

## CANCER

# Noncoding regions are the main source of targetable tumor-specific antigens

Céline M. Laumont<sup>1,2\*</sup>, Krystal Vincent<sup>1,2\*</sup>, Leslie Hesnard<sup>1,2</sup>, Éric Audemard<sup>1</sup>, Éric Bonneil<sup>1</sup>, Jean-Philippe Laverdure<sup>1</sup>, Patrick Gendron<sup>1</sup>, Mathieu Courcelles<sup>1</sup>, Marie-Pierre Hardy<sup>1</sup>, Caroline Côté<sup>1</sup>, Chantal Durette<sup>1</sup>, Charles St-Pierre<sup>1,2</sup>, Mohamed Benhammedi<sup>1,2</sup>, Joël Lanoix<sup>1</sup>, Suzanne Vobecky<sup>3</sup>, Elie Haddad<sup>3</sup>, Sébastien Lemieux<sup>1,4</sup>, Pierre Thibault<sup>1,5†</sup>, Claude Perreault<sup>1,2‡</sup>

Copyright © 2018  
The Authors, some  
rights reserved;  
exclusive licensee  
American Association  
for the Advancement  
of Science. No claim  
to original U.S.  
Government Works

Tumor-specific antigens (TSAs) represent ideal targets for cancer immunotherapy, but few have been identified thus far. We therefore developed a proteogenomic approach to enable the high-throughput discovery of TSAs coded by potentially all genomic regions. In two murine cancer cell lines and seven human primary tumors, we identified a total of 40 TSAs, about 90% of which derived from allegedly noncoding regions and would have been missed by standard exome-based approaches. Moreover, most of these TSAs derived from nonmutated yet aberrantly expressed transcripts (such as endogenous retroelements) that could be shared by multiple tumor types. Last, we demonstrated that, in mice, the strength of antitumor responses after TSA vaccination was influenced by two parameters that can be estimated in humans and could serve for TSA prioritization in clinical studies: TSA expression and the frequency of TSA-responsive T cells in the preimmune repertoire. In conclusion, the strategy reported herein could considerably facilitate the identification and prioritization of actionable human TSAs.

## INTRODUCTION

CD8<sup>+</sup> T cells are the main mediators of naturally occurring and therapeutically induced immune responses to cancer. Accordingly, the abundance of CD8<sup>+</sup> tumor-infiltrating lymphocytes (TILs) positively correlates with response to immune checkpoint inhibitors and favorable prognosis (1–3). Because CD8<sup>+</sup> T cells recognize major histocompatibility complex class I (MHC I)–associated peptides, the most important unanswered question is the nature of the specific peptides recognized by CD8<sup>+</sup> TILs (4). Knowing that the abundance of CD8<sup>+</sup> TILs correlates with the mutation load of tumors, the dominant paradigm holds that CD8<sup>+</sup> TILs recognize mutated tumor-specific antigens (mTSAs), commonly referred to as neoantigens (2, 5, 6). The superior immunogenicity of mTSAs is ascribed to their selective expression on tumors, which minimizes the risk of immune tolerance (7). Nonetheless, some TILs have been shown to recognize cancer-restricted nonmutated MHC peptides (8) that we will refer to as aberrantly expressed TSAs (aeTSAs). aeTSAs can derive from a variety of cis- or trans-acting genetic and epigenetic changes that lead to the transcription and translation of genomic sequences normally not expressed in cells, such as endogenous retroelements (EREs) (9–11).

Considerable efforts are being devoted to discovering actionable TSAs that can be used in therapeutic cancer vaccines. The most common strategy hinges on reverse immunology, in which exome sequencing is performed on tumor cells to identify mutations, and MHC-binding prediction software tools are used to identify which mutated peptides might be good MHC binders (12, 13). Although reverse immunology can enrich for TSA candidates, 90% of these candidates are false positives (6, 14) because available computational

methods may predict MHC binding, but they cannot predict other steps involved in MHC peptide processing (15, 16). To overcome this limitation, a few studies have included mass spectrometry (MS) analyses in their TSA discovery pipeline (17), thereby providing a rigorous molecular definition of several TSAs (18, 19). However, the yield of these approaches has been meager: In melanoma, one of the most mutated tumor types, an average of two TSAs per individual tumors has been validated by MS (20), whereas only a handful of TSAs has been found for other cancer types (15). The paucity of TSAs is puzzling because injection of TILs or immune checkpoint inhibitors would not cause tumor regression if tumors did not express immunogenic antigens (21). We surmised that approaches based on exonic mutations have failed to identify TSAs because they did not take into account two crucial elements. First, these approaches focus only on mTSAs and neglect aeTSAs, essentially because there is currently no method for high-throughput identification of aeTSAs. This represents a major shortcoming because, whereas mTSAs are private antigens (that is, unique to a given tumor), aeTSAs would be preferred targets for vaccine development because they can be shared by multiple tumors (8, 10). Second, focusing on the exome as the only source of TSAs is very restrictive. Of particular relevance to TSA discovery, 99% of cancer mutations are located in noncoding regions (22). Moreover, the exome (all protein-coding sequences) is only 2% of the human genome, whereas up to 75% of the genome can be transcribed and potentially translated (23). Hence, many allegedly noncoding regions are protein coding, and translation of noncoding regions has been shown to generate numerous MHC peptides (24, 25), some of which were retrospectively identified as targets of TILs and autoreactive T cells (26, 27).

With these considerations in mind, we developed a proteogenomic strategy designed to discover mTSAs and aeTSAs coded by all genomic regions. We used this approach to study two well-characterized murine cancer cell lines, CT26 and EL4, as well as seven primary human samples comprising four B-lineage acute lymphoblastic leukemias (B-ALLs) and three lung cancers. Our main objectives were to determine whether noncoding regions contribute to the TSA landscape and which parameters may influence TSA immunogenicity.

<sup>1</sup>Institute for Research in Immunology and Cancer (IRIC), Université de Montréal, Montréal, Quebec H3C 3J7, Canada. <sup>2</sup>Department of Medicine, Université de Montréal, Montréal, Quebec H3C 3J7, Canada. <sup>3</sup>CHU Sainte-Justine, Université de Montréal, Montréal, Quebec H3T 1C5, Canada. <sup>4</sup>Department of Computer Science and Operations Research, Université de Montréal, Montréal, Quebec H3C 3J7, Canada. <sup>5</sup>Department of Chemistry, Université de Montréal, Montréal, Quebec H3C 3J7, Canada.

\*These authors contributed equally to this work.

†These authors jointly supervised this work.

‡Corresponding author. Email: [claud.perreault@umontreal.ca](mailto:claud.perreault@umontreal.ca)

## RESULTS

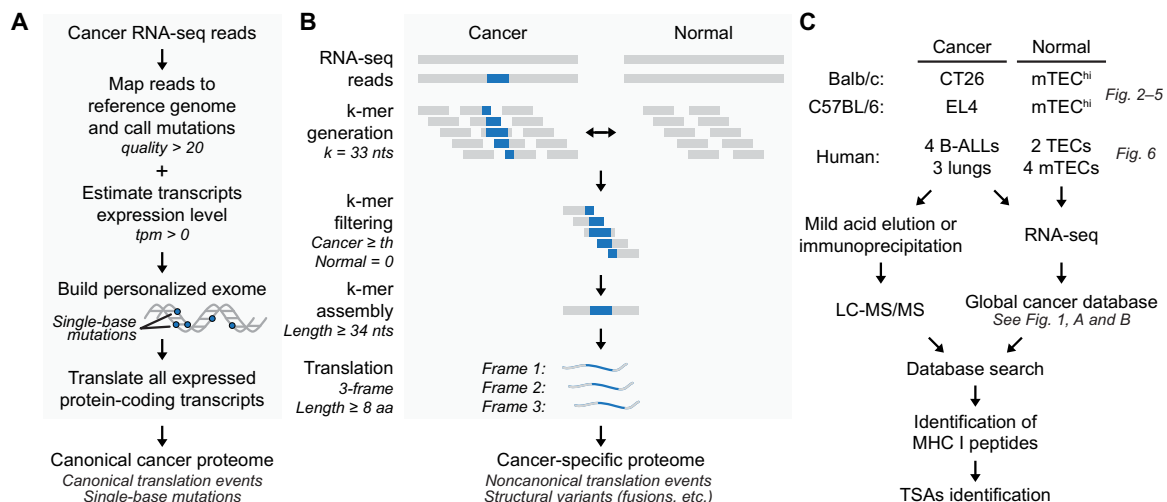
## Rationale and design of a proteogenomic method for TSA discovery

Attempts to computationally predict TSAs using various algorithms are fraught with exceedingly high false discovery rates (28). Hence, a system-level molecular definition of the MHC peptide repertoire may only be achievable by high-throughput MS studies (4). Current approaches use tandem MS (MS/MS) software tools, such as Peaks (29), which rely on a user-defined protein database to match each acquired MS/MS spectrum to a peptide sequence. Because the reference proteome does not contain TSAs, MS-based TSA discovery workflows must use proteogenomic strategies to build customized databases derived from tumor RNA sequencing (RNA-seq) data (30) that should ideally contain all proteins, even unannotated ones, expressed in the considered tumor sample. Because current MS/MS software tools cannot deal with the large search space created by translating all RNA-seq reads in all reading frames (31, 32), we devised a proteogenomic strategy enriching for cancer-specific sequences to comprehensively characterize the landscape of TSAs coded by all genomic regions. The resulting database, termed a global cancer database, is composed of two customizable parts. The first part, the canonical cancer proteome (Fig. 1A), was obtained by *in silico* translation of expressed protein-coding transcripts in their canonical frame; it therefore contains proteins coded by exonic sequences that are normal or contain single-base mutations. The second part, the cancer-specific proteome (Fig. 1B), was generated using an alignment-free RNA-seq workflow called k-mer profiling because current mappers and variant callers poorly identify structural variants. This second dataset enabled the detection of peptides encoded by any reading frame of any genomic origin (including structural variants), as long as they were cancer specific (that is, absent from normal cells). Here, we elected to use MHC II<sup>hi</sup> medullary thymic epithelial cells (mTEC<sup>hi</sup>) cells as a “normal control” because they express most known genes and orchestrate T cell selec-

tion to induce central tolerance to MHC peptides coded by their vast transcriptome (fig. S1A) (33). Thus, to identify RNA sequences that were cancer specific, we chopped cancer RNA-seq reads into 33-nucleotide-long sequences, called k-mers (34), from which we removed k-mers present in syngeneic mTEC<sup>hi</sup> cells (fig. S2, A and B). Redundancy inherent to the k-mer space was removed by assembling overlapping cancer-specific k-mers into longer sequences, called contigs, which were 3-frame translated *in silico* (Fig. 1B and fig. S2, C and D). We then concatenated the canonical and cancer-specific proteomes to create a global cancer database, one for each analyzed sample (table S1A). Using these optimized databases, we identified MHC peptides eluted from two well-characterized mouse tumor cell lines that we sequenced by MS, namely CT26, a colorectal carcinoma from a Balb/c mouse, and EL4, a T-lymphoblastic lymphoma from a C57BL/6 mouse (Fig. 1C and table S2A) (35, 36).

