OPEN ACCESS

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]

► Additional material is published online only. To view these files please visit the journal online (http://dx.doi.org/10.1136/gutjnl-2015-309453).

For numbered affiliations see end of article.

Correspondence to
Dr Paolo Dellabona, Division of Immunology, Transplantation and Infectious Diseases, San Raffaele Scientific Institute, Via Olgettina 58, Milano 20132, Italy; dellabona.paolo@hsr.it

DM and CM contributed equally.
GP, GC and PD are senior coauthors

Received 24 February 2015
Revised 4 November 2015
Accepted 6 November 2015
Published Online First
15 December 2015

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

To cite: Mennonna D, Maccalli C, Romano MC, *et al. Gut* 2017;**66**:454–463.

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 $SMAD4^{V370\text{-}WT}$ 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

	Tumour cell lines									
CAN genes	**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.

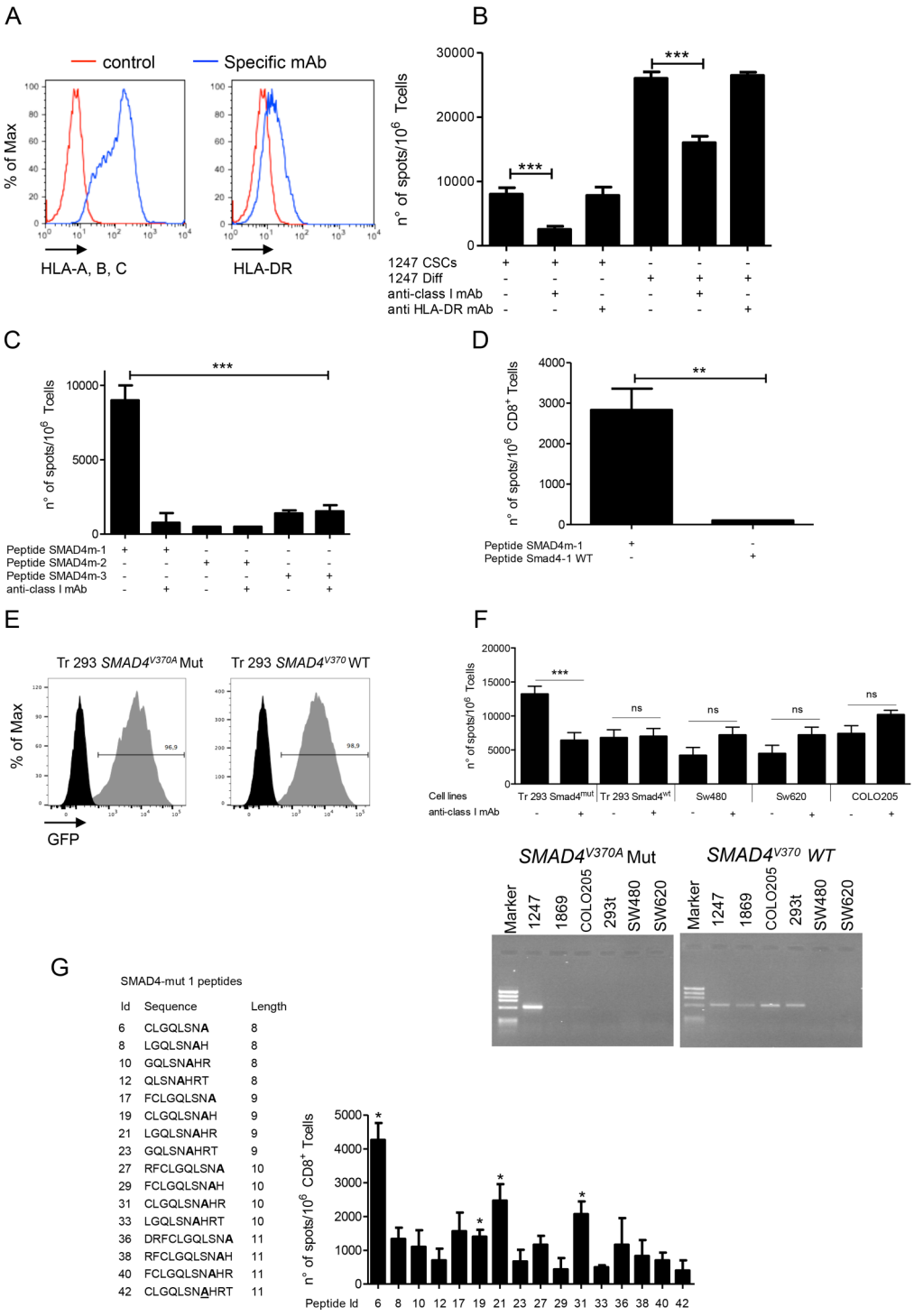

Figure 1 Patient T cells recognise a *SMAD4*^{V370A}-containing neoepitope presented by autologous cancer cells. (A) HLA class I and HLA-DR expression in 1247 cancer cells following 48 h IFNγ induction. (B–D) IFNγ ELISPOT of CD8$^+$ T cells from 1247 patient, stimulated with a peptide pool encompassing 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 HEK293t (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.

