ORIGINAL ARTICLE

T cell neoepitope discovery in colorectal cancer by high throughput profiling of somatic mutations in expressed genes

Daniele Mennonna,[1] Cristina Maccalli,[2] Michele C Romano,[1] Claudio Garavaglia,[1] Filippo Capocefalo,[2] Roberta Bordoni,[3] Marco Severgnini,[3] Gianluca De Bellis,[3] John Sidney,[4] Alessandro Sette,[4] Alessandro Gori,[5] Renato Longhi,[5] Marco Braga,[6] Luca Ghirardelli,[6] Ludovica Baldari,[6] Elena Orsenigo,[6] Luca Albarello,[7] Elisabetta Zino,[8] Katharina Fleischhauer,[8,9] Gina Mazzola,[10] Norma Ferrero,[10] Antonio Amoroso,[10] Giulia Casorati,[1] Giorgio Parmiani,[2] Paolo Dellabona[1]

ABSTRACT

Objective Patient-specific (unique) tumour antigens, encoded by somatically mutated cancer genes, generate neoepitopes that are implicated in the induction of tumour-controlling T cell responses. Recent advancements in massive DNA sequencing combined with robust T cell epitope predictions have allowed their systematic identification in several malignancies.

Design We undertook the identification of unique neoepitopes in colorectal cancers (CRCs) by using high-throughput sequencing of cDNAs expressed by standard cancer cell cultures, and by related cancer stem/initiating cells (CSCs) cultures, coupled with a reverse immunology approach not requiring human leukocyte antigen (HLA) allele-specific epitope predictions.

Results Several unique mutated antigens of CRC, shared by standard cancer and related CSC cultures, were identified by this strategy. CD8[+] and CD4[+] T cells, either autologous to the patient or derived from HLA-matched healthy donors, were readily expanded in vitro by peptides spanning different cancer mutations and specifically recognised differentiated cancer cells and CSC cultures, expressing the mutations. Neoepitope-specific CD8[+] T cell frequency was also increased in a patient, compared with healthy donors, supporting the occurrence of clonal expansion in vivo.

Conclusions These results provide a proof-of-concept approach for the identification of unique neoepitopes that are immunogenic in patients with CRC and can also target T cells against the most aggressive CSC component.

Significance of this study

What is already known on this subject?

► It is now well established that T lymphocytes play a critical role in controlling cancer progression.
► T lymphocytes recognised peptides, called epitopes, derived from tumour associated protein antigens.
► Epitopes derived from mutated cancer proteins are known to elicit strong antitumour T cell responses that correlate with clinical responses.
► Recent advancement in high throughput DNA sequencing techniques, in combination with the in silico prediction of T cell epitopes, have allowed the massive identification of mutated neoepitopes in melanoma, cholangiocarcionoma and chronic lymphocytic leukaemia (CLL).

What are the new findings?

► We have implemented a method to identify T cell mutated neoepitopes based on the massive parallel sequencing of expressed genes coupled with an immunology approach not requiring HLA allele-specific epitope predictions.
► This method allowed the identification of several mutated neoepitopes from colorectal cancer, the second cause of cancer death.
► We provide evidence supporting a spontaneous activation and expansion of patient's T cell specific for a mutated neoepitope expressed by the autologous tumour.
► Finally, our study also reveals that colon cancer stem/initiating cells, a subpopulation of cells that is supposed to drive tumour initiation, propagation and metastasis, express the mutated genes and are targeted by the neoepitope-specific T cells.

INTRODUCTION

Recent clinical results obtained with adoptive T cell therapy or immune checkpoint blockade by monoclonal antibody (mAbs) provide compelling evidence for spontaneous immunosurveillance and T cell mediated regression of human cancers.[1–4] T lymphocytes recognise epitopes derived from the processing of tumour-derived protein antigens and presented by major histocompatibility complex (MHC) molecules displayed on cancer cells.[5] Tumour associated antigens are encoded either by non-mutated genes, shared by different tumours, or by genes undergoing somatic mutations in cancer cells.[5] Somatically mutated cancer genes generate

Significance of this study

How might it impact on clinical practice in the foreseeable future?

▶ The systematic identification of mutated neoepitopes in colon cancer may provide new prognostic/predictive approaches based on the determination of specific T cell responses in patients with colorectal cancer, as well as prompt more efficacious immunotherapy strategy that can target T cells against the most aggressive cancer stem/ initiating cells component.

neoepitopes, unique to each tumour, which can induce tumour rejection in mice and appear to dominate the specificity of T cell responses to autologous mouse or human tumours.[6–8] The lack of suitable technologies for massive identification of unique cancer neoepitopes has prevented the systemic analyses of T cell responses specific for such epitopes and their exploitation in cancer immunotherapy. Recent advancements in high throughput DNA sequencing overcome these limits and provide powerful tools for the systemic identification of somatic mutations in cancer genes.[9 10] This information can be used to derive patients' specific mutated protein sequences, which predict synthetic peptides that bind patients' MHC and can be tested for T cell recognition in large-scale reverse immunology approaches.[11] This strategy has recently allowed a systematic definition of T cell responses specific for unique neoepitopes in mouse and human cancers, highlighting their relevant role in the tumour control achieved by active vaccination, adoptive T cell therapy or immune checkpoint blockade.[12–20]

Colorectal cancer (CRC) is the second cause of cancer death and responds poorly to current therapies. Average CRCs carry from 70 to more than 1000 non-synonymous exonic mutations per gene, depending on whether they are microsatellite stable or instable.[21] About 35 of such mutations recurrently affect expressed genes that are likely driving the oncogenic process (candidate cancer genes, CAN-gene).[22–24] T cell infiltration of CRC is a strong positive prognostic parameter,[25 26] implying that this cancer undergoes active immunosurveillance. The antigenic targets of CRC infiltrating T cells are not known and it is conceivable that they are formed, at least in part, by unique neoepitopes.

CRCs contain a small subpopulation of cells that display stem-cell like properties driving tumour initiation, propagation and metastasis.[27] Cancer stem/initiating cells (CSCs) are considered the critical targets for therapy, because their elimination is expected to completely halt cancer progression. CSCs exhibit immunosuppressive effects that may hamper the induction of T cell responses; however, they can be recognised and eliminated by activated T cells.[28]

In light of these considerations, hence, relevant questions are whether T cell recognition of unique epitopes occurs in CRC, and whether these epitopes can also target T cell responses against CSCs. To address these questions, we set up a proof of concept platform to systematically identify unique neoepitopes from somatically mutated CAN-genes expressed by CRC cells and in the derived CSC cultures. The tumour-derived cDNAs encoding the 20 most frequently mutated CAN-genes in CRC[22] were subjected to high throughput sequencing to identify mutations in the expressed genes. To avoid the need for precise

bioinformatic prediction and assay of all the possible mutated epitopes that can potentially bind each tumour HLA allele, we tested the ability of pools of long synthetic peptides, spanning the CAN-gene mutations, to elicit T cell responses that recognise the differentiated cancer cells and the CSCs expressing the targeted mutations. Following this approach, we identified unique immunogenic neoepitopes in CRCs and showed that they can target T cells against the CSC component.