## Noncoding regions as a major source of TSAs

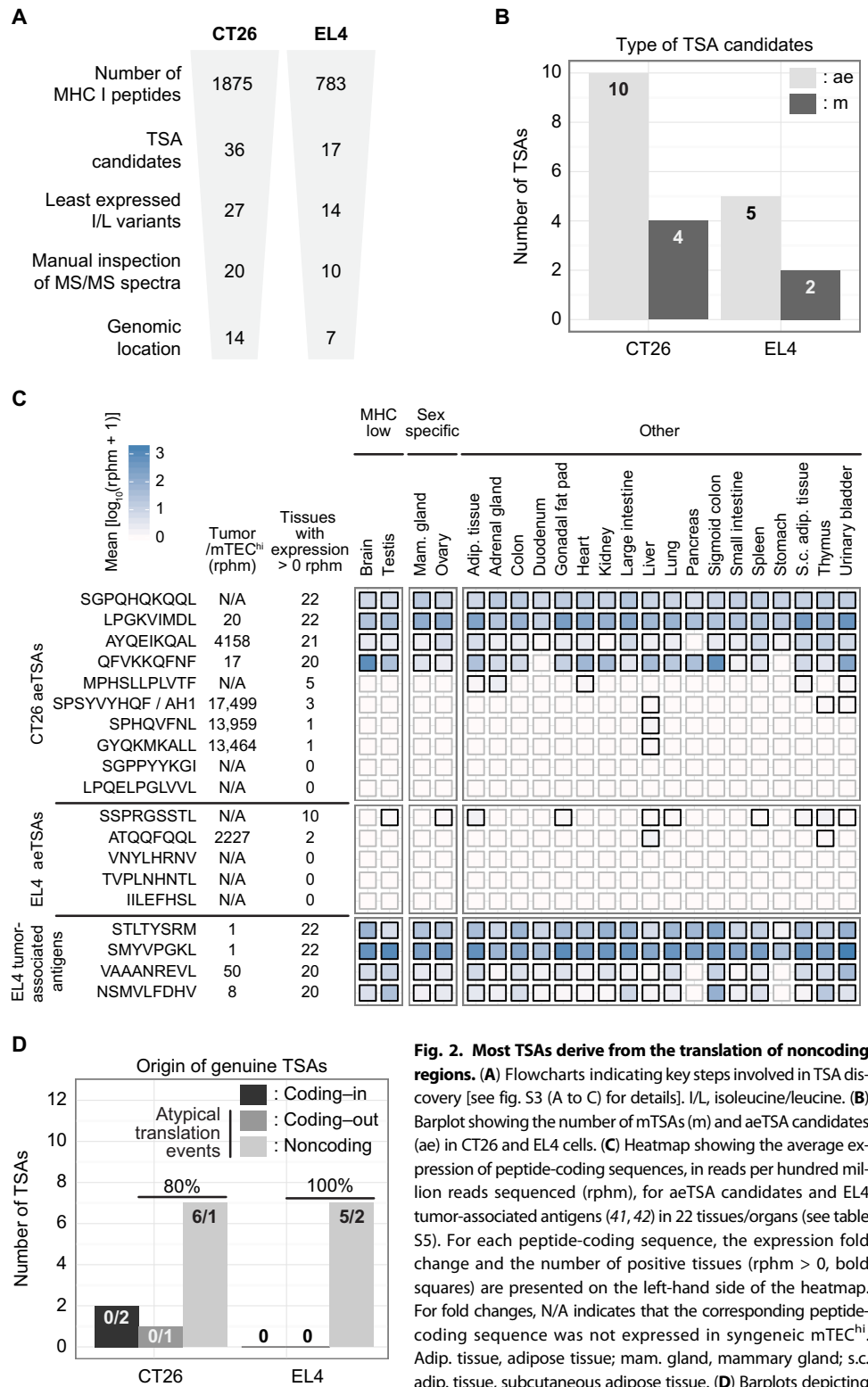
We identified 1875 MHC peptides on CT26 cells and 783 on EL4 cells (tables S3 and S4). Among these, peptides absent from the mTEC<sup>hi</sup> proteome were considered TSA candidates (i) if their 33-nucleotide-long peptide-coding sequence derived from a full cancer-restricted 33-nucleotide-long k-mer and was absent from the mTEC<sup>hi</sup> transcriptome or (ii) if their 24- to 30-nucleotide-long peptide-coding sequence, derived from a truncated version of a cancer-restricted 33-nucleotide-long k-mer, was overexpressed by at least 10-fold in the transcriptome of cancer cells versus mTEC<sup>hi</sup> cells (fig. S3A). Because no error estimation was used in our study, we manually validated the MS spectra of our TSA candidates. Before assigning peptides a genomic location, we also removed any indistinguishable isoleucine/leucine variants (figs. S3, B and C, and S4) and ended up with a total of 6 mTSAs and 15 aeTSA candidates: 14 presented by CT26 cells and 7 by EL4 cells (Fig. 2, A and B). MHC peptides that were both mutated and aberrantly expressed were included in the mTSA category. All of these peptides



**Fig. 1. Proteogenomic workflow for the identification of TSAs.** (A and B) Schematic detailing how the canonical cancer proteome (A) and cancer-specific proteome (B) were built for each analyzed sample. In (A), “quality” refers to the Phred score; a score of >20 means that the accuracy of the nucleobase call is at least 99%. (C) The combination of the above two proteomes, termed the global cancer database, was then used to identify MHC peptides, and more specifically TSAs, sequenced by liquid chromatography–MS/MS (LC-MS/MS). We analyzed two well-characterized murine cell lines, CT26 and EL4, and seven human primary samples, namely, four B-ALLs and three lung tumor biopsies ( $n = 2$  to 4 per sample). Statistics regarding each part of the global cancer database can be found in table S1, and implementation details of building the cancer-specific proteome by k-mer profiling are presented in fig. S2. aa, amino acids; nts, nucleotides; th, sample-specific threshold for k-mer occurrence; tpm, transcripts per million.

were absent from the Immune Epitope Database (37), except for one: the AH1 peptide (SPSYVYHQF), the sole aeTSA previously identified on CT26 cells using reverse immunology (10, 38).

To assess the stringency of our database-building strategy based on the removal of mTEC<sup>hi</sup> k-mers from cancer k-mers, we evaluated the peripheral expression of RNAs coding for aeTSAs across a panel of 22 tissues (table S5) (39, 40). Four of the 15 aeTSA candidates had an expression profile similar to that of previously reported “overexpressed” tumor-associated antigens (41, 42), as their peptide-coding sequences were expressed in most or all tissues (Fig. 2C). These four peptides were therefore excluded from the TSA list. In contrast, 11 peptides were considered genuine aeTSAs because their source transcripts were either totally absent or present at trace amounts in a few tissues (Fig. 2C). We note that detection of low transcript amounts is less relevant because MHC peptides preferentially derive from highly abundant transcripts (43, 44). This concept is illustrated by the AH1 TSA, which elicits strong anti-tumor responses devoid of adverse effects (10, 38), despite the weak expression of its peptide-coding sequence in the liver, thymus, and urinary bladder (Fig. 2C). These results demonstrate that subtracting mRNA sequences found in mTEC<sup>hi</sup> strongly enriches for cancer-restricted peptide-coding sequences. When we consider our entire murine TSA dataset (6 mTSAs and 11 aeTSAs), we find that most of them derive from atypical translation events: the out-of-frame translation of a coding exon or the translation of noncoding regions (Fig. 2D). We also noticed that any type of noncoding region can generate TSAs (table S6): intergenic and intronic sequences, noncoding exons, untranslated region (UTR)/exon junctions, and EREs, which appear to be a particularly rich source of TSAs (eight aeTSAs and one mTSA). Last, our approach efficiently captured at least one structural variant as we identified an antigen, VTPVYQHL, derived from a very large intergenic deletion (~7500 bp) in EL4 cells (table S6B). Together, these observations confirm that noncoding regions are the main source of TSAs and that they have the potential to considerably expand the TSA landscape of tumors.



**Fig. 2. Most TSAs derive from the translation of noncoding regions.** (A) Flowcharts indicating key steps involved in TSA discovery [see fig. S3 (A to C) for details]. I/L, isoleucine/leucine. (B) Barplot showing the number of mTSAs (m) and aeTSA candidates (ae) in CT26 and EL4 cells. (C) Heatmap showing the average expression of peptide-coding sequences, in reads per hundred million reads sequenced (rphm), for aeTSA candidates and EL4 tumor-associated antigens (41, 42) in 22 tissues/organs (see table S5). For each peptide-coding sequence, the expression fold change and the number of positive tissues (rphm > 0, bold squares) are presented on the left-hand side of the heatmap. For fold changes, N/A indicates that the corresponding peptide-coding sequence was not expressed in syngeneic mTEC<sup>hi</sup>. Adip. tissue, adipose tissue; mam. gland, mammary gland; s.c. adip. tissue, subcutaneous adipose tissue. (D) Barplots depicting the number of TSAs derived from the translation of noncoding regions (noncoding) and of coding exons in-frame (coding-in) or out-of-frame (coding-out). The number of aeTSAs/mTSAs is reported within bars. The proportion of TSAs derived from atypical translation events is shown above bars. Features of CT26 and EL4 TSAs can be found in table S6 (A and B, respectively).

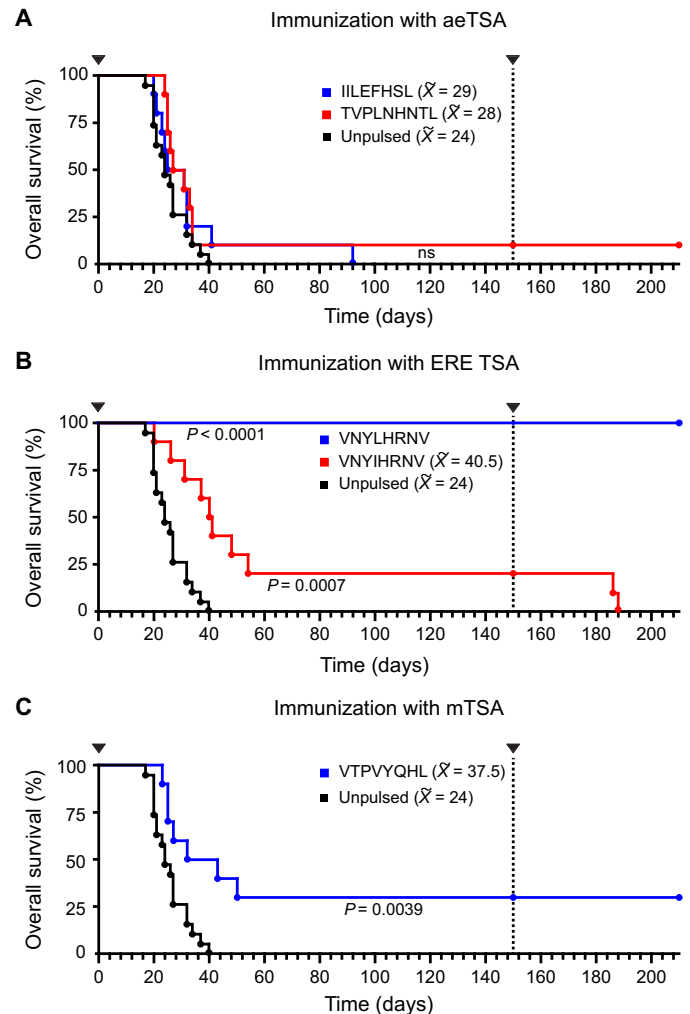
### Differential protection against EL4 cells after immunization against individual TSAs

We then performed detailed studies on some of the TSAs that seemed most therapeutically promising: those presented by EL4 cells and whose peptide-coding sequence was not expressed by any normal tissue (Fig. 2C and tables S6B and S7). To assess immunogenicity, C57BL/6 mice were immunized twice with either unpulsed (control group) or TSA-pulsed dendritic cells (DCs) before being challenged with live EL4 cells. Priming against IILEFHSL or TVPLNHNTL prolonged survival for 10% of mice, with only one TVPLNHNTL-immunized mouse surviving up to day 150 (Fig. 3A). In contrast, the other three TSAs showed superior efficacy, with day 150 survival rates of 20% (VNYIHRNV), 30% (VTPVYQHL), and 100% (VNYLHRNV) (Fig. 3, B and C). To evaluate the long-term efficacy of TSA vaccination, surviving mice were rechallenged with live EL4 cells at day 150 and monitored for signs of disease. The two VNYIHRNV-immunized survivors died of leukemia within 50 days, whereas all others (immunized against TVPLNHNTL, VTPVYQHL, or VNYLHRNV) survived the rechallenge (Fig. 3). We conclude that immunization against individual TSAs confers different degrees of protection against EL4 cells (0 to 100%) and that, in most cases, this protection is long-lasting.