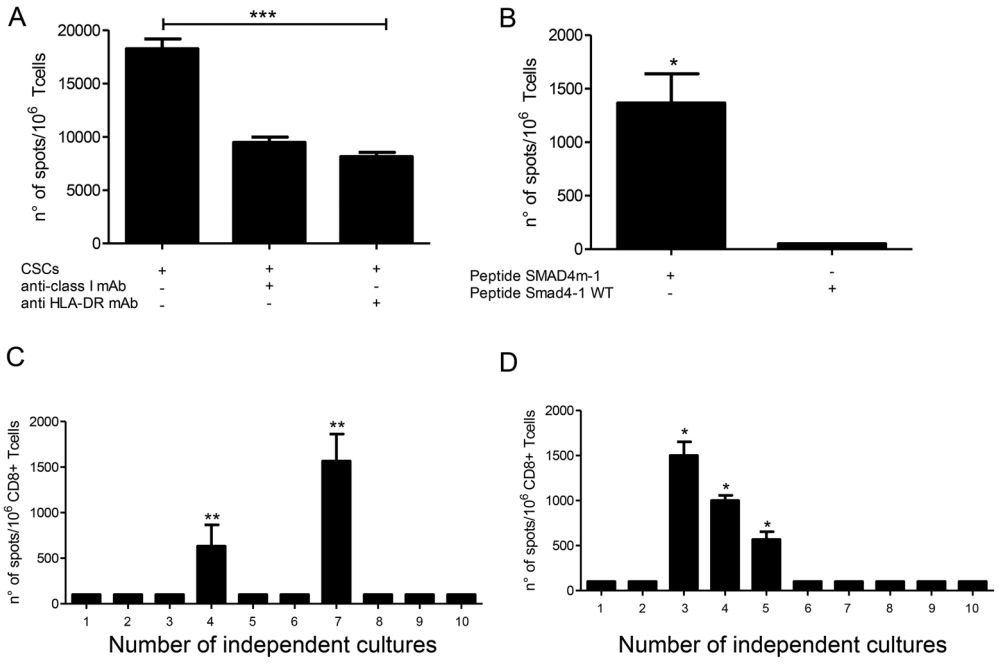

Figure 2 The *SMAD4*^{V370A} mutation of 1247 colorectal cancer (CRC) is spontaneously immunogenic for autologous CD8$^+$ T cells. (A) Mixed lymphocyte-tumour cell culture (MLTC) from patient 1247, induced by autologous cancer stem/initiating cell (CSC) cultures, assayed by IFNγ ELISPOT for the recognition 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

