The common neoantigens in colorectal cancer are predicted and validated to be presented or immunogenic

Zhaoduan Liang[a,b*], Lili Qin[a,b,c*], Lei Chen[a,b], Wenhui Li[a,b], Chao Chen[a,d,f], Yaling Huang[a,b], Le Zhang[a,d,f], Songming Liu[a,d,f], Si Qiu[a,f], Yuping Ge[a,b], Wenting Peng[a,b,d], Xinxin Lin[a,f], Xuan Dong[a,b], Xiuqing Zhang[a,b,f] and Bo Li[a,e,f#]

[a]BGI-Shenzhen, Shenzhen 518083, China

[b]China National GeneBank, BGI-Shenzhen, Shenzhen 518120, China

[c]College of Basic Medicine, Dali University, Dali 671000, China

[d]BGI Education Center, University of Chinese Academy of Sciences, Shenzhen, 518083, China

[e]BGI-GenoImmune, BGI-Shenzhen, Wuhan, 430079, China

[f]BGI-GenoImmune, BGI-Shenzhen, Shenzhen 518083, China

*these authors contributed equally to this article.

#Corresponding author: libo@genomics.cn; Tel: +86-1868-067-9919

ABSTRACT

Colorectal cancer (CRC) is a malignant cancer with high incidence and mortality in the world, as the result of the traditional treatments. Immunotherapy targeting neoantigens can induce durable tumor regression in cancer patients, but is almost limited to individual treatment, resulting from the unique neoantigens. Many shared oncogenic mutations are detected, but whether the common neoantigens can be identified in CRC is unknown. Using the somatic mutations data from 321 CRC patients combined with a filter standard and 7 predicted algorithms, we screened and obtained 25 HLA-A*11:01 restricted common neoantigens with high binding affinity (IC50<50 nM) and presentation score (EPIC>0.9). Except the positive epitope KRAS_G12V$_{8\text{-}16}$, 11 out of 25 common neoantigens were proved to be naturally processed and presented on constructed K562 cell surface by mass spectroscopy (MS), and 11 out of 25 common neoantigens specifically induced *in vitro* pre-stimulated cytotoxic lymphocyte (CTL) to secrete IFN-γ. However, only 2 out of 25 common neoantigens were simultaneously presented and immunogenic. Moreover, using cell-sorting technology combined with single-cell RNA sequencing, the immune repertoire profiles of C1orf170_S418G$_{413\text{-}421}$ and KRAS_G12V$_{8\text{-}16}$-specific CTL were clarified. Therefore, common neoantigens with presentation and immunogenicity could be found in CRC, which would be developed as the universal targets for CRC immunotherapy.

KEYWORDS: colorectal cancer; neoantigen; common; presentation; immunogenicity

Introduction

Colorectal cancer (CRC) is the third commonest diagnosed malignant cancer and the second leading cause of cancer death in the world [1]. In 2018, more than 1.8 million new cases of CRC and almost 881 thousand cases of CRC-interrelated death occurred in the world [1], and the global burden of CRC is estimated to reach over 2.2 million new cases and 1.1 million cancer deaths by 2030 [2]. Traditionally surgical resection can cure the early stage of CRC, but about 50% of patients ultimately die of distant metastases. While chemotherapy, radiation therapy and targeted therapy can extend overall survival, less than 15% of patients with metastatic CRC survive beyond 5 years [3]. Therefore, the novel and more effective therapeutic approaches for CRC are necessary to develop.

In the recent years, based on a better knowledge of the complex interactions between the immune system and the tumor microenvironment, immunotherapy has become a novel effective and promising therapeutic strategy for cancer, and its efficacy was widely tested by CRC model. The vast majority of CRC patients with deficient mismatch repair (dMMR) or highly microsatellite instable (MSI-H) benefit from immune checkpoint inhibitors, which is not effective in other CRC patients with proficient MMR (pMMR) or microsatellite stable (MSS) [4]. Patients with CRC do not respond to autologous tumour lysate DC (ADC) and peptide vaccines [4]. T cells, which are engineered to express an affinity-enhanced T-cell receptor (TCR) or an antibody-based chimeric antigen receptor (CAR) targeting tumor associated antigens (TAAs), such as carcinoembryonic antigen (CEA) [5, 6] and human epidermal growth factor receptor-2 (HER2) [7], regress metastatic CRC, but simultaneously mediate severe autoimmunity in patients. These results highlight the importance of identifying tumor specific antigens, which optimally discriminate tumor and normal tissues, for improving the efficacy and safety of adoptive T-cell therapy for CRC.

Compared with TAAs, which lowly express in some normal cells but overexpress in tumor cells [5-7], mutated tumor-specific antigens (TSAs) arise from the somatic mutations in protein-coding regions of tumor cells, and are exclusively

present in malignant cells and not produced by normal tissues [8]. The accumulated mutations in cancers include nonsynonymous, nonsense, indel and frame shift, and are classified into driver mutations, which involve in uncontrolled cell growth and tumor metastasis, and passenger mutations, which may not contribute to the tumorigenic phenotype, but increase immunogenicity [9]. Through the antigen presentation system, peptides, which contain the mutant site and are called neoantigens, can be presented on the surface of tumor cells by major histocompatibility complex (MHC) molecules, then recognized by T cells to lead robust anti-metastatic CRC activity [10]. Furthermore, T-cell responses elicited by neoantigens are not subject to host central tolerance in the thymus and also bring fewer toxicities deriving from autoimmune reactions to normal cells [11]. The highly individual neoantigens actualize the personalized cancer immunotherapies, but limited the development of “one fits all” pharmacologic solutions [11].

It has been shown that some cancers with high tumor mutational burden (TMB) possess a set of common neoantigens owing to the microsatellites [11]. The presence of microsatellite instability has been found in approximately 15-20% CRC, and dMMR CRC has a high TMB, which is far higher than the standard value [4]. Furthermore, driver mutations possess only 8% CD8 T-cell neo-epitopes, while passenger mutations possess 92% CD8 T-cell neo-epitopes and 100% CD4 T-cell neo-epitopes [9]. HLA-A*11:01 allele has high prevalence in US Caucasians, Asian-Americans and China [12] (http://www.allelefrequencies.net/). Therefore, we hypothesized that targeting the common neoantigens, comprising driver mutations and passenger mutations and restricted by HLA-A*11:01, not only improved the efficacy, safety and adoption of CRC immunotherapy, but also reduced the cost and time of CRC clinical treatment. In the present study, we predicted the HLA-A*11:01 restricted common neoantigens from the somatic mutations data of 321 patients with CRC and validated their presentation and immunogenicity, which would become the new targets for CRC immunotherapy.

Materials and methods

Cell lines

The TAP-deficient T2 cell line (CRL-1992), K562 cell line (CCL-243) and HEK-293 cell line (CRL-1573) were purchased from the American Type Culture Collection (ATCC), and respectively maintained in Iscove's Modified Dulbecco's Medium (IMDM, Gibco), RPMI-1640 medium (Gibco) and Dulbecco's Modified Eagle's Medium (DMEM, Gibco) with 10% fetal bovine serum (FBS; Hyclone) at 37 °C in a humidified 5% CO_2 incubator. T2 cell line and K562 cell line were retrovirally transduced with retrovirus encoding HLA-A*11:01. Cells were authenticated by HLA genotyping, tested for mycoplasma by PCR method, and maintained in medium no more than 2 months from each thaw. Human peripheral blood was obtained from anonymous healthy donors who had signed informed consents. Peripheral blood mononuclear cells (PBMCs) were isolated by Ficoll-Hypaque gradient centrifugation and maintained in RPMI-1640 medium supplemented with 10% FBS at 37 °C in a humidified 5% CO_2 incubator. The study was approved and conducted by Institutional Review Board of Beijing Genomics of Institute (BGI)-Shenzhen (No. BGI-IRB18142).

Mutation selection and epitope prediction

The somatic mutations data of 321 patients with CRC from China-Colorectal Cancer Project (COCA-CN, https://icgc.org/icgc/cgp/73/371/1001733) in ICGC (International Cancer Genome Consortium) database (http://icgc.org) were downloaded and further analyzed. In briefly, missense variants that caused amino acid changes in coding regions were filtered according to a standard, in which the frequency of single-nucleotide variants (SNVs) was over 5 out of 321 patients and insertions or deletions (InDels) was over 2 out of 321 patients. After obtaining the list of the tumor-specific mutant proteins, we extracted the peptide sequences around the mutated sites. As MHC class I molecules bind to peptides 9-10 amino acids in length with the highest affinity [13], peptides were extracted *in silico* from 19 amino acids sequences, with 9 amino acids upstream and 9 amino acids downstream of mutated

amino acids, and 19 sets of 9-mer or 10-mer sequence containing mutated site from each mutated protein were identified and predicted by algorithms. Wild-type peptides with the same length as the mutant peptides were extracted as references. The potential binding affinity between extracted peptides and HLA-A*11:01 allele was analyzed simultaneously by NetMHC-4.0 [14], NetMHCpan-3.0 [15], NetMHCpan-4.0 [16], PSSMHCpan-1.0 [13], PickPocket-1.0 [17] and SMM [18]. Results were exhibited as predicted equilibrium binding constants of IC50 (50% inhibitory concentration, nM), in which strong binders meant the predicted binding affinity IC50 values were less than 50 nM, and weak binders were that of 50-500 nM [19]. Moreover, we used our software Epitope Presentation Integrated prediCtion (EPIC) [20], with a fixed expression value 4 TPM as inputs, to predict the presentation of extracted peptides, in which the results were shown by the scores of highest presentation probability. These mutant peptides with strong binding capacity and high score of presentation probability were selected as potential neoantigens.

The construction of putative neoantigens transduced K562 cells (HLA-A*11:01^{+})

Six predicted neoantigens were linked into a tandem neoantigen as previously described [21]. Briefly, six predicted neoantigens, which each had 27 amino acids with the mutation at position 14, were connected by a start linker (GGSGGGGSGG), middle linkers (GGSGGGGSGG) and an end linker (GGSLGGGGSG). The N terminal of a tandem neoantigen was successively linked with the kozak sequence (GCCACC) and the signal peptide sequence (SPMRVTAPRTLILLLSGALALTET WAGS), and the C terminal was linked with the MHC class I trafficking signal (MITD) sequence (IVGIVAGLAVLAVVVIGAVVATVMCRRKSSGGKGGSYSQA ASSDSAQGSDVSLTA) and termination codon. The tandem minigene DNA fragment coding the above sequence was synthetized and cloned into lentiviral expressing vector pLVX (CMV-EF1a-ZsGreen-P2A-Bsd; provided by Viraltherapy Technologies Ltd, Wuhan, China). Lentiviral particles encoding tandem minigenes were produced from HEK-293 cells, which were simultaneously transduced with

packaging constructs (RRE, REV and VSVG (invitrogen)) and the expressing vector, and infected mono HLA-A*11:01 allelic K562 cells. Positive K562 cells were selected by Blasticidine S hydrochloride (5 μg/ml, Sigma) and detected through the percentage of reporter gene ZsGreen.

The preparation of MHC Class I bound peptides

MHC-I peptidomes were obtained from mono HLA-A*11:01 allelic K562 cells transduced with putative neoantigens as described previously [22]. In brief, 1×10^9 cells were dissociated using lysis buffer (0.25% sodium deoxycholate, 1% n-octyl glucoside, 100 mM PMSF and protease inhibitors cocktail in phosphate buffer saline (PBS)) at 4 ℃ for 60 min. Lysate were further cleared by centrifugation at 14,000 g for 30 min. Cleared lysate were purified with pan-HLA class I complexes antibody (clone W6/32) covalently bound Protein-A Sepharose CL-4B beads (GE Healthcare). Beads were washed with Tris-HCl buffer containing NaCl. The MHC-I molecules were eluted at room temperature using 0.1 N acetic acid. Eluate was loaded on Sep-Pak tC18 cartridges (Waters, 100 mg). The C18 cartridges were first washed with 0.1% TFA, then with 0.1% TFA containing 30% ACN to separate peptides from MHC-I complexes. Eluate was concentrated to 20 μl using vacuum centrifugation. Finally, 5 μl of sample was used for Parallel Reaction Monitoring (PRM) mass analysis.