MATERIALS AND METHODS

Establishment of tumour cells cultures
Peripheral blood mononuclear cells (PBMCs) were obtained from patients with CRC or HLA-matched healthy donors (HDs) by standard Ficoll separation (Ficoll-Paque PLUS, GE Healthcare Bio-Science). Differentiated and CSC cell lines were generated from surgical specimens as described in online supplementary methods. To collect tumour sample and peripheral blood, written informed consent in accordance with the Declaration of Helsinki was obtained from patients.

PCR amplification of CAN-gene cDNAs
cDNA synthesised from CRC cell poly(A) RNA was PCR amplified using primers specific for each CAN-gene (see online supplementary table S4). The PCR products were gel purified and equalised on Nanodrop before pooling and sequencing.

Massive parallel cDNA sequencing
Amplified cDNA pools (3 μg) were processed for massive sequencing according to the GS FLX Titanium protocol (454 Life Sciences, Roche, Branfort, Connecticut, USA), as detailed in online supplementary methods.

PCR assay
DNA extracted from PBMCs or B lymphoblastoid cell lines (LCLs) obtained from the patients with CRC was PCR amplified using specific primers designed around each autochthonous mutation. PCR products were gel purified and directly sequenced by Sanger method.

MHC-peptide binding analyses
Quantitative assays to measure the binding of peptides to purified HLA A*02:01 class I molecules were performed as described previously[29] and detailed in online supplementary methods.

Retroviral transduction of mutated and WT SMAD4 minigenes
Two 27 aa long minigenes encoding either the SMAD4^{V370A} mutation expressed by the 1247 CRC, or the corresponding SMAD4$^{V370-WT}$ residue, were cloned in the retroviral vector MSCV-IRES-GFP and transduced into HLA-A*02:01^{+} HEK293t human embryo kidney cells that were selected by cell sorting to express high levels of green fluorescence protein (GFP) (detailed in see online supplementary methods).

PCR typing of mutated and WT SMAD4
The indicated tumour cell lines were screened by RT-PCR typing for the expression of either SMAD4^{V370A} or SMAD4^{R361C} mutations, or the corresponding wild type (WT) sequence (see online supplementary methods).

Flow cytometry and CD8^{+} T cell enrichment
Cancer cells, pretreated with interferon (IFNγ) for 48 h, were stained with anti-HLA class I W6/32 and anti-HLA-DR L243

mAbs. T cell lines expanded from patients and HDs were stained with anti-CD3 fluorescein isothiocyanate (FITC), antihuman CD4 phycoerythrin (PE), antihuman CD8 antigen presenting cell (APC) mAbs (Becton Dickinson), 4',6-diamidino-2-phenylindole (DAPI) and acquired on a Canto II (Becton Dickinson). Results on viable cells were analysed using Flow-Jo software (Treestar).

T cell cultures

T cell lines and mixed lymphocyte-tumour cell culture (MLTC) were generated from PBMCs as described[28 30] and detailed in online supplementary methods.

ELISPOT assays

ELISPOT assay for IFNγ production by unique neoantigen specific T cells were performed as described[28] and detailed in online supplementary methods.

Statistical analysis

Comparisons between two groups were done with the two-tailed parametrical Student's t test for unpaired samples, multiple comparisons were done by one-way analysis of variance. Statistics were calculated using GraphPadTM Prism V.5.0 (GraphPad Software). Differences with a p value <0.05 were considered statistically significant. *p Value <0.05; **p value <0.01; ***p value <0.001.

RESULTS
Identification of somatic mutations in expressed genes of CRC and CSC cultures by massively parallel tumour cDNA sequencing

We first sought to identify unique antigens in cell lines that were established from primary surgical specimens of patients with CRC (see online supplementary table S1 and S2).[28] Single cell suspensions from cancer specimens were cultured in standard conditions to obtain 'differentiated' cancer cells and, when possible, also in serum-free conditions to support the generation of colon spheres displaying CSC characteristics.[28] The cDNAs encoding the 20 most frequently mutated CAN-genes in CRC[22] were PCR-amplified from eight differentiated CRC cell lines and two parallel CSC cultures and subjected to massively parallel sequencing. We found somatic mutations in 3–5 of the 20 expressed genes in all CRC cells (table 1).

The mutations found in the CRC cDNAs were lacking in the corresponding gene exons present in the DNA obtained from healthy cells (PBMCs or LCLs) of the same patients, confirming that they were somatically acquired (data not shown). cDNAs encoding oncogenes were mutated in about 50% of the obtained sequences, consistent with their dominant functions in the presence of a WT allele, with the exception of KRAS that was mutated in 100% of the reads in three of seven tumour samples. cDNAs encoding oncosuppressors were mutated in about 100% of the obtained sequences, consistent with the loss of heterozygosity state required for their functional loss. Two exceptions to this finding were the APC and SMAD4 cDNAs expressed by the 1247 CRC/CSC samples, which exhibited four (three non-sense, one miss-sense) and two (one missense, one nonsense) mutations, respectively, suggesting that both alleles of each oncosuppressor gene were expressed and carried mutations that either prevented the expression of the encoded APC protein, or that resulted in a non-functional SMAD4 protein.

Five genes (APC, KRAS, TP53, PIK3CA, FBXW7) were recurrently mutated in the majority of samples, whereas the other 15 genes were more rarely mutated, consistent with the published

data.[22 24] The identified mutations in the APC, KRAS, TP53, PIK3CA FBXW7, SMAD4 genes were already described in the catalogue of somatic mutations in cancer (COSMIC) (http://cancer.sanger.ac.uk) database of cancer gene mutations, with the exceptions of the: 1. frameshifts APC^{S139fs*2}, ^{V915fs*2}, ^{E1317fs*3} mutations in CRCs 1039, 1076 and 1869, respectively; 2. missense PIK3CA^{R770Q} and SMAD4^{V370A} mutations in CRC 1247. All the additional somatic mutations in the 15 more rarely mutated genes were apparently newly identified and unique to the expressing CRC samples. The great majority (30/38, 79%) of somatic mutations found in the CRC samples produced modified amino acid sequences of the encoded proteins, as a result of missense mutations generating a new amino acid residue, or frameshift mutations introducing novel open reading frames at the C-terminal protein sequence. A few nonsense mutations introduced a stop codon in the APC (R1114*, R1450*, R2204*) gene, in one SMAD4 (E41*) allele of the 1247CRC/CSC samples, and in the APC (E893*) and SMAD2 (G457*) genes of the 21 052 CRC. Finally, each pair of differentiated and CSC cultures from either 1076 or 1247 CRCs harboured the same mutations.

