARLLPAAFAAARCRPPS
6	65	PRR36	L760Q	ASAPLTTPPLENQPSLAPPPLQTAS
6	66	OLIG2	S88A	KSSSSSTSSSTSAAAASSTKKDKKQ
6	67	NUTM2F	P559H	VETSPPQTAAQDHQGQGRVRTGMAR

6	68	C1orf21	T75S	LEKSASSNVRLKSNKEVPGLVHQPR
6	69	PTGS2	Q39H	QNRGVCMSVGFDHYKCDCTRTGFYG
6	70	CKAP4	Q57P	PPPAPHPQQHPQPHPQNQAHGKGGH
6	71	WDHD1	R369H	DEDDEDLMMASGHPRQRSHILEDDE
6	72	ITGAM-1	T486P	LVLIGAPHYYEQPRGGQVSVCPLPR
7	73	ITGAM-2	T486P	PGPHRGPPLLRAAPRGPGVRVPLAQ
7	74	RAVER1-1	G65S	PESLPEPTQPAPSQQPGGSSKAFQL
7	75	RAVER1-2	G65S	PESLPEPTQPAPSQQPGGPWRWQQQ
7	76	EYA2	S100I	PPPAQAYGIPSYIIKTEDSLNHSPG
7	77	TRANK1	V257L	LHVVAAHSPGYLLKRQTEDVQMLLR
7	78	ZKSCAN7-1	G195R	PPYDPGTHHLPSRDFGYCGI
7	79	ZKSCAN7-2	G195R	PPYDPGTHHLPSRDFAQCTSPVPTL
7	80	DUX4L4	Q378L	QGSPWWGWGRGPLVAGAAWEPQAGE
7	81	TNXB	R1081H	TVTDRTSDSLLLHWTVPEGEFDSFV
7	82	MUC12	E2160D	SPGSTTALSFGQDSTTFHSSPGSTH
7	83	KCNQ4	A348V	EKRRMPAANLIQVAWRLYSTDMSRA
7	84	MTERFD3	Q192 frameshift	PVEKNKQMVRILRELSRCRWL
8	85	DMXL2	T1644M	AMGIGWWVRNINMLRRCIEKVAKAS
8	86	OR2C1	Y252C	TCLSHLLVVFLFCGSASYGYLLPAK
8	87	UNC13D	G1035S	REVPGLSGSEEPSEVPQTRLPLTYP
8	88	SAMSN1	R235H	DDDGPYSGPFCGHARVHTDFTPSPY
8	89	AF035281	T192M	TPTTARTSGRSPMSAPSSLAACPRC
8	90	FBXL18	T546M	RHLTLAQLPSVLMGSGLVNIGLQCQ
8	91	OCRL	G147E	KLDTKDKPSVFSELLGFEDNFSSMN
8	92	ZNF182	A6E	MTPASESGEDSGSFYSWQ
8	93	DTX3	P2L	MLILSSSGSKMAAC
8	94	MAG	A404T	CVAENQYGQRATTFNLSVEFAPVLL
8	95	PCDHGC5-1	R816W	EPDAIRSRSNTLWERSQQAPPNTDW
8	96	PCDHGC5-2	R816W	EPDAIRSRSNTLWERSQVRGSAPPR
9	97	RXFP3	T179M	YASVFFLTAMSVMRYHSVASALKSH
9	98	KCNA10	G49R	PKGRPGGSSFSNRKILISESTNHET
9	99	TRH	R226W	DDLSRSQGAEEKWQHPGRRAAWVRE
9	100	C4orf48	V85A	GARARAEPAGSAAPAQSRPCVDCHA
9	101	FPGS	E429K	CLEHQQHWNHLDKEQASPDLWSAPS
9	102	FXN	C50S	GRRGLRTDIDATSTPRRASSNQRGL
9	103	HPSE2-1	E262D	QFSNTYSNLILTDPNNYRTMHGRAV
9	104	HPSE2-2	E262D	PGPDYYLKNYEDDPNNYRTMHGRAV
9	105	HPSE2-3	E262D	SKKYNISWELGNDPNNYRTMHGRAV
9	106	RAD9A-1	T93M	GQDLLRCKILMKMVEKCCISLNGRS
9	107	RAD9A-2	T93M	LSVFRSLAMLEKMVEKCCISLNGRS
9	108	OGFOD1-1	V39L	SDAVTEETLKKQLAEAWSRRTPFKV
10	109	OGFOD1-2	V39L	SDAVTEETLKKQLAEAWSRRTPFSH
10	110	MUC12	T1035A	TTSHSSPGSTDTALLPASTTTSGPS
10	111	NRK	Q1252H	DITKLIRRRPFRHIQVLEPLNLLIT
10	112	ITIH6	R523C	PGKQELGIHLAACGPKDQLLVAHHS
10	113	TFAP2E	Q97P	GGLAPLAQPQPPPAAWAAPRAAARA
10	114	ATPAF1	D84G	VRPGSGRPEGGAGSSGVGAEAELQA
10	115	MUC5B	T2445P	PGTTWILTELTTPATTTESTGSTAT
10	116	POU5F1P3	G59R	IGPGFGPGSEEWRIPPCPPPYEFCG
10	117	FAM71E2	R480L	PSQKAPAIPAPSLKASAASASPRKA
10	118	FAM134A	L67R	TLWLRLRGWEAVRAAAQRLLVWEKP
10	119	KCNIP3	A210T	MGRHTYPILREDTPAEHVERFFEKM
10	120	PCDH10	R230H	GGGGAGLPPQQQHTGTALLTIRVLD
11	121	ATPAF1	E93A	GGADSSGVGAEAALQANPFYDRYRD
11	122	SERPING1	D30A	AGSEAGWLRRSAAVAAQMASRLTLL
11	123	LEPREL2	L60R	YAAGAWAPAVALRREALRSQAALGR
11	124	SLC25A35	V114G	PARSAAAGAMAGGMGAYLGSPIYMV
11	125	HIPK4	R603P	LPPRRSHQHGPPPGATSFLQHVTGH
11	126	IQCJ	K146T	QLARPTGFIHTLTEPQIERLGFLTL
11	127	CIZ1	Q578L	FCYICKASCSSQLEFQDHMSEPQHQ
11	128	VCX	M180K	SQVEEPPSQESEKEELPSV
11	129	CROCC	L1066Q	LLAESEKQQALSQKESEKTALSEKL
11	130	GJB4	D142V	TYLLSLIFKAAVVAGFLYIFHRLYK
11	131	TNNT3-1	V26A	EEEEEAQEEAAEAHEEEEKPRPKLT
11	132	TNNT3-2	V26A	EEEEEAQEEAAEAHEEDTAEEDAEE
12	133	TNNT3-3	V26A	EEEEEAQEEAAEAHEEGMRTQDFFH
12	134	HOXA3	L90Q	PPSQPPSLGEPPQHPPPPQAAPPAP
12	135	FEM1A	D257A	AGGEAQPGLPQEAPSTSQGCAQPQG
12	136	DMD-1	I106R	GNHKLTLGLIWNRILHWQVKNVMKN
12	137	DEAF1	A388V	NTAQQLKTLFEQVKHASTYREAATN

12	138	ZNF705E	F271Y	DNSGKAFSQSSGYRGNKIIHTGEKP
12	139	CABP1	H88P	AAAASGGSRAPRPGPARDPGLPSRR
12	140	CCER1	D10A	MTQTLDTREAPLNLGGGGGGGC
12	141	SMARCE1	M103L	LKLWEIGKIIGGLWRDLTDEEKQEY
12	142	UBE2Z	D107G	RTAPQCLLRIKRGIMSIYKEPPPGM
12	143	ARHGAP15	I159N	AKEKSSRKNVFQNTTVSGNEFLLQS
12	144	PCNT	S2667R	LESEQGKGRALQRQLEEEQLRHLQR
13	145	ZNF451	S853T	ERKLKQAINYSKTLDMEKGVENDLS
13	146	TNNT3-4	V26A	EEEEEAQEEAAEAHEEVHEPEEVQE
13	147	GLI3	P942A	KAKYAAATGGPPATPLPNMERMSLK