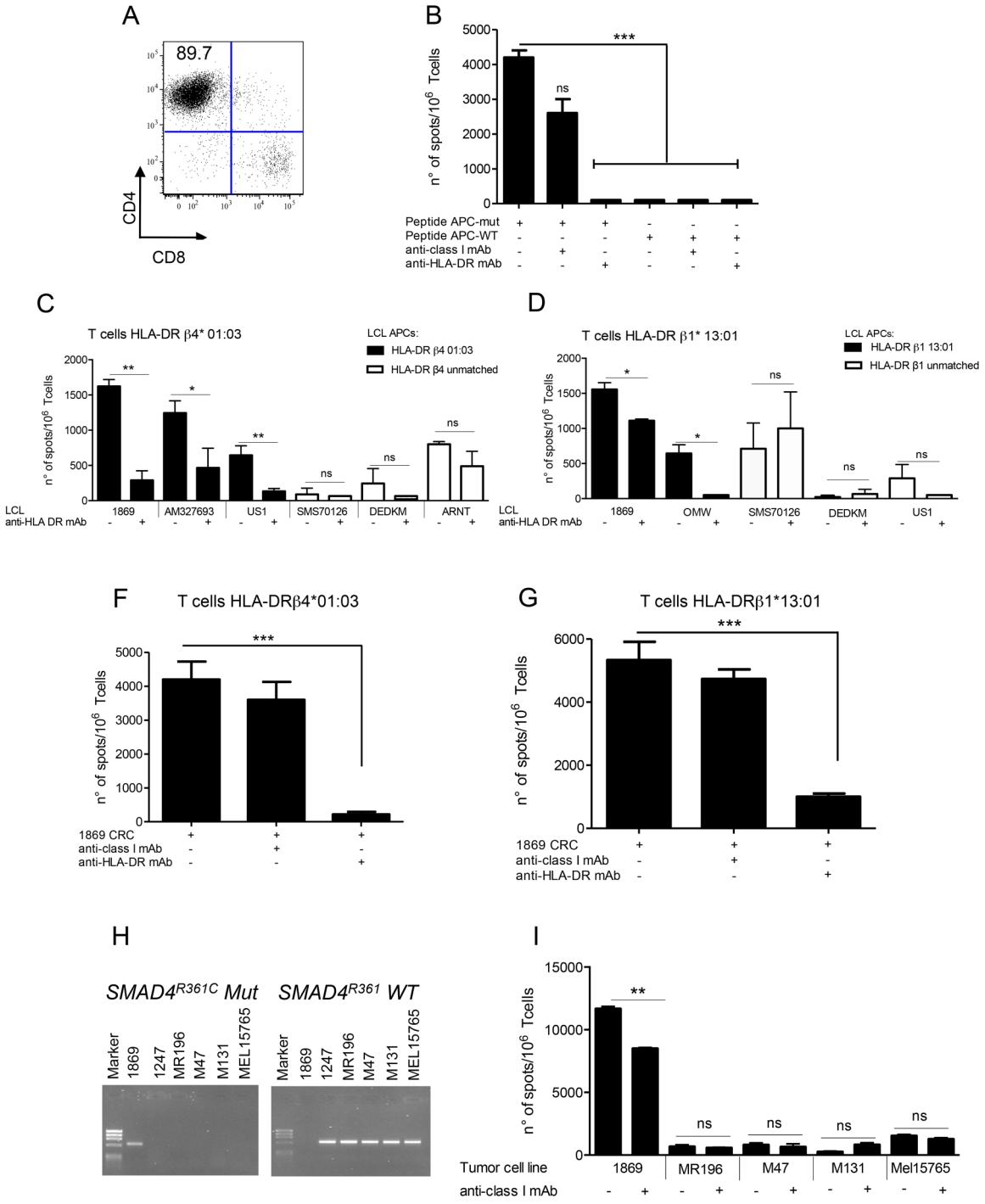

Figure 3 The 1869 colorectal cancer (CRC) presents neoepitopes from mutated genes to allogeneic CD4$^+$ and CD8$^+$ T cells. (A) CD4$^+$ phenotype of a T cell line from a HLA-DR*β4 01:03 healthy donor (HD) expanded with the 30 aa long antigen presenting cell (APC)^{E1317KfsX4} synthetic peptide. (B) 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 frameshift 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

Acknowledgements The authors thank the members of the San Raffaele programme of Immunology and Bio-Immunotherapy of Cancer for suggestions and criticisms throughout the study.

Contributors DM, CM, MCR, FC designed and performed experiments, analysed data and wrote the manuscript; RB, MS, GDB performed and analysed deep sequencing; JS, AS performed peptide binding studies and epitope prediction analysis; AG, RL synthesised peptides; MB, LG, LB, EO, LA took care of the clinical cases and pathology; EZ, KF provided HLA-matched healthy donors; GM, NF, AA typed tumour samples and provided HLA-matched healthy donors; GC, GP, PD envisaged the study, supervised experiments and wrote the manuscript.

Funding Study supported by Associazione Italiana per la Ricerca sul Cancro grant AIRC-IG11524.

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.

SUPLEMENTAL METHODS

Establishment of tumor cells cultures

PBMCs were obtained from CRC patients or HLA-matched healthy donors by standard Ficoll separation (Ficoll-Paque PLUS, GE Healthcare Bio-Science). To generate differentiated CRC cell lines, surgical specimens were cut into fragments, filtered through sterile gauze and cultured in RPMI medium supplemented with 5% FCS. CSCs were isolated by mechanical processing of primary tumor tissue and cultured as spheres in the presence of stem cell permissive medium (DMEM/F12) as described [1]. EBV-transformed B-cell lines were established from patients PBMCs by immortalization with EBV containing culture supernatant. HLA-A*0201-positive T2 cells were obtained from the American Type Culture Collection (ATCC, Manassas, USA).

Massive parallel cDNA sequencing

Amplified cDNA pools (3µg) were processed for massive sequencing according to the GS FLX Titanium protocol (454 Life Sciences, Roche, Branfort, CT, USA). Single strand (sstDNA) libraries were quantitated by RiboGreen RNA Quantitation Kit (Invitrogen Inc., Carlsbad, CA, USA), checked for quality by capillary electrophoresis (Agilent Bioanalyzer 2100 with the RNA Pico 6000 LabChip kit; Agilent Technologies, Palo Alto, CA, USA) and amplified in emulsion-based clonal reaction. Each enriched sample was separately loaded onto a PicoTiterPlate (PTP) and pyrosequenced. From each sequencing run, we obtained an average of 126824±42030 high quality sequencing reads (approximately 50Mbases of raw sequence), with an optimal average 1180X depth per sample for all the 20 genes selected in the

panel. Each gene had a depth ≥100X in all the samples. cDNA sequences for the 20 selected genes were retrieved from NCBI (http://www.ncbi.nlm.nih.gov/gene) and used as reference. Raw reads were aligned by using gsMapper software (v. 2.5, Roche) using default mapping parameters. Candidate mutations were first screened keeping in the analysis only those having more than 10 varied reads and an incidence of varied reads on the total of at least 10%. The alignments of the reads supporting all remaining variants were then manually checked in order to filter out false positives in error-prone regions (e.g.: homopolymers). After this step, the analysis was prioritized on the variants having an incidence of ≥40%. Peptide sequences were generated on a sample-by-sample basis introducing in the cDNA sequence the mutations deriving from the previous analysis and, then, traducing the sequence into peptides using EMBOSS [2] function “transeq”; finally, open-reading-frames (ORFs) were predicted by using the function “checktrans” of the EMBOSS suite.

MHC-peptide binding analyses

Quantitative assays to measure the binding of peptides to HLA A*02:01 class I molecules are based on the inhibition of binding of a radiolabeled standard peptide (HBV core 18-27 analog, FLPSDYFPSV). MHC molecules were purified by affinity chromatography from the EBV transformed homozygous cell line JY, and assays performed as described previously [3]. Peptides were tested at six different concentrations covering a 100,000-fold dose range in three or more independent assays, and the concentration of peptide yielding 50% inhibition of the binding of the radiolabeled probe peptide (IC_{50}) was

calculated. Predictions of the HLA A*02:01 binding capacity of peptides 8 to 11 residues in length were performed using the command-line version of the consensus prediction tool available at the Immune Epitope Database (IEDB; www.iedb.org) [4].