Peptide validation by mass spectroscopy (MS) analysis with PRM

Peptides were separated by a nanoflow HPLC (15 cm long, 75 μm inner diameter column with ReproSil-Pur C18-AQ 1.9 μm resin) and coupled on-line to a Fusion Lumos mass spectrometer (Proxeon Biosystems, Thermo Fisher Scientific) with a nanoelectrospray ion source (Proxeon Biosystems). Peptides were eluted with a linear gradient of 5-80% buffer B (98% ACN and 0.1% FA) at a flow rate of 500 nl/min over 3 hours. Data of each injection was acquired using a corresponding transition list (Supplemental Table 1-6). Full scan MS spectra were acquired at a resolution of 6,000 at 350-1,400 m/z with a target value of 4×10^5 ions. MS/MS resolution was 60,000 at

150-2,000 m/z, and higher collisional dissociation (HCD) was employed for ion fragmentation. The interpretation of MS data was performed with Skyline. To validate a peptide which could be presented by MHC-I complex, the following criteria were considered: i) the variation of retention time between precursor ions was less than 3 min; ii) the pattern and retention time were matched between synthetic peptide and target peptide for no less than 5 product ions.

The preparation of tetramer of peptide-MHC complex

Peptides (Table 1) and HLA-A*11:01-restricted KRAS G12V$_{8-16}$ (VVGAVGVGK) as positive peptide were synthesized from GenScript (Nanjing, China), with purity greater than 98% by mass spectroscopy. Peptide-MHC tetramers were generated as previously described [23]. In briefly, Peptides (400 μM) were mixed with Flex-T™ HLA-A*11:01 Monomer UVX (Biolegend), then subjected to UV light for 30 min on ice. The MHC monomers exchanged with peptides were tetramerized in the presence of allophycocyanin (APC) conjugated streptavidin (BD Biosciences) for 30 min at 37 °C, then the reaction was stopped by PBS containing D-Biotin (500 μM) and NaN_3 (10%), then kept at 4 °C overnight for use.

The generation of mature dendritic cells (mDCs)

Monocytes (CD14 positive) were positively selected using CD14 MicroBeads (Miltenyi Biotec) from PBMCs of healthy donors as the manufacturer's protocol, and cultured in CellGenix™ DC media (CellGenix) supplemented with 2% human serum albumin (HSA, CSL Behring L.L.C.), granulocyte-macrophage colony stimulating factor (GM-CSF, 100 ng/ml; PeproTech) and interleukin 4 (IL-4, 100 ng/ml; PeproTech) for 5 days. On day 6, immature DCs were stimulated to mature by TNF-α (10 ng/ml; PeproTech), IL-6 (50 ng/ml; PeproTech), IL-1β (10 ng/ml; PeproTech), Prostaglandin E2 (500 ng/ml; Sigma) and Poly(I:C) (10 μg/ml; InvivoGen) for 2 days. After harvest, mDCs were pulsed with peptides (1 μg/ml, Table 1) in FBS-free RPMI-1640 medium for 4 hours at 37 °C, and used as the antigen presented cells. The maturity of DC was determined through the morphology and the phenotype of the

expressions of CD80, CD83, CD86, CD11c and HLA-DR.

The induction of neoantigen-specific cytotoxic lymphocyte (CTL)

$CD8^+$T cells were positively enriched using CD8 MicroBeads (Miltenyi Biotec) from PBMCs of healthy donors as the manufacturer's protocol, stimulated by mDCs pre-loaded with peptides at a 4:1 ratio, and maintained in HIPP™-T009 medium (Bioengine) in the presence of 2% autoserum and IL-21 (30 ng/ml, PeproTech) in a 37°C 5% CO_2 incubator for 12 days. On day3, the co-culture system was supplemented with IL-2 (5 ng/ml, PeproTech), IL-7 (10 ng/ml, PeproTech) and IL-15 (10 ng/ml, PeproTech), which were repeated every 2-3 days. After 12-day culture, the pre-stimulated $CD8^+$T cells were re-stimulated with same peptide-pulsed mDCs as above and incubated for another 12 days.

Enzyme-linked immunospot (ELISPOT) assay for IFN-γ

IFN-γ ELISPOT assay strip plate which was pre-coated with anti-human IFN-γ mAb (1-D1K, Mabtech) was washed with PBS and blocked with RPMI-1640 containing 10% FBS. CTLs were co-cultured with T2 cells pre-pulsed with or without peptides (10 μg/ml) in the above pre-treated ELISPOT plate for 24 hours. Each sample was set with repetition. Plate was rinsed with PBS, then added with alkaline phosphatase (ALP) labeled anti-human IFN-γ mAb (7-B6-1-ALP, 1:200; Mabtech) for 2 hours. After rinsing, 5-Bromo-4-chloro-3-indolyl phosphate/Nitro blue tetrazolium (BCIP/NBT, Mabtech) was used to develop the immune-spot according to the manufacturer's protocol. Spots were imaged and counted by an ELISPOT Reader (BioReader 4000, BIOSYS). Positive response was judged according to that the number of specific spots was more than 10 and at least two-fold greater than that of negative control [24].

Fluorescence-activated cell sorting (FACS)

Cells were collected, washed and resuspended in PBS containing 2% FBS

(FACS buffer). Cells were stained with fluorescent dye conjugated antibodies for 15 min at 4 °C. PE conjugated anti-CD8 antibody, APC conjugated pMHC tetramer, APC conjugated anti-CD86 antibody, PE conjugated anti-CD83 antibody, PE conjugated anti-CD80 antibody, APC conjugated anti-CD11c antibody, PE conjugated anti HLA-DR antibody, and isotype matched antibodies were used in this study and purchased from BD Biosciences. After washing twice in FACS buffer, Cells were analyzed using a FACSAria Ⅱ (BD Biosciences) with live cell gating based on 4',6-diamidino-2-phenylindole (DAPI) exclusion, and $CD8^+$pMHC $tetramer^+$ cells were sorted for single-cell RNA sequencing. The data were analyzed using FlowJo software (Tree Star).

The analysis of neoantigen-specific T-cell receptor repertoire by single-cell RNA sequencing

According to the manufacturer's protocol of Chromium™ Single Cell V(D)J Reagent Kits (10x Genomics, Inc.), sorted T cells were partitioned and captured into the Gel Bead in Emulsion (GEM) through the rapid and efficient microfluidics technology of the Chromium™ single-cell controller (10x Genomics, Inc.). Single cell and the Gel Bead were lysed in the GEM, then the contents of the GEM were incubated in the Reverse Transcription-Polymerase Chain Reaction (RT-PCR) to generate full-length, barcoded cDNA from the poly A-tailed mRNA transcripts. Barcoded cDNA molecules were pooled after GEMs being broken, and full-length V(D)J segments from TCR cDNA were enriched by PCR amplification and constructed as a library for Illumina®-ready sequencing. TCR repertoire and paired TCR were analyzed by the Cell Ranger™ analysis pipelines. The gene usage was assigned using the IMGT nomenclature.

Results

The selection of mutant candidate peptides

In order to analyze the potential common neoantigens of CRC in China, which

derived from driver mutations or passenger mutations, we ultimately collected 3,500 SNVs, of which the frequency was over 5 out of 321 patients, and 191 InDels, of which the frequency was over 2 out of 321 patients, from the somatic mutations data of 321 CRC patients from COCA-CN in ICGC database on August 29, 2018. Both 9-10-mer mutant epitope candidates and reference peptides were extracted from 19 amino acids length, with 9 amino acids upstream and 9 amino acids downstream of the 3,691 mutant sites. A total of 60,169 epitopes of SNV and 6,891 epitopes of Indels were generated and predicted the binding affinity with HLA-A*11:01 allele, by NetMHC-4.0, NetMHCpan-3.0, NetMHCpan-4.0, PSSMHCpan-1.0, PickPocket-1.0 and SMM simultaneously. As a result, 56 mutant epitopes were selected as the IC50 value of predicted binders was less than 50 nM by at least three software packages, and the smallest affinity predicted value during three softwares was taken as the affinity predicted score of the peptide-MHC complex (Table 1 and data not shown). Furthermore, the algorithm EPIC, which was an effective, flexible and publicly available HLA-I presented epitope prediction method, was used to evaluate the probability of presentation of 56 mutant epitopes with strong binding affinity. The EPIC score value of 25 out of 56 mutant epitopes was more than 0.9, which meant that the probability of predicted epitope being presented by MHC was absolutely high (Table 1). Finally, 25 mutant epitopes related to 25 somatic mutations of 21 genes were selected as peptide candidates, and assessed their presentation and immunogenicity (Table 1). Moreover, during the 21 genes, we found that *RNF43* gene was a tumor suppressor gene, *CTNNB1* gene was an oncogene, and the remaining genes had not been identified as tumor-associated genes, which also encoded potential neoantigens by CRC, in Cancer Gene Census database (https://cancer.sanger.ac.uk/census) (Table 1).

The expression of predicted peptides from constructed K562 cells

T cells attacking targeted cells mainly depends on that TCR recognize T-cell epitopes, which are expressed, naturally processed and presented by MHC molecules

on the cell surface [24]. In order to improve the probability of the predicted epitopes being presented by MHC class I, the predicted peptides were linked into the tandem minigenes, and fused with N-terminal leader peptide and C-terminal MITD, which have been proved to strongly improve the presentation of MHC class I and class II epitopes [25]. The assembled base sequences were cloned into the multiple clone site (MCS) of lentiviral vector (Figure 1(a)), which were operated by CMV promoter and tracked by a reporter gene ZsGreen. As a result, we constructed five tandem-minigene stably transfected K562 cell lines with mono HLA-A*11:01 allele. CRC-1-K562 cells contained the minigenes of GLCE_V533I, C1orf170_S418G, MUC3A_I29T, CCRL2_F179Y, KIAA1683_M313T and KLHL40_N345S; CRC-2-K562 cells contained the minigenes of MUC3A_S175P, RNF43_I47V, SYNE2_A2395T, TLR10_I369L, ANKRD36C_N1571S and IYD_F231I; CRC-3-K562 cells contained the minigenes of SSX5_E19Q, MUC3A_S326T, ARHGEF11_H1427R, CTNNB1_T41A, LILRB5_L605F and EIF2A_T92S; CRC-4-K562 cells contained the minigenes of TMPRSS15_P732S, MUC6_P2049L, TMEM185B_A42G, UNC93A_M403T, MUC3A_Q31H and FSIP2_R1288Q; CRC-5-K562 cells contained the minigenes of TMEM185B_A42G, UNC93A_M403T, MUC3A_Q31H, FSIP2_R1288Q, FSIP2_T184NX and KRAS_G12V (Table 2 and Figure 1(b)). In present study, we selected KRAS_G12V, which is accepted as a common oncogenic mutation and has been proved to be presented by HLA-A*11:01 allele on the basis of T-cell response to KRAS_G12V positive target cells [12], as a positive control of presentation and immunogenicity. In Figure 1(b), the vast majority of tandem-minigene stably transfected K562 cell lines expressed reporter gene ZsGreen, and the range of the expression rate was 85-93%, which was determined by FACS and indirectly reflected the constructed tandem minigenes were expressed in K562 cells.

The identification of MHC Class I presented epitopes from constructed K562 cells by MS

In order to verify the predicted epitopes could be naturally processed and

presented by cells, we employed an immunoproteomics approach to enrich the immunopeptidome of constructed K562 cells with mono HLA-A*11:01 allele. In this approach, MHC-I restricted peptides were isolated and analyzed with MS. Targeted MS assays with PRM were developed and characterized using stringent search criteria, and resulted in 12 epitopes, including KRAS_G12V$_{8\text{-}16}$, GLCE_V533I$_{526\text{-}535}$, MUC3A_I29T$_{28\text{-}36}$, KLHL40_N345S$_{341\text{-}349}$, MUC3A_S175P$_{172\text{-}181}$, RNF43_I47V$_{46\text{-}54}$, IYD_F231I$_{229\text{-}237}$, MUC3A_S326T$_{319\text{-}327}$, ARHGEF11_H1427R$_{1427\text{-}1435}$, CTNNB1_T41A$_{41\text{-}49}$, FSIP2_R1288Q$_{1285\text{-}1293}$ and FSIP2_T184NX, were confirmed to be presented by constructed K562 cells (Figure 2(a-l)).