13	148	TNRC18	E60A	GLYPSYLHLNHLAPPSSGSPLLSQL
13	149	DISP1	E473D	GESMMNIYLDNFDNWNSSDGVTTIT
13	150	PRR35	S358P	LPKASPSLTRFCPRSSLPTGSSVML
13	151	ADAT1	A354 frameshift	QRALIGRCQNVSLLYQKASEFKN
13	152	RSAD1	N26T	AARAAQRRRRVETAGGSPSPEPAGR
13	153	TEX14	L105F	VHAAAFSGNQWIFSKLLDAGGDLRL
13	154	KCNH6	V110G	YYRKDASSFRCLGDVVPVKNEDGAV
13	155	USP36	P234A	KACLNGCAKCVLAHPPSQGFGHLFY
13	156	CTDP1	V43G	RLLEWRVAAGAAGRIGSVLAVFEAA
14	157	RPS15	R16Q	EQLMQLYSARQRQRLNRGLRRKQHS
14	158	TNNT3-5	V26A	EEEEEAQEEAAEAHEEVHEPEEKPR
14	159	SFT2D3	L153R	APSRPALLYMAARGATLFAALGLRS
14	160	CCM2L	G210A	ICSLDWRMGWGGAAAEARAGGGGGG
14	161	PKDREJ	L139Q	TWSVRLPRSPGRQAWAFRLRLLGPG
14	162	RPL34	Q114L	IVVKVLKAQAQSLKAK
14	163	TRAPPC3L	K92 frameshift	IGITFLKKRDEKNI
14	164	AKAP12	E24A	PEQPPEGSSTPAAPEPSGGGPSAEA
14	165	PEX6	L81R	PGPPQLLVSRALRRLLALGSGAWVR
14	166	WDR86	Q407P	AAGLIPRGPCRRPPRHPAAPRAPRP
14	167	TNNT3-6	V26A	EEEEEAQEEAAEAHEEVHEPEDDLK
14	168	DMD-2	I106R	GNHKLTLGLIWNRILHWQVLGDRWA
15	169	GDPD2	L441Q	DLPLLDKDRFLQPAQAGLKLLASS
15	170	AIM1L	Q562L	KGPGAPAASSPTLKEVVQGSGAPAA
15	171	SERINC2	D343A	LLCTLFISLRSSAHRQVNSLMQTEE
15	172	C11orf65	K110M	ANSPRNYAKLPAMHTSHNKNDHLQE
15	173	UHRF1BP1L	D844V	ESLILLSENLRKVVEAVTGSPASQT
15	174	PKP2	D26G	RTVLGQQILGQLGSSSLALPSEAKL
15	175	SMARCD1	Q46H	PGPPVRMGPAPGHGLYRSPMPGAAY
15	176	KCNK13	D384H	NGCPHQTSTLARHNEFSGGVGAFAI
15	177	SLC12A1	L553P	NNEPLRGYILTFPIAMAFILIAELN
15	178	SMAD6	F2L	MLRSKRSGLVRRLW
15	179	SSTR5-AS1	P39T	MRLRTPKPGRPRTRPHAAADPRPRP
15	180	YBX3P1	Q41P	PATKSRVGSGAPPAAAPAPAALVAG
16	181	MSLNL	L326Q	GANLWASANCSLQQGFWCQPASQLP
16	182	SARM1	C518S	TAAREMLHSPLPSTGGKPSGDTPDV
16	183	RUNDC3A-1	F73L	FAAILEQILSHRLKACAPAGPVSWF
16	184	RUNDC3A-2	F73L	FAAILEQILSHRLKGPVSWFSSDGQ
16	185	HID1	S675R	WREQRRPSTSSARGQWSPTPEWVLS
16	186	OSBPL1A	A216D	LKNKNDQKPLDLDQGAEMKHILVGN
16	187	NLRP13	C270Y	FQQRFSYVFYLSYHKIRYMKETTFA
16	188	KANK3	L224R	EQVRALRAEKARRLAGRAQPEPDGE
16	189	GRB14	R48Q	AHDLAPAPWLHAQALLPLPDGTRGC
16	190	SPEG	L2525Q	AQAGATTPSAESQGSEASATSGSSA
16	191	TRPC4AP	Q56L	GQLTGRGLVRAVLFTETFLTERDKQ
16	192	MPST	S9P	MAEPGSREPETRARSPSVAAM
17	193	ZNF662	E17A	ALASGTRLGLVLALLPGQPALPRAR
17	194	ADAMTS2	23-24 deletion	LLCPALLLLLLLPPPLLPPPPPPAN
17	195	AMZ1	E287A	LRCLMQGALSLDAALRRPLDLCPIC
17	196	PDLIM2	H284P	PASRALATPPKLPTCEKCSTSIANQ
17	197	NKX3-1	E61A	QRQRDPEPEPEPAPEGGRSRAGAQN
17	198	CEBPD	S191R	EKSAGKRGPDRGRPEYRQRRERNNI
17	199	SGK223	L1188H	PCSSAAPPAGGTHSPAAGPASPEGP
17	200	ZNF462	P582T	PPHQVPPQPQTQTPPTQQPQPPTQA
17	201	ARID1A	L1051 frameshift	TNLPAVGRKPLDLLSPLCVCEGDWWIDSGQQEQ KMAGTCNQPQCGHIKQCCQLLEKAVYPVSLCL

Supplementary Table 2. 4095 tandem minigene (TMG) library

TMG ID	Minigene ID	Gene name	Position and change of mutated amino acid	Mutated amino acid sequence
1	1	KLF5	P309A	ATYFPPSPPSSEAGSPDRQAEMLQN
1	2	ZC3H3	T840P	HGPRKPSASQRPPRQTPSSAALTAA
1	3	WRNIP1	L12I	MEVSGPEDDPFISQLHQVQCPVCQ
1	4	SENP2	G262R	LKTKGWGEEQNHRVKTTQFVPKQYR
1	5	UGT8	K128N	LMVGNHALIQGLNKEKFDLLLVDPN
1	6	WWC1	K671N	ATRIQIALKYDENNKQFAILIIQLS
1	7	PEX1	Q79E	SDQGENVAEINREVGQKLGLSNGGQ
1	8	KRAS	G12D	MTEYKLVVVGADGVGKSALTIQLI
1	9	ROBO2	P929T	GDPSYPWLADSWTATSLPVNNSNSG
1	10	PGBD1	E656K	VRATGTIRENRTKKCPLMNVEHMKK
1	11	PLEC	L3602M	KGLVLRQHGIRLMEAQIATGGIIDP
1	12	ESRRA	P173L	YKRRPEVDPLPFLGPFPAGPLAVAG
2	13	LIX1L	G121W	QMKQSRGADLKNWALVVYEMVPSNS
2	14	TACC2	A677T	EEDPVLPPVPDGTGEPTVPEGAIWE
2	15	KDM2A	A507S	LLEELANSDPKLSLTGVPIVQWPKR
2	16	KCNH2	E435D	VFTPYSAAFLLKDTEEGPPATECGY
2	17	PCSK5	R1302H	FFFLLRSKGECHHSCPDHYYVEQST
2	18	KCNN3	D711Y	SVAVGTTHTPISYSPIGVSSTSFPT
2	19	DNAJB6	N248T	DALAEERMRRGQTALPAQPAGLRPP
2	20	FAM109A	Q151P	QSHPCLLPQPRSPPCPCPAGPVPSR
2	21	DDX60	L101I	DAEYAYFNFPELISLRTALILHLQK
2	22	DUOX2	K529N	YWFENTRNGLFSNKEIEDIRNTTLR
2	23	COL17A1	W455G	GAGGGPWGPAPAGCPCGSCCSWWKW
2	24	WISP1	M32K	VLATALSPAPTTKDFTPAPLEDTSS
3	25	AGAP3	H576P	SSSPKLDPPPSPPSNRKKHRRKKST