### Frequency of TSA-responsive T cells in naïve and immunized mice

In various models, the strength of *in vivo* immune responses is regulated by the number of antigen-reactive T cells (45, 46). We therefore assessed the frequency of TSA-responsive T cells in naïve and immunized mice using a tetramer-based enrichment protocol (47, 48), for which the gating strategy and one representative experiment can be found in fig. S5 (A to C). As positive controls, we used tetramers to detect CD8<sup>+</sup> T cells specific for three immunodominant viral epitopes (gp-33, M45, and B8R). We confirmed that these T cells had a high abundance and that their frequency was similar to that observed in previous studies (Fig. 4A) (46). In naïve mice, CD8<sup>+</sup> T cells specific for TVPLNHNTL, VTPVYQHL, and IILEFHSL were rare (less than one tetramer<sup>+</sup> cell per 10<sup>6</sup> CD8<sup>+</sup> T cells), whereas CD8<sup>+</sup> T cells specific for the ERE TSAs (VNYIHRNV and VNYLHRNV) displayed frequencies similar to those of our viral controls (Fig. 4A and fig. S6A). Accordingly, in mice immunized with TSA-pulsed DCs, we found that the T cell frequencies against the two ERE TSAs, as assessed by tetramer staining or interferon- $\gamma$  (IFN- $\gamma$ ) enzyme-linked immunospot (ELISpot) assays (figs. S1B, S5, C and D, and S6A), were higher than that of TVPLNHNTL, VTPVYQHL, and IILEFHSL (Fig. 4, B and C). Moreover, in both naïve and immunized mice, results obtained with tetramer staining and IFN- $\gamma$  ELISpot correlated with each other (fig. S7). Last, we estimated that the functional avidity of T cells specific for VNYIHRNV and VNYLHRNV was similar to that of T cells specific for two highly immunogenic nonself antigens: the minor histocompatibility antigens H7<sup>a</sup> and H13<sup>a</sup> (Fig. 4D). Hence, these TSAs, derived from allegedly noncoding regions, were recognized by highly abundant T cells with a high functional avidity. This is particularly noteworthy for the VNYLHRNV aeTSA because it has an unmutated germline sequence.

Together, our results show that the frequency of TSA-responsive T cells was a crucial parameter for TSA immunogenicity. However, VTPVYQHL was an outsider: It afforded the second-best protection against EL4 challenge although its cognate T cells were present at a very low frequency (Figs. 3 and 4, A to C). To better evaluate the importance of T cell expansion in leukemia protection, we estimated the frequency of tetramer<sup>+</sup> CD8<sup>+</sup> T cells in long-term survivors after



**Fig. 3. Immunization against individual TSAs confers different degrees of protection against EL4 cells.** C57BL/6 mice were immunized twice with DCs pulsed with individual TSAs: (A) two aeTSAs, (B) two ERE TSAs (one aeTSA or one mTSA), or (C) one mTSA. Mice were injected intravenously with  $5 \times 10^5$  live EL4 cells (arrowheads) on day 0, and all surviving mice were rechallenged on day 150. Control groups were immunized with unpulsed DCs (solid black line).  $\bar{X}$  represents the median survival. Statistical significance of immunized group versus control group was calculated using a log-rank test, where ns stands for not significant ( $P > 0.05$ ).  $n = 10$  mice per group for peptide-specific immunization,  $n = 19$  mice for control group.

rechallenge with EL4 cells on day 150 (Fig. 3). These analyses were performed on day 210 or at the time of sacrifice (in the case of VNYIHRNV-primed mice). Two points can be made from these analyses (fig. S6, B and C). First, all long-term survivors, including VTPVYQHL-immunized mice, showed a discernable population of TSA-responsive (tetramer<sup>+</sup>) CD8<sup>+</sup> T cells. Second, although VNYIHRNV was recognized by a particularly large population of tetramer<sup>+</sup> cells, it was the only TSA that did not protect mice upon rechallenge. Together, our results suggest that expansion of TSA-responsive T cells was necessary for protection against EL4 cells but was insufficient in the case of VNYIHRNV.

### The importance of antigen expression for protection against EL4 cells

Next, we evaluated the impact of antigen expression on immunogenicity by assessing the abundance of TSAs at the RNA level in the EL4 cell

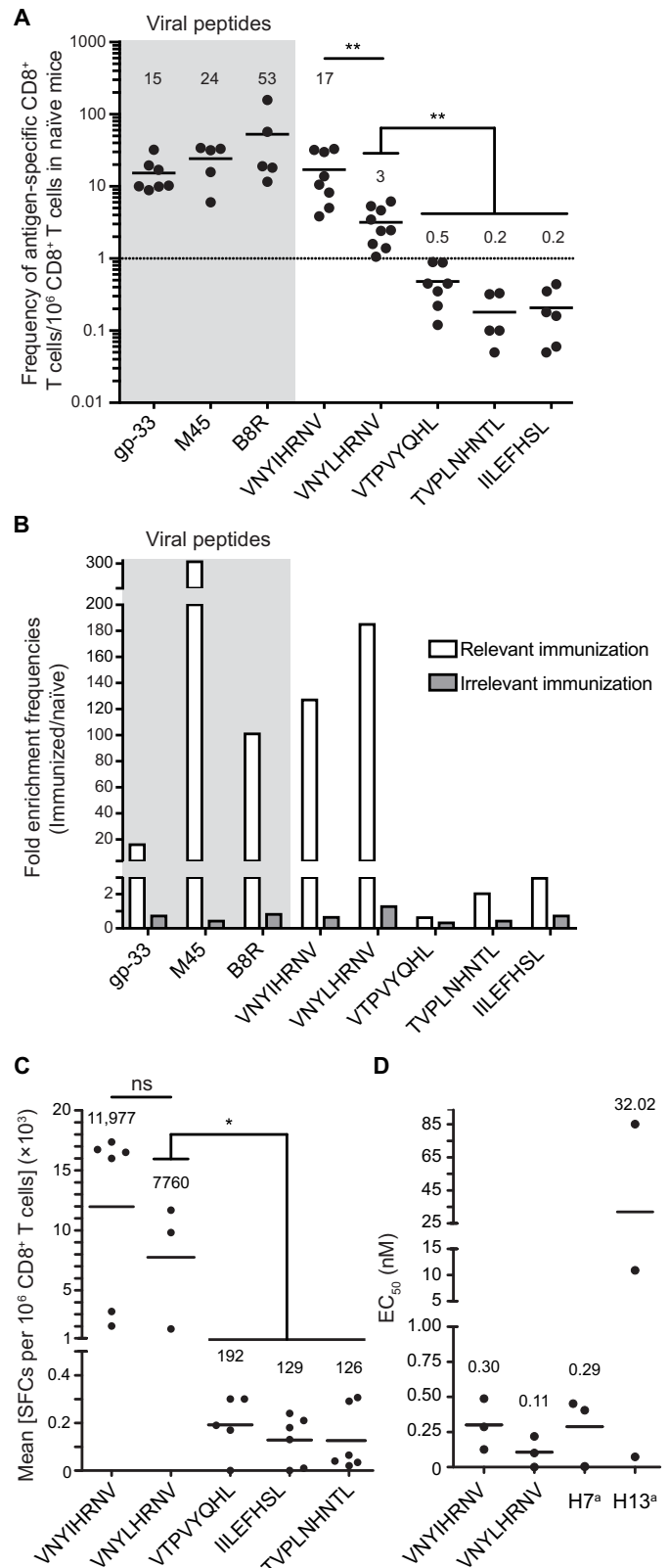


**Fig. 4. Frequency of and IFN- $\gamma$  secretion by TSA-responsive T cells in naïve and immunized mice.** (A) Number of tetramer<sup>+</sup> CD8<sup>+</sup> T cells per 10<sup>6</sup> CD8<sup>+</sup> T cells in naïve mice. Circles, one mouse ( $n = 5$  to 9 mice per group); dotted line, frequency of 1 tetramer<sup>+</sup> T cell per 10<sup>6</sup> CD8<sup>+</sup> T cells. (B) Fold enrichment of tetramer<sup>+</sup> CD8<sup>+</sup> T cells after immunization with relevant (white bars) or irrelevant (gray bars) peptides ( $\frac{\text{Mean frequency}_{\text{immunized}}}{\text{Mean frequency}_{\text{naive}}}$ ). (C) The number of spot-forming cells (SFCs), measured by an IFN- $\gamma$  ELISpot assay, averaged across technical replicates (circles) after being converted to SFCs per 10<sup>6</sup> CD8<sup>+</sup> T cells:  $[\frac{\text{SFCs}_{\text{immunized}} - \text{SFCs}_{\text{naive}}}{\text{Number of T cells plated}}] \times 10^6$ . (D) The functional avidity of T cells recognizing specific TSAs and two previously reported positive controls [H7<sup>a</sup> and H13<sup>a</sup> (42)] was estimated by calculating a half maximal effective concentration (EC<sub>50</sub>), corresponding to the peptide concentration where half of plated antigen-specific T cells secreted IFN- $\gamma$ . (B to D) Three independent experiments. On relevant panels, full horizontal lines and numbers above each condition represent mean values. Viral peptides used as control are highlighted in gray. \* $P \leq 0.05$  and \*\* $P \leq 0.01$  (two-sided Wilcoxon rank sum test with the Benjamini-Hochberg correction).