The analysis of predicted neoantigens activating CTL to secrete IFN-γ in vitro

T cells targeting mutations can be detected from tumor infiltrating lymphocytes (TILs), peripheral memory lymphocytes of cancer patients and peripheral naïve lymphocytes of healthy donors [12, 26]. However, except mouse model, PBMCs from healthy donors were easier obtained to determine the immunogenicity of the predicted neoantigens, comparing to that from patients. To test whether the predicted neoantigens had the characteristic of stimulating CTL to secrete IFN-γ, we used peptide-pulsed mDCs to co-culture with the bulk CD8^{+}T cells isolated from HLA-A*11:01^{+} healthy donors for twice in the presence of cytokines, which was supposed to expand neoantigen-reactive CTL. Then CTL were re-stimulated by the corresponding peptide-pulsed T2 cells in the IFN-γ ELISPOT plate. Under inverted phase contrast microscope of 40× object lens, monocytes isolated from PBMCs were observed to display as small round cells, and gradually stretch and adhere the plastic surface of the culture plate on day1(Figure 3(a)). On day8 after the stimulation of a cytokine cocktail for two days, mature DCs were exhibited as irregularly large round and suspension cells with blunt, elongate dendritic processes as previously reported (Figure 3(a)) [27, 28]. Furthermore, the phenotype of mature DCs was analyzed by FACS testing the expression of co-stimulatory molecules and maturation markers. Figure 3(b) showed that more than 98% of DCs expressed CD86, CD80, CD11c and

HLA-DR, but a relatively low proportion of mature DC cells expressed CD83 (71.5%), of which the maturation level of DCs was basically reached the international criteria level [29]. After being re-stimulated with T2 cells, which were respectively pre-loaded with epitopes of KRAS_G12V$_{8-16}$, C1orf170_S418G$_{413-421}$, KIAA1683_M313T$_{311-319}$, SSX5_E19Q$_{12-20}$, TMEM185B_A42G$_{42-51}$ and UNC93A_M403T$_{402-410}$, CTL largely secreted IFN-γ, which was manifested by the formation of more than 10 times spots on IFN-γ ELISPOT plate compared with the negative control (Figure 4(a) and (b)). In addition, these epitopes of GLCE_V533I$_{526-535}$, CCRL2_F179Y$_{174-183}$, ANKRD36C_N1571S$_{1571-1579}$, MUC3A_S326T$_{319-327}$, ARHGEF11_H1427R$_{1427-1435}$ and MUC6_P2049L$_{2044-2053}$ also respectively and positively activated CTL to produce low level of IFN-γ, in which the spot number was more than 10 and two-fold greater than that of negative control (Figure 4(a) and (b)). The number of spot mediating by the remainder peptides almost equated that of the negative control, which signified that those peptides did not stimulate CTL to express IFN-γ (Figure 4(a) and (b)). Totally, except the positive control peptide KRAS_G12V$_{8-16}$, 11 out of 25 predicted peptides could activate CTL to secrete IFN-γ *in vitro* and had immunogenicity, which was proved by PBMCs from only one healthy donor, and the positive rate was 44% (11/25).

The immune repertoire profiles of epitope-specific CTL from single-cell RNA sequencing

To determine whether epitope-specific CTL were expanded by peptide-pulsed mDCs to the degree of sorting, we selected 6 peptides of KRAS_G12V$_{8-16}$, C1orf170_S418G$_{413-421}$, KIAA1683_M313T$_{311-319}$, SSX5_E19Q$_{12-20}$, TMEM185B_A42G$_{42-51}$ and UNC93A_M403T$_{402-410}$, which strongly activated CTL to secrete IFN-γ (Figure 4(a) and (b)), to prepare tetramer-APC, and stained the bulk CTL with tetramer-APC and CD8-PE simultaneously. We found that the range of the frequency of the epitope-specific CTL after being co-cultured was 0.22-7.08% (Figure 5(a)). Based on the ability of inducing T cells to secrete IFN-γ, we selected to sort

KRAS_G12V$_{8-16}$ -specific CTL and C1orf170_S418G$_{413-421}$-specific CTL by FACS (Figure 5(b)), of which the sorted cells were single-cell sequenced by the Chromium Single-Cell Immune Profiling Solution to determine T-cell receptor repertoires.

For T-cell receptor repertoires recognizing HLA-A*11:01 presented C1orf170_S418G$_{413-421}$, a total of 9,243 cells (exactly GEM) were finally estimated from 17,800 sorted cells, and 14,042 TCR α chain and 9,346 TCR β chain amino acid sequences were respectively obtained, of which 8,826 pairs of TCR were produced. TCRα repertoire was preferentially biased toward usage of TRAV29DV5 gene (55.38%) and TRAV35 gene (37.35%) (Figure 6(a)), which corresponding mainly rearranged with TRAJ54 (99.88%) and TRAJ29 (Figure 6(b), and data not shown). The length distribution of CDR3α was preferentially restricted to 12 mer (55.64%) and 15 mer (39.09%) (Figure 6(c)), and the motif of 12-mer and 15-mer CDR3α was respectively highly conserved as CAASGGAQKLVF (Figure 6(g)) and CAGLLYNSGNTPLVF (Figure 6(h)), which was consistent with that of TRAV29DV5-TRAJ54 and TRAV35-TRAJ29. The gene usage of TCRβ repertoire was highly biased TRBV27 gene (87.47%) (Figure 6(d)), which mainly rearranged with TRBJ1-5 (99.91%) (Figure 6(e)). The length distribution of CDR3β was highly restricted to 15 mer (88.04%) (Figure 6(f)), and the motif of 15-mer CDR3β was highly conserved as CASSRDRGSNQPQHF (Figure 6(i)), which was consistent with that of TRBV27-TRBJ1-5. There were 298 diversities in the 8,826 TCRα/TCRβ pairs, but two clones accounted for a large proportion in the repertoire. 53.2% (4,694/8,826) T cells (named as Clonotype1) expressed TRAV29DV5-TRAJ54 and TRAV35-TRAJ29 containing TCRα and TRBV27-TRBJ1-5 containing TCRβ, and 31.6% (2,786/8,826) T cells (named as Clonotype2) expressed TRAV29DV5-TRAJ54 containing TCRα and TRBV27-TRBJ1-5 containing TCRβ (Table 3).

For T-cell receptor repertoires recognizing HLA-A*11:01 presented KRAS_G12V$_{8-16}$, a total of 12,530 cells (exactly GEM) were finally estimated from 21,000 sorted cells, and 11,137 TCR α chain and 13,126 TCR β chain amino acid sequences were respectively obtained, of which 10,559 pairs of TCR were produced.

TCRα and β repertoires respectively used immunodominant TRAV8-3 gene (81.69%) (Figure 7(a)), which mainly rearranged with TRAJ20 (99.69%) (Figure 7(b)), and TRBV11-2 gene (83.35%) (Figure 7(d)), which mainly rearranged with TRBJ2-7 (99.81%) (Figure 7(e)). The length distribution of CDR3α and CDR3β was respectively highly restricted to 10 mer (81.88%) (Figure 7(c)) and 14 mer (86.4%) (Figure 7(f)). The motif of 10-mer CDR3α and 14-mer CDR3β was respectively conserved as CASNDYKLSF (Figure 7(g)), which was consistent with that of TRAV8-3-TRAJ20, and CASSLDGVSYEQYF (Figure 7(h)), which was consistent with that of TRBV11-2-TRBJ2-7. There were 1,610 diversities in the 10,559 TCRα/TCRβ pairs, but two clones accounted for a large proportion in the repertoire. 80.5% (8,499/10,559) T cells (named as Clonotype1) expressed TRAV8-3-TRAJ20 containing TCRα and TRBV11-2-TRBJ2-7 containing TCRβ, and 14.3% (1,515/10,559) T cells (named as Clonotype2) expressed TRBV11-2-TRBJ2-7 containing TCRβ but were not detected TCRα (Table 3). Our results revealed that the dominant clone represented the usage genes and the CDR3 motifs of the TCR repertoire.

The data that support the findings of this study have been deposited in the CNSA (https://db.cngb.org/cnsa/) of CNGBdb with accession code CNP0000518.

Discussion

Cancer immunotherapy emerges as a very promising therapeutic approach for tumors. Different from chemotherapy, radiotherapy and targeted therapy, which directly target tumors, targeting the immune system offers the potential for durable activity and long-lasting survival outcomes [30]. It is widely accepted that anti-tumor immunity is especially mediated by the responses of tumor-specific T cells, which can effectively delete the primary tumor lesions and protest against metastases [30, 31]. T cells utilize TCR to recognize the short-peptide antigens bound in the groove of MHC molecules, and discriminate self and non-self, in which the short and cytosol-derived peptides mainly determine the specificity of the T cell-dependent immune response

[32, 33].

During their carcinogenesis and progression, tumors usually obtain numerous somatic mutations. Mutant genes are translated into proteins and presented on cell surface by MHC, which result in arising neoantigens [11]. Neoantigens are uniquely produced by tumor cells, not totally found in normal tissues, and unparalleled tumor biomarkers [11]. Based on T cells recognizing neoantigens are not subject to thymic selection and central tolerance, neoantigen specific T cell with high-avidity is very likely to exist in the human body [8, 11]. Whole exome sequencing and RNA sequencing combined with bioinformatic pipelines make the reality of disclosing tumor-specific alterations with single nucleotide resolution and predicting neoantigens for cancer immunotherapy [34]. After pinpointing missense mutations and gene expression levels, peptides are assessed using various algorithms to predict binding affinity to MHC or presentation on MHC [20]. However, the vast majority of predicted neoantigens fail to turn up in tumors, and a handful is found to elicit T-cell responses [35]. Using five kinds of cancer patient-derived PBMCs, Chizu et al. only identified one immunogenic peptide from 26 mutant epitopes, which were predicted by NetMHC4.0 algorithm to have strong-binding capacity to HLA-A*24:02 [9]. Using four kinds of healthy donor-derived T cells, Stronen et al. showed T-cell reactivity toward 3-5 of 20 neoantigens, which were predicted by NetMHC3.2 or netMHCpan2.0 algorithm to have high predicted binding to HLA-A*02:01[36]. In a small group of patients with stages III and IV melanoma, Ott et al. demonstrated only 16% neoantigens, which were predicted by NetMHCpan-2.4 algorithm, were recognized by $CD8^+$ T cell [37]. Zhang et al. found a significant T-cell response in two of nine neoantigens for one breast cancer patient, and one of eight neoantigens for the other two patients with breast cancer, in which neoantigens were predicted by NetMHC-3.2 algorithm [38]. Therefore, we considered that none of the current algorithms was perfect, and it was necessary to simultaneously use multiple algorithms to increase the accuracy of peptide binding affinity prediction. In the present study, we respectively used peptide-MHC binding-affinity prediction

algorithms, including NetMHC-4.0, NetMHCpan-3.0, NetMHCpan-4.0, PSSMHCpan-1.0, PickPocket-1.0 and SMM, and EPIC algorithm that predicted epitope presentation to evaluate the extracted peptides. Finally, we selected 25 candidate peptides, which simultaneously met the conditions of the frequency of SNV being over 5 out of 321 patients and InDel being over 2 out of 321 patients, IC50 value being less than 50 nM by at least three software packages and EPIC score value being more than 0.9 (Table 1), to assay their characteristics of presentation and inducing cytotoxic T cells. We found that 11 out of 25 (44%) predicted epitopes were proved to be presented by HLA-A*11:01 allele through MS (Figure 2), and 11 out of 25 (44%) predicted epitopes induced specific CTL to secrete IFN-γ through ELISPOT assay (Figure 4), and 20 out of 25 (80%) predicted epitopes could either be presented or have immunogenicity (Figure 2 and Figure 4). However, it was a pity that, except the positive epitope (KRAS_G12V$_{8\text{-}16}$), only 2 out of 25 (8%) predicted epitopes were analyzed not only to be endogenously expressed in tumor cells, but also to induce a T-cell response (Figure 2 and Figure 4).

