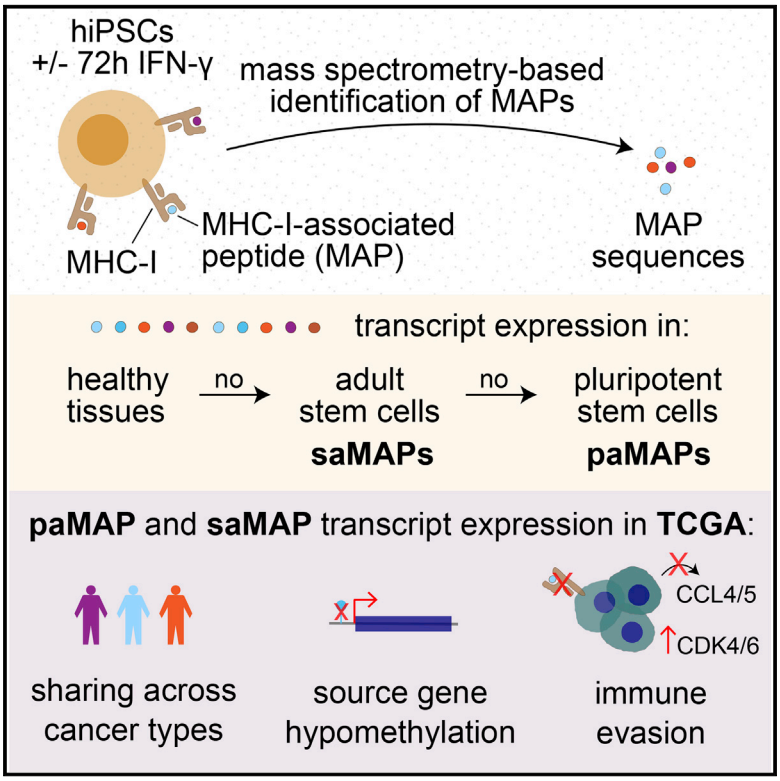
Cell Reports

Article

Induced pluripotent stem cells display a distinct set of MHC I-associated peptides shared by human cancers

Graphical abstract



Highlights

- Human iPSCs express a distinct set of pluripotency-associated MAPs (paMAPs)
- High-stemness cancers acquire paMAP expression
- paMAP expression is associated with immune evasion
- paMAPs shared between iPSCs and cancer cells are immunogenic

Authors

Anca Apavaloaei, Leslie Hesnard, Marie-Pierre Hardy, ..., Mick Bhatia, Pierre Thibault, Claude Perreault

Correspondence

pierre.thibault@umontreal.ca (P.T.), claude.perreault@umontreal.ca (C.P.)

In brief

Using immunopeptidomic analyses, Apavaloaei et al. show that human iPSCs express a set of pluripotency-associated MHC I peptides (paMAPs) that are shared by multiple cancer types. The presence of paMAPs is associated with a pan-cancer rewiring of the epigenome, metabolome, and signaling networks instrumental in immune escape.

Apavaloaei et al., 2022, Cell Reports *40*, 111241
August 16, 2022
https://doi.org/10.1016/j.celrep.2022.111241

CellPress

Cell Reports

CellPress
OPEN ACCESS

Article

Induced pluripotent stem cells display a distinct set of MHC I-associated peptides shared by human cancers

Anca Apavaloaei,[1,2] Leslie Hesnard,[1] Marie-Pierre Hardy,[1] Basma Benabdallah,[3] Gregory Ehx,[1] Catherine Thériault,[1] Jean-Philippe Laverdure,[1] Chantal Durette,[1] Joël Lanoix,[1] Mathieu Courcelles,[1] Nandita Noronha,[1,2] Kapil Dev Chauhan,[4] Sébastien Lemieux,[1,5] Christian Beauséjour,[3,6] Mick Bhatia,[4,7] Pierre Thibault,[1,8,*] and Claude Perreault[1,2,9,*]

[1]Institute for Research in Immunology and Cancer (IRIC), University of Montreal, Montreal, QC H3T 1J4, Canada
[2]Department of Medicine, University of Montreal, Montreal, QC H3T 1J4, Canada
[3]CHU Sainte-Justine Research Center, Montreal, QC H3T 1C5, Canada
[4]Faculty of Health Sciences, Michael G. DeGroote School of Medicine, McMaster University, Hamilton, ON L8N 3Z5, Canada
[5]Department of Biochemistry and Molecular Medicine, University of Montreal, Montreal, QC H3T 1J4, Canada
[6]Department of Pharmacology and Physiology, University of Montreal, Montreal, QC H3T 1J4, Canada
[7]Department of Biochemistry and Biomedical Sciences, McMaster University, Hamilton, ON L8N 3Z5, Canada
[8]Department of Chemistry, University of Montreal, Montreal, QC H3T 1J4, Canada
[9]Lead contact
*Correspondence: pierre.thibault@umontreal.ca (P.T.), claude.perreault@umontreal.ca (C.P.)
https://doi.org/10.1016/j.celrep.2022.111241

SUMMARY

Previous reports showed that mouse vaccination with pluripotent stem cells (PSCs) induces durable anti-tumor immune responses via T cell recognition of some elusive oncofetal epitopes. We characterize the MHC I-associated peptide (MAP) repertoire of human induced PSCs (iPSCs) using proteogenomics. Our analyses reveal a set of 46 pluripotency-associated MAPs (paMAPs) absent from the transcriptome of normal tissues and adult stem cells but expressed in PSCs and multiple adult cancers. These paMAPs derive from coding and allegedly non-coding (48%) transcripts involved in pluripotency maintenance, and their expression in The Cancer Genome Atlas samples correlates with source gene hypomethylation and genomic aberrations common across cancer types. We find that several of these paMAPs were immunogenic. However, paMAP expression in tumors coincides with activation of pathways instrumental in immune evasion (WNT, TGF-β, and CDK4/6). We propose that currently available inhibitors of these pathways could synergize with immune targeting of paMAPs for the treatment of poorly differentiated cancers.