Hence, the sequencing of the 20 most frequently mutated genes in CRC provided four to five somatic mutations per tumour, which were a potential source of unique T cell neoepitopes.

Patients' CD8$^+$ T cells induced by a mutated SMAD4 peptide recognise autologous cancer cells and CSCs

We collected enough PBMCs from patients 1247 and 1039 to investigate T cell recognition of epitopes derived from the mutated gene products. PBMCs from the two patients were stimulated at least twice in vitro with pools of synthetic peptides consisting, for each mutated protein, of three 15 aa long peptides spanning the mutated residues and overlapping by 11 residues (see online supplementary table S3). This approach is based on the evidence that 15 aa long peptides are naturally processed by APCs into epitopes that are presented by MHC class I and II molecules to autologous T cells, without prior knowledge of the exact HLA allele-specific epitope structure.[31] CD8$^+$ T cells isolated from the stimulated PBMCs were tested for the recognition of the autologous cancer cells, which expressed HLA-A, B, C and HLA-DR upon IFNγ treatment (figure 1A). In patient 1247, peptide pools covering the PIK3CA^{R770Q} and C10orf127^{S168L} mutations did not elicit CD8$^+$ T cells able to recognise the autologous cancer cells (data not shown). Remarkably, however, we found specific recognition of differentiated and CSC cultures by CD8$^+$ T cells induced with the peptide pool encompassing the SMAD4^{V370A} mutation (figure 1B). In preliminary experiments, we found that CD8$^+$ T cells induced with the SMAD4-mut peptide pool were specific for the SMAD4mutated-1 (SMAD4m-1) peptide. Because patient 1247 expressed HLA-A*0201 (see online supplementary table S2), we assessed whether the SMAD4m-1 peptide was presented by this HLA allele. The CD8$^+$ T cells recognised the SMAD4m-1 peptide presented by HLA-A*02:01$^+$ T2 cells (figure 1C), but not the other two mutated peptides, or the peptide spanning the WT SMAD4 sequence corresponding to the SMAD4m-1 sequence (figure 1C,D). To further confirm the specificity of the peptide-induced CD8$^+$ T cells for the SMAD4^{V370A}-containing neoepitope, we transduced HLA-A*0201$^+$ HEK293t cells with two minigenes encompassing the WT or the SMAD4^{V370A} sequences, respectively (figure 1E). The SMAD4m-1 specific CD8$^+$ T cells specifically recognised HEK293t cells transduced with the SMAD4^{V370A}-mutated

Table 1 Mutated CAN-genes found in eight massively sequenced CRC cell lines

CAN genes	Tumour cell lines									
	1039	1076 Diff CSC		1247 Diff CSC		1546	1869	1872	21 052	23 873
APC	S1395fs*2	V915fs*2	V915fs*2	R1114*d (44%)e R1450* (52%) R2204* (33%) A2650V (58%)	R1114* (44%) R1450* (52%) R2204* (33%) A2650V (58%)	Q1328*	E1317fs*3		E893*	
KRAS	G12D	G12D	G12D			Q61L	G13D	G12D	G12C	G13D
TP53	Y107D	R248W	R248W			R175H	R175H	R248Q	R158L	
PIK3CA	Q546K			R770Q	R770Q					E545K
FBXW7							R479Q			H379Y
CSMD3									L2424fs*9	
TNN										
NAV3										
SMAD4				V370A/E41*	V370A/E41*		R361C			S325P
EPHA3										
MAP2K7										
EPHRB6	173insSS	C316Y	C316Y							
PTEN										
ADAMTSL3										
GUCY1A2										
SMAD2									G457*	
OR51E1										
LAMA1										
C10orf137				S168L	S168L					
TCF7L2						P477T	P477T			