At present, MS-based approach is the relatively unbiased methodology to identify the repertoire of peptides, which are naturally processed and presented by MHC molecules *in vivo*, from human cancer cell lines, tumors and healthy tissues and body fluids [39]. Although the use of MS-based immunopeptidomics would reduce the false positive number of predicted *in silico* neoantigens, and ensure highly accurate and reliable assignment of neoantigen's sequences, neoantigens have not been regularly and sensitively disclosed by MS compared with tumor associated antigens (TAAs) [34, 39]. Based on the limited sensitivity, the false negative neoantigens that are naturally presented but not detected by MS are expected for immunopeptidomics. Due to the reports that relatively large biological samples, abundance of proteins containing specific sequences, expression in mono-allelic cells, proteasomal processing and the MITD trafficking signal for siting in endolysosomal compartments are important facts for discovering the presented peptides [22, 25, 39, 40], Five tandem minigenes, which each encoded six neo-antigenic peptides, was

operated by CMV promoter and linked with the MITD trafficking signal at the C terminal (Figure 1(a)), were constructed and transfected into HLA-A*11:01 mono-allelic K562 cells. We found that over 85% HLA-A*11:01 mono-allelic K562 cells highly expressed the tandem minigenes (Figure 1(b)), and used 1×10^9 cells to extract and purify the peptides, which were followed by analysis with a mass spectrometer. Finally, except the positive epitope (KRAS_G12V$_{8\text{-}16}$), we confirmed that 11 out of 25 predicted neoantigens were naturally processed and presented by HLA-A*11:01 allele, in which the positive rate was up to 44% (11/25) (Figure 2). However, the remaining neo-peptides were not detected in our system. Recently, Muhammad Ali et al. proved that minimizing the formation of irrelevant immunogenic peptides could increase the targeted epitopes to bind HLA, and the order of arranged epitopes in the tandem minigene were involved in the efficiency of antigen presentation [26]. In our present study, the predicted neoantigens in the tandem minigene each had 27 amino acids with the mutation at position 14 (Table 2), which may result in the formation of other high-affinity irrelevant and competitive presented peptides, and were randomly combined in the tandem minigene. Therefore, we considered that the proper length and the appropriate order of desired neoantigens in the tandem minigene may improve the probability of the remaining neo-peptides being detected by MS.

The immunogenicity of predicted neoantigens is usually detected through the response of TIL [11, 12]. Moreover, although the frequency of the neoantigen-reactive T cells in the peripheral blood is quiet low compared with that in the tumor sample, neoantigen-specific T cells, which derive from circulating $CD8^+$ memory T cells of cancer patients or circulating $CD8^+$ naïve T cells of healthy donors, can be enriched after being *in vitro* co-cultured with the cognate DCs pre-loaded neoantigens, and their response to recognize the corresponding neoantigens can be detected by conventional experimental methods [12, 26, 41]. Therefore, we generally adopted one of the proven methods to detect the immunogenicity of our predicted neoantigens, in which the circulating bulk $CD8^+$ T cells from one healthy donor were stimulated for two rounds by cognate DCs pre-loaded neoantigens. Except the

positive epitope (KRAS_G12V$_{8-16}$), we found 11 out of 25 (44%) predicted neoantigens that were pre-loaded on T2 cells could induce co-cultured CTL to secrete IFN-γ (Figure 4), and the frequencies of the specific CTL, that produced over 10 folds IFN-γ spots than the negative control, were estimated to range between 0.22% and 7.08% (Figure 5(a)). However, most of the presented neoantigens were not detected to be immunogenic in our study. As approximately only one in 10^5-10^6 T cells is specific for a given antigen in the lymph node, and circulating CD8$^+$ memory T cells from healthy donors do not contribute to the production of neoantigen-responsive T cells, to enrich circulating CD8$^+$ naïve T cells from healthy donors before priming can enhance the probability of specific T cells encountering DCs presenting the cognate neoantigen [26, 42]. Based on the donor-dependent variability and sufficient diversity of the human TCR repertoire, several donors need to be screened for identifying the immunogenicity of a novel candidate neoantigen [26]. Therefore, we considered that CTL, which were expanded from CD8$^+$ naïve T cells of other healthy donors, may respond to the remaining immunogenic negative neo-peptides.

In conclusion, except the positive epitope (KRAS_G12V$_{8-16}$), our results revealed several common actual T-cell neo-epitopes of CRC, which would be developed as the universal targets for CRC immunotherapy in the form of vaccines based on peptide, RNA, DNA and DCs and therapies based on adoptive TCR transgenic T cells.

Abbreviations

CRC, colorectal cancer; dMMR, deficient mismatch repair; MSI-H, highly microsatellite instable; pMMR, proficient mismatch repair; MSS, microsatellite stable; ADC, autologous tumour lysate DC; CAR, chimeric antigen receptor; TAA, tumor associated antigen; CEA, carcinoembryonic antigen; HER2, human epidermal growth factor receptor-2; TSA, tumor-specific antigen; MHC, major histocompatibility complex; TMB, tumor mutational burden; TCR, T-cell receptor; PRM, parallel reaction monitoring; TIL, tumor infiltrating lymphocyte; MS, mass

spectrometry; ATCC, American Type Culture Collection; PBMC, peripheral blood mononuclear cell; ICGC, International Cancer Genome Consortium; COCA-CN, China-Colorectal Cancer Project; InDel, insertions or deletion; EPIC, Epitope Presentation Integrated prediCtion; MITD, MHC class I trafficking signal; CTL, cytotoxic lymphocyte; ELISPOT, enzyme-linked immunospot; FACS, fluorescence-activated cell sorting; GEM, Gel Bead in Emulsion.

Acknowledgments

We thank Weipeng Hu from BGI-GenoImmune for technical assistance on this study.

Funding

This work was supported by the National Natural Science Foundation of China under Grant No. 81702826, Science, Technology and Innovation Commission of Shenzhen Municipality under Grant No. JCYJ20170303151334808, and Science, Technology and Innovation Commission of Shenzhen Municipality under Grant No. JCYJ20170817145845968.

Disclosure of Interest

The authors report no conflict of interest.

References

1. Bray, F., et al., *Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries.* CA Cancer J Clin, 2018. **68**(6): p. 394-424.
2. Arnold, M., et al., *Global patterns and trends in colorectal cancer incidence and mortality.* Gut, 2017. **66**(4): p. 683-691.
3. Magee, M.S., et al., *GUCY2C-directed CAR-T cells oppose colorectal cancer metastases without autoimmunity.* Oncoimmunology, 2016. **5**(10): p. e1227897.
4. Ciardiello, D., et al., *Immunotherapy of colorectal cancer: Challenges for therapeutic efficacy.* Cancer Treat Rev, 2019. **76**: p. 22-32.
5. Parkhurst, M.R., et al., *T cells targeting carcinoembryonic antigen can mediate regression of metastatic colorectal cancer but induce severe transient colitis.* Mol Ther, 2011. **19**(3): p. 620-6.
6. Zhang, C., et al., *Phase I Escalating-Dose Trial of CAR-T Therapy Targeting CEA(+)*

604 *Metastatic Colorectal Cancers.* Mol Ther, 2017. **25**(5): p. 1248-1258.
605 7. Morgan, R.A., et al., *Case report of a serious adverse event following the administration*
606 *of T cells transduced with a chimeric antigen receptor recognizing ERBB2.* Mol Ther,
607 2010. **18**(4): p. 843-51.
608 8. Zhou, Z., et al., *TSNAD: an integrated software for cancer somatic mutation and*
609 *tumour-specific neoantigen detection.* R Soc Open Sci, 2017. **4**(4): p. 170050.
610 9. Nonomura, C., et al., *Identification of a neoantigen epitope in a melanoma patient with*
611 *good response to anti-PD-1 antibody therapy.* Immunol Lett, 2019. **208**: p. 52-59.
612 10. Tran, E., et al., *T-Cell Transfer Therapy Targeting Mutant KRAS in Cancer.* N Engl J Med,
613 2016. **375**(23): p. 2255-2262.
614 11. Wirth, T.C. and F. Kuhnel, *Neoantigen Targeting-Dawn of a New Era in Cancer*
615 *Immunotherapy?* Front Immunol, 2017. **8**: p. 1848.
616 12. Cafri, G., et al., *Memory T cells targeting oncogenic mutations detected in peripheral*
617 *blood of epithelial cancer patients.* Nat Commun, 2019. **10**(1): p. 449.
618 13. Liu, G., et al., *PSSMHCpan: a novel PSSM-based software for predicting class I peptide-*
619 *HLA binding affinity.* Gigascience, 2017. **6**(5): p. 1-11.
620 14. Nielsen, M., et al., *Reliable prediction of T-cell epitopes using neural networks with novel*
621 *sequence representations.* Protein Sci, 2003. **12**(5): p. 1007-17.
622 15. Nielsen, M. and M. Andreatta, *NetMHCpan-3.0; improved prediction of binding to MHC*
623 *class I molecules integrating information from multiple receptor and peptide length*
624 *datasets.* Genome Med, 2016. **8**(1): p. 33.
625 16. Jurtz, V., et al., *NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions*
626 *Integrating Eluted Ligand and Peptide Binding Affinity Data.* J Immunol, 2017. **199**(9): p.
627 3360-3368.
628 17. Zhang, H., O. Lund, and M. Nielsen, *The PickPocket method for predicting binding*
629 *specificities for receptors based on receptor pocket similarities: application to MHC-*
630 *peptide binding.* Bioinformatics, 2009. **25**(10): p. 1293-9.
631 18. Peters, B. and A. Sette, *Generating quantitative models describing the sequence*
632 *specificity of biological processes with the stabilized matrix method.* BMC Bioinformatics,
633 2005. **6**: p. 132.
634 19. Creaney, J., et al., *Strong spontaneous tumor neoantigen responses induced by a natural*
635 *human carcinogen.* Oncoimmunology, 2015. **4**(7): p. e1011492.
636 20. Hu, W., et al., *EPIC: MHC-I epitope prediction integrating mass spectrometry derived*
637 *motifs and tissue-specific expression profiles.* bioRxiv, 2019: p. 567081.
638 21. Sahin, U., et al., *Personalized RNA mutanome vaccines mobilize poly-specific*
639 *therapeutic immunity against cancer.* Nature, 2017. **547**(7662): p. 222-226.
640 22. Abelin, J.G., et al., *Mass Spectrometry Profiling of HLA-Associated Peptidomes in Mono-*
641 *allelic Cells Enables More Accurate Epitope Prediction.* Immunity, 2017. **46**(2): p. 315-
642 326.
643 23. Rodenko, B., et al., *Generation of peptide-MHC class I complexes through UV-mediated*
644 *ligand exchange.* Nat Protoc, 2006. **1**(3): p. 1120-32.
645 24. Yamamiya, D., et al., *Immune responses of human T lymphocytes to novel hepatitis B*
646 *virus-derived peptides.* PLoS One, 2018. **13**(6): p. e0198264.
647 25. Kreiter, S., et al., *Increased antigen presentation efficiency by coupling antigens to MHC*