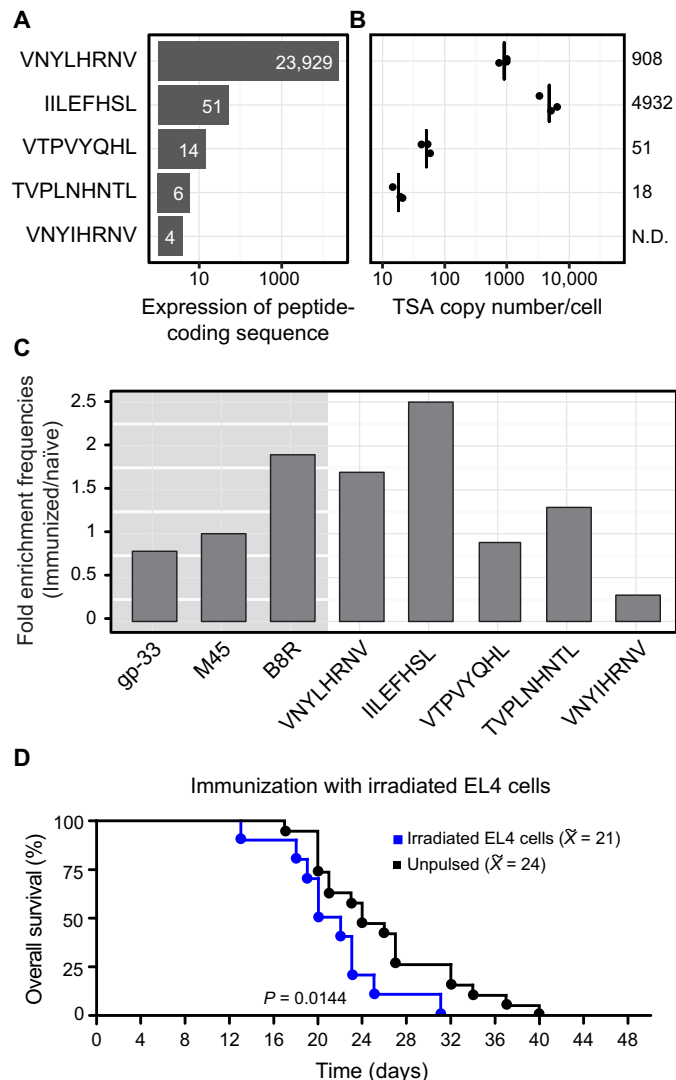
population that was injected on day 0 (Fig. 3). The sequence encoding the TSA conferring the best protection against EL4 cells (VNYLHRNV) was expressed more than the other TSA-coding sequences (Fig. 5A). This suggests that VNYLHRNV is likely “clonal” (expressed by all EL4 cells) and highly expressed, whereas the other TSAs are subclonal and/or expressed at low amounts. Next, using parallel reaction monitoring (PRM) MS, we analyzed the TSA copy number per cell in the EL4 cell population used for rechallenge (day 150; Fig. 5B). As expected (41), there was no linear relationship between TSA abundance at the RNA and peptide levels (Fig. 5, A and B). The most protective TSA, VNYLHRNV, was one of the two most abundant TSAs (>500 copies per cell), whereas VNYIHRNV, which offered no protection upon rechallenge (Fig. 3B), was no longer detected on EL4 cells. This observation suggests that VNYIHRNV was a subclonal TSA and that antigen loss most likely explained the lack of protection upon rechallenge. Last, we noted that TSAs were immunogenic when presented by DCs but not when presented by EL4 cells: Injection of live EL4 cells without prior immunization did not induce an expansion of TSA-responsive T cells (Fig. 5C and fig. S6D), and immunization with irradiated EL4 cells did not confer any protection against live EL4 cells (Fig. 5D). This suggests that, in the absence of immunization, highly immunogenic TSAs (such as VNYLHRNV) were ignored likely because they were not efficiently cross-presented by DCs, highlighting the importance of efficient T cell priming in cancer immunotherapy.

**Impact of noncoding regions on the TSA landscape of human primary tumors**

Having established that noncoding regions are a major source of TSAs in two murine cell lines, we applied our proteogenomic approach to seven human primary tumor samples: four B-ALLs and three lung cancers (fig. S8 and tables S8 to S15). Rather than using RNA-seq data from murine syngeneic mTEC<sup>hi</sup>, we sequenced the transcriptome of total TECs ( $n = 2$ ) and purified mTECs ( $n = 4$ ) from six unrelated donors undergoing corrective cardiovascular surgery (table S2B). We found minimal interindividual differences and demonstrated that this cohort size was sufficient to cover almost the full breadth of the mTEC transcriptomic landscape, as computing the cumulative number of detected transcripts showed that minimal gains would be achieved by adding more samples (fig. S9). Using these RNA-seq data as the repertoire of normal k-mers to generate global cancer databases (table S1B), as described in Fig. 1, we identified 2 mTSAs and 27 aeTSA candidates (Fig. 6A). After validating their assignment to a single genomic location and the quality of their MS spectra, we also



ensured that mTSAs did not intersect with known germline polymorphisms (figs. S3, 10, and 11). To further validate the status of aeTSA candidates, as we did for murine aeTSAs (Fig. 2C), we analyzed the



**Fig. 5. High expression of EL4 TSAs is necessary but not sufficient to induce antileukemic responses.** (A and B) Analysis of TSA expression at the RNA and peptide levels was performed on EL4 cells injected into mice at day 0 or day 150, respectively. (A) The number of RNA-seq reads fully overlapping the RNA sequences encoding each TSA. (B) TSA copy number per cell was estimated by PRM MS using  $^{13}\text{C}$ -synthetic peptide analogs of the TSAs (three replicates). Black lines represent the mean TSA copy number per cell (also indicated on the left-hand side of the graph). N.D., not detected. (C) Fold enrichment for tetramer $^{+}$  CD8 $^{+}$  T cells after injection with live EL4 cells without prior immunization ( $\frac{\text{Mean frequency}_{\text{EL4-injected}}}{\text{Mean frequency}_{\text{naive}}}$ ). Fold enrichment for T cells recognizing viral peptides is shown as negative controls and is highlighted in gray. Three independent experiments were performed. (D) Overall survival of C57BL/6 female mice immunized twice with irradiated (10,000 cGy) EL4 cells (blue line,  $n = 10$  mice) or unpulsed DCs (black line,  $n = 19$  mice) and then injected intravenously with  $5 \times 10^5$  live EL4 cells.  $\bar{X}$  represents the median survival. Statistical significance of immunized group versus control group was calculated using a log-rank test.

expression of aeTSA-coding sequences in RNA-seq data from 28 tissues (6 to 50 individuals per tissue; Fig. 6B and table S16). On the basis of these data, we excluded six aeTSA candidates: three that were widely expressed, like most previously reported overexpressed tumor-associated antigens (49), and three that were expressed at substantial amounts in a

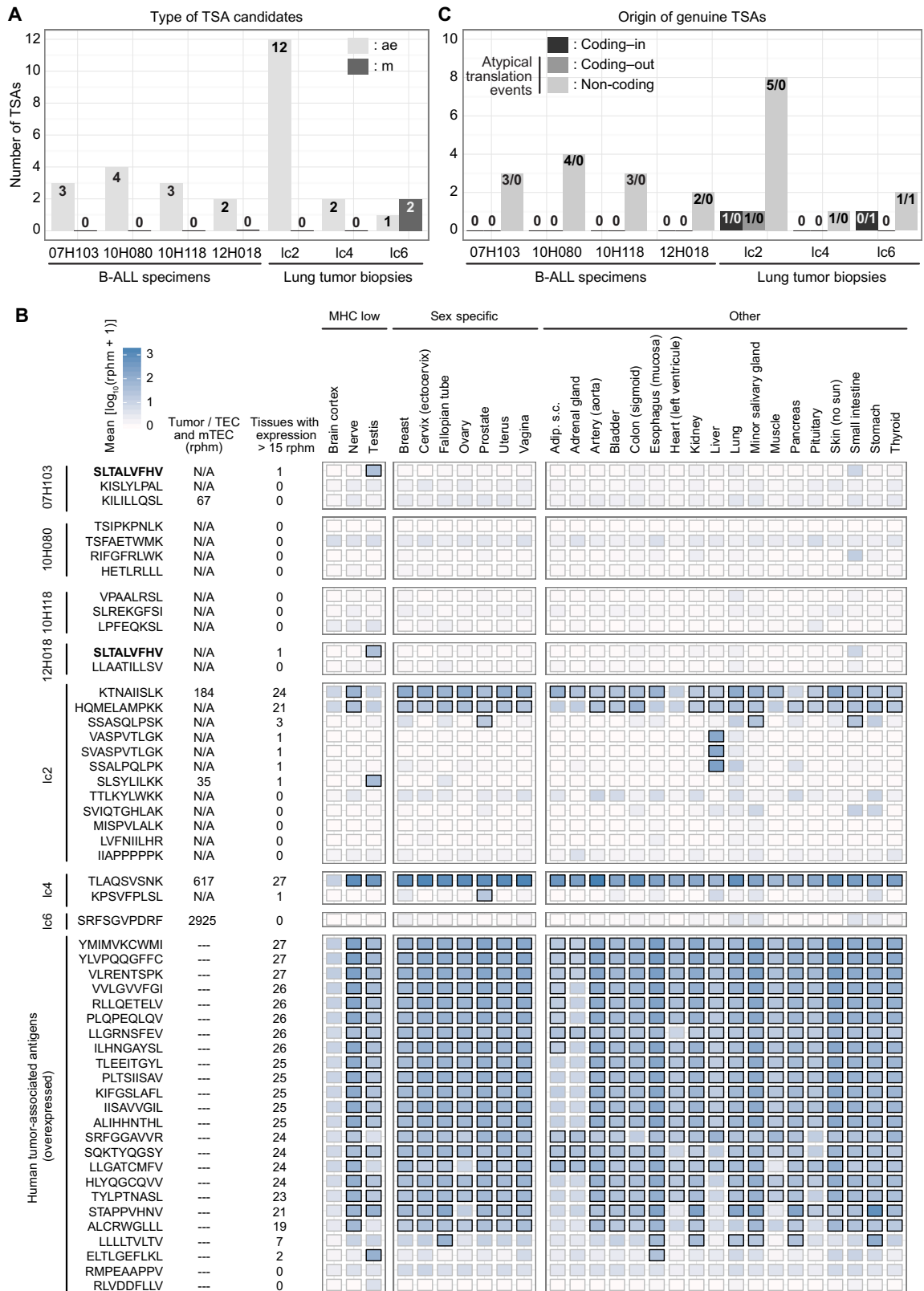
single organ, the liver (Fig. 6B). We therefore ended up with a total of two mTSAs and 20 nonredundant aeTSAs (Fig. 6C and tables S17 and S18). Of note, the SLTALVFHV aeTSA was shared by our two HLA-A\*02:01-positive B-ALLs (table S17). This aeTSA derives from the 3' UTR of *TCL1A*, a gene implicated in lymphoid malignancies. Together, our results show that our proteogenomic approach can characterize the repertoire of mTSAs and aeTSAs on individual tumors in about 2 weeks.

## DISCUSSION

To explore the global landscape of TSAs, we developed a proteogenomic approach that incorporates two features in the construction of databases for MS analyses: alignment-free k-mer profiling of RNA-seq data and subtraction of mTEC k-mers. In a context where TSA discovery is a critical unmet medical need, our approach led us to discover that the TSA landscape is much larger than previously anticipated. Thirty-five of the 39 nonredundant TSAs reported here derived from atypical translation events: 2 from the out-of-frame translation of coding exons and 33 from allegedly noncoding regions. Hence, ~90% of our TSAs would have been missed by standard approaches focusing on exonic mutations. In addition to MHC peptides derived from RNAs containing single-base mutations, our approach efficiently captured peptides generated by complex structural variants, as exemplified by VTPVYQHL, which derived from a large intergenic deletion (~7500 bp) in EL4 cells. Subtraction of mTEC k-mers was critical for the high-throughput identification of immunogenic aeTSAs including unmutated peptides that are not constitutively presented by mTEC $^{\text{hi}}$  to thymocytes during the establishment of central tolerance. This is well-illustrated by VNYLHRNV, an unmutated TSA absent from mTEC $^{\text{hi}}$  and other peripheral tissues, although strongly expressed in EL4 cells, that is, recognized by highly abundant CD8 $^{+}$  T cells with a high functional avidity. Nonetheless, a few peptide-coding transcripts undetectable in mTEC $^{\text{hi}}$  were detected in peripheral tissues, suggesting that k-mers from both mTECs and peripheral tissues should be used to identify genuine TSAs. In the present study, we chose to use peripheral expression as an a posteriori validation step. An alternative approach would be to remove all k-mers expressed in peripheral tissues when building the database for MS.