*d, nonsense mutation introducing stop codon; e (percentage of variation), indicated the frequency of cDNA reads containing the mutation encoding the reported protein sequence, determined on all the APC cDNA sequences obtained.
APC, antigen presenting cells; CSC, cancer stem/initiating cell cultures; CRC, colorectal cancers; Diff, differentiated cancer cell lines; fs*, frameshift mutation followed by nucleotide number before stop codon.
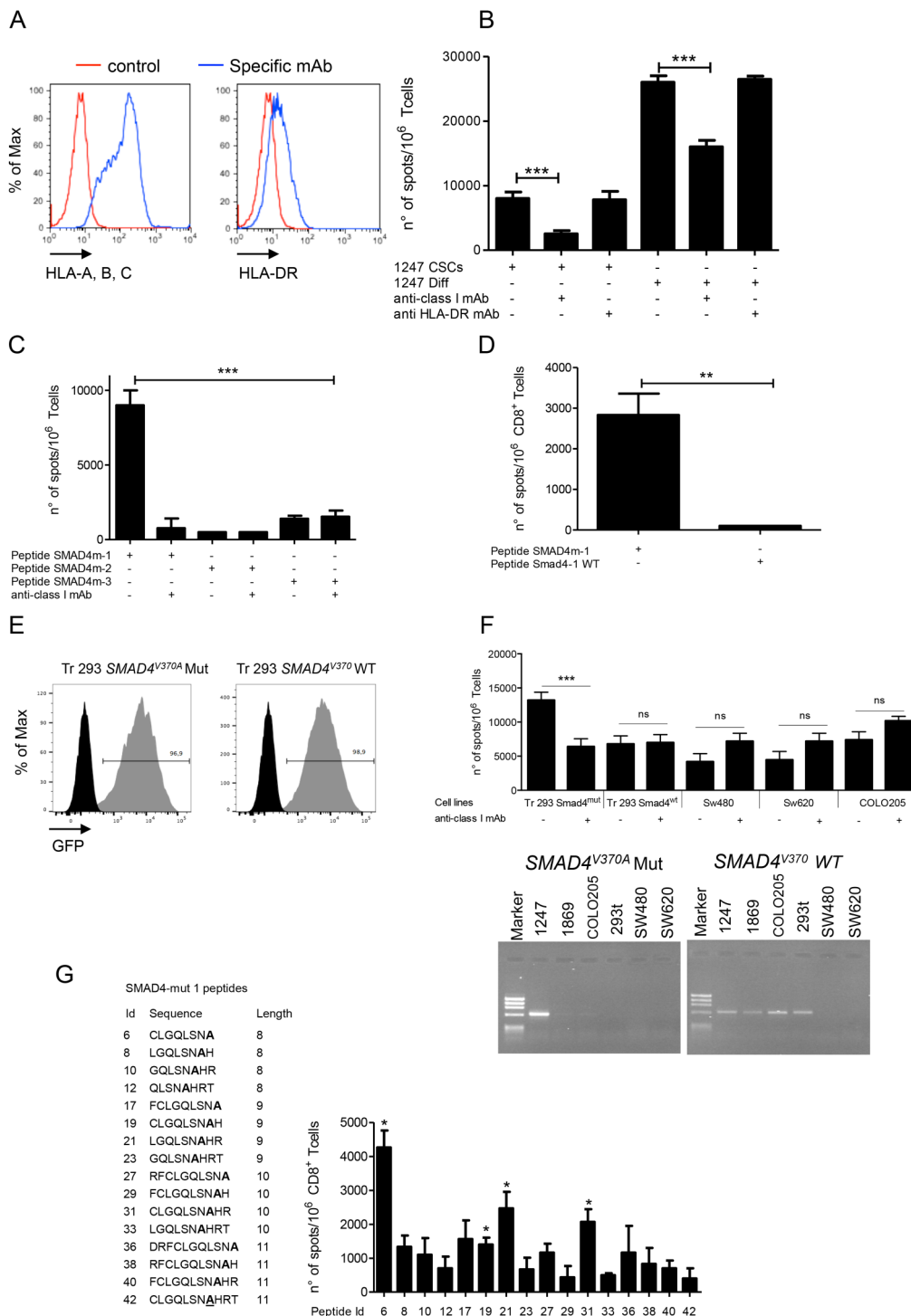
Figure 1 Patient T cells recognise a *SMAD4^V370A*-containing neoepitope presented by autologous cancer cells. (A) HLA class I and HLA-DR expression in 1247 cancer cells following 48 h IFNγ induction. (B–D) IFNγ ELISPOT of CD8+ T cells from 1247 patient, stimulated with a peptide pool encompassing the *SMAD4^V370A* mutation, assayed for the recognition of: 1247 cancer stem/initiating cells (CSCs) and differentiated colorectal cancer (CRC) cells (B), of SMAD4 mutated peptides presented by HLA-A*0201+ T2 cells (C), of SMAD4m-1 and SMAD41 WT peptides presented by T2 cells (D), ±anti-class I or HLA-DR mAbs. (E) Percentage of GFP expression by sorted HEK293t cells transduced with retroviral vectors encoding the 1247 SMAD4^V37A mutated or the corresponding SMAD4^V37 WT minigenes; (F) Upper panel. SMAD4^V37A-specific CD8+ T cells assayed by IFNγ ELISPOT for the recognition of HEK293t cells transduced with the SMAD4 minigenes, or of three HLA-A*02:01+ CRC cell lines negative for the SMAD4^V37A mutation, ±anti-class I mAb. Lower panel. PCR typing for the expression of the SMAD4^V37A mutation, or the corresponding SMAD4^V37 WT sequence in the CRC cell lines shown in the upper panel, and in untransduced HEK293 (293t) cells. The 1247 and 1869 CRC cell lines are positive and negative controls for the PCR, respectively. (G) SMAD4m-1 peptide-stimulated CD8+ T cells assayed by IFNγ ELISPOT (right panel) for the specific recognition of the peptide epitopes of different lengths (left panel), presented by T2 cells. All IFNγ ELISPOT data are triplicate mean±SD, subtracted of the background spots produced by T cells alone, and are representative of three independent experiments performed with independently induced CD8+ T cell lines. Only the experiment in panel F was performed twice with the same CD8+ T cell line. *p≤0.05; **p≤0.01; ***p≤0.001; ns, non-statistically significant.
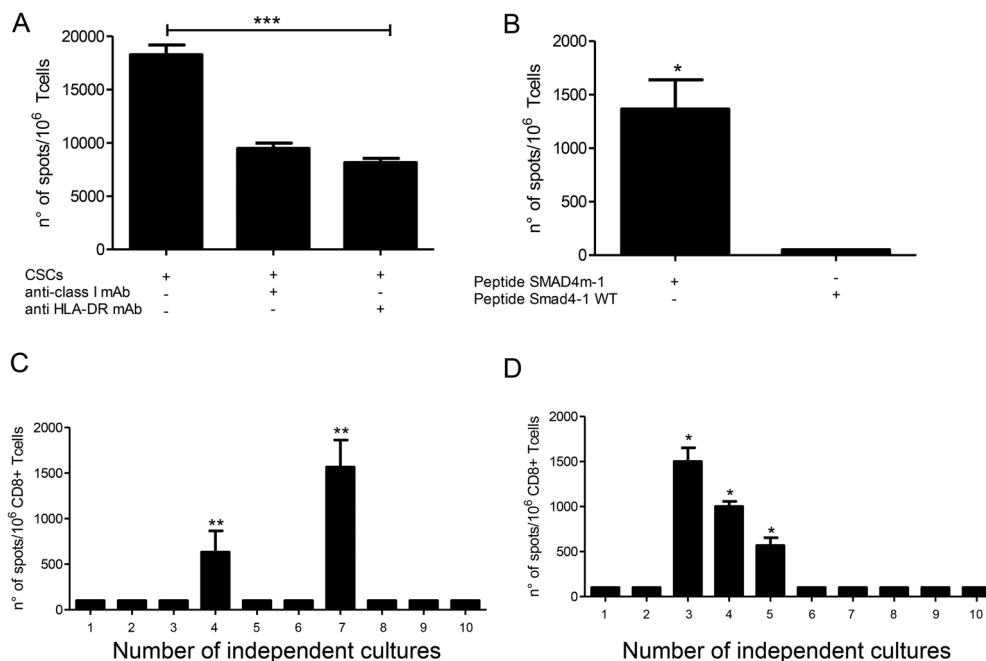
Figure 2 The *SMAD4*^{V370A} mutation of 1247 colorectal cancer (CRC) is spontaneously immunogenic for autologous CD8$^+$ T cells. (A) Mixed lymphocyte-tumour cell culture (MLTC) from patient 1247, induced by autologous cancer stem/initiating cell (CSC) cultures, assayed by IFNγ ELISPOT for the recognition of the autologous CSCs, ±anticlass I mAb or anti-HLA-DR mAb. (B) CD8$^+$ T cells, enriched from the previous MLTC, assayed by IFNγ ELISPOT for the recognition of the SMAD4m-1 or SMAD4-1 WT peptides presented by T2 cells; (C) Primary CD8$^+$ T cells purified from patient 1247, stimulated twice with the SMAD4-1 peptide in the presence of autologous PBMCs in two series of 10 wells, each containing 5×10^4 (left panel) or 10^5 (right panel) cells/well for a total of 1.5×10^6 precursors, and assayed by IFNγ ELISPOT for the recognition of the 1247 CRC cells. IFNγ ELISPOT data are represented as triplicate mean±SD, subtracted of the background spots produced by T cells alone, and are representative of three (A) and two (B) independent experiments performed. *p≤0.05; **p≤0.01; ***p≤0.001.

but not with the SMAD4 WT-minigene, nor three different HLA-A*0201$^+$ CRC cell lines that were all negative for the SMAD4^{V370A} mutation, and were either positive (COLO293) or negative (SW480, SW620) for the corresponding SMAD4 WT sequence (figure 1F). This result confirmed that the induced CD8$^+$ T cells were specific for a naturally processed SMAD4^{V370A}-containing neoepitope presented by HLA-A*0201.