class I trafficking signals. J Immunol, 2008. **180**(1): p. 309-18.
26. Ali, M., et al., *Induction of neoantigen-reactive T cells from healthy donors.* Nat Protoc, 2019. **14**(6): p. 1926-1943.
27. Setum, C.M., J.R. Serie, and O.D. Hegre, *Dendritic cell/lymphocyte clustering: morphologic analysis by transmission electron microscopy and distribution of gold-labeled MHC class II antigens by high-resolution scanning electron microscopy.* Anat Rec, 1993. **235**(2): p. 285-95.
28. Steinman, R.M. and Z.A. Cohn, *Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution.* J Exp Med, 1973. **137**(5): p. 1142-62.
29. Tanyi, J.L., et al., *Personalized cancer vaccine effectively mobilizes antitumor T cell immunity in ovarian cancer.* Sci Transl Med, 2018. **10**(436).
30. Zacharakis, N., et al., *Immune recognition of somatic mutations leading to complete durable regression in metastatic breast cancer.* Nat Med, 2018. **24**(6): p. 724-730.
31. Ge, Y., et al., *Blockade of PD-1/PD-L1 immune checkpoint during DC vaccination induces potent protective immunity against breast cancer in hu-SCID mice.* Cancer Lett, 2013. **336**(2): p. 253-9.
32. Glatzova, D. and M. Cebecauer, *Dual Role of CD4 in Peripheral T Lymphocytes*. Front Immunol, 2019. **10**: p. 618.
33. Koning, D., et al., *CD8+ TCR repertoire formation is guided primarily by the peptide component of the antigenic complex.* J Immunol, 2013. **190**(3): p. 931-9.
34. Lancaster, E.M., D. Jablons, and J.R. Kratz, *Applications of Next-Generation Sequencing in Neoantigen Prediction and Cancer Vaccine Development.* Genet Test Mol Biomarkers, 2019.
35. *The problem with neoantigen prediction.* Nat Biotechnol, 2017. **35**(2): p. 97.
36. Stronen, E., et al., *Targeting of cancer neoantigens with donor-derived T cell receptor repertoires.* Science, 2016. **352**(6291): p. 1337-41.
37. Ott, P.A., et al., *An immunogenic personal neoantigen vaccine for patients with melanoma.* Nature, 2017. **547**(7662): p. 217-221.
38. Zhang, X., et al., *Breast Cancer Neoantigens Can Induce CD8(+) T-Cell Responses and Antitumor Immunity.* Cancer Immunol Res, 2017. **5**(7): p. 516-523.
39. Bassani-Sternberg, M. and G. Coukos, *Mass spectrometry-based antigen discovery for cancer immunotherapy.* Curr Opin Immunol, 2016. **41**: p. 9-17.
40. Boehm, K.M., et al., *Predicting peptide presentation by major histocompatibility complex class I: an improved machine learning approach to the immunopeptidome.* BMC Bioinformatics, 2019. **20**(1): p. 7.
41. Cohen, C.J., et al., *Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes.* J Clin Invest, 2015. **125**(10): p. 3981-91.
42. Lin, Q., et al., *Visualizing DC morphology and T cell motility to characterize DC-T cell encounters in mouse lymph nodes under mTOR inhibition.* Sci China Life Sci, 2019.

691 Table 1. The list of the common predicted peptide candidates of CRC

Putative neo-epitope	Gene	Mutant amino acid site	Mutant frequency	Peptide with mutant site*	Affinity (nM)	EPIC score	Role in cancer
GLCE_V533I$_{526-535}$	*GLCE*	V533I	6 out of 321	STIDESP**I**FK	2.7	0.989077	---
C1orf170_S418G$_{413-421}$	*C1orf170*	S418G	5 out of 321	SVAGP**G**PNK	21.7	0.974471	---
MUC3A_I29T$_{28-36}$	*MUC3A*	I29T	6 out of 321	S**T**SQVPFPR	13.8	0.970609	---
CCRL2_F179Y$_{174-183}$	*CCRL2*	F179Y	5 out of 321	ATLPE**Y**VVYK	4.7	0.968889	---
KIAA1683_M313T$_{311-319}$	*KIAA1683*	M313T	6 out of 321	ST**T**TTTPPK	7.9	0.964324	---
KLHL40_N345S$_{341-349}$	*KLHL40*	N345S	5 out of 321	ASLS**S**QVPK	9.4	0.956861	---
MUC3A_S175P$_{172-181}$	*MUC3A*	S175P	6 out of 321	STY**P**MTTTEK	5.8	0.952998	---
RNF43_I47V$_{46-54}$	*RNF43*	I47V	7 out of 321	A**V**IRVIPLK	7.8	0.949232	TSG
SYNE2_A2395T$_{2389-2398}$	*SYNE2*	A2395T	5 out of 321	STQESA**T**VEK	28.9	0.946256	---
TLR10_I369L$_{367-375}$	*TLR10*	I369L	5 out of 321	RT**L**QLPHLK	13.2	0.945037	---
ANKRD36C_N1571S$_{1571-1579}$	*ANKRD36C*	N1571S	5 out of 321	**S**TMEKCIEK	6.6	0.94484	---
IYD_F231I$_{229-237}$	*IYD*	F231I	6 out of 321	QV**I**GKIILK	21.4	0.940062	---
SSX5_E19Q$_{12-20}$	*SSX5*	E19Q	5 out of 321	RVGSQIP**Q**K	35.7	0.936871	---
MUC3A_S326T$_{319-327}$	*MUC3A*	S326T	7 out of 321	TTLPTTI**T**R	16.1	0.936033	---
ARHGEF11_H1427R$_{1427-1435}$	*ARHGEF11*	H1427R	5 out of 321	**R**TIEQLTLK	10.8	0.933929	---
CTNNB1_T41A$_{41-49}$	*CTNNB1*	T41A	6 out of 321	**A**TAPSLSGK	6.9	0.932574	Oncogene
LILRB5_L605F$_{603-611}$	*LILRB5*	L605F	5 out of 321	RS**F**PLTLPR	6.9	0.931333	---
EIF2A_T92S$_{86-95}$	*EIF2A*	T92S	5 out of 321	ATWQPY**S**TSK	12.9	0.931289	---
TMPRSS15_P732S$_{732-740}$	*TMPRSS15*	P732S	5 out of 321	**S**TDGGPFVK	12	0.930946	---
MUC6_P2049L$_{2044-2053}$	*MUC6*	P2049L	9 out of 321	GTVPP**L**TTLK	16.5	0.917963	---
TMEM185B_A42G$_{42-51}$	*TMEM185B*	A42G	5 out of 321	**G**VFAPIWLWK	5.7	0.904248	---
UNC93A_M403T$_{402-410}$	*UNC93A*	M403T	6 out of 321	S**T**FLCVHVK	4.5	0.903415	---
MUC3A_Q31H$_{28-36}$	*MUC3A*	Q31H	5 out of 321	SIS**H**VPFPR	14.1	0.901598	---
FSIP2_R1288Q$_{1285-1293}$	*FSIP2*	R1288Q	5 out of 321	SSL**Q**SQLSK	13.8	0.900572	---
FSIP2_T184NX	*FSIP2*	T184NX	2 out of 321	TTLPKFNKK	10.2	0.961067	---

*: the mutant site was shown in bold and red; TSG: Tumor Suppressor Gene

721 Table 2. The list of the amino acid sequence of mutant minigenes

Constructed K562 cell line	Mutant gene	Amino acid sequence*
CRC-1-K562 cells	GLCE_V533I	NQLQLLSTIDESP**I**FKEFVKRWKSYLK
	C1orf170_S418G	PRKKKVRFSVAGP**G**PNKPGSGQASARP
	MUC3A_I29T	SPWATGTLSTATS**T**SQVPFPRAEAASA
	CCRL2_F179Y	LAWVTAILATLPE**Y**VVYKPQMEDQKYK
	KIAA1683_M313T	VSVTLPQTYPAST**T**TTTPPKTSPVPKV
	KLHL40_N345S	DPAANECYCASLS**S**QVPKNHVSLVTKE
CRC-2-K562 cells	MUC3A_S175P	VTQKPVTTVTSTY**P**MTTTEKGTSAMTS
	RNF43_I47V	AAVESERSAEQKA**V**IRVIPLKMDPTGK
	SYNE2_A2395T	ATSDVQESTQESA**T**VEKLEEDWEINKD
	TLR10_I369L	ANNILTDELFKRT**L**QLPHLKTLILNGN
	ANKRD36C_N1571S	QMKDIEKMYKSGY**S**TMEKCIEKQERFC
	IYD_F231I	SIACGILLAALQV**I**GKIILKELALISF
CRC-3-K562 cells	SSX5_E19Q	AFVRRPRVGSQIP**Q**KMQKHPWRQVCDR
	MUC3A_S326T	PLSTLVTTLPTTI**T**RSTPTSETTYTTS
	ARHGEF11_H1427R	PPSLALRDVGMIF**R**TIEQLTLKLNRLK
	CTNNB1_T41A	QSYLDSGIHSGAT**A**TAPSLSGKGNPEE
	LILRB5_L605F	SPGPQASPPPPRS**F**PLTLPRCRLLHLK
	EIF2A_T92S	SPKNTVLATWQPY**S**TSKDGTAGIPNLQ
CRC-4-K562 cells	TMPRSS15_P732S	LGLGSGNSSKPIF**S**TDGGPFVKLNTAP
	MUC6_P2049L	ASIHSTPTGTVPP**L**TTLKATGSTHTAP
	TMEM185B_A42G	PLRLDGIIQWSYW**G**VFAPIWLWKLLVV
	UNC93A_M403T	WEALGFVIAFGYS**T**FLCVHVKLYILLG
	MUC3A_Q31H	WATGTLSTATSIS**H**VPFPRAEAASAVL
	FSIP2_R1288Q	ICPKLHMGFKSSL**Q**SQLSKYTAKIVNI
CRC-5-K562 cells	TMEM185B_A42G	PLRLDGIIQWSYW**G**VFAPIWLWKLLVV
	UNC93A_M403T	WEALGFVIAFGYS**T**FLCVHVKLYILLG
	MUC3A_Q31H	WATGTLSTATSIS**H**VPFPRAEAASAVL

	FSIP2_R1288Q	ICPKLHMGFKSSL**Q**SQLSKYTAKIVNI
	FSIP2_T184NX	LKTRSKITTLPKFNKKNTLRTECC
	KRAS_G12V	MTEYKLVVVGA**V**GVGKSALTIQLIQ

*: the mutant site was shown in bold and red

Table 3. Top 2 TCR α and β chain pairs from two individual TCR repertoire

neo-epitope	ID of paired TCR	CDR3 sequence	TRV gene	J gene	proportion
C1orf170_S418G$_{413-421}$	Clonotype1	TRA: CAASGGAQKLVF	TRAV29DV5	TRAJ54	53.2%
		TRA: CAGLLYNSGNTPLVF	TRAV35	TRAJ29	
		TRB: CASSRDRGSNQPQHF	TRBV27	TRBJ1-5	
	Clonotype2	TRA: CAASGGAQKLVF	TRAV29DV5	TRAJ54	31.6%
		TRB: CASSRDRGSNQPQHF	TRBV27	TRBJ1-5	
KRAS_G12V$_{8-16}$	Clonotype1	TRA: CASNDYKLSF	TRAV8-3	TRAJ20	80.5%
		TRB: CASSLDGVSYEQYF	TRBV11-2	TRBJ2-7	
	Clonotype2	TRA: ---	---	---	14.3%
		TRB: CASSLDGVSYEQYF	TRBV11-2	TRBJ2-7	