T cell cultures

T cell lines and mixed lymphocyte-tumor cell culture (MLTC) from patient 147 were generated from PBMCs as described [5]. To induce MLTCs, 1247 CSC cultures were pre-treated o.n.with hrIFN-γ (1000 IU/ml; PeproTech), before culturing with autologous PBMCs. To induce peptide-specific T cell lines, PBMCs were cultured with either mutated or WT peptides (10 μg/ml) in 96-well flat bottom plates in X-VIVO 15 medium (Lonza) supplemented with 5% normal human serum (EuroClone). After two days, recombinant human IL-2 (150 IU/mL; R&D) and IL-7 (10 ng/mL; R&D) were added to the cultures. At day 7, T cell cultures were re-stimulated with autologous irradiated (50 Gy) PBMCs pulsed with peptides and then re-stimulated every 7 days. At day 14, T cells were assayed for antigen recognition by ELISPOT, which was repeated twice in the next 14 days to confirm the results. Effector $CD8^+$ T cells were enriched from the cultures by negative immunomagnetic sorting using $CD4^+$ T cell Isolation Kit (Miltenyi Biotec). Mixed lymphocyte-tumor cell culture (MLTC) between 1247 PBMCs and autologous CSC were induced as described [1]. PBMCs (10^6 cells/well) of CRC patients were co-cultured with autologous irradiated (150 Gy) 1247 CSC cells in X-VIVO 15 medium (Lonza) supplemented with 10% normal human serum, 120 IU/mL rhIL-2, 5 ng/mL rhIL-7 and 10 μg/ml anti–IL-4 neutralizing mAb 3007 (all from R&D System).

MLTC were stimulated weekly with irradiated autologous CRC/CSC cells. T cell reactivity was tested after the third restimulation.

Retroviral transduction of mutated and WT SMAD4 minigenes

Two minigenes enconding 27 aa-long polipeptides encompassing either the SMAD4^{V370A} mutation expressed by the 1247 CRC, or the corresponding SMAD4$^{V370-WT}$ residue, flanked on each side by 14 and 12 aminoacids of the SMAD4 WT sequence, were cloned using BglII and XhoI restriction sites in the retroviral vector MSCV-IRES-GFP upstream of an IRES sequence preceding a GFP tag (kindly provided by Dr. K. Murphy, Washington University). Each minigene sequences contain a 5' Kozak consensus box and a 3' STOP codon:

```
Minigene SMAD4 Mut 1247
ATGtctggaggagatcgcttttgtttgggtcaactctccaatgCccacaggacagaagccattgagagagcaaggttgcacTGA
M  S  G  G  D  R  F  C  L  G  Q  L  S  N  A  H  R  T  E  A  I  E  R  A  R  L  H  STOP

Minigene SMAD4 WT
ATGtctggaggagatcgcttttgtttgggtcaactctccaatgTccacaggacagaagccattgagagagcaaggttgcacTGA
M  S  G  G  D  R  F  C  L  G  Q  L  S  N  V  H  R  T  E  A  I  E  R  A  R  L  H  STOP
```

Each retrovirus was produced upon transfection into human embryo kidney HEK293t packaging cells and transduced into different HEK293t cells, which were used as recipient cells because they are HLA-A*02:01^{+} and negative for the SMAD4^{V370A} mutation. Transduced HEK293t cells were sorted for the expression of EGFP (90%) and tested for the capacity to stimulate T cells from the 1247 patient that were specific for the SMAD4^{V370A} mutation.

PCR typing of mutated and WT SMAD4

The indicated tumor cell lines were screened for the expression of mRNA encoding either the SMAD4^{V370A} or SMAD4^{R361C} mutation, expressed by the

CRC cell lines 1247 and 1869, respectively. Total RNA was extracted from each tumor cell line, reverse transcribed into cDNA typed by PCR with upstream oligonucleotide primers ending exactly on the mutated nucleotide present in each sequence, coupled with downstream primers complementary to the WT sequence. The expression of the WT gene segment corresponding to either mutation was detected by similar PCR typing, utilizing upstream oligonucleotide primers ending exactly on the WT nucleotide present in each sequence, coupled with downstream primers complementary to the WT sequence. The upstream primer sequences are listed in Supplemental Table 4. PCR conditions were:

- SMAD4 mut and WT 1247: 35 cycles, 30" 94°C, 20" 64°C, 30" 72°C;
- SMAD4 mut and WT 1869: 30 cycles, 30" 94°C, 30" 65°C, 30" 72°C

PCR cycles were preceded by 10' denaturation at 94°C and terminated by 10' at 72°C.

ELISPOT assays

ELISPOT assay for IFNγ production by unique neo-antigen specific T cells were performed as described [1]. T cells were plated in duplicates at 5×10^3 cells/well in 96-well plates (MAIPS4510; Millipore, Bedford, Mass., USA) precoated with anti-IFNγ capture mAb (Mabtech) and stimulated for 4h with: CRC or CSC cell, LCL or T2 cell pre-loaded with each mutated or wild-type peptides (10 μg/ml), ±10 μg/ml of anti-HLA class I and/or anti-HLA-DR mAb. Biotinylated anti-IFNγ mAb (Thermo Scientific Pierce, Rockford, IL, USA), streptavidin–alkaline phosphatase conjugate (Amersham Pharmacia Biotech Europe GmbH, Freiburg, Germany,) and chromogenic substrate (nitroblue

tetrazolium-BCIP [5-bromo-4-chloro-3-indolylphosphate] (Sigma, St. Louis, Mo., USA) were subsequently added. The individual spots were counted using an automated ELISA-Spot Assay Video Analysis System (Eli-Scan) with the software Eli.Analyse V4.2 (A.EL.VIS, Hannover, Germany). As positive control, effector T cells were stimulated with PHA (1 µg/ml) (Sigma, St. Louis, Mo., USA). Unstimulated T lymphocytes represented the negative control.