3	26	ZC3H3	T493P	ALRGKSSPVLKKPPNKGLVQVTTHR
3	27	C8orf82	V165G	RLYHPAPERAGGGGLVRSALAFELS
3	28	LINC00115	L16H	VFHRPHPRPSWPHRAALGFGRRQSS
3	29	TRIM11	H123L	CAACERSGEHWALRVRPLQDAAEDL
3	30	CPEB3	R544L	DPRKTIFVGGVPLPLRAVELAMIMD
3	31	PRB1	R558L	PQGPPPQGGNRPLGPPPPGKPQGPP
3	32	PRG4	L1016F	AGGAEGETPHMLFRPHVFMPEVTPD
3	33	HLA-DQB1	S229R	TVEWRAQSESAQRKMLSGVGGFVLG
3	34	ZNF717	V114L	GICHRSLVELQEL
3	35	MUC4	D1125Y	VSTGHTTPLPVTYTSSASTGHATSL
3	36	MCHR2	N116T	PLCTIITSLDTCTQFACSAIMTVMS
3	37	OBSCN	R4668W	WHKGMERIQPGGWFEVVSQGRQQML
4	38	DSPP	S845R	SNSSDSSSDSSNSRDSSDSSDSSDGS
4	39	ADAMTSL2	R420W	EQAGGGACEGPPWGKGFRDRNVTGT
4	40	RSAD2	F248L	QIKALNPVRWKVLQCLLIEGENCGE
4	41	AEBP1	P387R	EEWTPTEKVKCPRIGMESHRIEDNQ
4	42	AKAP6	D1033Y	EALKKGGVLLPNYLLEKVDSINEKW
4	43	TMEM132D	T394M	SSSMGLMEGRGTMTDRSILQKKKGQ
4	44	DPY19L2	S598P	FEKVIFGILTVMPIQGYANLRNQWS
4	45	ARL14EPL	A65S	DFNPETRQQKKKSRMSKMNEYFSTK
4	46	SPTB	V445M	MRETWLSENQRLMAQDNFGYDLAAV
4	47	PPM1K	V194M	NCAWSAALDLEPMDTICGASVEREI
4	48	TTN	D4157E	SWFKDGKEIAASERYRIAFVEGTAS
4	49	IL17F	R72H	GIINENQRVSMSHNIESRSTSPWNY
5	50	FRMPD1	T1562M	SVKLLARQCTALMAAVFCLTQKFRA
5	51	LRFN5	R516H	GCIQFTTEQDYVHCHFMQSQFLGGT
5	52	ANKRD18B	S397L	NEEMITKKVAQYLQQLNDLKAENAR
5	53	NPC1L1	R242C	GSGIQPLNEGVACCNESQGDDVATC
5	54	PCDHB3	E320K	NFEAINSYEVDIKAKDGGGLSGKST
5	55	LRRC4	S147Y	LTVIPSGAFEYLYKLRELWLRNNPI
5	56	MAGI2	R136C	ELQQIIRDNLYLCTVPCTTRPHKEG
5	57	LMOD1	P521T	PKPSPQPSPKPSTKNSPKKGGAPAA
5	58	ZNF853	V457M	ELMVLPAVAAPAMVAIPGPAGSAAL
5	59	TENM3	V17F	APDPGNLAALGSFPHGHSEGAPRQE
5	60	MYO1G	D488V	GTITDRIFLQTLVMHHRHHLHYTSR
5	61	SYT7	M256L	TTSQSLGQLQAHLASAPGPNPRAYG

Supplementary Table 3. 4112 tandem minigene (TMG) library

TMG ID	Minigene ID	Gene name	Position and change of mutated amino acid	Mutated amino acid sequence
1	1	STS	V215 frameshift	LNCLGLLHVPLGVFFQPSLPSSPNPDPFLGLPSLLP APELLHDEELRDHSAAHVL
1	2	HEPH	S881R	TLFTVFSRTEHLRPLTVITKETEKA
1	3	POLA1	D105Y	EIFDDDLEDDALYADEKGKDGKARN
1	4	COL4A5	G902R	GFPGTKGEMGMMRPPGPPGPLGIPG
1	5	SEC23B	Q152 frameshift	EEDDLQALKESLRCP
1	6	USP34	Q1495K	SWSCKFVAAGGLKQLLEIFNSGILE
1	7	RBM41	G210D	MKKRLEEFQLMRDEPFASHSLVSAT
1	8	NAA35	D139Y	QTVFTCLYIHNPYFIEDPAMKAFAL
1	9	VPS13A	G1333V	SGVTTNASHHSGVATVVTAAVVEVH
1	10	TJP2	G904V	DSRLISDFEDTDVEGGAYTDNELDE
1	11	KIAA2026	V1464A	TRSEATAATNGDAISGTPVQKLMLV
1	12	SETX	G2567R	LWDPQPSSPQHPRATPPTGEPGFPV
1	13	DNM1	P756L	DINTTTVSTPMPLPVDDSWLQVQSV
2	14	C5	G836R	FLEMNIPYSVVRREQIQLKGTVYNY
2	15	STK3	V220L	TVIGTPFWMAPELIQEIGYNCVADI
2	16	CRISPLD1	T12S	MKCTAREWLRVSTVLFMARAIPAM
2	17	PRKDC	A1942S	NQLLERRRLYHCSAYNCAISVICCV
2	18	OPLAH	P869S	ADIGGITPGSMPSHSTMLQQEGAVF
2	19	OPLAH	P143L	RRHRGHHTRLHALPLHHAATGGCRL
2	20	PLEC	E4084K	AQIATGGIIDPEKSHRLPVEVAYKR
2	21	RIMS2	R479T	RKTKREKMETMLTNDSLSSDQSESV
2	22	DCAF13	E192G	YDPALHPFEVPRGYIRALNATKLER
2	23	PTCD1	I625N	MKQNRVPVNEVVNRQLEFAAQYPPT
2	24	ARPC1A	S201I	MPFGQLMSEFGGIGTGGWVHGVSFS
2	25	TRRAP	A2809T	MDKAKKEHERSNTSPAIFPEYQLWE
2	26	DLX5	E204Q	IKKIMKNGEMPPQHSPSSSDPMACN
3	27	CALCR	R454P	QFKIQWNQRWGRPPSNRSARAAAAA
3	28	AKAP9	V1990D	RQKEAMKAEAGPDEQQLLQETEKLM
3	29	STEAP2	P260T	YKIPIEIVNKTLTIVAITLLSLVYL
3	30	DMTF1	R569K	TSDNVTVQCHTPKVIIQTVATEDIT
3	31	SUMF2	K308N	SAIPSSRASASGNNFPFPVSHPSVA
3	32	ADCY1	Q1019H	SRMDSTGVQGRIHVTEEVHRLLRRC
3	33	TAX1BP1	M527T	DFDIVTKGQVCETTKEIADKTEKYN
3	34	CDCA7L	A216T	MPDFFPVRTPTSTSRKKTVRRAFSE
3	35	REPIN1	G285V	PRGRPAVTAPRPVGDAVDRPFQCAC
3	36	ZNF282	WD339-340CY	LSRIKQEEHQCVCYQQDLADRDIPTD
3	37	KIAA1244	C1357F	QVFANAATSYIMFLMKFVKGLGEVD
3	38	MET	G643V	NKHFNMSIIISNVHGTTQYSTFSYV
3	39	TRIP6	G210V	PHFPLPGRGEVWVPGYRSQREPGPG
4	40	ORC3	G477D	YYLKNEALKSEEDCIPNIAPDICIA
4	41	SYNCRIP	R412T	RGGRGARGAAPSTGRGAAPPRGRAG
4	42	DST	D31A	LQAYEDVLERYKAERDKVQKKTFTK
4	43	ZNF445	G491 frameshift	HTVGVSFKCSDCEGLSVIAPILRIIRDFTLKRKHLN VGCVGKPSGGVPTVRGMRKFTLE
4	44	PSTPIP2	S330C	PDDPNYSLVDDYCLLYQ