Figure 1

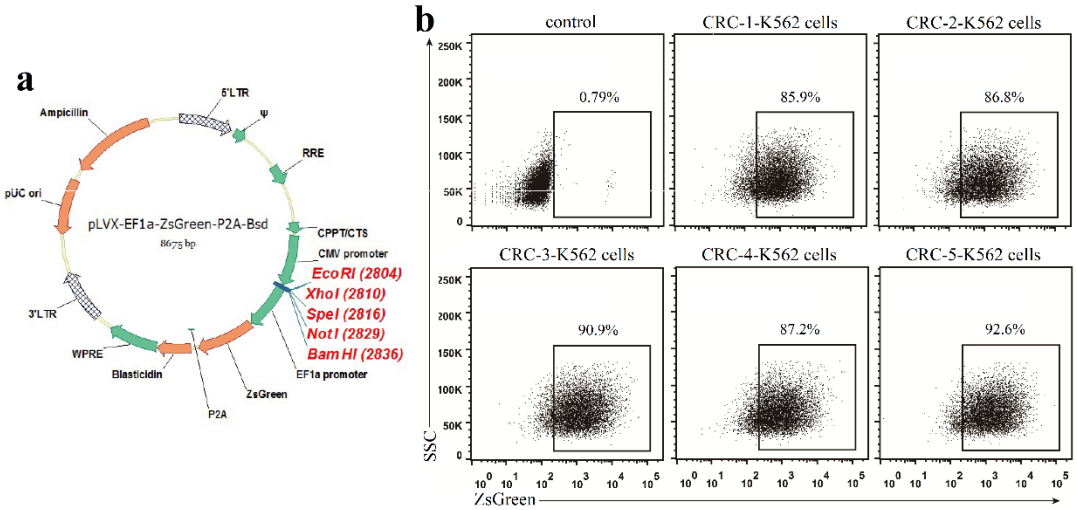

a
5'LTR
Ψ
RRE
CPPT/CTS
CMV promoter
EcoRI (2804)
XhoI (2810)
SpeI (2816)
NotI (2829)
BamHI (2836)
EF1a promoter
ZsGreen
P2A
Blasticidin
WPRE
3'LTR
pUC ori
Ampicillin
pLVX-EF1a-ZsGreen-P2A-Bsd
8675 bp
b
control
CRC-1-K562 cells
CRC-2-K562 cells
CRC-3-K562 cells
CRC-4-K562 cells
CRC-5-K562 cells
0.79%
85.9%
86.8%
90.9%
87.2%
92.6%
SSC
ZsGreen

Figure 2

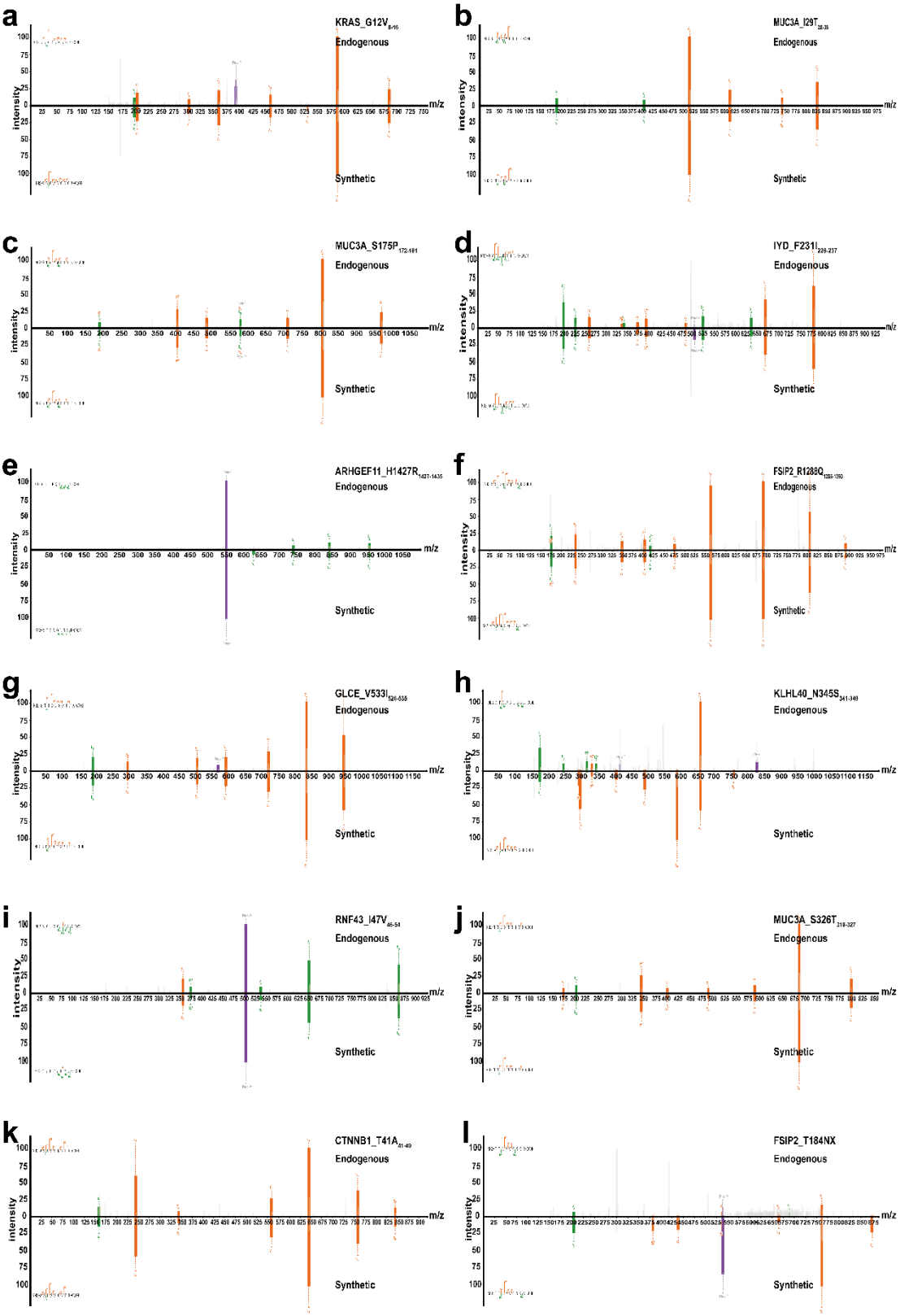

a
KRAS_G12V
Endogenous
Synthetic
b
MUC3A_I29T
Endogenous
Synthetic
c
MUC3A_S175P
Endogenous
Synthetic
d
IYD_F231I
Endogenous
Synthetic
e
ARHGEF11_H1427R
Endogenous
Synthetic
f
FSIP2_R1288Q
Endogenous
Synthetic
g
GLCE_V533I
Endogenous
Synthetic
h
KLHL40_N345S
Endogenous
Synthetic
i
RNF43_I47V
Endogenous
Synthetic
j
MUC3A_S326T
Endogenous
Synthetic
k
CTNNB1_T41A
Endogenous
Synthetic
l
FSIP2_T184NX
Endogenous
Synthetic
intensity
m/z

Figure 3

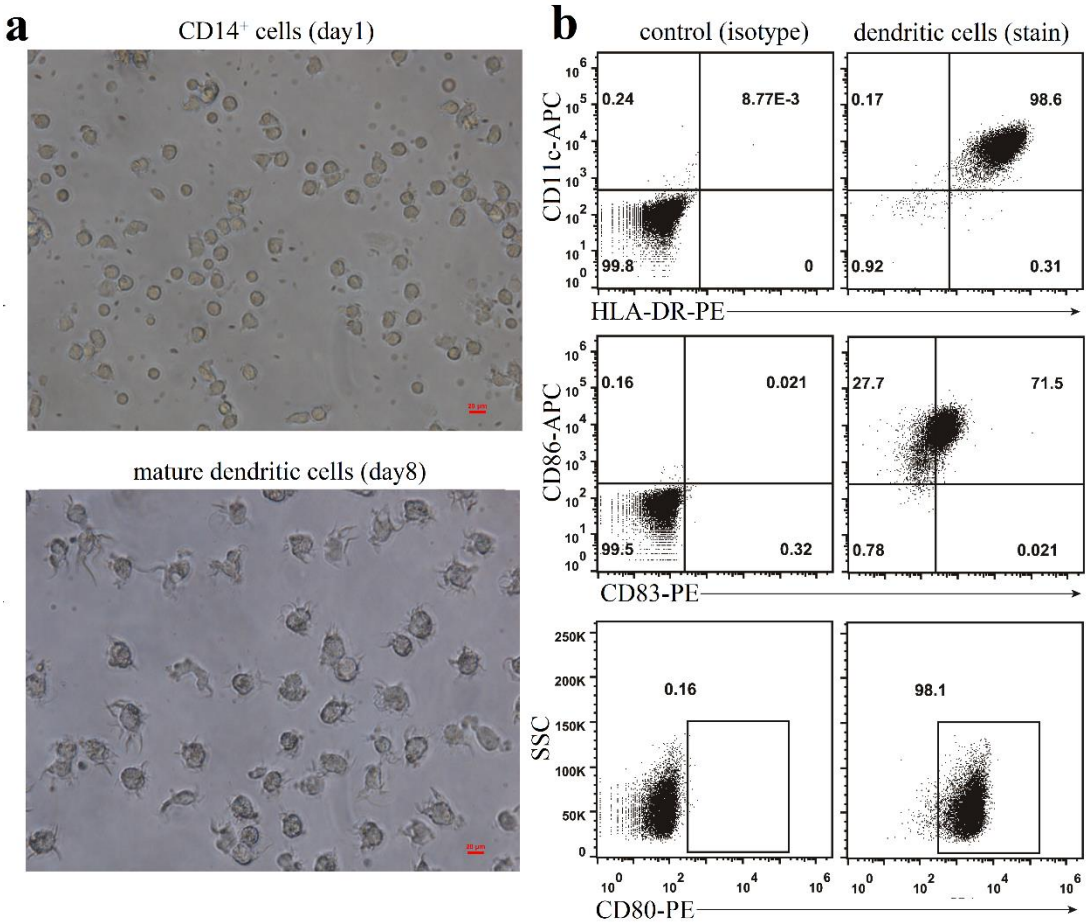
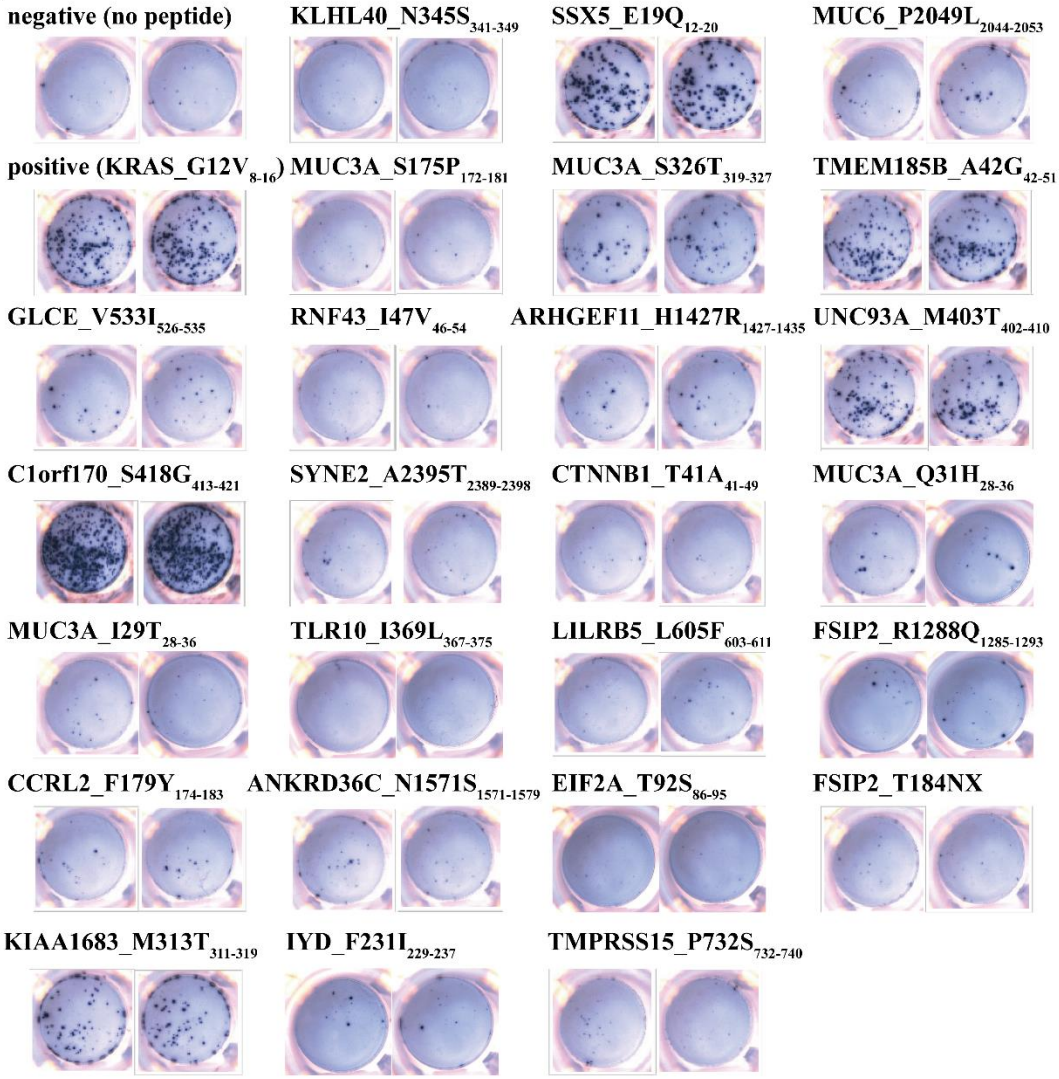

a
CD14+ cells (day1)
mature dendritic cells (day8)
b
control (isotype)
dendritic cells (stain)
CD11c-APC
HLA-DR-PE
0.24
8.77E-3
99.8
0
0.17
98.6
0.92
0.31
CD86-APC
CD83-PE
0.16
0.021
99.5
0.32
27.7
71.5
0.78
0.021
SSC
CD80-PE
0.16
98.1