REFERENCES

1 Volonte A, Di Tomaso T, Spinelli M, Todaro M, Sanvito F, Albarello L, *et al.* Cancer-initiating cells from colorectal cancer patients escape from T cell-mediated immunosurveillance in vitro through membrane-bound IL-4. Journal of immunology 2014;192:523-32.

2 Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000;16:276-7.

3 Sidney J, Southwood S, Moore C, Oseroff C, Pinilla C, Grey HM*, et al.* Measurement of MHC/peptide interactions by gel filtration or monoclonal antibody capture. Curr Protoc Immunol 2013;Chapter 18:Unit 18 3.

4 Vita R, Overton JA, Greenbaum JA, Ponomarenko J, Clark JD, Cantrell JR*, et al.* The immune epitope database (IEDB) 3.0. Nucleic acids research 2015;43:D405-12.

5 Campi G, Crosti M, Consogno G, Facchinetti V, Conti-Fine BM, Longhi R*, 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.

Supplemental Table 1. Clinical characteristics of CRC patients

Patients						
	1076	1247	21052	1546	23873	1039
Sex	M	F	F	M	F	M
date of birth	15/12/22	31/01/47	09/04/39	07/09/36	06/02/63	18/12/55
diagnosis age	87	63	70	74	45	53
anatomic stage	IIA	IIA	IIIB	IVA	IIA	IIIC
Neo adjuvant therapy	no	no	no	no	no	no
treatment	surgical	surgical	surgical	surgical/ RT	surgical	surgical
pTNM	pT3pN0 (0/17)	pT3pN0 (0/20)	pT3pN1 (2/31)	pT4apN2b (12/27)	pT3pN0 (0/8)	pT3pN2b (18/23)
histotype	adenoCA G3	adenoCA G3	adenoCA G2	adenoCA G2	adenoCA G2	adenoCA
DNA mismatch repair complex expression	Expressed MSH2,100,N,3 MLH1,100,N,3 PMS2,90,N,3 MSH6,90,N,3	Expressed MSH2,100,N,2 MLH1,100,N,2 PMS2,90,N,2 MSH6,80,N,2	Expressed MSH2,100,N,3 MLH1,100,N,3 PMS2,60,N,2 MSH6,90,N,3	Expressed MSH2,100,N,3 MLH1,100,N,3 PMS2,80,N,2 MSH6,100,N,2	Expressed MSH2,100,N,2 MLH1,80,N,2 PMS2,80,N,1 MSH6,60,N,2	Expressed MSH2,80,N,2 MLH1,80,N,2 PMS2,90,N,2 MSH6,80,N,2
adjuvant therapy	CT	CT: FOLFOX	CT: FOLFOX	CT: XELOX	CT: FOLFOX	CT
follow up	positive	negative	negative	positive	positive	positive
local recurrence	no	no	no	yes	no	no
treatment	-	-	-	RT	-	-
distant recurrence	liver	-	-	liver	liver	liver; lung
Treatment	CT	-	-	CT:CPT11, Cetuximab	surgical	surgical CT
DFS12	6 m.	4 y. and 6 m.	4 y. and 4 m.	4 m.	1 y. and 6 m.	6 m.
OS13	10 m.	4 y. and 6 m.	4 y. and 8 m.	1 y. and 2 m.	2 y. and 6 m.	1 y. and 4 m.
Death	yes	no	no	yes	yes	yes
cause of death	HF	-	-	HF	HF	RF
current disease stage	deceased	disease free	disease free	deceased	deceased	deceased

RT: radiotherapy; adenoCA: adenocarcinoma; CT: chemotherapy; FOLFOX: folinic acid + fluorouracil + oxaliplatin; XELOX: capecitabine + oxaliplatine; DFS: disease free survival; OS: overall survival; HF: hepatic failure; RF: respiratory failure; DNA mismatch repair complex expression, shown as the percentage of tumor cells stained in immunohistochemistry by each Ab specific for the indicated protein of the complex. Indicated are the nuclear immunoreactivity (N) and the intensity of immunoreactivity on a 0-3 scale.

Supplemental Table 2. HLA typing of CRC cells investigated for T cell recognition of unique tumor antigens.

Patient	**HLA typing**								
	HLA-A	**HLA-B**	**HLA-Cw**	**HLA-DRβ1**	**HLA-DRβ3**	**HLA-DRβ4**	**DQA1**	**DQB1**	**DPB1**
1039	02:01, 33:01	10:24, 51:01	01:02, 08:02	01:02, 11:01	02:02		01:01, 05:05	03:01, 05:01	04:01, 11:01
1247	02, 03	15:01, 15:03	03:03, 04:01	01:01, 1101	02:02		01:01, 05:05	03:01, 05:01	03:01, 09:02
1869	03:01, 24:03	35:01, 38:01	0401, 1203	04:02, 13:01	01:01	01:03	0103, 0301	0302, 0603	02:01, 04:01

Supplemental Table 3. Synthetic peptides encompassing the somatic mutations found in the CRC cDNAs

