**Supplemental Information**

**CD8+ T cell immune surveillance against a tumor antigen encoded by an oncogenic long non-coding RNA, *PVT1***

Yasuhiro Kikuchi1, Serina Tokita1,2, Tomomi Hirama1,2, Vitaly Kochin1,3, Munehide Nakatsugawa1,4, Tomoyo Shinkawa1, Yoshihiko Hirohashi1, Tomohide Tsukahara1, Fumitake Hata2, Ichiro Takemasa5, Noriyuki Sato1,2, Takayuki Kanaseki1\*, Toshihiko Torigoe1

1 Department of Pathology, Sapporo Medical University, Japan

2 Sapporo Dohto Hospital, Japan

3 Department of Immunology, Nagoya University, Japan

4 Department of Pathology, Tokyo Medical University Hachioji Medical Center, Japan

5 Department of Surgery, Surgical Oncology and Science, Sapporo Medical University, Japan

**Figure S1.**

Six formalin-fixed paraffin embedded (FFPE) CRC tissues were mounted and stained with hematoxylin-eosin (HE), anti-MSH6 (EP49, DAKO), anti-PMS2 (EP51, DAKO), and anti-pan HLA class I (EMR8-5, Hokudo), along with corresponding secondary antibodies on DAKO autostainers (magnification, ×100). All tissues were histologically diagnosed as adenocarcinoma, and classified as a mismatch repair proficient (pMMR) type according to intact MSH6 and PMS2 expression in cancer cells. HLA class I staining showed moderate to strong protein expression on the surface of cancer cells in all cases.

**Figure S2.**

**A**, Length distribution (left), and sequence logos of 9-mer peptides (right), of the HLA-A24 ligands identified in 6 CRC and 4 matched normal tissues. Both canonical and cryptic ligands showed similar sequence profiles: 9-mer peptides were dominant, and Y/F at P2 and F/L/I at P9 were strongly conserved across ligands. **B**, Kernel density estimation of genes expressed in the indicated sample groups. The black lines represent the whole expressed genes (TPM > 0), and the red lines represent the genes encoding canonical or cryptic HLA-A24 ligands. In contrast to the whole gene, the source genes of both HLA-A24 ligand types were biased toward high expression values. **C**, Length (top) and position (bottom) distribution of the ORFs encoding canonical and cryptic HLA-A24 ligands identified in CRC tissues. With regard to the length distribution, violin plots show that the lengths of ORFs encoding cryptic ligands are significantly shorter (313 nucleotides on average) compared with canonical ORFs (2,770 nucleotides on average). ORF positions represent the order of the ATG codon appearing within a corresponding transcript in the 5’ to 3’ direction. In the position distribution, violin plots indicate that translation does not always start from the first ATG but begins with the 2nd or later ATG. These profiles unique to cryptic ligands, are constantly observed across all CRC tissues.

**Figure S3.**

**A**, Normalized abundance distribution of HLA-A24 ligands across CRC and matched normal tissues measured by the LFQ method. **B**, Heatmap showing the differential scaled abundance of HLA-A24 ligands in CRC and matched normal tissues. Each horizontal line represents an HLA-A24 ligand. Ligands with a scaled abundance of more than 100 are indicated in red, and those either below 100 or undetected in the corresponding sample are indicated in blue (scaled abundance, 0-1,000). Hierarchical clustering (Euclidean distance measure and Ward algorithm) stratifies the ligands based on the scaled abundance across the samples, and a CRC-enriched group is included in green.

**Figure S4.**

**A**, Composition of immune cells defined by z-scores of MCP-counter on RNA-seq data[1](#_ENREF_1). CD8+ T cells or cytotoxic lymphocytes are enriched in CRC tissues compared with matched normal mucosa. **B**, Immunohistochemistry of CRC tissues stained with anti-CD8 (C8/114B, DAKO) (magnification, ×200). Red-dotted lines represent tumor-invasive margins. **C,** The number of CD8+ TILs observed along tumor invasive margins and inside tumor parenchyma were separately counted in 10 high-power fields. Distribution patterns indicate excluded-type CD8+ T-cell infiltration in the pMMR CRC tissues[2](#_ENREF_2).

**Figure S5.**

**A**, MS/MS spectra of endogenous and synthetic HWNDTRPAHF (HF10) peptide, detected in the HLA-A24 ligandome of Colo320 at an FDR of 0.01. The corresponding b- and y-fragment ions are indicated. **B**, Pie charts showing the proportion of cryptic HLA ligands identified in Colo320 (left), and composition of cryptic ligands classified according to the GENCODE biotypes of the source genes (right). Numbers in parentheses indicate the numbers of nonredundant ligands (left) and corresponding gene types (right). **C**, **D**, Length (top) and position (bottom) distribution of the ORFs encoding canonical and cryptic HLA-A24 ligands identified in Colo320. ORF positions represent the order of the ATG codon appearing within a corresponding transcript in the 5’ to 3’ direction.

**Figure S6.**

**A**, Schematic representation of full-length *PVT1* transcript variants encoding HF10. The transcript variants were detected in Colo320 using Iso-seq. Black and red boxes indicate the positions of the ORFs and HF10, respectively. A unique ORF sequence encoding HF10 is shared across all variants. The shared ORF ends with TGA, which is located 130 nucleotides upstream of an exon-exon junction and regarded as a premature termination codon (PTC, red arrows). Black arrows indicate the positions of a primer pair used for RT-PCR in order to evaluate the expression of *PVT1* transcripts harboring the HF10-encoding ORF. **B**, Sequence of the HF10-encoding ORF.

**Figure S7**

**A**, Quantitative-PCR of Colo320 transfected with siRNAs targeting the exon 2 of *PVT1* transcripts (Silencer Select Pre-designed siRNA, siRNA IDs: n545814 (siRNA *PVT1* #2), n272515 (siRNA *PVT1* #3), n272519 (siRNA *PVT1* #4), all obtained from Thermo), or a negative control siRNA (Silencer Select Negative Control #1 siRNA, 4390843 (siRNA ctrl #2), Thermo). Data are representative of three independent experiments. **B**, IFNγ ELISPOT assay of clone H3 cultured with transfected Colo320. **C**, *In vitro* cell proliferation of transfected Colo320. Transfected Colo320 cells were cultured *in vitro* for 72 h, and cell growth was measured by absorbance at 450 nm using WST-8 reagent. Data in B and C represent means with SEM (n = 3 for B and n = 10 for C); and p-values were calculated using a two-tailed t-test (\*\*\* p<0.001). **D**, *In vivo* tumor development of *PVT1*-knock down Colo320 cells. Briefly, 5x106 Colo320 cells were subcutaneously injected into NSG mice, 48 h after transfection. Data represent means with SEM (n = 5), and p-values were calculated using a two-tailed t-test (\* p < 0.05; \*\* p < 0.01).

**Reference**

1 Becht, E. *et al.* Estimating the population abundance of tissue-infiltrating immune and stromal cell populations using gene expression. *Genome Biol* **17**, 218, doi:10.1186/s13059-016-1070-5 (2016).

2 Hegde, P. S., Karanikas, V. & Evers, S. The Where, the When, and the How of Immune Monitoring for Cancer Immunotherapies in the Era of Checkpoint Inhibition. *Clin Cancer Res* **22**, 1865-1874, doi:10.1158/1078-0432.CCR-15-1507 (2016).