Figure 4

a


negative (no peptide)
KLHL40_N345S341-349
SSX5_E19Q12-20
MUC6_P2049L2044-2053
positive (KRAS_G12V8-16)
MUC3A_S175P172-181
MUC3A_S326T319-327
TMEM185B_A42G42-51
GLCE_V533I526-535
RNF43_I47V46-54
ARHGEF11_H1427R1427-1435
UNC93A_M403T402-410
C1orf170_S418G413-421
SYNE2_A2395T2389-2398
CTNNB1_T41A41-49
MUC3A_Q31H28-36
MUC3A_I29T28-36
TLR10_I369L367-375
LILRB5_L605F603-611
FSIP2_R1288Q1285-1293
CCRL2_F179Y174-183
ANKRD36C_N1571S1571-1579
EIF2A_T92S86-95
FSIP2_T184NX
KIAA1683_M313T311-319
IYD_F231I229-237
TMPRSS15_P732S732-740

b

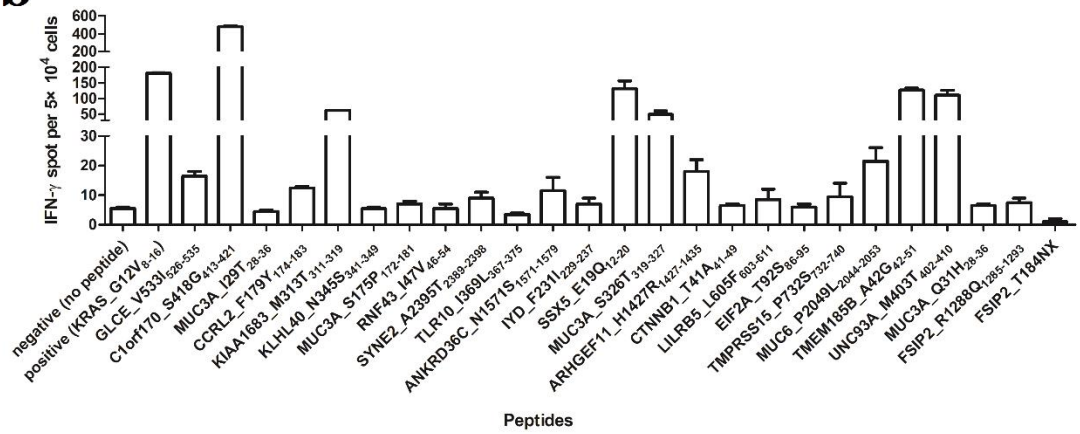
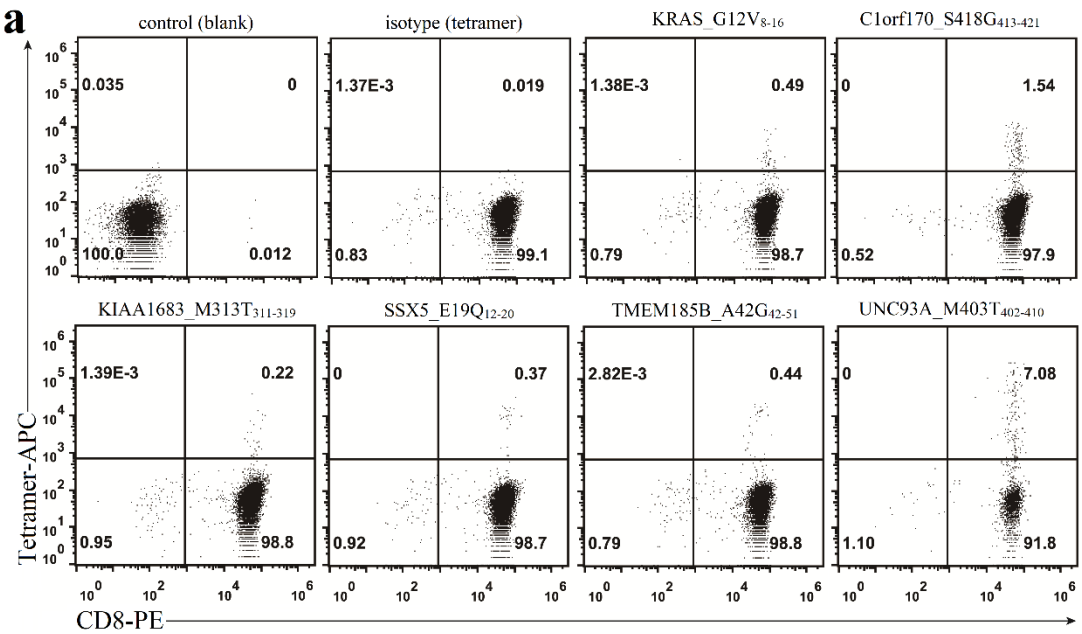

IFN-γ spot per 5× 10^4 cells
600
400
200
200
150
100
50
30
20
10
0
negative (no peptide)
positive (KRAS_G12V8-16)
GLCE_V533I526-535
C1orf170_S418G413-421
MUC3A_I29T28-36
CCRL2_F179Y174-183
KIAA1683_M313T311-319
KLHL40_N345S341-349
MUC3A_S175P172-181
RNF43_I47V46-54
SYNE2_A2395T2389-2398
TLR10_I369L367-375
ANKRD36C_N1571S1571-1579
IYD_F231I229-237
SSX5_E19Q12-20
MUC3A_S326T319-327
ARHGEF11_H1427R1427-1435
CTNNB1_T41A41-49
LILRB5_L605F603-611
EIF2A_T92S86-95
TMPRSS15_P732S732-740
MUC6_P2049L2044-2053
TMEM185B_A42G42-51
UNC93A_M403T402-410
MUC3A_Q31H28-36
FSIP2_R1288Q1285-1293
FSIP2_T184NX
Peptides

Figure 5


a
control (blank)
isotype (tetramer)
KRAS_G12V8-16
C1orf170_S418G413-421
KIAA1683_M313T311-319
SSX5_E19Q12-20
TMEM185B_A42G42-51
UNC93A_M403T402-410
0.035
0
100.0
0.012
1.37E-3
0.019
0.83
99.1
1.38E-3
0.49
0.79
98.7
0
1.54
0.52
97.9
1.39E-3
0.22
0.95
98.8
0
0.37
0.92
98.7
2.82E-3
0.44
0.79
98.8
0
7.08
1.10
91.8
Tetramer-APC
CD8-PE

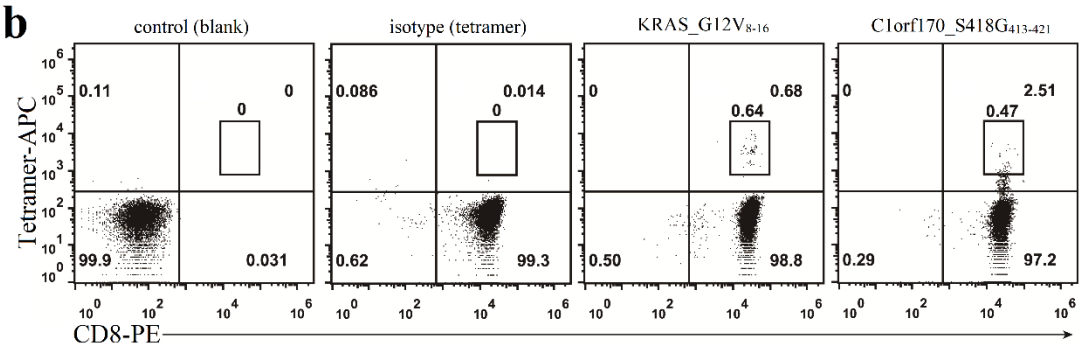

b
control (blank)
isotype (tetramer)
KRAS_G12V8-16
C1orf170_S418G413-421
0.11
0
0
99.9
0.031
0.086
0.014
0
0.62
99.3
0
0.68
0.64
0.50
98.8
0
2.51
0.47
0.29
97.2
Tetramer-APC
CD8-PE

Figure 6

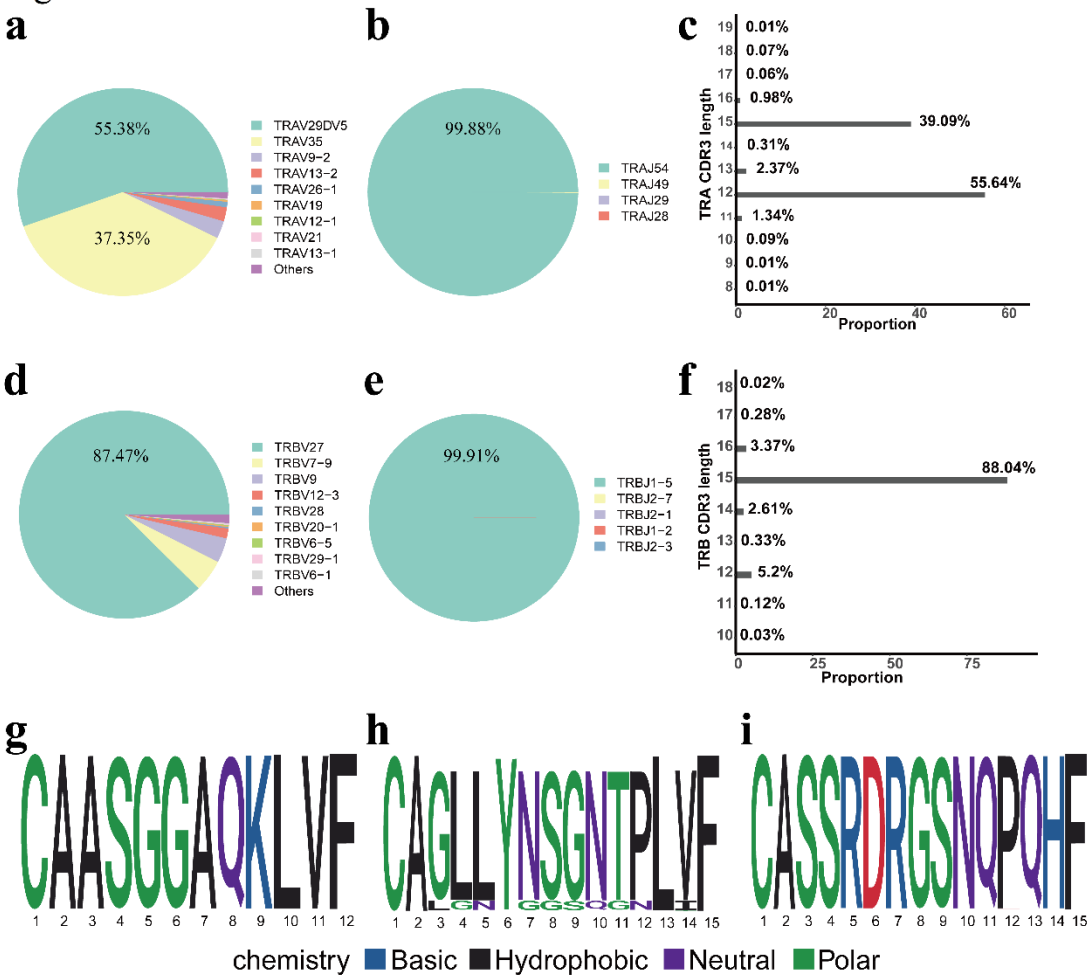

a
55.38%
37.35%
TRAV29DV5
TRAV35
TRAV9-2
TRAV13-2
TRAV26-1
TRAV19
TRAV12-1
TRAV21
TRAV13-1
Others
b
99.88%
TRAJ54
TRAJ49
TRAJ29
TRAJ28
c
TRA CDR3 length
19 0.01%
18 0.07%
17 0.06%
16 0.98%
15 39.09%
14 0.31%
13 2.37%
12 55.64%
11 1.34%
10 0.09%
9 0.01%
8 0.01%
0 20 40 60
Proportion
d
87.47%
TRBV27
TRBV7-9
TRBV9
TRBV12-3
TRBV28
TRBV20-1
TRBV6-5
TRBV29-1
TRBV6-1
Others
e
99.91%
TRBJ1-5
TRBJ2-7
TRBJ2-1
TRBJ1-2
TRBJ2-3
f
TRB CDR3 length
18 0.02%
17 0.28%
16 3.37%
15 88.04%
14 2.61%
13 0.33%
12 5.2%
11 0.12%
10 0.03%
0 25 50 75
Proportion
g
CAASGGAQKLVF
1 2 3 4 5 6 7 8 9 10 11 12
h
CAGLLYNSGNTPLVF
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
i
CASSRDRGSNQPQHF
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15
chemistry Basic Hydrophobic Neutral Polar