4	45	DST	L1606F	DTSATHREVQRKFDHATDRFRSLYS
4	46	GCLC	R128L	EFNTVEANMRKRLKEATSILEENQA
4	47	TMEM63B	N186Y	LPVNFSGDLLENYAYSFGRTTIANL
4	48	CNPY3	V55L	VRLPSKCEVCKYLAVELKSAFEETG
4	49	ULBP3	A223S	PPTMAPGLAQPKSIATTLSPWSFLI
4	50	HIVEP2	D20N	LGQKATSRSGETNKASGRWRQEQSA
4	51	SLC37A3	R9I	MAWPNVFQIGSLLSQFSHHHV
4	52	SLC39A1	T130A	MGFFLVLVMEQIALAYKEQSGPSPL
5	53	PAK1IP1	D304H	TNARLTCLGVWLHKVADMKESLPPA
5	54	AIM1	E1034G	GKVVIYSEPDVSGKCIEVFSDIQDC
5	55	ZFYVE16	F1118L	NEDTIPKDIFRLLITIYKDALKGKY
5	56	SKP2	H171L	DGTLQLLKEALPLLQINCSHFTTIA
5	57	RAI14	G402C	TQTDLGPSLGKPCETSPPDSKSSPS
5	58	TARS	V587L	YVSHDGDDKKRPLIVHRAILGSVER
5	59	GOLPH3	G32V	RNAADKERAAGGVAGSSEDDAQSRR
5	60	KDM3B	G334V	GPWKGGNASGEPVLDQRAKQPPSTF
5	61	SNX2	F369L	EVGRFEKERVKDLKTVIIKYLESLV
5	62	DMXL1	R1069K	LAVAYKQPASNSKSSQDFVMHVSIF
5	63	MARCH6	R690S	AACGLYVCWLTISAVTVMVAWMPQG
5	64	NFXL1	C486F	RDCQKHQCRRKCFPGNCPPCDQNCG
5	65	ATP10D	E853A	TLCIAKKVMSDTAYAEWLRNHFLAE

6	66	ATP8A1	K27N	AEGYEKTDDVSENTSLADQEEVRTI
6	67	HTT	D1056Y	GWHCGVPPLSASYESRKSCTVGMAT
6	68	GPR125	G320R	ALTISNIQAGSTRNWGCHVQTKRGN
6	69	FAM198B	K399E	RLDTNCCGFRPREEDACVQNGLRPK
6	70	KIAA0922	P1439L	ENGVPCVIQESALVHNSFIDWSATC
6	71	NAA15	I658V	KLAKVETPLEEAVKFLTPLKNLVKN
6	72	FAT4	S1402P	ITAKDQGRPPRSPTMSVVIHVRDFN
6	73	ALPK1	N1103S	YVTEFNKRLYEQSIPTQIFYIPSTI
6	74	ALPK1	E554D	SDAFRVSLDQDVDTETEPSDYSNGE
6	75	SETD5	E650K	AQQAELSQAALEKGGSNSLVTPTEA
6	76	PPM1M	G128E	RLLGTLAVSRGLEDHQLRVLDTNIQ
6	77	COL7A1	R185L	VGIKNADPEELKLVASQPTSDFFFF
6	78	DST	D209A	LDPAERAVLRIAAERDKVQKKTFTK
7	79	PARL	S163G	FRKEINKWWNNLGDGQRTVTGIIAA
7	80	ITSN1	G783S	GEWVDESQTGEPSWLGGELKGKTGW
7	81	PARL	A284S	GMILGWKFFDHASHLGGALFGIWYV
7	82	TTC14	S134I	DIERGDIVIGRIISIREFGFFMVLI
7	83	TBC1D5	G363C	VWDALFADGLSLCLVDYIFVAMLLY
7	84	ZBTB38	Y709F	SNSSENAASVISFSGSAPSVIVHSS
7	85	DNAJC13	L877I	PKVNMKCLCLQAIAIVYGRCHEEIG
7	86	GOLGB1	V1076L	EIYLKQTISEKELELQHIRKDLEEK
7	87	GOLGB1	V3175M	LRKLLEEERDQRMAAENALSVAEEQ
7	88	WDR52	E584K	KPHTACVTALAYKRDGEILATGSKD
7	89	RIBC2	T204S	DETAKHLQKLESSTRKAVCASVKDF
7	90	MKL1	S318F	PPVHSLSTTNSSFSSGAPGPCGLAR
7	91	SF3A1	A771S	YEGIFIKDSNSLSYYNMANGAVIHL
8	92	MTMR3	R728S	EGKEDPLLEKESSRKTPEASAIGLH
8	93	GUCD1	L118M	DSEFERALQKLQMTRSIWTIDLAYL
8	94	BCR	K293T	SIYVGGMMEGEGTGPLLRSQSTSEQ
8	95	TRPM2	V1477L	HACDSGASIRWQLVDRRIPLYANHK
8	96	C2CD2	G43W	PEMAVNIQPKALWEDQVAETSAMSD
8	97	PARL	S163G	FRKEINKWWNNLGDGQRTVTGLC
8	98	ITSN1	G741S	IVMVDESQTGEPSWLGGELKGKTGW
8	99	ATP5J	F17L	GGISMILQRLFRLSSVIRSAVSVHL
8	100	CSE1L	L340M	PHYKNLFEDQNTMTSICEKVIVPNM
8	101	SAMHD1	D435Y	TDNIFLEILYSTYPKLKDAREILKQ
8	102	NDRG3	G128V	SLKSIIGIGVGAVAYILSRFALNHP
8	103	AAR2	S187C	GQNLPRCGIECKCYQEGLARLPEMK
8	104	RBM12	H325Y	FSAMENDVRDFFYGLRVDAVHLLKD
9	105	CEP250	L1519P	EENHHKMECQQKPIKELEGQRETQR
9	106	PSMD10	Stop loss	IQDTEGNTPLDTWPVMRREWKKQNCWCPKEQV FTLRIKKKRHPCKWPKVAWV
9	107	PTCD3	G213V	QEPSTDYHFQQTVQSEALEEENDET
9	108	CDTN1	A419T	QQRERLQEELSQTESTIDELKEQVD
9	109	EHBP1	T165P	SLSCIFLREGKAPDEDMQSLASLMS
9	110	MID2	C734S	PQESPYVSGMKTSH
9	111	BIRC6	M2155I	DNLLSPLQPQLPIHRRTEGVLDIPM
9	112	LRRFIP1	T437A	EAAVTQVEEQAGAVASCPLGHSDDT
9	113	USP40	G1060E	RTDRQPLREYKLERRIEICLEPLQK
9	114	CHPF	E313K	SHLELSPGEPVQKGDPHFRSALTAH
9	115	TNS1	Y366F	VGHTQGPLDGSLFAKVKKKDSLHGS
9	116	ABCA12	L1290I	TYGMAAPWYFPIIPSYWKERFGCAE
9	117	ZDBF2	G1986W	QDDRKTKKKVKIWTVEFPASCTKVL
10	118	ALS2	M689I	KDSYLALVDKNIIGYIASLHELATT
10	119	TRAK2	R795Q	HSPCPSPLPFEPQVHLSENFLASRP
10	120	SPATS2L	T27N	SGTLLWIPRAYSNRSKMAELNTHVN
10	121	WDR75	R228C	DCIASGHMDGKICLWRNFYDDKKYT
10	122	DHRS9	D286G	SLFPKTHYAAGKGAKIFWIPLSHMP
10	123	TTC21B	A1012S	GKLEDVPRFFSMSEKRNSRAKLEPG
10	124	GCA	A62P	AGDSVYTYFSAVPGQDGEVDAEELQ
10	125	NBAS	C144S	KPQWRRVAWSYDSTLLAYAESTGTV
10	126	NBAS	V2177F	ESSHHEAEFQHLFLLLQAWPPMKSE
10	127	NEB	D3056G	QLGHHIGARNIEGDPKMMWSMHVAK
10	128	NEB	K6202M	VNSELKYKETYEMQKGHYLAGKVIG