TSAs derived from noncoding regions present a number of peculiar and highly relevant features. First, it is evident that EREs are a rich source of TSAs; they generated 9 of the 17 TSAs found in murine cell lines and 4 of the 23 TSAs in human tumors. The difference in the proportion of ERE TSAs that we identified in mouse cell lines versus primary human tumors might be ascribed to the fact that in vitro culture conditions do not recapitulate the immune pressure exerted on developing tumors, that ERE expression greatly varies across tumor types (50), or that human EREs are more degenerated and therefore less likely to be translated than murine EREs (9). Nonetheless, ERE TSAs remain particularly relevant to the development of cancer vaccines because both oncogenic viruses and viral-like sequences in the human genome appear to be particularly immunogenic (51, 52). Second, most TSAs derived from noncoding regions do not overlap mutations and are therefore, by definition, aeTSAs. Such aeTSAs, which include EREs, present a major advantage over mTSAs (of both coding or noncoding origin): Whereas mTSAs are private antigens, aeTSAs can be shared by multiple tumors (8). We were able to identify such shared aeTSA (SLTALVFHV) in humans, whereas Probst *et al.* (10) showed that mice immunized against the AH1 aeTSA (SPSYVYHQF), which we identified

**Fig. 6. Most TSAs detected in human primary tumors derive from the translation of noncoding regions.** (A) Barplot showing the number of aeTSAs candidates (ae) and mTSAs (m) in each primary sample analyzed. (B) Heatmap showing the average expression of peptide-coding sequences, in rphm, for aeTSAs and overexpressed tumor-associated antigens obtained from the Cancer Immunity Peptide database (49) across a panel of 28 tissues (see table S16). For each peptide-coding sequence, the expression fold change (tumor/TEC and mTEC) and the number of positive tissues (rphm > 15, bold squares) are shown on the left-hand side of the heatmap. For fold changes, N/A, and—indicate that the corresponding peptide-coding sequence was not detected in TEC/mTEC samples or not computed, respectively. Adip. s.c., adipose subcutaneous. (C) Barplots depicting the number of TSAs derived from the translation of noncoding regions (non-coding) or from coding exons translated in-frame (coding-in) or out-of-frame (coding-out). The number of aeTSAs/mTSAs is shown within bars. Features of human TSAs identified in each sample can be found in tables S17 and S18.



by MS on CT26 cells, survived the challenge with three different cancer cell lines: the WEHI-164 fibrosarcoma and the C51 and CT26 colorectal cancers.

Together, our MS-based discovery of TSAs in primary human B-ALL and lung cancer demonstrates that the impetus to develop TSA vaccines should not be limited to cancers with a high mutational

load. Because B-ALLs harbor very few exonic mutations, it was presumed that they might not express any TSA (5). Our data show that TSAs can be found in B-ALL provided that the search strategy encompasses aeTSAs. Moreover, two elements argue that TSAs derived from atypical translation events represent promising targets for T cell-based cancer immunotherapy: (i) They outnumber TSAs derived from coding regions and (ii) they are mostly unmutated, which increases their potential to be shared between patients.

We acknowledge that our study presents several limitations for which solutions can be envisioned. First, because our approach is not compatible with the computation of classical false discovery rates, TSAs must undergo meticulous validation by manual inspection of MS spectra and, ideally, confirmation using synthetic peptide analogs. Second, our approach relies on shotgun MS, which suffers from a limited dynamic range. Consequently, we only detected the most abundant TSAs and are likely underestimating the extent of aeTSA sharing between patients. Because shared aeTSAs may represent promising actionable targets for cancer immunotherapy (8), aeTSA frequency across patients and/or tumor types could be further evaluated by targeted MS analyses that have a more limited scope but are 10 times more sensitive than shotgun MS and can provide quantitative data such as the number of TSA copies per cell (53). Third, because TSA immunogenicity cannot be predicted (54), it has to be tested experimentally for each TSA. This issue is being addressed by several research groups that are currently developing artificial platforms requiring less material than the conventional IFN- $\gamma$  ELISpot assays used for such testing (8, 55, 56).

In practice, how could we prioritize TSAs for clinical trials? In our EL4 tumor model, the efficacy of TSA immunization was largely determined by two criteria: TSA abundance and the frequency of TSA-responsive T cells. TSA abundance can be assessed by targeted MS analyses, and the frequency of TSA-responsive T cells in peripheral blood mononuclear cells of a cohort of subjects could be estimated using MHC peptide multimers or functional assays. Widely shared, highly abundant TSAs recognized by high-frequency T cells could then be selected for clinical trials. These optimal aeTSAs could then potentially be combined in a single vaccine using already available synthesis and delivery platforms (13, 57).

## MATERIALS AND METHODS

### Study design

The purpose of this study was to develop a proteogenomic approach that would enable the identification of TSAs derived from any region of the genome and to identify features that influenced TSA immunogenicity. To do so, we first characterized the TSA landscape of two murine cell lines, the EL4 T-lymphoblastic lymphoma cell line and the CT26 colorectal cancer cell line that were both obtained from the American Type Culture Collection. As a normal control, we used thymi isolated from 5- to 8-week-old C57BL/6 or Balb/c mice obtained from the Jackson Laboratory. Mice were housed under specific pathogen-free conditions, and all experimental protocols were approved by the Comité de Déontologie de l'Expérimentation sur des Animaux of Université de Montréal. We also applied our approach to seven human primary cancer samples from treatment-naïve patients. These included three lung tumor biopsies (lc2, lc4, and lc6) purchased from Tissue Solutions and four primary leukemic samples (B-ALL specimens 07H103, 10H080, 10H118, and 12H018) that were collected and cryopreserved at the Banque de Cellules Leucémiques du Québec at Hôpital Maisonneuve-Rosemont. The project was approved by the

Comité d'éthique de la recherche de l'Hôpital Maisonneuve-Rosemont (CÉR 12100). NOD Cg-Prkd<sup>scid</sup>Il2rg<sup>tm1Wjl</sup>/SzJ (NSG) mice were used to expand our B-ALL specimens. These mice were purchased and housed as described for C57BL/6 and Balb/c mice. As a normal control, we used thymi obtained from 3-month-old to 7-year-old patients undergoing corrective cardiovascular surgery (CHU Sainte-Justine Research Ethic Board, protocol and biobank no. 2126). No statistical method was used to predetermine sample size. One replicate was sequenced for all RNA-seq experiments. For MS, at least two replicates were analyzed. To assess the immunogenicity of EL4-derived TSAs, we measured the frequency and antigen avidity of T cells recognizing TSAs. In addition, we estimated the survival of 8- to 12-week-old C57BL/6 female mice that were immunized or not with individual TSAs. Investigators were not blinded during sample preparation or during data collection and analysis. For all in vitro and in vivo experiments described in this manuscript, at least three replicates were analyzed and found to be concordant with each other. No data were excluded from the analyses, and values are reported in table S7. The number of mice used, numbers of replicates, and statistical values (where applicable) are provided in the figure legends. For information regarding original RNA-seq and MS data, see the Data and materials availability section in Acknowledgements and table S2.

### Statistical analysis

Procedures to evaluate statistical significance are described in the relevant figure legends. Overall, a log-rank test was used for survival curves, a Wilcoxon rank sum test with the Benjamini-Hochberg correction for multiple testing was used to compare T cell frequencies as estimated by tetramer and ELISpot, and a one-sided Wilcoxon rank sum test was used to compare T cell frequencies between immunized and rechallenge mice.  $P \leq 0.05$  was considered significant.

## SUPPLEMENTARY MATERIALS

[www.sciencetranslationalmedicine.org/cgi/content/full/10/470/eaa05516/DC1](http://www.sciencetranslationalmedicine.org/cgi/content/full/10/470/eaa05516/DC1)  
Materials and Methods

Fig. S1. Gating strategies for cells isolated by fluorescence-activated cell sorting.

Fig. S2. Architecture of the codes used for our k-mer profiling workflow.

Fig. S3. TSA validation process.

Fig. S4. MS validation of CT26 and EL4 TSA candidates using synthetic analogs.

Fig. S5. Detection of antigen-specific CD8<sup>+</sup> T cells in naïve and immunized mice.

Fig. S6. Frequencies of antigen-specific T cells.

Fig. S7. Correlation between antigen-specific T cell frequencies in naïve and immunized mice.

Fig. S8. Purity of the 10H080 B-ALL sample after expansion in NSG mice.

Fig. S9. Overview of the human TEC and mTEC transcriptomic landscapes.

Fig. S10. MS validation of B-ALL TSA candidates using synthetic analogs.

Fig. S11. MS validation of lung cancer TSA candidates using synthetic analogs.

Table S1. Statistics related to the generation of the global cancer databases.

Table S2. Information about samples used in this study.

Table S3. List of CT26 MHC class I-associated peptides.

Table S4. List of EL4 MHC class I-associated peptides.

Table S5. Accession numbers of the ENCODE datasets used in this study.

Table S6. Features of murine TSAs.

Table S7. Experimental values obtained in analyses of mouse TSA immunogenicity.

Table S8. List of 07H103 MHC class I-associated peptides.

Table S9. List of 10H080 MHC class I-associated peptides obtained by mild acid elution.

Table S10. List of 10H080 MHC class I-associated peptides obtained by immunoprecipitation.

Table S11. List of 10H118 MHC class I-associated peptides.

Table S12. List of 12H018 MHC class I-associated peptides.

Table S13. List of lc2 MHC class I-associated peptides.

Table S14. List of lc4 MHC class I-associated peptides.

Table S15. List of lc6 MHC class I-associated peptides.

Table S16. Accession numbers of the Genotype-Tissue Expression (GTEx) datasets used in this study.



Table S17. Features of human TSAs detected in B-ALL specimens.

Table S18. Features of human TSAs detected in lung tumor biopsies.