Figure 7

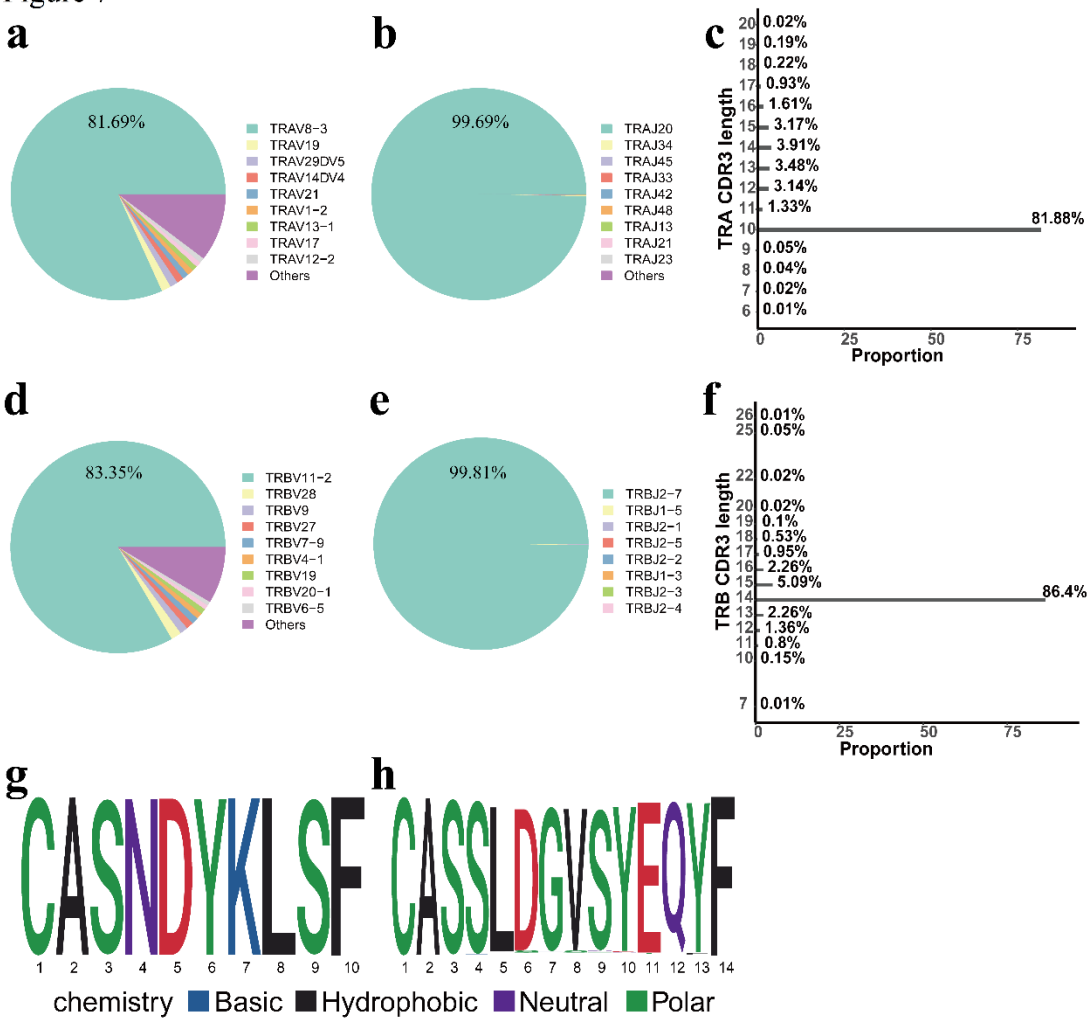

a
81.69%
TRAV8-3
TRAV19
TRAV29DV5
TRAV14DV4
TRAV21
TRAV1-2
TRAV13-1
TRAV17
TRAV12-2
Others
b
99.69%
TRAJ20
TRAJ34
TRAJ45
TRAJ33
TRAJ42
TRAJ48
TRAJ13
TRAJ21
TRAJ23
Others
c
TRA CDR3 length
20 0.02%
19 0.19%
18 0.22%
17 0.93%
16 1.61%
15 3.17%
14 3.91%
13 3.48%
12 3.14%
11 1.33%
10 81.88%
9 0.05%
8 0.04%
7 0.02%
6 0.01%
0 25 50 75
Proportion
d
83.35%
TRBV11-2
TRBV28
TRBV9
TRBV27
TRBV7-9
TRBV4-1
TRBV19
TRBV20-1
TRBV6-5
Others
e
99.81%
TRBJ2-7
TRBJ1-5
TRBJ2-1
TRBJ2-5
TRBJ2-2
TRBJ1-3
TRBJ2-3
TRBJ2-4
f
TRB CDR3 length
26 0.01%
25 0.05%
22 0.02%
20 0.02%
19 0.1%
18 0.53%
17 0.95%
16 2.26%
15 5.09%
14 86.4%
13 2.26%
12 1.36%
11 0.8%
10 0.15%
7 0.01%
0 25 50 75
Proportion
g
CASNDYKLSF
1 2 3 4 5 6 7 8 9 10
h
CASSLDGVSYEQYF
1 2 3 4 5 6 7 8 9 10 11 12 13 14
chemistry Basic Hydrophobic Neutral Polar

Figure legends

Figure 1. The construction and expression of tandem minigenes encoding predicted neoantigens in K562 cells. (a) The map of the adopted lentiviral vector. The lentiviral vector was gifted and possessed two promoters, of which CMV promoter operated the inserted genes in MCS, and EF1a promoter operated the reporter gene ZsGreen and a screening gene Blasticidine. (b) FACS detected the expression of tandem minigenes in K562 cells. 25 predicted peptides and KRAS_G12V were constructed into five tandem minigenes, packaged into lentivirus and transfected into mono HLA-A*11:01 allelic K562 cells, which resulted in five K562 cell lines, including CRC-1-K562 cells, CRC-2-K562 cells, CRC-3-K562 cells, CRC-4-K562 cells and CRC-5-K562 cells, were obtained (Table 2). The blasticidine-resistance K562 cells expressed ZsGreen, which indirectly reflected the expression of predicted peptides and could be directly tested by FACS.

Figure 2. Validation of naturally presented neo-epitopes from constructed K562 cells. The mirror plot of mass spectrometry (MS/MS) chromatographs of the synthetic peptides (bottom) versus theirs experimentally identified analogs (top), including (a) KRAS_G12V$_{8\text{-}16}$, (b) MUC3A_I29T$_{28\text{-}36}$, (c) MUC3A_S175P$_{172\text{-}181}$, (d) IYD_F231I$_{229\text{-}237}$, (e) ARHGEF11_H1427R$_{1427\text{-}1435}$, (f) FSIP2_R1288Q$_{1285\text{-}1293}$, (g) GLCE_V533I$_{526\text{-}535}$, (h) KLHL40_N345S$_{341\text{-}349}$, (i) RNF43_I47V$_{46\text{-}54}$, (j) MUC3A_S326T$_{319\text{-}327}$, (k) CTNNB1_T41A$_{41\text{-}49}$, and (l) FSIP2_T184NX, were exhibited.

Figure 3. The morphology and the phenotype of mature DCs. (a) The morphology of monocytes and mature dendritic cells. CD14^{+} monocytes on day1 and mature dendritic cells on day8 were observed using inverted phase contrast microscope and imaged by camera (Nikon ECLIPSE TS100) under 40× object lens (scale bar=20 μm). (b) The phenotype of mature dendritic cells. Mature dendritic cells on day8 were

collected and stained with fluorescein conjugated antibodies targeting CD11c, HLA-DR, CD86, CD83 and CD80, and the fluorescence signal was detected by FACS.

Figure 4. The immune response of CTL to predicted peptide-pulsed T2 cells. (a) The immune-spot diagram of IFN-γ. CTL, which had been co-cultured with peptide-pulsed mDCs, were re-stimulated by peptide-pulsed T2 cells in the IFN-γ ELISPOT plate for 24 hours. The secreted IFN-γ from epitope specific CTL was exhibited as immune spot. There were 25 predicted peptides (Table 1) and a reported positive peptide KRAS_G12V$_{8-16}$ to be tested, and each was in duplicate wells. (b) The statistical number of IFN-γ immune spot from (a). Spots were imaged and counted by an ELISPOT Reader (BioReader 4000, BIOSYS).

Figure 5. The proportion of epitope-specific CTL. Top 6 immunogenic peptides of KRAS_G12V$_{8-16}$, C1orf170_S418G$_{413-421}$, KIAA1683_M313T$_{311-319}$, SSX5_E19Q$_{12-20}$, TMEM185B_A42G$_{42-51}$ and UNC93A_M403T$_{402-410}$ were used to prepare pMHC tetramer. CTL, which had been co-cultured with peptide-loaded mDCs and confirmed to respond to peptide re-stimulation, were stained with CD8-PE and tetramer-APC. The fluorescence signal from cells was analyzed by FACS (a), and the CD8^{+}pMHC tetramer^{+} cells that respectively responded to KRAS_G12V$_{8-16}$ and C1orf170_S418G$_{413-421}$ were sorted for single-cell RNA sequencing (b).

Figure 6. The characteristics of TCR repertoire recognized HLA-A*11:01 presenting C1orf170_S418G$_{413-421}$. The frequency of usage gene of TRAV (a) and TRBV (d). The pie charts showed the top 9 usage genes, and the remaining usage genes were marked as “others”. The frequency of all the usage gene of TRAJ rearranged with TRAV29DV5 (b) and TRBJ rearranged with TRBV27 (e). The frequency of length distribution of CDR3α (c) and CDR3β (f). The sequence motif of 12-mer CDR3α (g), 15-mer CDR3α (h) and 15-mer CDR3β (i).

Figure 7. The characteristics of TCR repertoire recognized HLA-A*11:01 presenting KRAS_G12V$_{8-16}$. The frequency of usage gene of TRAV (a) and TRBV (d). The pie charts showed the top 9 usage genes, and the remaining usage genes were marked as "others". The frequency of all the usage gene of TRAJ rearranged with TRAV8-3 (b) and TRBJ rearranged with TRBV11-2 (e). The frequency of length distribution of CDR3α (c) and CDR3β (f). The sequence motif of 10-mer CDR3α (g) and 14-mer CDR3β (h).