To define the minimal HLA-A*02:01-restricted SMAD4^{V370A}-containing neoepitope recognised by the T cells of patient 1247, we searched the public prediction database Immune Epitopes Database (http://www.iedb.org) for progressively shorter epitopes from either SMAD4m-1 or SMAD4–1 WT 15mers that were predicted to bind HLA-A*0201. Synthetic peptides corresponding to the predicted epitopes were then tested for recognition by CD8$^+$ T cell lines induced with the SMAD4m-1 15mer. T cells recognised the 8, 9 and 10 aa long SMAD4m-1-derived peptides number 6, 19, 21 and 31 (figure 1G), but not the corresponding non-mutated peptides (data not shown), defining the recognised minimal CLGQLSNA mutated epitope. Binding assays performed with the three recognised SMAD4 mutated peptides established a very low binding affinity (>500 nM) for HLA-A*0201, in the range of the corresponding non-mutated peptides (not shown), suggesting that the antigenicity of the mutated SMAD4 peptide epitope was not due to an increased binding affinity for HLA, compared with the non-mutated epitope.

In contrast to patient 1247, CD8$^+$ T cells from patient 1039 that were induced with peptide pools spanning the autologous *KRAS*^{G12D}, *TP53*^{Y107D} and *PIK3CA*^{Q546K} CRC mutations did

not recognise autologous cancer cells, suggesting that the protein encoded by these three mutated genes could not generate naturally processed neoepitopes (data not shown).

Together, these findings indicated that the *SMAD4*^{V370A} somatic mutation expressed by the colon cancer 1247 generated a naturally processed neoepitope recognised by autologous CD8$^+$ T cells on differentiated and CSC cultures.

The mutated SMAD4-1 epitope is immunogenic for autologous CD8$^+$ T cells

To investigate the spontaneous immunogenicity of the *SMAD4*^{V370A} somatic mutation, we used T cell lines obtained by stimulating PBMCs from patient 1247 with autologous CSCs in MLTC,[28] performed by neutralising the immunosuppressive interleukin (IL) 4 produced by CSCs.[28] The MLTC contained CD4$^+$ and CD8$^+$ T cells that specifically recognised the autologous CSC cultures (figure 2A). CD8$^+$ T cells, enriched from these MLTCs, were also specifically stimulated by T2 cells loaded with the SMAD4-m1 but not with the SMAD4-1 WT peptide (figure 2B), suggesting that T cell precursors specific for the *SMAD4*^{V370A} somatic mutation had been naturally expanded by autologous CSCs in the MLTC.

To assess the frequency of CD8$^+$ T cell precursors specific for the *SMAD4*^{V370A} epitope in the PBMCs of patient 1247 and in two HLA-A*0201 matched HDs, a total of 1,5×10^6 CD8$^+$ T cells from each individual were distributed in two series of 10 wells, containing 5×10^4 or 10^5 cells/well, respectively (figure 2C,D), and stimulated twice with the SMAD4m-1 peptide in the same wells. The cells contained in each well were then independently assayed for the recognition of differentiated 1247
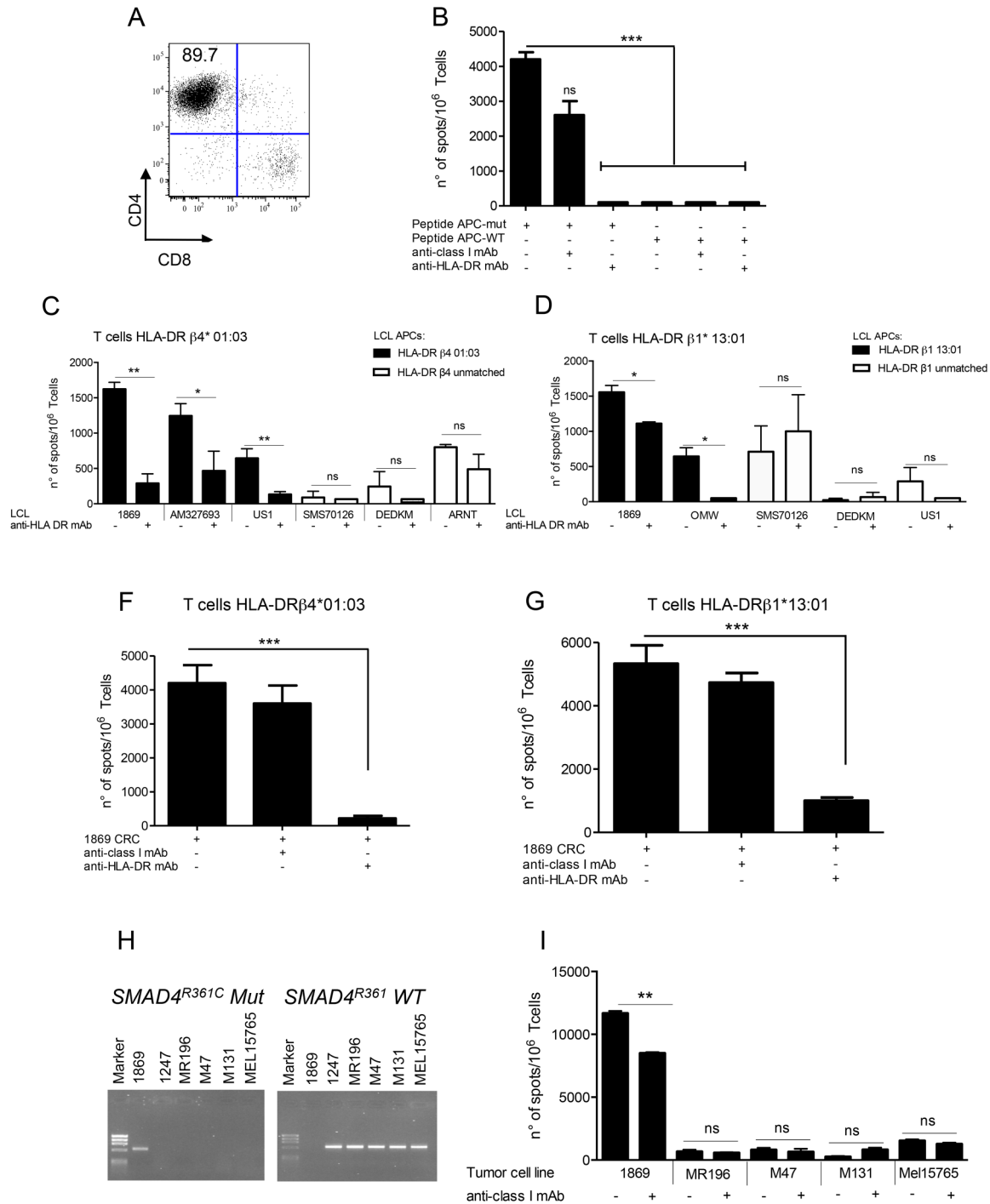
Figure 3 The 1869 colorectal cancer (CRC) presents neoepitopes from mutated genes to allogeneic CD4+ and CD8+ T cells. (A) CD4+ phenotype of a T cell line from a HLA-DR*β4 01:03 healthy donor (HD) expanded with the 30 aa long antigen presenting cell (APC)^{E1317KfsX4} synthetic peptide. (B) T cells from the HLA-DR*β4 01:03 HD assayed by IFNγ ELISPOT for the recognition of 1869 B lymphoblastoid cell lines (LCLs) loaded with the APC^{E1317KfsX4} or the WT peptide, ±anti-HLA-DR mAb. Similar results were obtained with CD4+ T cell lines elicited by the APC^{E1317KfsX4} peptide from HLA-DR*β1 13:01 HDs. (C–D) APC^{E1317KfsX4}-specific CD4+ T cells from either HLA-DR*β4 01:03 or HLA-DR*β113:01 donor assayed by IFNγ ELISPOT for the recognition of LCL cells homozygous for either HLA-DR*β4 01:03 (C) or HLA-DR*β113:01 (D), or negative for these alleles, loaded with the APC^{E1317KfsX4} peptide±anti-HLA-DR mAb. (F–G) APC^{E1317KfsX4}-specific CD4+ T cells induced from HLA-DR*β4 01:03- or HLA-DR*β1 13:01-matched HDs, assayed by IFNγ ELISPOT for the recognition of 1869 CRC cells, ±anti-HLA-DR mAbs. (H) PCR typing for the expression of the SMAD4^{R361C} mutation, or the corresponding SMAD4^{V37} WT sequence, in the HLA-B*35:01+ cancer cell lines used for the recognition assay. The 1869 and 1247 cell lines are positive and negative control of the PCR, respectively. (I) CD8+ T cells elicited from HLA-B*35:01 HD with the peptide pool encompassing the SMAD4^{R361C} mutation assayed by IFNγ ELISPOT for the recognition of 1869 CRC cells, or of HLA-B*35:01+ and SMAD4^{R361C}-negative kidney cancer cell line MR196 and melanoma cell lines M47, M131 and Mel15765,±anti-class I mAb. IFNγ ELISPOT data are represented as triplicate mean±SD, subtracted of the background spots produced by T cells alone, and are representative of three independent experiments. *p≤0.05; **p≤0.01; ***p≤0.001; ns, non-statistically significant.