CRC	peptides
1039	KRAS^{G12D}
	peptide mut-1 MTEYKLVVVGADGVG
	peptide mut-2 ---------KLVVVGADGVGKSAL
	peptide mut-3 ----------VGADGVGKSALTIQL
	peptidie WT MTEYKLVVVGA**G**GVG
	TP53^{Y107D}
	peptide mut-1 SVPSQKTYQGSDGFR
	peptide mut-2 ---------QKTYQGSDGFRLGFL
	peptide mut-3 ----------QGSDGFRLGFLHSGT
	peptide WT SVPSQKTYQGS**Y**GFR
	PIK3CA^{Q546K}
	peptide mut-1 STRDPLSEITEKEKD
	peptide mut-2 --------PLSEITEKEKDFLWS
	peptide mut-3 ---------- ITEKEKDFLWSHRHY
	peptide WT STRDPLSEITE**Q**EKD
1247	PIK3CA^{R770Q}
	peptide mut-1 HQLGNLRLEECQIMS
	peptide mut-2 --------NLRLEECQIMSSAKR
	peptide mut-3 ----------EECQIMSSAKRPLWL
	peptide WT HQLGNLRLEEC**R**IMS
	SMAD4^{V370A}
	peptide mut-1 GDRFCLGQLSN<u>A</u>HRT
	peptide mut-2 ---------CLGQLSN<u>A</u>HRTEAIE
	peptide mut-3 ---------- LSN<u>A</u>HRTEAIERARL
	peptide WT-1 GDRFCLGQLSN**V**HRT
	C10orf137^{S168L}
	peptide mut-1 DELDIQELFMRLSQT
	peptide mut-2 ---------IQELFMRLSQTGDWT
	peptide mut-3 ----------FMRLSQTGDWTWLKE
	peptide WT DELDIQELFMR**S**SQT
1869	APC E1317KfsX4 (1288-1318)
	peptide mut NQTTQEADSANTLQIAEIKEKIGTRSA<u>KIL</u>
	peptide WT NQTTQEADSANTLQIAEIKEKIGTRSA**EDP**
	KRAS^{G12D}
	peptide mut-1 MTEYKLVVVGADGVG
	peptide mut-2 ---------KLVVVGADGVGKSAL
	peptide mut-3 ----------VGADGVGKSALTIQL
	peptide WT MTEYKLVVVGA**G**GVG
	TP53^{R175H}
	peptide mut-1 KQSQHMTEVVR<u>H</u>CPH
	peptide mut-2 -------- HMTEVVR<u>H</u>CPHHERC
	peptide mut-3 ----------VVR<u>H</u>CPHHERCSDSD
	peptide WT-1 KQSQHMTEVVR**R**CPH
	SMAD4^{R361C}
	peptide mut-1 VDGYVDPSGGD<u>C</u>FCL
	peptide mut-2 ----------VDPSGGD<u>C</u>FCLGQLS
	peptide mut-3 ----------GGD<u>C</u>FCLGQLSNVHR
	peptide WT VDGYVDPSGGD**R**FCL