References (60–68)

## REFERENCES AND NOTES

- B. Mlecnik, G. Bindea, A. Kirilovsky, H. K. Angell, A. C. Obenauf, M. Tosolini, S. E. Church, P. Maby, A. Vasaturo, M. Angelova, T. Fredriksen, S. Mauger, M. Waldner, A. Berger, M. R. Speicher, F. Pagès, V. Valge-Archer, J. Galon, The tumor microenvironment and Immunocore are critical determinants of dissemination to distant metastasis. *Sci. Transl. Med.* **8**, 327ra326 (2016).
- P. Charoentong, F. Finotello, M. Angelova, C. Mayer, M. Efremova, D. Rieder, H. Hackl, Z. Trajanoski, Pan-cancer immunogenomic analyses reveal genotype-immunophenotype relationships and predictors of response to checkpoint blockade. *Cell Rep.* **18**, 248–262 (2017).
- S. C. Wei, J. H. Levine, A. P. Cogdill, Y. Zhao, N.-A. S. Anang, M. C. Andrews, P. Sharma, J. Wang, J. A. Wargo, D. Pe'er, J. P. Allison, Distinct cellular mechanisms underlie anti-CTLA-4 and anti-PD-1 checkpoint blockade. *Cell* **170**, 1120–1133.e17 (2017).
- W. Shao, P. G. A. Pedrioli, W. Wolski, C. Scurtescu, E. Schmid, J. A. Vizcaino, M. Courcelles, H. Schuster, D. Kowalewski, F. Marino, C. S. L. Arlehamn, K. Vaughan, B. Peters, A. Sette, T. H. M. Ottenhoff, K. E. Meijgaard, N. Nieuwenhuizen, S. H. E. Kaufmann, R. Schlapbach, J. C. Castle, A. I. Nesvizhskii, M. Nielsen, E. W. Deutsch, D. S. Campbell, R. L. Moritz, R. A. Zubarev, A. J. Ytterberg, A. W. Purcell, M. Marcilla, A. Paradelo, Q. Wang, C. E. Costello, N. Ternet, P. A. van Veelen, C. A. C. M. van Els, A. J. R. Heck, G. A. de Souza, L. M. Sollid, A. Admon, S. Stevanovic, H.-G. Rammensee, P. Thibault, C. Perreault, M. Bassani-Sternberg, R. Aebersold, E. Caron, The Systemic MHC Atlas project. *Nucleic Acids Res.* **46**, D1237–D1247 (2018).
- S. D. Martin, G. Coukos, R. A. Holt, B. H. Nelson, Targeting the undruggable: Immunotherapy meets personalized oncology in the genomic era. *Ann. Oncol.* **26**, 2367–2374 (2015).
- R. Marty, S. Kaabinejad, D. Rossell, M. J. Sliker, J. van de Haar, H. B. Engin, N. de Prisco, T. Ideker, W. H. Hildebrand, J. Font-Burgada, H. Carter, MHC-I genotype restricts the oncogenic mutational landscape. *Cell* **171**, 1272–1283.e15 (2017).
- M. Yarchoan, B. A. Johnson III, E. R. Lutz, D. A. Laheru, E. M. Jaffe, Targeting neoantigens to augment antitumor immunity. *Nat. Rev. Cancer* **17**, 209–222 (2017).
- M. H. Gee, A. Han, S. M. Lofgren, J. F. Beausang, J. L. Mendoza, M. E. Birnbaum, M. T. Bethune, S. Fischer, X. Yang, R. Gomez-Eerland, D. B.ingham, L. V. Sibener, R. A. Fernandes, A. Velasco, D. Baltimore, T. N. Schumacher, P. Khatri, S. R. Quake, M. M. Davis, K. C. Garcia, Antigen identification for orphan T cell receptors expressed on tumor-infiltrating lymphocytes. *Cell* **172**, 549–563.e16 (2018).
- G. Kassiotis, J. P. Stoye, Immune responses to endogenous retroelements: Taking the bad with the good. *Nat. Rev. Immunol.* **16**, 207–219 (2016).
- P. Probst, J. Kopp, A. Oxenius, M. P. Colombo, D. Ritz, T. Fugmann, D. Neri, Sarcoma eradication by doxorubicin and targeted TNF relies upon CD8<sup>+</sup> T-cell recognition of a retroviral antigen. *Cancer Res.* **77**, 3644–3654 (2017).
- S. Goel, M. J. DeCristo, A. C. Watt, H. BrinJones, J. Sceneay, B. B. Li, N. Khan, J. M. Ubellacker, S. Xie, O. Metzger-Filho, J. Hoog, M. J. Ellis, C. X. Ma, S. Ramm, I. E. Krop, E. P. Winer, T. M. Roberts, H.-J. Kim, S. S. McAllister, J. J. Zhao, CDK4/6 inhibition triggers anti-tumor immunity. *Nature* **548**, 471–475 (2017).
- P. A. Ott, Z. Hu, D. B. Keskin, S. A. Shukla, J. Sun, D. J. Bozym, W. Zhang, A. Luoma, A. Giobbie-Hurder, L. Peter, C. Chen, O. Olive, T. A. Carter, S. Li, D. J. Lieb, T. Eisenhaure, E. Gjini, J. Stevens, W. J. Lane, I. Javeri, K. Nellaippan, A. M. Salazar, H. Daley, M. Seaman, E. I. Buchbinder, C. H. Yoon, M. Harden, N. Lennon, S. Gabriel, S. J. Rodig, D. H. Barouch, J. C. Aster, G. Getz, K. Wucherpfennig, D. Neuberg, J. Ritz, E. S. Lander, E. F. Fritsch, N. Hacohen, C. J. Wu, An immunogenic personal neoantigen vaccine for patients with melanoma. *Nature* **547**, 217–221 (2017).
- U. Sahin, E. Derhovanessian, M. Miller, B.-P. Kloke, P. Simon, M. Löwer, V. Bukur, A. D. Tadmor, U. Luxemburger, B. Schrörs, T. Omokoko, M. Vormehr, C. Albrecht, A. Paruzynski, A. N. Kuhn, J. Buck, S. Heesch, K. H. Schreeb, F. Müller, I. Ortseifer, I. Vogler, E. Godehardt, S. Attig, R. Rae, A. Breitkreuz, C. Tolliver, M. Suchan, G. Martic, A. Hohberger, P. Sorn, J. Diekmann, J. Ciesla, O. Waksman, A.-K. Brück, M. Witt, M. Zillgen, A. Rothermel, B. Kasemann, D. Langer, S. Bolte, M. Diken, S. Kreiter, R. Nemecek, C. Gebhardt, S. Grabbe, C. Höller, J. Utikal, C. Huber, C. Loqui, Ö. Türeci, Personalized RNA mutanome vaccines mobilize poly-specific therapeutic immunity against cancer. *Nature* **547**, 222–226 (2017).
- A.-H. Capietto, S. Jhunjhunwala, L. Delamarre, Characterizing neoantigens for personalized cancer immunotherapy. *Curr. Opin. Immunol.* **46**, 58–65 (2017).
- The problem with neoantigen prediction. *Nat. Biotechnol.* **35**, 97 (2017).
- H. Pearson, T. Daouda, D. P. Granados, C. Durette, E. Bonneil, M. Courcelles, A. Rodenbrock, J.-P. Laverdure, C. Côté, S. Mader, S. Lemieux, P. Thibault, C. Perreault, MHC class I-associated peptides derive from selective regions of the human genome. *J. Clin. Invest.* **126**, 4690–4701 (2016).
- M. Di Marco, J. K. Peper, H.-G. Rammensee, Identification of immunogenic epitopes by MS/MS. *Cancer J.* **23**, 102–107 (2017).
- M. Yadav, S. Jhunjhunwala, Q. T. Phung, P. Lupardus, J. Tanguay, S. Bumbaca, C. Franci, T. K. Cheung, J. Fritsche, T. Weinschenk, Z. Modrusan, I. Mellman, J. R. Lill, L. Delamarre, Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).
- M. Bassani-Sternberg, G. Coukos, Mass spectrometry-based antigen discovery for cancer immunotherapy. *Curr. Opin. Immunol.* **41**, 9–17 (2016).
- M. Bassani-Sternberg, E. Bräunlein, R. Klar, T. Engleitner, P. Sinitcyn, S. Audehm, M. Straub, J. Weber, J. Slotta-Huspenina, K. Specht, M. E. Martignoni, A. Werner, R. Hein, D. H. Busch, C. Peschel, R. Rad, J. Cox, M. Mann, A. M. Krackhardt, Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat. Commun.* **7**, 13404 (2016).
- M. M. Gubin, X. Zhang, H. Schuster, E. Caron, J. P. Ward, T. Noguchi, Y. Ivanova, J. Hundal, C. D. Arthur, W.-J. Krebber, G. E. Mulder, M. Toebes, M. D. Vesely, S. S. K. Lam, A. J. Korman, J. P. Allison, G. J. Freeman, A. H. Sharpe, E. L. Pearce, T. N. Schumacher, R. Aebersold, H.-G. Rammensee, C. J. Melief, E. R. Mardis, W. E. Gillanders, M. N. Artyomov, R. D. Schreiber, Checkpoint blockade cancer immunotherapy targets tumour-specific mutant antigens. *Nature* **515**, 577–581 (2014).
- E. Khurana, Y. Fu, D. Chakravarty, F. Demicheli, M. A. Rubin, M. Gerstein, Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108 (2016).
- S. Djebali, C. A. Davis, A. Merkel, A. Dobin, T. Lassmann, A. Mortazavi, A. Tanzer, J. Lagarde, W. Lin, F. Schlesinger, C. Xue, G. K. Marinov, J. Khatun, B. A. Williams, C. Zaleski, J. Rozowsky, M. Röder, F. Kokocinski, R. F. Abdelhamid, T. Alioto, I. Antoshechkin, M. T. Baer, N. S. Bar, P. Batut, K. Bell, I. Bell, S. Chakraborty, X. Chen, J. Chrast, J. Curado, T. Derrien, J. Drenkow, E. Dumais, J. Dumais, R. Duttagupta, E. Falconnet, M. Fastuca, K. Fejes-Toth, P. Ferreira, S. Foissac, M. J. Fullwood, H. Gao, D. Gonzalez, A. Gordon, H. Gunawardena, C. Howald, S. Jha, R. Johnson, P. Kapranov, B. King, C. Kingswood, O. J. Luo, E. Park, K. Persaud, J. B. Preall, P. Ribeca, B. Risk, D. Robyr, M. Sammeth, L. Schaffer, L.-H. See, A. Shahab, J. Skancke, A. M. Suzuki, H. Takahashi, H. Tilgner, D. Trout, N. Walters, H. Wang, J. Wrobel, Y. Yu, X. Ruan, Y. Hayashizaki, J. Harrow, M. Gerstein, T. Hubbard, A. Reymond, S. E. Antonarakis, G. Hannon, M. C. Giddings, Y. Guan, B. Wold, P. Carninci, R. Guigó, T. R. Gingeras, Landscape of transcription in human cells. *Nature* **489**, 101–108 (2012).
- C. M. Laumont, T. Daouda, J.-P. Laverdure, É. Bonneil, O. Caron-Lizotte, M.-P. Hardy, D. P. Granados, C. Durette, S. Lemieux, P. Thibault, C. Perreault, Global proteogenomic analysis of human MHC class I-associated peptides derived from non-canonical reading frames. *Nat. Commun.* **7**, 10238 (2016).
- C. M. Laumont, C. Perreault, Exploiting non-canonical translation to identify new targets for T cell-based cancer immunotherapy. *Cell. Mol. Life Sci.* **75**, 607–621 (2018).
- S. A. Rosenberg, P. Tong-On, Y. Li, J. P. Riley, M. El-Gamil, M. R. Parkhurst, P. F. Robbins, Identification of BING-4 cancer antigen translated from an alternative open reading frame of a gene in the extended MHC class II region using lymphocytes from a patient with a durable complete regression following immunotherapy. *J. Immunol.* **168**, 2402–2407 (2002).
- M. J. L. Kracht, M. van Lummel, T. Nikolic, A. M. Joosten, S. Laban, A. R. van der Slik, P. A. van Veelen, F. Carlotti, E. J. P. de Koning, R. C. Hoeben, A. Zaldumbide, B. O. Roep, Autoimmunity against a defective ribosomal insulin gene product in type 1 diabetes. *Nat. Med.* **23**, 501–507 (2017).
- Y. Simoni, E. Becht, M. Fehlings, C. Y. Loh, S.-L. Koo, K. W. W. Teng, J. P. S. Yeong, R. Nahar, T. Zhang, H. Kared, K. Duan, N. Ang, M. Poidinger, Y. Y. Lee, A. Larbi, A. J. Khng, E. Tan, C. Fu, R. Mathew, M. Teo, W. T. Lim, C. K. Toh, B.-H. Ong, T. Koh, A. M. Hillmer, A. Takano, T. K. H. Lim, E. H. Tan, W. Zhai, D. S. W. Tan, I. B. Tan, E. W. Newell, Bystander CD8<sup>+</sup> T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* **557**, 575–579 (2018).
- J. Zhang, L. Xin, B. Shan, W. Chen, M. Xie, D. Yuen, W. Zhang, Z. Zhang, G. A. Lajoie, B. Ma, PEAKS DB: De novo sequencing assisted database search for sensitive and accurate peptide identification. *Mol. Cell. Proteomics* **11**, M111.010587 (2012).
- J. A. Alfaro, A. Sinha, T. Kislinger, P. C. Boutros, Onco-proteogenomics: Cancer proteomics joins forces with genomics. *Nat. Methods* **11**, 1107–1113 (2014).
- A. I. Nesvizhskii, Proteogenomics: Concepts, applications and computational strategies. *Nat. Methods* **11**, 1114–1125 (2014).
- W. S. Noble, Mass spectrometrists should search only for peptides they care about. *Nat. Methods* **12**, 605–608 (2015).
- Y. Takahama, I. Ohigashi, S. Baik, G. Anderson, Generation of diversity in thymic epithelial cells. *Nat. Rev. Immunol.* **17**, 295–305 (2017).
- G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**, 764–770 (2011).
- J. C. Castle, M. Loewer, S. Boegel, J. de Graaf, C. Bender, A. D. Tadmor, V. Boisguerin, T. Bukur, P. Sorn, C. Paret, M. Diken, S. Kreiter, Ö. Türeci, U. Sahin, Immunomic, genomic