CRC cells. Tumour-specific CD8$^+$ T cells were detectable in this condition only in the T cell cultures derived from the patient (figure 3C,D), with an estimated $SMAD4^{V370A}$-specific precursor frequency of about 1 in 2.5×10^5 CD8$^+$ T cells. No specific T cell response could be detected in the cultures established from the HDs suggesting that, in health conditions, $SMAD4^{V370A}$-reactive T cell precursors were either not expanded, or present at a frequency below that determined in the patient with cancer, patient 1247.

Hence, the $SMAD4^{V370A}$ somatic mutation generates a neoepitope that is naturally immunogenic for autologous T cells, resulting in the expansion of specific CD8$^+$ T cell precursors in vivo.

Induction of CD4$^+$ and CD8$^+$ T cells from HLA-matched HDs specific for somatically mutated CRC gene products

We next investigated the recognition of the potential unique neoepitopes derived from the somatically mutated gene products expressed by the 1869 CRC. The 1869 CRC cell lines expressed class I and HLA-DR upon IFNγ pretreatment (not shown) and could be tested for recognition by the peptide-induced T cell lines. We first sought to specifically investigate the CD4$^+$ T cells response against the mutated APC^{E1317KfsX4} gene product expressed by the 1869 CRC. Because there were not enough autologous T cells, the PBMCs of two HDs sharing the HLA-DR*β4 01:03 and HLA-DR*β1 13:01 alleles, respectively, with the 1869 CRC were stimulated with a 30 aa long peptide (APCmut) incorporating at the C-terminus the three substituted residues encoded by the frame-shift mutated APC^{E1317KfsX4} gene (see online supplementary table S1). Such long peptides, at least in vitro, are taken up by APCs contained in PBMCs, processed and presented mainly in class II, selectively expanding CD4$^+$ T cells. The resulting CD4$^+$ T cell lines (figure 3A) from either donors recognised the APC-mut peptide, but not the APC-WT one, loaded on LCL cells from the 1869 patient (figure 3B). Each CD4$^+$ T cell line was also specifically stimulated by LCL cell lines either homozygous for HLA-DR*β4 01:03 (figure 3C) or for HLA-DR*β1 13:01 (figure 3D), but not from LCL cells homozygous for different HLA-DR alleles, confirming their respective HLA-DR restrictions. The two CD4$^+$ T cell lines were also specifically stimulated in an HLA-DR-restricted manner by the 1869 cancer cells (figure 3F,G), indicating the presentation of naturally processed class II neoepitopes containing the APC mutation by CRC cells.

We finally induced T cells lines from a different HD matched for HLA-B*35:01 with the 1869 cancer cells, using a pool of synthetic 15-mer peptides encompassing the SMAD4^{R361C} mutation (see online supplementary table S1). The induced CD8$^+$ T cells specifically recognised the target 1869 CRC cell line, which expresses the SMAD4^{R361C} mutation but not the SMAD4 WT sequence, whereas they did not recognise HLA-B35$^+$ kidney cancer (MR196) and melanoma (M47, M131, Mel15765) cell lines, which were all negative for the SMAD4^{R361C} mutation and positive for the corresponding SMAD4^{WT} sequence (figure 3H,I). Hence, the induced CD8$^+$ T cells were specific for a naturally processed neoepitope derived from the SMAD4^{R361C} mutation uniquely expressed and presented by the 1869 CRC cells.