Name	Description	Sequence (5'->3')	Strand on template	Length	Start	Stop	Tm	Product length
APCforward	APC and APC(1) amplification forward primer	GAATGGAGGTGCTGCCGGACTCGGAAATGG	plus	30	9	39	70,9	
APCrev1	APC whole orf amplification reverse primer	GCATGTATCTCCATTGTTTATGGAAGCCTGG	minus	31	10701	10670	90	10692
APCrev2	APC(1) amplification reverse primer	TCCTCCTTGAGCCTCATCTGTACTTCTGC	minus	29	5167	5138	68	5158
up2 APC	APC(2) amplification forward primer	GGCACAGTCAGGTGAATTTG	plus	20	5091	5110	57,3	
rev2 APC	APC(2) amplification reverse primer	GGCTTCCAGAACAAAAACCCTC	minus	22	8731	8710	60,2	3640
up KRAS	KRAS amplification forward primer	CATTTCGGACTGGGAGCGAG	plus	20	85	104	61,4	
down KRAS	KRAS amplification reverse primer	CTAACAGTCTGCATGGAGCAG	minus	21	872	852	59,8	787
up EPH Receptor B6	EPHRB6 amplification forward primer	GAGTCTTGCAAAAGCTGCAG	plus	20	711	730	57,3	
down EPH Receptor B6	EPHRB6 amplification reverse primer	CTGAGTCACGGGTATCGTC	minus	19	3876	3858	58,8	3165
up PIK3CA	PIK3CA amplification forward primer	CGTTTCTGCTTTGGGACAAC	plus	20	81	100	57,3	
down PI3CA	PIK3CA amplification reverse primer	GCAGTGTGGAATCCAGAGTG	minus	20	3408	3390	62,7	3327
up TP53	TP53 amplification forward primer	CAGACTGCCTTCCGGGTC	plus	18	227	245	60,5	
down TP53	TP53 amplification reverse primer	CTGTCAGTGGGGAACAAGAAG	minus	21	1463	1441	59,8	1236
up FBXW7	FBXW7 amplification forward primer	CTAGCCAAGGTCCAAGAAGTAGC	plus	23	111	133	62,4	
down FBXW7	FBXW7 amplification reverse primer	GGCAGGGAGTATATCGTCTACAC	minus	23	2324	2302	62,4	2213
up SMAD4	SMAD4 amplification forward primer	CCTTGCAACGTTAGCTGTTG	plus	20	441	460	57,3	
down SMAD4	SMAD4 amplification reverse primer	GTCCACCATCCTGATAAGGTTAAGG	minus	25	2241	2217	62,9	1800
up 1 CSMD3	CSMD3(1) amplification forward primer	ACAACAGCAACAACTCCACTGC	plus	22	99	120	62,7	
down 1 CSMD3	CSMD3(1) amplification reverse primer	CCACCAAGACAGATGATCTCTG	minus	22	3823	3812	60,2	3724
up 2 CSMD3	CSMD3(2) amplification forward primer	GACCTTCTCATGCTCTTCGG	plus	20	3771	3790	59,3	
down 2 CSMD3	CSMD3(2) amplification reverse primer	GGCAGGACAGAGCACTTGAC	minus	20	7533	7514	61,4	3762
up 3 CSMD3	CSMD3(3) amplification forward primer	GAACGACTGCAGATGGATGG	plus	20	7480	7499	62,7	
down 3 CSMD3	CSMD3(3) amplification reverse primer	CTGATAGTGGATGCTCCTCC	minus	20	11591	11571	59,3	4111
up TNN	TNN amplification forward primer	GGTCTGCTCCCTGTCTTTCC	plus	20	92	109	61,4	
down TNN	TNN amplification reverse primer	TGTCTCCTGCGAGGACTGC	minus	19	4044	4026	60,9	3952
up EPH A3	EPH A3 amplification forward primer	GAGATATGCTCCTCTCACTGCC	plus	22	189	2110	62,1	
down EPH A3	EPH A3 amplification reverse primer	TTCCGTCCAGAAGCACTTCC	minus	20	3207	3189	59,3	3018
up 1 NAV3	NAV3(1) amplification forward primer	GTCTACCAGACTGAGGTTAGAAGC	plus	24	112	135	62,7	
down 1 NAV3	NAV3(1) amplification reverse primer	GCTGGGCATTACTGAGGAAG	minus	20	4004	3985	62,7	3892
up 2 NAV3	NAV3(2) amplification forward primer	CTGCCTCTGCACCTAATACTGAG	plus	23	3952	3974	62,4	
down 2 NAV3	NAV3(2) amplification reverse primer	CACCCTCTAGAGGGTAGATTCAAG	minus	24	7271	7248	62,7	3319
up 1 LAMA1	LAMA1(1) amplification forward primer	CTTTCTCCCCAGACCCACCGAG	plus	22	57	78	65,8	
down 1 LAMA1	LAMA1(1) amplification reverse primer	TGAGTGTGTGGGGCAGTCACAGG	minus	22	3105	3084	63,9	3048
up 2 LAMA1	LAMA1(2) amplification forward primer	GTGTCACTGTGTCCCAGGTGTGG	plus	23	3020	2958	65,9	
down 2 LAMA1	LAMA1(2) amplification reverse primer	GGTTTGCCTGGTAATTTCAACAGC	minus	24	6064	6041	61	3044
up 3 LAMA1	LAMA1(3) amplification forward primer	GCAGGAAGCTTCCAGGTATTGC	plus	22	5970	5991	62,1	
down 3 LAMA1	LAMA1(3) amplification reverse primer	GCAATGATTCCAACTGAGGATTCTGC	minus	26	9355	9330	63,2	3385
up C10orf.137	C10orf.137 amplification forward primer	GTATCGCCTGCCCTGGATCG	plus	20	269	288	63,9	
down C10orf.137	C10orf.137 amplification reverse primer	TCTGGGACACGGCTCTGTGC	minus	20	3922	3911	63,4	3653
down MAP2K7	MAP2K7 amplification reverse primer	CTCTCTGTCCTCAGTCCTAGGTG	minus	23	1456	1434	59,3	
down MAP2K7	MAP2K7 amplification reverse primer	CTCTCTGTCCTCAGTCCTAGGTG	minus	23	1456	1434	59,3	1393
up PTEN	PTEN amplification forward primer	GCTTCTGCCATCTCTCTCCTCC	plus	22	983	1004	63,9	
down PTEN	PTEN amplification reverse primer	TCTGACACAATGTCCTATTGCC	minus	22	2343	2322	58,3	1360
up ADAMTSL3	ADAMTSL3 amplification forward primer	GCTACAACTGAGACCGGAGG	plus	21	192	212	59,3	
down ADAMTSL3	ADAMTSL3 amplification reverse primer	CTTCACAGCAGCATCATGACC	minus	21	5322	5312	59,8	5130
up GUCY1A2	GUCY1A2 amplification forward primer	CGCTTAACGTTGTCGCTTGC	plus	20	293	313	59,3	