10	129	PTPN18	G317L	GRVPADQSPAGSLAYEDVAGGAQTG
10	130	POTEF	WD779-780CY	LKYPMEHGIITNCYDMEKIWHHTFYN
11	131	ZNF584	C244F	HQKVHTGIKPFKFSDCGKTFNRKDA
11	132	ZNF587B	R412G	KGNLILHQHGHTGKRPYMCWECGKL
11	133	PTPRH	S44I	PGRNLTVETQTTISISLSWEVPDGL
11	134	POLD1	L546V	EMARVTGVPLSYVLSRGQQVKVVSQ
11	135	NR1H2	Q164L	PLGLGLPLVDLRLAARAPGKARVSS

11	136	CYTH2	A81S	ENELLQNTPEEISRFLYKGEGLNKT
11	137	ZNF114	Y342N	EECGKVIRESSKNTHIRSHTGEKPY
11	138	KCNN4	G50V	GLMVLHAEMLWFVGCSWALYLFLVK
11	139	PALMD	A45T	EDKLKHQHLKKKTLREKWLLDGISS
11	140	PHLDB3	A232S	QEMREQLDVAQRSYEDLEFQQLERE
11	141	C19orf47	G234V	LQYAGVLKKLGRVPAKASPQPALTV
11	142	ACTN4	G68S	FTAWCNSHLRKASTQIENIDEDFRD
11	143	YIF1B	DP19-20EL	KRRIPVSQPGMAELHQLFDDTSSAQS
12	144	SUGP2	E905D	EEEDEDDEDGGEDAPAPGGAGKSEG
12	145	COLGALT1	M233T	RKRDRRGCFAVPTVHSTFLIDLRKA
12	146	ANKRD12	T1727I	KVELEENAEDDKIENQIPQRMTRNK
12	147	RBFA	G232V	LWSTKGGKIKGSVAWCGRGRWLS
12	148	ZNF516	R424L	PVNSYQAWQLATLGKVAEPAEYLKY
12	149	PIGN	P579L	MLTAGLTAFAAWLFLTRLWTRAKMT
12	150	CCDC53	V47A	VHTVQFLNRFSTACEETGSHFVTHA
12	151	PIK3C3	R572K	QRESGNRKKKNEKLQALLGDNEKMN
12	152	ROCK1	E524V	VENEVSTLKDQLVDLKKVSQNSQLA
12	153	RNF213	A1254S	EPLSEPKEDQEASELLSEPEEESER
12	154	EIF4A3	D169Y	QHVVAGTPGRVFYMIRRRSLRTRAI
12	155	POLR2A	G816V	HRTLPHFIKDDYVPESRGFVENSYL
12	156	TTYH2	Y297C	TEGQISTEVTRYCLYCSQSGSSPFQ
13	157	WIPI1	R212L	KLASASEKGTVILVFSVPDGQKLYE
13	158	RABEP1	K427N	GSLQSKALGYNYNAKSAGNLDESDF
13	159	SRCIN1	V151L	MREQVGGWTVDPLCLLSSLCSHLHG
13	160	AP2B1	Q59H	VSSLFPDVVNCMHTDNLELKKLVYL
13	161	ZNF830	W365C	DEGELQDLLSQDCRVKGALL
13	162	SSH2	H487Y	ASKQRHNKLWRSYSDSDLSDHHEPI
13	163	SPECC1	E745D	VVANDIKCEAQQDLRTVKRKLLEEE
13	164	ALDH3A2	K447 insertion KV	ANKLRYPPNSQSKVVDWGKFFLLKR
13	165	FAM83G	A767S	LPDPGSPRLAQNSRPMTDGRATEEH
13	166	FANCA	E1293K	LPKAFHVCAAILKCLEKRKISWLAL
13	167	ZNF778	G641V	SSHLIVHIRTHTVEKPYICKECGKA
13	168	NUDT7	V111L	LRPHQVEVVCCLLPCLIDTDTLITP
13	169	DST	D24A	WNAKLVGLMCCMAERDKVQKKTFTK
14	170	NUDT7	V111L	LRPHQVEVVCCLLPCLIDVRVS
14	171	AMFR	V440A	RIASWLPSFSVEAMHTTNILGITQA
14	172	TAOK2	R555P	ARRHQAIGEKEAPAAQAEERKFQQH
14	173	ETFA	D236Y	GLKSGENFKLLYYLADQLHAAVGAS
14	174	UBL7	D352Y	PQLQQLRDMGIQYDELSLRALQATG
14	175	SPG11	303-304 deletion	QHPGHLLCERILDLPIQGPKGVDED
14	176	SPG11	A340T	AKFSFQIDRSWKTQLSSLNETIKNS
14	177	MGA	R2543S	QEFLPKKISGDMSGIQYKWKESESR
14	178	ZNF770	C16S	LCKLNPSYLKITSGKRSKQITPTYY
14	179	YLPM1	G1480W	EQKEQLQKMKDFWSEPQMADHLPPQ
14	180	NUMB	S288P	SFRGFPALSQKMPPFKRQLSLRINE
14	181	SLC39A9	A87S	SETHNVIASDKASEKSVVHEHEHSH
14	182	SYNE2	R3216T	CSIKAVTAIEKQTEENSSEASDVET
15	183	DLGAP5	K156N	KAIPSSVRITRSNAKDQMEQTKIDN
15	184	C1orf43	E80D	YRMKALDAIRTSDIPFHSEGRHPRS
15	185	SLC22A15	Q517L	GEEALSLQALDPLQCVDKESSLGSE
15	186	AHNKA2	P588N	TEGQIRMPKFKINSLGWSPSKHTKT
15	187	AHNKA2	G844W	KFKMPKFKMPSFWVSAPGKSMEDSV
15	188	DYNC1H1	V1762A	VVLSAQIAWSENAETALSSMGGGGD
15	189	DYNC1H1	R223L	PMITNVAKQCYELGEKPKVTDFGDK
15	190	DIAPH3	E507D	ICIDQAKLEEFEDKASELYKKFEKE
15	191	HSPH1	G204E	DEKPRIVVFVDMEHSAFQVSACAFN
15	192	CRL1	C78F	LSVEEQLSLISGFPNIQEAVEGAMH
15	193	TMTC2	Q634R	YEEALSVYKEAIRKMPRQFAPQSLY
15	194	RAB21	Y89C	GQERFHALGPIYCRDSNGAILVYDI
15	195	CDCA3	P175R	WKPNSSKVLGRSRLTILQDDNSPGT
16	196	PIGV	M227I	HSQCQGFFSSLTILNPLRQLFKLMA
16	197	TENC1	P1059H	SQMPWLVASPEPHQSSPTPAFPLAA
16	198	FGD4	Q686L	LYMYGAPQDVRALATIPLLGYVVDE
16	199	ETNK1	F202L	QAHGCAPQLYCTLNNGLCYEFIQGE
16	200	ZNF268	L195I	TFGKLCLLSTKYISRQKPHKCGTHG
16	201	RILPL1	G219R	DLRHRVTVVEAQRKALIEQKVELEA
16	202	TCP11L2	G477C	GLAVIQQELEALCSQYANIVNLNKQ
16	203	CCDC53	V47A	VHTVQFLNRFSTACEEKLADLSLRI
16	204	NUDT7	V111L	LRPHQVEVVCCLLPCLIDRWGSRYV
16	205	WRAP73	V208L	WDTCLEVRILNHLTWKMITEFGHPA

16	206	GAS2L3	C110F	EESGNFPMRKVPFKKDAASGSFFAR
16	207	ANKS1B	K362N	HTISDHYLDNLSNISEEELGKNGSQ
16	208	PCF11	A926S	GLRFEGGHGPSGSAIRFDGPHGQPG
17	209	RSF1	D681N	LKEDSEFTKVEMNNLDNAQTSGIEE