DISCUSSION

Recent publications have highlighted the power of combining next-generation sequencing of cancer DNA with reverse immunology to identify T cell epitopes from unique tumour antigens involved in the control of mouse (melanoma and chemically induced sarcomas) and human (melanoma, cholangiocarcinoma and CLL) cancers.[12 18 20] Our study extends those findings in several ways. First, we have focused on CRC, which is a frequent epithelial cancer and a big killer never investigated for the expression of unique tumour antigens by this approach. For a proof of concept, we generated primary tumour cell lines to assess direct tumour recognition by mutated peptide-specific T cells, implying the natural processing and presentation of the somatically mutated epitope. Second, we sequenced the expressed genome (cDNA) from CRC cells, which confirms that the mutated genes are actually expressed by the malignant cells. This approach might well be replaced by advanced RNA sequencing techniques,[32] when considering a clinical setting in which primary tumour samples must be directly sequenced to speed up the possible therapeutic application of neoepitope-based vaccines. Third, we elicited tumour-specific CD8$^+$ and CD4$^+$ T cell responses in vitro using small pools of long peptides (>15aa) encompassing the somatic mutation, with no need for precise HLA allele-specific epitope prediction and the extensive synthesis and testing of the defined peptides. Finally, and importantly, because a critical issue concerning cancer therapy is the ability to target the CSC component to actually eradicate the tumour,[33 34] we were able to show that CD8$^+$ T cells induced with a mutated SMAD4 peptide recognised autologous CSCs expressing the same mutation. Although we did not sequence the whole expressed genome, the two pairs of CSCs and differentiated CRCs derived from each common surgical sample shared the same somatic cDNA mutations, implying the possibility that T cells directed against the corresponding mutated epitopes might effectively target the tumour initiating compartment in vivo for therapeutic purposes. With regards to this, it has been shown that CSCs from CRC are endowed with immunosuppressive mechanisms that inhibit the induction of T cell responses.[28 35] However, effector T cells can efficiently recognise CSCs from CRCs suggesting that, once they are activated by strategies that counteract such suppression in vitro or in vivo, T cells specific for unique tumour antigens might be able to therapeutically target the CSC component in vivo.

On average, we find that 4 of 20 sequenced genes are somatically mutated in CRCs, representing potential T cell antigens. Even though parsimonious, our sequencing approach proved efficacious in identifying antigenic somatic cancer mutations. In the first three CRCs tested, two were found able to process and present somatically mutated epitopes to CD4$^+$ and CD8$^+$ T cells. However, in one CRC sample (1039), the missense mutations found in three genes did not generate antigenic epitopes recognised by autologous CD8$^+$ T cells. It is conceivable that sequencing the whole RNA from each CRC cancer would significantly increase the likelihood to identify somatically mutated tumour antigens in all patients.

We sequenced CAN-genes that are clearly drivers involved in the oncogenic transformation of colon epithelium.[22] These genes and proteins are ideal targets of T cell immunotherapy because they are less likely to be lost by cancer immune escape mutants. Data obtained in mouse and human tumours, however, suggest that tumour-specific T cells recognise neoepitopes derived mostly from passenger rather than driver somatic mutations,[5 12–18] implying a possible immune-escape mechanism. It is conceivable that a more extensive sequencing of additional expressed genes from our CRC samples would have identified passenger mutations also, generating immunogenic neoepitopes for autologous T cells.

The unique $SMAD4^{V370A}$ epitope identified in the CRC 1247 displays a very low binding affinity for HLA-A*0201, well

above 500 nM that is considered to be the threshold for productive epitope binding to MHC and presentation to cognate T cells. The mutation acquired in the 1247 CRC cells modifies the predicted putative MHC anchor of the epitope, but it does not increase the binding affinity for HLA compared with the WT sequence. Part of the very low binding affinity may be due to the fact that the peptides bear cysteine, which does not behave well in solution, in a position that generally has an appreciable influence on binding capacity. Nevertheless, we cannot readily explain the immunogenicity of the low affinity $SMAD4^{V370A}$ neoepitope. One possibility is that the new anchor residue introduced by the somatic $SMAD4^{V370A}$ mutation might generate a new agretope for HLA-A*02:01, which binds the cognate TCRs with increased C-terminal stability compared with the WT peptide, sufficient to lead to tolerance break and T cell activation. This possibility has been suggested by a recent study that identified 8/10 mutated neoepitopes from two chemically induced mouse sarcomas displaying extremely low affinity (over 500 nM), yet strongly immunogenic and able to induce potent T cell dependent tumour rejection upon immunisation in vivo.[17] The mutations found in most of the immunogenic neoepitopes identified in the mouse sarcoma modify their C-terminal anchor residues.[17] In support of this hypothesis, we indeed find that the $SMAD4^{V370A}$ epitope is immunogenic in vivo, as suggested by the increased frequency of specific T cell precursors found in the patient, compared with the nearly undetectable frequency of T cell precursors specific for the same epitope found in HLA-A*02:01-matched HD.

Of the two patients in whom we have investigated autologous peripheral T cell responses, one presented expanded circulating T cells specific for the $SMAD4^{V370A}$ tumour mutation and is still alive after almost 5 years from surgery. In contrast, the other patient in whom no T cell responses specific for unique antigens were detected developed fatal cancer progression 6 months post surgery. The possibility that the T cell response specific for the unique antigens may have contributed to the postsurgery survival warrants future investigations in more patients with CRC, including also the analysis of the tumour infiltrating lymphocytes and of the tumour immune microenvironment. Recent clinical results, reporting that mismatch repair-deficient CRCs are strikingly more responsive to anti-PD-1 mAb (pembrolizumb) therapy than mismatch repair-proficient tumours, suggest to extend such investigations also to patients treated with immune checkpoint blockade.[36] This difference, in fact, correlated with a greater mean number of somatic mutations found in mismatch repair-deficient (1782) compared with mismatch repair-proficient (73) CRCs, which resulted in the prediction of 20 times more theoretical neoepitopes available for potential T cell responses in the former tumours.[36]

Collectively, our study shows a new strategy for the quantitative identification of mutated neoepitopes in CRC. Because the progression of this tumour is critically controlled by the immune system, particularly by the degree and quality of CD8$^+$ and CD4$^+$ T cell infiltration within the tumour tissue,[37 38] this approach can be easily scaled up to thoroughly characterise the protecting immune responses in patients undergoing surgery, as well as to define neoepitopes of tumour-specific antigens for effective cancer vaccines.