down GUCY1A2	GUCY1A2 amplification reverse primer	GAGGAGTCTTTGATCTGTAGCAGG	minus	24	2615	2592	62,7	2322
up SMAD2	SMAD2 amplification forward primer	AGAGGCTGTTTTCCTAGCGTGG	plus	22	206	227	62,1	
down SMAD2	SMAD2 amplification reverse primer	CCATAGGGACCACACACAATGC	minus	22	1729	1707	62,1	1523
up OR51E1	OR51E1 amplification forward primer	GGTGCTGGTCACAGTTCAGC	plus	20	119	138	62,1	
down OR51E1	OR51E1 amplification reverse primer	CTGAATCAGAGGACTCTGAATGG	minus	23	1147	1125	60,6	1028
up TCF7L2	TCF7L2 amplification forward primer	CCAAAATTGCTGCTGGTGGG	plus	20	478	497	59,3	
down TCF7L2	TCF7L2 amplification reverse primer	AACAAAACCTTTCAAGGTGGGG	minus	22	2387	2366	58,3	1909
up MAP2K7	MAP2K7 amplification forward primer	GAAGATGGCGGCGTCCTC	plus	18	63	79	60,5	
CAN01 TNN for	TNN amplification forward primer	TCTGCTCCCTGTCTTTCCAAGG	plus	22	92	113	62,1	
CAN02 TNN rev	TNN amplification reverse primer	TCTCCTGCGAGGACTGCTCAC	minus	21	4044	4026	63,7	3950
CAN03 EPHA3 for	EPH A3 amplification forward primer	TGGAGATATGCTCCTCTCACTGC	plus	23	187	209	62,4	
CAN04 EPHA3 rev	EPH A3 amplification reverse primer	TCTGTCTGCAGGATGATGCTACG	minus	23	3247	3225	62,4	3060
CAN05 MAP2K7 for	MAP2K7 amplification forward primer	TCCTCCCTGGAACAGAAGCTGTC	plus	23	75	97	64,2	1381
CAN06 ADAMTSL3 for	ADAMTSL3 amplification forward primer	ATTCCGCACGAGGTGTTGACG	plus	21	86	106	62,1	
CAN07 ADAMTSL3 sv rev	ADAMTSL3 splice variant amplification reverse primer	TTTGGTGTGGATGTCCTTCTG	minus	22			58,4	1154
CAN08 GUCY1A2 for	GUCY1A2 amplification forward primer	TTAACGTTGTCGCTTGCCGGTCC	plus	23	296	318	64,2	2319
CAN09 OR51E1 for	OR51E1 amplification forward primer	ACATTCCTTCCATACGGTTGAGC	plus	23	84	106	60,6	1064
CAN10 TCF7L2 for	TCF7L2 amplification forward primer	TCTTCCAAAATTGCTGCTGGTGG	plus	23	474	492	60,6	
CAN11 TCF7L2 rev	TCF7L2 amplification reverse primer	ACATCAACTAATGTAGCCACATGG	minus	24	2432	2409	62,1	1958
CAN12 qRT TNN for	semi quantitative RT-PCR forward primer	ATATGGCCACTGTCTCCTGG	Plus	20	2809	2828	59.95	364
CAN13 qRT TNN rev	semi quantitative RT-PCR reverse primer	CCCAGCTGCATCTCCTTAAC	Minus	20	3172	3153	59.84	
CAN14 qRT NAV3 for	semi quantitative RT-PCR forward primer	TTCAGCAAATGCTCACCTTG	Plus	20	4949	4968	59.99	412
CAN15 qRT NAV3 rev	semi quantitative RT-PCR reverse primer	ATGTGATGAAGGAGGCTTGG	Minus	20	5360	5341	60.07	
CAN16 qRT EPHA3 for	semi quantitative RT-PCR forward primer	CCACCCCAATATCATTCGAC	Plus	20	2259	2278	60.01	329
CAN17 qRT EPHA3 rev	semi quantitative RT-PCR reverse primer	TTGGGATCTTCCCTCCTCTT	Minus	20	2587	2568	60.01	
CAN18 qRT ADAMTSL3 for	semi quantitative RT-PCR forward primer	ACAGGGAGTGTGTCCCAAAG	Plus	20	3687	3706	60.00	394
CAN19 qRT ADAMTSL3 rev	semi quantitative RT-PCR reverse primer	TGACAGGTGCCTCTGCATAC	Minus	20	4080	4061	59.86	
CAN20 qRT GUCY1A2 for	semi quantitative RT-PCR forward primer	TGGAAGACCGATTCAGATGAG	Plus	21	2211	2231	60.20	357
CAN21 qRT GUCY1A2 rev	semi quantitative RT-PCR reverse primer	AACATGGTGCCGATGTTGTA	Minus	20	2567	2548	59.85	
CAN22 qRT OR51E1 for	semi quantitative RT-PCR forward primer	TTCCATACGGTTGAGCCTCT	Plus	20	91	110	59.69	382
CAN23 qRT OR51E1 rev	semi quantitative RT-PCR reverse primer	AGGAGTGGATGGCAAACATC	Minus	20	472	453	59.93	
CAN24 APC all splvar for	APC(1) amplification all known spliced variant, forward primer	GGCTGCAGCTTCATATGATCAGTTGTT	Plus	27	88	114	58.08	5107
CAN25 APC exonanal rev	APC analysis of first 15 exons reverse primer	CCACAAAGTTCCACATGCATTACTGAC	Minus	27	2248	2222	57.11	2053
CAN26 bACT for	beta ACTIN human amplification forward primer	AGCCTCGCCTTTGCCGATCC	Plus	20	34	53	59.49	
CAN27 bACT rev	beta ACTIN human amplification reverse primer	CCCAGGGAGACCAAAAGCCTTCA	Minus	23	1735	1713	58.49	1701
CAN28 BRAF for	BRAF amplification forward primer	GTTATAAGATGGCGGCGCTGAG	Plus	22	54	75	56.01	
CAN29 BRAF rev	BRAF amplification reverse primer	TTTGTTGCTACTCTCCTGAACTCTC	Minus	25	2400	2376	54.56	2347
up-1247-SMAD4 WT	PCR typing expression SMAD4 WT in tumor cell lines	GCTTTTGTTTGGGTCAACTCTCCAATGT	Plus	28			64	629
up-1247-SMAD4 Mut	PCR typing expression SMAD4 Mut in tumor cell lines	GCTTTTGTTTGGGTCAACTCTCCAATGC	Plus	28			65	629
up1869-SMAD4 WT	PCR typing expression SMAD4 WT in tumor cell lines	CGTGGACCCTTCTGGAGGAGATC	Plus	23			66	640
up1869-SMAD4 Mut	PCR typing expression SMAD4 Mut in tumor cell lines	CGTGGACCCTTCTGGAGGAGATT	Plus	23			64	640
up1869-TP53 WT	PCR typing expression TP53 WT in tumor cell lines	CACATGACGGAGGTTGTGAGGCG	Plus	23			66	729
up1869-TP53 Mut	PCR typing expression TP53 Mut in tumor cell lines	CACATGACGGAGGTTGTGAGGCA	Plus	23			64	729