17	210	DGAT2	I157M	VKTHNLLTTRNYMFGYHPHGIMGLG
17	211	BBS1	A11S	MAAASSSDSDSCGAESNEANSKW
17	212	SF3B2	Q873 frameshift	SDMVAEHAAKQKQKKTESSAPGQPWGQQEI
17	213	EIF1AD	E123V	EKHNNRNRQTQPVLPAEPQLSGEES
17	214	TMEM216	I41T	WYNATYFLLELFTFLYKGVLLPYPT
17	215	TMEM109	A191S	VPDPSTRALLLLSLLILYALLSRLT
17	216	OSBP	R320T	LEQLAKQHNHLETAFRGATVLPANT
17	217	PTPRJ	G75V	STAESFHKQNGTVTPQVETNTSEDG
17	218	ZNF408	G505V	CGRAFRQRGNLRVHLRLHTGERPYR
17	219	KIF18A	E202Q	HGLTLHQPKSSEQILHLLDNGNKNR
17	220	ARHGAP32	D827N	GASFLDSPGYSKNKPSANKKDAETG
17	221	MUC5B	L1907I	ATPSSTPGTTWIITKPTTTATTTAS
18	222	DIXDC1	I23F	SPSPIHSAKSESFITQSEEKADFVI
18	223	FDX1	D132H	EDHIYEKLDAITHEENDMLDLAYGL
18	224	ADM	H122Q	CRFGTCTVQKLAQQIYQFTDKDKDN
18	225	PFKFB3	V371F	RQENVLVICHQAFLRCLLAYFLDKS
18	226	FAM107B	R43S	QDLHRELLMNQKSGLAPQNKPELQK
18	227	DNTTIP2	A303T	ETKQNCKDLDEDTNGITDEGKEINE
18	228	LRRC8B	L545F	GFQDLKNLRTLYFKSSLSRIPQVVT
18	229	PKN2	D123Y	DPEDITDCPRTPYTPNNDPRCSTSN
18	230	HS2ST1	E334 frameshift	EQFQFIRAHAVREKRWRPLHPRTKLFL
18	231	NOL9	T343S	DYLECDLGQTEFSPPGCISLLNITE
18	232	INADL	R1506M	RNSSHEEAITALMQTPQKVRLVVYR
18	233	WRAP73	V253L	GSYDGKVRILNHLTWKMITEFGHPA
18	234	CCDC53	V47A	VHTVQFLNRFSTACEETGSCCITEA
19	235	MFSD5	T549I	VRHDAELRVPSPIEEPYAPEL
19	236	ASAP3	A244P	AAQSLFPFIEKLPASVHALHQAQED
19	237	HEATR1	M619I	VVINNDDTESAEIKIAIYLSKSGIC
19	238	TARBP1	V544L	CYLLQTAMNLLDLEKVSLSDVSTFL
19	239	OBSCN	G4333R	FTQDLKTKEASERATATLQCELSKV
19	240	OBSCN	A2103T	QGTATMEVQLSHTDVDGSWTRDGLR
19	241	CCDC42BPA	V1277L	QLVAVISGRNRHLRLFPMSALDGRE
19	242	IARS2	V443F	QNKAVLEEGTDVFIKMLQTAKNLLK
19	243	GPATCH2	K416N	QLLRDNRAERGHNKNCSVRTASRQT
19	244	ZBED6	S355I	LSDTLHGEKSTGIQDLTAEDLSDSD
19	245	OTUD3	H146D	AGNDAIVAFARNDQLNVVIHQLNAP
19	246	SWT1	S257Y	QKLVEENVFNIDYNNSKTKQEEREY
19	247	CACNA1E	R1297L	VDHEKNKMEVKGLEWKRHEFHYDNI
19	248	CEP350	R1167L	QHSSGAQSAASSLSSTSSKGKKGKK
20	249	CRP	G97V	WSKDIGYSFTVGVSEILFEVPEVTV
20	250	DNAJC16	G93I	SNEEKRSNYDQYIDAGENQGYQKQQ
20	251	ARHGEF2	V500M	KDVLVLLMTDVLMFLQEKDQKYIFP
20	252	ASH1L	E2901K	ATANVSEGEKKTKESSQEPQSTCTP
20	253	UBAP2L	G640V	QLQTTQSVEGATVSAVKSDSPSTSS
20	254	NIN	K1192 frameshift	NPSGTMNPTEQEN
20	255	C1orf43	K62N	ARLLQLETQGNQNIPFHSEGRHPRS
20	256	TPM3	V59L	NLRARVDINCSPLGTFWALINKVSA
20	257	SIRT5	H286Q	PCGTTLPEALACQENETVS
20	258	SLC39A1	H86R	HGLLPGPGDGADRTGLQGAVRAVTS
20	259	SELENBP1	P188S	IVQTLSLKDGLISLEIRFLHNPDAA
20	260	TBX15	P453T	PSNGAFGERQYLTSGMEHSMHMISP
20	261	FKBP3	N158 frameshift	DTNIQTSAKKKKKCQAFKF
20	262	RHOC	G144C	AKMKQEPVRSEECRDMANRISAFGY
20	263	KCNN4	G50V	GLMVLHAEMLWFVGCSTHFG

Supplementary Table 4. Predicted high-affinity peptides from 4112 TMG-9.

Short peptide pool (SPP) ID	Peptide ID	Peptide sequence	Allele	IEDB		NetMHC	
				% Rank	IC50 (nM)	% Rank	Affinity (nM)
1	1	AVAYKQPASNSK	HLA-A*03:01	0.9	129		
1	2	AWSYDSTLL	HLA-C*04:01	0.4	1862		
1	3	CAPQLYCTL	HLA-C*12:03			1.1	611
1	4	DNIFLEILYSTY	HLA-B*18:01	0.8	362		
1	5	DSYLALVDKNIIGY	HLA-A*25:01/B*18:01	0.5/0.8	1202/318		
1	6	DVLMFLQEKDQKY	HLA-A*25:01	0.2	267		
1	7	DVLVLLMTDVLMF	HLA-A*25:01	0.8	2258		
2	8	DVPRFFSMS	HLA-A*25:01			1.7	11595
2	9	EILYSTYPK	HLA-A*03:01			2	1050
2	10	EILYSTYPKL	HLA-A*25:01	1	4079		
2	11	ETAFRGAT	HLA-A*25:01	0.7	4127		
2	12	ETAFRGATV	HLA-A*25:01			0.5	4192
2	13	FADGLSLCL	HLA-C*04:01			0.7	6237
2	14	FADGLSLCLV	HLA-C*12:03	0.4	12		
3	15	FADGLSLCLVDYI	HLA-C*12:03	0.2	8		
3	16	FFSMSEKRNSRAKL	HLA-C*04:01	0.4	2262		
3	17	FLEILYSTY	HLA-A*25:01/C*12:03			0.8 / 0.7	6350 / 330
3	18	GEGTGPLL	HLA-C*04:01	0.7	3951		
3	19	GEGTGPLLR	HLA-C*04:01			1.2	7571
3	20	GMMEGEGTGPLL	HLA-C*04:01	0.6	3228		
3	21	ILYSTYPKL	HLA-C*04:01			0.4	4756
4	22	ILYSTYPKLK	HLA-A*03:01	0.15	12		
4	23	KLEDVPRFF	HLA-C*04:01/C*12:03			1.7 / 1.8	8834 / 1516
4	24	LALVDKNIIGYI	HLA-C*12:03	0.8	34		
4	25	LEILYSTYPKL	HLA-C*04:01	0.9	4417		
4	26	LFADGLSLC	HLA-C*04:01			1.1	7466