Author affiliations
[1]Division of Immunology, Transplantation and Infectious Diseases, San Raffaele Scientific Institute, Milan, Italy
[2]Division of Experimental Oncology, San Raffaele Scientific Institute, Milan, Italy
[3]Institute for Biomedical Technologies, National Research Council, Segrate, Italy
[4]La Jolla Institute for Allergy & Immunology, La Jolla, California, USA
[5]Institute of Molecular Recognition Chemistry, National Research Council, Milan, Italy
[6]Department of Surgery, San Raffaele Scientific Institute, Milan, Italy
[7]Department of Pathology, San Raffaele Scientific Institute, Milano, Italy
[8]Unit of Molecular and Functional Immunogenetics, San Raffaele Scientific Institute, Milan, Italy
[9]Institute for Experimental Cellular Therapy, University Hospital Essen, Essen, Germany
[10]Department of Medical Sciences, Center for Transplantation Biology and Immunogenetics, University of Turin, Turin, Italy

Competing interests None declared.

Ethics approval Institutional Review Board of San Raffaele Scientific Institute, Milano, Italy (study CAN-GENES 01).

Provenance and peer review Not commissioned; externally peer reviewed.

REFERENCES

1 Rosenberg SA, Yang JC, Sherry RM, et al. Durable complete responses in heavily pretreated patients with metastatic melanoma using T-cell transfer immunotherapy. *Clin Cancer Res* 2011;17:4550–7.
2 Hodi FS, O'Day SJ, McDermott DF, et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N Engl J Med* 2010;363:711–23.
3 Topalian SL, Hodi FS, Brahmer JR, et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* 2012;366:2443–54.
4 Brahmer JR, Tykodi SS, Chow LQ, et al. Safety and activity of anti-PD-L1 antibody in patients with advanced cancer. *N Engl J Med* 2012;366:2455–65.
5 Coulie PG, Van den Eynde BJ, van der Bruggen P, et al. Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. *Nature reviews Cancer* 2014;14:135–46.
6 Mumberg D, Wick M, Schreiber H. Unique tumor antigens redefined as mutant tumor-specific antigens. *Semin Immunol* 1996;8:289–93.
7 Dudley ME, Roopenian DC. Loss of a unique tumor antigen by cytotoxic T lymphocyte immunoselection from a 3-methylcholanthrene-induced mouse sarcoma reveals secondary unique and shared antigens. *J Exp Med* 1996;184:441–7.
8 Lennerz V, Fatho M, Gentilini C, et al. The response of autologous T cells to a human melanoma is dominated by mutated neoantigens. *Proc Natl Acad Sci USA* 2005;102:16013–18.
9 Ronaghi M, Uhlen M, Nyren P. A sequencing method based on real-time pyrophosphate. *Science* 1998;281:363, 5.
10 Margulies M, Egholm M, Altman WE, et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 2005;437:376–80.
11 Segal NH, Parsons DW, Peggs KS, et al. Epitope landscape in breast and colorectal cancer. *Cancer Res* 2008;68:889–92.
12 Castle JC, Kreiter S, Diekmann J, et al. Exploiting the mutanome for tumor vaccination. *Cancer Res* 2012;72:1081–91.
13 Robbins PF, Lu YC, El-Gamil M, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med* 2013;19:747–52.
14 van Rooij N, van Buuren MM, Philips D, et al. Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J Clin Oncol* 2013;31:439–42.

15 Rajasagi M, Shukla SA, Fritsch EF, *et al*. Systematic identification of personal tumor-specific neoantigens in chronic lymphocytic leukemia. *Blood* 2014;124:453–62.

16 Matsushita H, Vesely MD, Koboldt DC, *et al*. Cancer exome analysis reveals a T-cell-dependent mechanism of cancer immunoediting. *Nature* 2012;482: 400–4.

17 Duan F, Duitama J, Al Seesi S, *et al*. Genomic and bioinformatic profiling of mutational neoepitopes reveals new rules to predict anticancer immunogenicity. *J Exp Med* 2014;211:2231–48.

18 Gubin MM, Zhang X, Schuster H, *et al*. Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* 2014;515:577–81.

19 Linnemann C, van Buuren MM, Bies L, *et al*. High-throughput epitope discovery reveals frequent recognition of neo-antigens by CD4+ T cells in human melanoma. *Nat Med* 2015;21:81–5.

20 Cohen CJ, Gartner JJ, Horovitz-Fried M, *et al*. Isolation of neoantigen-specific T cells from tumor and peripheral lymphocytes. *J Clin Invest* 2015;125:3981–91.

21 Vogelstein B, Papadopoulos N, Velculescu VE, *et al*. Cancer genome landscapes. *Science* 2013;339:1546–58.

22 Wood LD, Parsons DW, Jones S, *et al*. The genomic landscapes of human breast and colorectal cancers. *Science* 2007;318:1108–13.

23 Lawrence MS, Stojanov P, Mermel CH, *et al*. Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 2014;505:495–501.

24 Network CGA. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487:330–7.

25 Pagès F, Berger A, Camus M, *et al*. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med* 2005;353:2654–66.

26 Galon J, Costes A, Sanchez-Cabo F, *et al*. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006;313:1960–4.

27 Ricci-Vitiani L, Lombardi DG, Pilozzi E, *et al*. Identification and expansion of human colon-cancer-initiating cells. *Nature* 2007;445:111–15.

28 Volonte A, Di Tomaso T, Spinelli M, *et al*. Cancer-initiating cells from colorectal cancer patients escape from T cell-mediated immunosurveillance in vitro through membrane-bound IL-4. *J Immunol* 2014;192:523–32.

29 Sidney J, Southwood S, Moore C, *et al*. Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. *Curr Protoc Immunol* 2013; Chapter 18:Unit 18.3.

30 Campi G, Crosti M, Consogno G, *et al*. CD4(+) T cells from healthy subjects and colon cancer patients recognize a carcinoembryonic antigen-specific immunodominant epitope. *Cancer Res* 2003;63:8481–6.

31 Sedegah M, Kim Y, Peters B, *et al*. Identification and localization of minimal MHC-restricted CD8+ T cell epitopes within the Plasmodium falciparum AMA1 protein. *Malar J* 2010;9:241.

32 Hashimshony T, Wagner F, Sher N, *et al*. CEL-Seq: single-cell RNA-Seq by multiplexed linear amplification. *Cell Rep* 2012;2:666–73.

33 Visvader JE, Lindeman GJ. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer* 2008;8:755–68.

34 Clevers H. The cancer stem cell: premises, promises and challenges. *Nat Med* 2011;17:313–19.

35 Maccalli C, Volontè A, Cimminiello C, *et al*. Immunology of cancer stem cells in solid tumours. A review. *Eur J Cancer* 2014;50:649–55.

36 Le DT, Uram JN, Wang H, *et al*. PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* 2015;372:2509–20.

37 Bindea G, Mlecnik B, Tosolini M, *et al*. Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 2013;39:782–95.

38 Fridman WH, Pagès F, Sautes-Fridman C, *et al*. The immune contexture in human tumours: impact on clinical outcome. *Nature reviews Cancer* 2012;12:298–306.