- and transcriptomic characterization of CT26 colorectal carcinoma. *BMC Genomics* **15**, 190 (2014).
36. P. Fontaine, G. Roy-Proulx, L. Knafo, C. Baron, D.-C. Roy, C. Perreault, Adoptive transfer of minor histocompatibility antigen-specific T lymphocytes eradicates leukemia cells without causing graft-versus-host disease. *Nat. Med.* **7**, 789–794 (2001).
  37. R. Vita, J. A. Overton, C. J. Mungall, A. Sette, B. Peters, FAIR principles and the IEDB: Short-term improvements and a long-term vision of OBO-foundry mediated machine-actionable interoperability. *Database (Oxford)* **2018**, bax105 (2018).
  38. A. Y. Huang, P. H. Gulden, A. S. Woods, M. C. Thomas, C. D. Tong, W. Wang, V. H. Engelhard, G. Pasternack, R. Cotter, D. Hunt, D. M. Pardoll, E. M. Jaffee, The immunodominant major histocompatibility complex class I-restricted antigen of a murine colon tumor derives from an endogenous retroviral gene product. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 9730–9735 (1996).
  39. F. Yue, Y. Cheng, A. Breschi, J. Vierstra, W. Wu, T. Ryba, R. Sandstrom, Z. Ma, C. Davis, B. D. Pope, Y. Shen, D. D. Pervouchine, S. Djebali, R. E. Thurman, R. Kaul, E. Rynes, A. Kirilusha, G. K. Marinov, B. A. Williams, D. Trout, H. Amrhein, K. Fisher-Aylor, I. Antoshechkin, G. DeSalvo, L.-H. See, M. Fastuca, J. Drenkow, C. Zaleski, A. Dobin, P. Prieto, J. Lagarde, G. Bussotti, A. Tanzer, O. Denas, K. Li, M. A. Bender, M. Zhang, R. Byron, M. T. Groudine, D. McCleary, L. Pham, Z. Ye, S. Kuan, L. Edsall, Y.-C. Wu, M. D. Rasmussen, M. S. Bansal, M. Kellis, C. A. Keller, C. S. Morrissey, T. Mishra, D. Jain, N. Dogan, R. S. Harris, P. Cayting, T. Kawli, A. P. Boyle, G. Euskirchen, A. Kundaje, S. Lin, Y. Lin, C. Jansen, V. S. Malladi, M. S. Cline, D. T. Erickson, V. M. Kirkup, K. Learned, C. A. Sloan, K. R. Rosenbloom, B. Lacerda de Sousa, K. Beal, M. Pignatelli, P. Flicek, J. Lian, T. Kahveci, D. Lee, W. J. Kent, M. Ramalho Santos, J. Herrero, C. Notredame, A. Johnson, S. Yong, K. Lee, D. Bates, F. Neri, M. Diegel, T. Canfield, P. J. Sabo, M. S. Wilken, T. A. Reh, E. Giste, A. Shafer, T. Kutayavin, E. Haugen, D. Dunn, A. P. Reynolds, S. Neph, R. Humbert, R. S. Hansen, M. De Bruijn, L. Sella, A. Rudensky, S. Josefowicz, R. Samstein, E. E. Eichler, S. H. Orkin, D. Levasseur, T. Papayannopoulou, K.-H. Chang, A. Skoultschi, S. Gosh, C. Disteche, P. Treuting, Y. Wang, M. J. Weiss, G. A. Blobel, X. Cao, S. Zhong, T. Wang, P. J. Good, R. F. Lowdon, L. B. Adams, X.-Q. Zhou, M. J. Pazin, E. A. Feingold, B. Wold, J. Taylor, A. Mortazavi, S. M. Weissman, J. A. Stamatoyannopoulos, M. P. Snyder, R. Guigo, T. R. Gingeras, D. M. Gilbert, R. C. Hardison, M. A. Beer, B. Ren; The Mouse ENCODE Consortium, A comparative encyclopedia of DNA elements in the mouse genome. *Nature* **515**, 355–364 (2014).
  40. C. A. Sloan, E. T. Chan, J. M. Davidson, V. S. Malladi, J. S. Strattan, B. C. Hitz, I. Gabdank, A. K. Narayanan, M. Ho, B. T. Lee, L. D. Rowe, T. R. Dreszer, G. Roe, N. R. Podduturi, F. Tanaka, E. L. Hong, J. M. Cherry, ENCODE data at the ENCODE portal. *Nucleic Acids Res.* **44**, D726–D732 (2016).
  41. M.-H. Fortier, É. Caron, M.-P. Hardy, G. Voisin, S. Lemieux, C. Perreault, P. Thibault, The MHC class I peptide repertoire is molded by the transcriptome. *J. Exp. Med.* **205**, 595–610 (2008).
  42. K. Vincent, M.-P. Hardy, A. Trofimov, C. M. Laumont, D. Sriranganadane, S. Hadj-Mimoune, I. Salem Fourati, H. Soudeyns, P. Thibault, C. Perreault, Rejection of leukemic cells requires antigen-specific T cells with high functional avidity. *Biol. Blood Marrow Transplant.* **20**, 37–45 (2014).
  43. M. Bassani-Sternberg, S. Pletscher-Frankild, L. J. Jensen, M. Mann, Mass spectrometry of human leukocyte antigen class I peptidomes reveals strong effects of protein abundance and turnover on antigen presentation. *Mol. Cell. Proteomics* **14**, 658–673 (2015).
  44. D. P. Granados, C. M. Laumont, P. Thibault, C. Perreault, The nature of self for T cells—A systems-level perspective. *Curr. Opin. Immunol.* **34**, 1–8 (2015).
  45. W. Chen, L. C. Antón, J. R. Bennink, J. W. Yewdell, Dissecting the multifactorial causes of immunodominance in class I-restricted T cell responses to viruses. *Immunity* **12**, 83–93 (2000).
  46. M. K. Jenkins, J. J. Moon, The role of naive T cell precursor frequency and recruitment in dictating immune response magnitude. *J. Immunol.* **188**, 4135–4140 (2012).
  47. J. J. Moon, H. H. Chu, M. Pepper, S. J. McSorley, S. C. Jameson, R. M. Kedl, M. K. Jenkins, Naive CD4<sup>+</sup> T cell frequency varies for different epitopes and predicts repertoire diversity and response magnitude. *Immunity* **27**, 203–213 (2007).
  48. F. P. Legoux, J. J. Moon, Peptide:MHC tetramer-based enrichment of epitope-specific T cells. *J. Vis. Exp.* 4420 (2012).
  49. N. Vigneron, V. Stroobant, B. J. Van den Eynde, P. van der Bruggen, Database of T cell-defined human tumor antigens: The 2013 update. *Cancer Immunol.* **13**, 15 (2013).
  50. M. S. Rooney, S. A. Shukla, C. J. Wu, G. Getz, N. Hacohen, Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).
  51. V. P. Balachandran, M. Łuksza, J. N. Zhao, V. Makarov, J. A. Moral, R. Remark, B. Herbst, G. Askán, U. Bhanot, Y. Senbabaoglu, D. K. Wells, C. I. O. Cary, O. Grbovic-Huezo, M. Attiyeh, B. Medina, J. Zhang, J. Loo, J. Saglimbeni, M. Abu-Akeel, R. Zappasodi, N. Riaz, M. Smoragiewicz, Z. L. Kelley, O. Basturk; Australian Pancreatic Cancer Genome Initiative, M. Gönen, A. J. Levine, P. J. Allen, D. T. Fearon, M. Merad, S. Gnjatic, C. A. Iacobuzio-Donahue, J. D. Wolchok, R. P. DeMatteo, T. A. Chan, B. D. Greenbaum, T. Merghoub, S. D. Leach, Identification of unique neoantigen qualities in long-term survivors of pancreatic cancer. *Nature* **551**, 512–516 (2017).
  52. M. J. Welters, T. C. van der Sluis, H. van Meir, N. M. Loof, V. J. van Ham, S. van Duikeren, S. J. Santegoets, R. Arens, M. L. de Kam, A. F. Cohen, M. I. van Poelgeest, G. G. Kenter, J. R. Kroep, J. Burggraaf, C. J. Melief, S. H. van der Burg, Vaccination during myeloid cell depletion by cancer chemotherapy fosters robust T cell responses. *Sci. Transl. Med.* **8**, 334ra352 (2016).
  53. E. Caron, D. J. Kowalewski, C. Chiek Koh, T. Sturm, H. Schuster, R. Aebersold, Analysis of major histocompatibility complex (MHC) immunopeptidomes using mass spectrometry. *Mol. Cell. Proteomics* **14**, 3105–3117 (2015).
  54. D. Gfeller, M. Bassani-Sternberg, Predicting antigen presentation—What could we learn from a million peptides? *Front. Immunol.* **9**, 1716 (2018).
  55. Y.-C. Lu, Z. Zheng, P. F. Robbins, E. Tran, T. D. Prickett, J. J. Gartner, Y. F. Li, S. Ray, Z. Franco, V. Bliskovsky, P. C. Fitzgerald, S. A. Rosenberg, An efficient single-cell RNA-seq approach to identify neoantigen-specific T cell receptors. *Mol. Ther.* **26**, 379–389 (2018).
  56. Z. Hu, A. J. Anandappa, J. Sun, J. Kim, D. E. Leet, D. J. Bozym, C. Chen, L. Williams, S. A. Shukla, W. Zhang, D. Tabbaa, S. Steelman, O. Olive, K. J. Livak, H. Kishi, A. Muraguchi, I. Guleria, J. Stevens, W. J. Lane, U. E. Burkhardt, E. F. Fritsch, D. Neuberger, P. A. Ott, D. B. Keskin, N. Hacohen, C. J. Wu, A cloning and expression system to probe T-cell receptor specificity and assess functional avidity to neoantigens. *Blood* **132**, 1911–1921 (2018).
  57. L. M. Kranz, M. Diken, H. Haas, S. Kreiter, C. Loquai, K. C. Reuter, M. Meng, D. Fritz, F. Vascotto, H. Hefesha, C. Grunwitz, M. Vormehr, Y. Hüseman, A. Selmi, A. N. Kuhn, J. Buck, E. Derhovanessian, R. Rae, S. Attig, J. Diekmann, R. A. Jabulowsky, S. Heesch, J. Hassel, P. Langguth, S. Grabbe, C. Huber, Ö. Türeci, U. Sahin, Systemic RNA delivery to dendritic cells exploits antiviral defence for cancer immunotherapy. *Nature* **534**, 396–401 (2016).
  58. J. A. Vizcaíno, A. Csordas, N. del-Toro, J. A. Dianes, J. Griss, I. Lavidas, G. Mayer, Y. Perez-Riverol, F. Reisinger, T. Ternent, Q.-W. Xu, R. Wang, H. Hermjakob, 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, 11033 (2016).
  59. J. Lanoix, C. Durette, M. Courcelles, E. Cossette, S. Comtois-Marotte, M.-P. Hardy, C. Côté, C. Perreault, P. Thibault, Comparison of the MHC I immunopeptidome repertoire of B-cell lymphoblasts using two isolation methods. *Proteomics* **18**, e1700251 (2018).
  60. M.-J. Kim, C. M. Miller, J. L. Shadrach, A. J. Wagers, T. Serwold, Young, proliferative thymic epithelial cells engraft and function in aging thymuses. *J. Immunol.* **194**, 4784–4795 (2015).
  61. C. Stoekle, I. A. Rota, E. Tolosa, C. Haller, A. Melms, E. Adamopoulou, Isolation of myeloid dendritic cells and epithelial cells from human thymus. *J. Vis. Exp.* e50951 (2013).
  62. A. Dobin, C. A. Davis, F. Schlesinger, J. Drenkow, C. Zaleski, S. Jha, P. Batut, M. Chaisson, T. R. Gingeras, STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
  63. T. Daouda, C. Perreault, S. Lemieux, pyGeno: A Python package for precision medicine and proteogenomics. *F1000Res.* **5**, 381 (2016).
  64. D. R. Zerbino, P. Achuthan, W. Akanni, M. R. Amode, D. Barrell, J. Bhai, K. Billis, C. Cummins, A. Gall, C. G. Girón, L. Gil, L. Gordon, L. Haggerty, E. Haskell, T. Hourlier, O. G. Izuogu, S. H. Janacek, T. Juettemann, J. K. To, M. R. Laird, I. Lavidas, Z. Liu, J. E. Loveland, T. Maurel, W. McLaren, B. Moore, J. Mudge, D. N. Murphy, V. Newman, M. Nuhn, D. Ogeh, C. K. Ong, A. Parker, M. Patricio, H. S. Riat, H. Schuilenburg, D. Sheppard, H. Sparrow, K. Taylor, A. Thormann, A. Vullo, B. Walts, A. Zadissa, A. Frankish, S. E. Hunt, M. Kostadima, N. Langridge, F. J. Martin, M. Muffato, E. Perry, M. Ruffier, D. M. Staines, S. J. Trevanion, B. L. Aken, F. Cunningham, A. Yates, P. Flicek, Ensembl 2018. *Nucleic Acids Res.* **46**, D754–D761 (2018).
  65. E. Caron, K. Vincent, M.-H. Fortier, J.-P. Laverdure, A. Bramoullé, M.-P. Hardy, G. Voisin, P. P. Roux, S. Lemieux, P. Thibault, C. Perreault, The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. *Mol. Syst. Biol.* **7**, 533 (2011).
  66. M. Andreatta, M. Nielsen, Gapped sequence alignment using artificial neural networks: Application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2016).
  67. J. T. Robinson, H. Thorvaldsdóttir, W. Winckler, M. Guttman, E. S. Lander, G. Getz, J. P. Mesirov, Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
  68. M. S. Bereman, J. Beri, V. Sharma, C. Nathe, J. Eckels, B. MacLean, M. J. MacCoss, An automated pipeline to monitor system performance in liquid chromatography–tandem mass spectrometry proteomic experiments. *J. Proteome Res.* **15**, 4763–4769 (2016).