4	27	LLMTDVLMF	HLA-C*04:01			0.8	6544
4	28	LLMTDVLMFLQEK	HLA-A*03:01	0.7	78		
5	29	LMTDVLMFL	HLA-C*04:01			1.2	7671
5	30	LVDKNIIGY	HLA-A*25:01/C*04:01			1.4 / 1.1	10282 / 7427
5	31	LWDPQPSSPQHPR	HLA-C*04:01	0.8	3948		
5	32	LWDPQPSSPQHPRA	HLA-C*04:01	0.4	2578		
5	33	LYSTYPKL	HLA-C*04:01	0.8	4171		
5	34	LYSTYPKLK	HLA-C*04:01			1.9	9145
5	35	MEGEGTGPL	HLA-B*18:01/C*04:01			0.3 / 0.25	98 / 4161
6	36	MEGEGTGPLL	HLA-C*04:01	0.4	2251		
6	37	MEGEGTGPLLRSQ	HLA-B*18:01	0.8	207		
6	38	MMEGEGTGPLL	HLA-C*04:01	0.5	2899		
6	39	NIFLEILYSTY	HLA-A*25:01	0.3	613		
6	40	NIIGYIASL	HLA-A*25:01/C*12:03			0.03 / 1.5	270 / 1102
6	41	NIIGYIASLHEL	HLA-A*25:01	0.6	1701		
6	42	QWRRVAWSYDSTLL	HLA-C*04:01	0.2	967		
7	43	RATPPTGEP	HLA-C*12:03			1.8	1571
7	44	RATPPTGEPGFPV	HLA-C*12:03	0.7	24		
7	45	RFFSMSEKR	HLA-C*04:01			0.3	4630
7	46	RFFSMSEKRNSRAK	HLA-A*03:01	1	152		
7	47	RVAWSYDSTLLAY	HLA-A*25:01	0.5	1158		
7	48	SEPDVSGKC	HLA-C*04:01			0.8	6443
7	49	SEPDVSGKCIEVF	HLA-B*44:03/C*04:01	0.8 / 0.4	461 / 2314		
8	50	STYPKLDAREILK	HLA-A*03:01	0.3	40		
8	51	SYDSTLLAY	HLA-B*18:01			1.9	2617
8	52	TAFRGATV	HLA-C*12:03	0.2	5		
8	53	TAFRGATVL	HLA-C*04:01			0.8	6513
8	54	VAWSYDSTL	HLA-C*12:03			0.4	144
8	55	VAYKQPASNSK	HLA-A*03:01	0.9	750		
8	56	VLMFLQEK	HLA-A*03:01	0.9	57		
9	57	VWDALFADGLSLCL	HLA-C*04:01	0.4	2384		
9	58	WRRVAWSYDSTLL	HLA-C*04:01	0.4	1789		
9	59	WSYDSTLL	HLA-C*12:03	0.9	36		
9	60	WSYDSTLLA	HLA-C*12:03			1.3	917
9	61	WSYDSTLLAY	HLA-A*25:01/C*12:03	0.5 / 0.9	1535 / 35		
9	62	YLALVDKNIIGY	HLA-A*25:01	0.8	2969		
9	63	YSEPDVSGK	HLA-C*12:03			1.9	1687
10	64	YSEPDVSGKCI	HLA-C*12:03	0.6	25		
10	65	YSEPDVSGKCIEV	HLA-C*12:03	0.7	24		
10	66	YSEPDVSGKCIEVF	HLA-C*12:03	0.7	30		
10	67	YVGGMMEGEGTGPL	HLA-A*25:01	0.6	1753		

Supplementary Table 5. 4171 peptide library

Peptide pool ID	Peptide ID	Gene name	Position and change of mutated amino acid	Mutated amino acid sequence
1	1	KCNH6	R362H	MVAAIPFDLLIFHTGSDETTTLIGL
1	2	CD226	T157M	GKNVTLTCQPQMMWPVQAVRWEKIQ
1	3	BRCA1	N1384S	LCLPQSIYRSELSVYAFGEHILQIS
1	4	ZNF335	L26H	LAGCPAPYPAAKHFPSLFHAPQEEV
1	5	TP53	R209Q	YMCNSSCMGGMNQRPILTIITLEDS
1	6	MKNK1	G4C	MILCHCSLDLLGSSNP
1	7	ZHX3	R720C	SLEMPSSHILAECKVSPIKINLKNL
1	8	POLD1	P815L	EAADWVSGHFPSLIRLEFEKVYFPY
1	9	KRAS	G12D	MTEYKLVVVGADGVGKSALTIQLI
1	10	PIK3CA	R93W	EFFDETRRLCDLWLFQPFLKVIEPV
1	11	LCE1F	R87H	CCLSHHRRRRSHHHRPQSSDCCSQP
1	12	CTAG2	A60T	AGAARASGPRGGTPRGPHGGAASAQ
1	13	PDLIM3	A158V	PIGLYSTSNIQDVLHGQLRGLIPSS
1	14	ZNF181	I111N	FHSKSTLSEPQKNSAEGNSHKYDIL
2	15	TEX13B	R88H	VRFAHRQGQLQNHRVQWLQGFAKLH
2	16	UGT3A1	R377W	VRVVAKNYGVSIWLNQVTADTLTLT
2	17	PIK3CA	Q546E	ISTRDPLSEITEEEKDFLWSHRHYC
2	18	TMEM232	L76F	KEELLELARKIIFRCKRKLGLKTLG
2	19	TATDN3	L8F	MRAAGVGFVDCHCHLSAPDF
2	20	CHST11	T88M	VLHQMRRDQVTDMCRANSATSRKRR
2	21	SYNE1	R471W	LSLSLAQPLRSEWSGRDTPASVDSI
2	22	NME5	M5V	MEISVPPPQIYVEKTLA
2	23	SLAIN1	S292T	SFLQPPKPLSSLTTLRDGNWRDGCY
2	24	SCN5A	R1644H	RILRLIRGAKGIHTLLFALMMSLPA
2	25	PLCD3	R173C	QRWVRGLTKLRACLDAMSQRERLDH
2	26	VCAN	I663T	TEIELFPYSGDKTLVEGISTVIYPS
2	27	NOTCH3	H2227R	PGHGEEYPVAGARSSPPKARFLRVP
2	28	TRIM43	L73S	KMDFKTNILLKNSVTIARKASLWQF
3	29	LRP2	I4371V	FIEGSTTECDAAVELPINLPPPCRC
3	30	APOB	L1238F	VGSKLIVAMSSWFQKASGSLPYTQT
3	31	NHSL1	R436W	VIAIPTAQSAGQWESKSSGSSHARI
3	32	LOR	C103W	SGGGGSSGGGSGWFSSGGGGSGCFS
3	33	SIN3A	N520I	LGKFPELFNWFKIFLGYKESVHLET
3	34	CPAMD8	R1270H	LTAFVLKSFAQAHSFIFVDPRELAA
3	35	APC2	G322D	SGCLPLLLQILHDTEAAAGGRAGAP
3	36	ATP8B3	T1104M	VNFFMTLWISRDMAGPASFSDHQSF
3	37	AMY2A	R267Q	EPIKSSDYFGNGQVTEFKYGAKLGT
3	38	CROT	G565A	FTDPLFSKSGGGANFVLSTSLVGYL
3	39	KIF5C	R859C	VHKQLVRDNADLCCELPKLEKRLRA
3	40	GFRA2	F138V	RTDHLCRSRLADVHANCRASYQTVT
3	41	ARHGAP15	T53I	SKSMILTDVGKVIEPISRHRRNHSQ
3	42	APBB1IP	P270T	LEKEEKYAVFKNTQNFYLDNRGKKE
4	43	UBR7	P39 frameshift	NSKFKNLECKLLLTKQR