**Acknowledgments:** We thank the following members of IRIC core facilities for sound advice and technical assistance: J. Huber and F. Guilloteau from the genomic platform; S. Comtois-Marotte and É. Cossette from the proteomic platform; G. Dulude, D. Gagné, and A. Gosselin from the flow cytometry platform; and I. Caron from the animal care facility. We acknowledge the dedicated work of C. Rondeau from the Banque de Cellules Leucémiques du Québec (BCLQ). We also thank the NIH Tetramer Core Facility for providing all the tetramers used in this study. Furthermore, we thank the ENCODE consortium, especially the laboratories of T. Gingeras (Cold Spring Harbor Laboratory) and M. Snyder (Stanford University) for generating the murine tissue datasets used in this study. Last, we thank the GTEx Project for providing RNA-seq data from human tissues used in this study. **Funding:**

This work was supported by grants from the Canadian Cancer Society (grant 701564 to C.P. and P.T.), the Terry Fox Research Institute (grant TRP 1060/32-iTNT to C.P.), and the Quebec Breast Cancer Foundation (grant 19579 to C.P.). C.M.L. is supported by a Cole Foundation fellowship. IRIC receives infrastructure support from Genome Canada, the Canadian Center of Excellence in Commercialization and Research, the Canadian Foundation for Innovation, and the Fonds de Recherche du Québec-Santé (FRQS). **Author contributions:** C.M.L., K.V., and C.P. designed the study. C.M.L., K.V., L.H., É.A., É.B., J.-P.L., P.G., M.C., M.-P.H., C.C., C.D., C.S.P., M.B., and J.L. performed the experiments or bioinformatics analysis. S.V. and E.H. provided human thymi for TEC/mTEC extraction. C.M.L., K.V., and L.H. analyzed the data. C.M.L., K.V., L.H., É.A., É.B., S.L., P.T., and C.P. discussed the results. C.M.L., K.V., and C.P. wrote the first draft of the manuscript, and L.H. contributed to the writing. All authors edited and approved the final manuscript. **Competing interests:** C.M.L., S.L., P.T., and C.P. are named inventors in the patent application 782-15691.134-US PROV APPLICATION (pending) filed by Université de Montréal on 22 December 2017. This patent application covers the method used for TSA discovery described in Fig. 1 and the TSAs listed in tables S6, S17, and S18. The remaining authors declare that they have no competing interests. **Data and materials availability:** In-house scripts used in this study are available on Zenodo at DOI: 10.5281/zenodo.1484486. pyGeno is available on GitHub (<https://github.com/tariqdaouda/pyGeno>). Information regarding all samples used in this study is listed in table S2. The databases for human TECs and mTECs are available on Zenodo using the following DOIs: 10.5281/zenodo.1484261 ( $k = 24$  nucleotides) and 10.5281/zenodo.1484490 ( $k = 33$  nucleotides). All other sequencing and expression data have been deposited to the NCBI Sequence Read Archive and

GEO under accession code GSE113992, containing the GSE111092 and the GSE113972 sets of murine and human sequencing and expression data, respectively. MS raw data and associated databases are deposited to the ProteomeXchange Consortium via the PRIDE (58) partner repository with the following dataset identifiers: PXD009065 and 10.6019/PXD009065 (CT26 cell line), PXD009064 and 10.6019/PXD009064 (EL4 cell line), PXD009749 and 10.6019/PXD009749 (07H103), PXD009753 and 10.6019/PXD009753 (10H080, mild acid elution), PXD007935—assay no. 81756 and 10.6019/PXD007935 (10H080, immunoprecipitation) (59), PXD009750 and 10.6019/PXD009750 (10H118), PXD009751 and 10.6019/PXD009751 (12H018), PXD009752 and 10.6019/PXD009752 (lc2), PXD009754 and 10.6019/PXD009754 (lc4), and PXD009755 and 10.6019/PXD009755 (lc6).

Submitted 21 June 2018

Accepted 15 November 2018

Published 5 December 2018

10.1126/scitranslmed.aau5516

**Citation:** C. M. Laumont, K. Vincent, L. Hesnard, É. Audemard, É. Bonneil, J.-P. Laverdure, P. Gendron, M. Courcelles, M.-P. Hardy, C. Côté, C. Durette, C. St-Pierre, M. Benhammadi, J. Lanoix, S. Vobecky, E. Haddad, S. Lemieux, P. Thibault, C. Perreault, Noncoding regions are the main source of targetable tumor-specific antigens. *Sci. Transl. Med.* **10**, eaau5516 (2018).

## Noncoding regions are the main source of targetable tumor-specific antigens

Céline M. Laumont, Krystel Vincent, Leslie Hesnard, Éric Audemard, Éric Bonneil, Jean-Philippe Laverdure, Patrick Gendron, Mathieu Courcelles, Marie-Pierre Hardy, Caroline Côté, Chantal Durette, Charles St-Pierre, Mohamed Benhammedi, Joël Lanoix, Suzanne Vobecky, Elie Haddad, Sébastien Lemieux, Pierre Thibault and Claude Perreault

*Sci Transl Med* 10, eaau5516.  
DOI: 10.1126/scitranslmed.aau5516

### Expanding the landscape of immunotherapy targets

Most searches for druggable tumor-specific antigens (TSAs) start with an examination of peptides derived from protein-coding exons. Laumont *et al.* took a different approach and found numerous TSAs aberrantly expressed from noncoding sequences in murine cell lines and in B-lineage acute lymphoblastic leukemia and lung cancer patient samples, but not in cells responsible for T cell selection. The authors validated the immunogenicity and efficacy of TSA vaccination for select antigens in mouse models of cancer. The finding that noncoding regions are a potentially rich source of TSAs could greatly expand the number of targetable antigens across different cancers, including those with low mutational burdens.

#### ARTICLE TOOLS

<http://stm.sciencemag.org/content/10/470/eaau5516>

#### SUPPLEMENTARY MATERIALS

<http://stm.sciencemag.org/content/suppl/2018/12/03/10.470.eaau5516.DC1>

#### RELATED CONTENT

<http://stm.sciencemag.org/content/scitransmed/10/450/eaar3342.full>  
<http://stm.sciencemag.org/content/scitransmed/10/461/eaat1445.full>  
<http://stm.sciencemag.org/content/scitransmed/10/433/eaar1916.full>  
<http://stm.sciencemag.org/content/scitransmed/10/436/eaao5931.full>

#### REFERENCES

This article cites 66 articles, 12 of which you can access for free  
<http://stm.sciencemag.org/content/10/470/eaau5516#BIBL>

#### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)