4	44	UBR7	P115 frameshift	NSKFKNLECKLLLFQMR
4	45	DLGAP4	D944N	KSKPAVSRDKASNASDKQRQEARKR
4	46	GRID2IP	T349M	VVSMLQGSGAMPMLVVEEGLVPFAS
4	47	AP4E1	A284T	SPKINLKYLGLKTLTYVIQQDPTLA
4	48	ZNF707	R167Q	QRCGRRPGRRERQKQRAVELSFICG
4	49	SVEP1	G1021R	NLEHFTCESCRIRSYQDEEGQLECK
4	50	ZDBF2	R12H	MQKRQGYCSYCHVQYNNLEQHLFS
4	51	VWDE	I68F	DHSLSPGWYRFLFLDRPAEMPTKCV
4	52	PPAPDC2	S21R	EGRPLGVSASSSRSSPGSPAHGGGG
4	53	FAM71A	S519F	KSGRSLWTTSSGFSKGLGRVSSFLR
4	54	TRPC4	L227I	DPFLTAFQLSWEIQELSKVENEFKS
4	55	SERTAD4	G64E	GAGPPLAGSHYREISNPITTSKITY
4	56	FANCM	A1052 frameshift	SGASCSKSRPHLLGHILLLDLRRKE
5	57	FANCM	A1052 frameshift	HLLGHILLLDLRRKEKEPVFL
5	58	UNC79	S516T	VQEQALLWLHVLTELDIMVPLQLLI
5	59	SIGLEC8	R68W	WTDSDPVHGYWFWAGDRPYQDAPVA
5	60	GALNT10	R30C	YEISFKVWMCGGCMEDIPCSRVGHI
5	61	GALNT10	R200C	FLLAMQVWMCGGCMEDIPCSRVGHI
5	62	OR2B3	F310L	KEAFKRLMPRIFLCKK
5	63	PTPLA	D142G	PPRPINFCIFGRGRVSPCWPGWSRT
5	64	PRMT3	I380S	SDLEFSSDFTLKSTRTSMCTAIAGY
5	65	MYT1	A359T	SCSSSPGVKSPDTSQRHSSTSAPSS
5	66	NADSYN1	I109M	SGGVDSAATACLMYSMCCQVCEAVR
5	67	ATR	N811I	ETDVKAVLGTLLILMEDPDKDVRVA

5	68	OR13C8	R122C	TECMILGTMALDCYVAICYPLRYPV
5	69	MAOA	R96W	RKFVGGSGQVSEWIMDLLGDQVKLN
5	70	LILRB4	R188W	RSPMDTFLLIKEWAAHPLLHLRSEH
6	71	OR5B17	A103P	SYSACAAQMFFCPVFATVENYLLSS
6	72	KCNS2	R418Q	ITLIFNKFSHFYQRQKQLESAMRSC
6	73	ZFAND4	S539R	NLQHFQEENFRKRSPQLEHTGVFLS
6	74	TMEM132B	A80T	KKGRGCSLQYQHTTVRVLTQFVAES
6	75	NBEAL2	P386L	YPHLQEVLQSHGLPTHRLLQELLNM
6	76	SLCO6A1	L376F	GLVLIPGGALGQFLGGVIVSTLEMS
6	77	GTF3C1	R777S	KKSDNKMGITPLSNYHPIVVPGLGR
6	78	MAGIX	N102K	HRPQVGDLVLHIKGESTQGLTHAQA
6	79	MAGIX	N178K	GRLEVGDLVLHIKGESTQGLTHAQA
6	80	STAT5B	H588N	GVMEVLKKHLKPNWNDGAILGFVNK
6	81	SPAG17	R225Q	VKKEDTIVPPNLQSRSWETFPSVEK
6	82	DNAH5	L3563V	KARKIPFGKNLNVSEMLIDAPTISE
6	83	PCDHB2	K74Q	AVRGARVVSKGKQMHLQFDRQTGDL
6	84	ATP2B2	V194M	LQSRIEQEQKFTMVRAGQVVQIPVA
7	85	FBXO11	L87V	LPDEVVLKIFSYVLEQDLCRAACVC
7	86	NUP107	E681D	KFLASKKHEAAKDVFVKIPQDSIAE
7	87	BPIFB3	E23D	LLLLWGLATPCQDLLETVGTLARID
7	88	PCDHB12	L327R	YSIIIQATDGGGRFGKSTVRIQVMD
7	89	GUCY2F	G434R	YTVDMEMELLRFRGTPIHFPGGRPP
7	90	KRT73	K137T	IQKVRAQEREQITVLNNKFASFIDK
7	91	GNAS	R314Q	PLMPRREEKYPLQGTDPLPPGQPQR
7	92	KMT2D	R5500W	FDKEDKIIIISSWRIPKGEELTYDY
7	93	DCHS2	I1155F	DRRLRSLTAQIVFLDVNDHNPTFIS
7	94	ARMC8	W355C	QQLRTSFQDHAVCKPLMKVLQNAPD
7	95	CCDC151	E73A	GAGKPSVHSQVAALHKKIQLLEGDR
7	96	CCDC151	E73A	GAGKPSVHSQVAALHKKIQLLESHS
7	97	OR52N5	F268L	AIIITYVPAFFTLFAHRFGGHTIPP
7	98	MICALL2	K439N	LGPDPAFGLGLGNGLLSLQGACGQQ
8	99	PLG	N428K	YPNAGLTMNYCRKPDADKGPWCFTT
8	100	C1orf101	I224M	LSDDERRSVAHVMLSRDGIVFLING
8	101	DDX21	K2N	HISGATSVDQRSVINSNVGFVTMIL
8	102	PASD1	L561V	MNMRGEKRRDKVNP
8	103	PCDHGA4	L798R	EKSEPLLITQDLRETKGDPNLQVSQ
8	104	PCDHGA4	L798R	EKSEPLLITQDLRETKGDPNLQQAP
8	105	ATF7	E345D	KLWVSSLEKKAEDLTSQNIQLSNEV
8	106	SMAD6	S133I	RTRSKIGFGILLIKEPDGVWAYNRG
8	107	CYFIP1	L253P	MKRLESKYAPLHPVPLIERLGTPQQ
8	108	MED15	Q241P	QQQQQQQQALQAPPPIQQPPMQQPQ
8	109	ZKSCAN4	G146S	EDIAQIPTHAEASEQEGRLQRKQKN
8	110	TERF1	Q275H	VVESKRTRTITSHDKPSGNDVEMET
8	111	CYFIP1	K606R	LESLIADKSGSKRTLRSSLEGPTIL
8	112	CD40	A17T	PLQCVLWGCLLTTVHPEPPTACREK
8	113	SLC11A1	R206C	SALVKSREIDRACRADIREANMYFL
9	114	PRR34	T53P	RARCPQSAHPAPPRGALTFWAPGSW
9	115	ZNF729	F416Y	YKCEECGKAFSQYSTLKKHKIIHTG
9	116	NRCAM	G612R	HLVVADVSDDDSRTYTCVANTTLDS
9	117	KIAA0020	K4N	MEVNGKKQFTGKSTKT
9	118	CILP2	E504G	EPLRFARILLGQGPIGFTAYQGDFT
9	119	ATM	W579C	NRSFSLKESIMKCLLFYQLEGDLEN
9	120	WNK2	H758P	PQPVVPLQPVPPPLPPYLAPASQVG
9	121	DLGAP2	S848T	QNMDPSAMPRPTTQDLAGYWDMLQL
9	122	TCF3	A162S	YPSYSGSSRRRASDGSLDTQPKKVR
9	123	OPRM1	S83R	NASNCTDALAYSRCSPAPSPGSWVN
9	124	LEFTY1	F78I	QRSHGDRSRGKRISQSFREVAGRFL
9	125	THSD7B	S993Y	CSSSCGIGVRIRYKWLKEKPYNGGR
9	126	CCDC88A	S583C	ERENRKLKKTLDCFKNLTFQLESLE
9	127	KRTAP10-12	G232V	SCQPSCGRLASCVSLLCRPTCSRLA
9	128	SYK	E26K	PFFFGNITREEAKDYLVQGGMSDGL