INTRODUCTION

CD8 T cell-based anti-tumor responses rely on the recognition of MHC class I-associated peptides (MAPs) on cancer cells. Both spontaneously and following cancer immunotherapy, initial priming of naive CD8 T cells against tumor MAPs dictates cancer cell killing (Ochsenbein, 2005; Salmon et al., 2016). CD8 T cell priming is achieved primarily by professional antigen-presenting cells, which take up and cross-present stable and abundant tumor antigens from neighbor cells. However, a significant limitation of the dependency on antigen transfer to professional antigen-presenting cells is that the immune system remains ignorant of many tumor MAPs derived from unstable tumor proteins since they are not cross-presented (Apavaloaei et al., 2020; Yewdell, 2010). Hence, vaccines using dendritic cells (DC) loaded with desired tumor antigens have the potential to broaden the repertoire of anti-tumor T cells by eliciting responses against tumor MAPs that are ignored when presented solely by cancer cells (Apavaloaei et al., 2020). One key hurdle of this approach is our rudimentary understanding of the structure of actionable MAPs that mediate cancer regression.

MAPs overexpressed on tumors relative to normal cells, i.e., tumor-associated antigens, have led to disappointing results in vaccination studies due to the low affinity of specific T cells (Haen et al., 2020). In contrast, MAPs arising from somatic cancer mutations, commonly known as mutated tumor-specific antigens (mTSAs) or neoantigens, can elicit potent anti-tumor responses because they do not induce central immune tolerance (Schumacher et al., 2019). However, mTSAs are lowly abundant in most cancer types and highly patient specific (Apavaloaei et al., 2020; Ehx and Perreault, 2019; Löffler et al., 2019). Other MAPs called aberrantly expressed TSAs (aeTSAs) can arise from epigenetic, transcription, or translation events specific to cancer. One such family of aeTSAs, namely MAPs derived from cancer-germline antigens (CGAs), is coded by non-mutated canonical exons normally restricted to germ cells and re-expressed in multiple cancers due to epigenetic alterations (Coulie et al., 2014). DC-based vaccines against these antigens have shown encouraging results in recent clinical trials in advanced melanoma and lymphoma (Sahin et al., 2020; Vasileiou et al., 2021). In addition, we and others have also shown that allegedly non-protein-coding regions of the

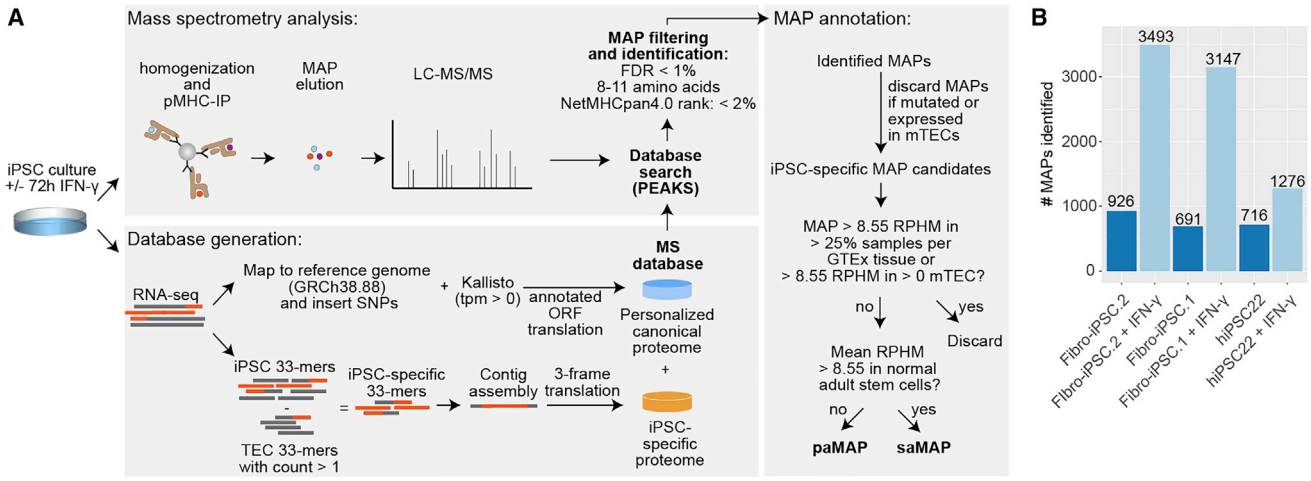


Figure 1. MS-based identification of paMAPs using human iPSCs. See also Figure S1

(A) Workflow for paMAP identification using iPSCs, based on the proteogenomic approach from (Laumont et al., 2018). pMHC-IP, peptide-MHC I immunoprecipitation; MAP, MHC I-associated peptide; TEC, thymic epithelial cells; LC-MS/MS, liquid chromatography with tandem mass spectrometry; FDR, false discovery rate; RPHM, reads per hundred million.

(B) Total number of MAPs identified per iPSC sample before MAP annotation.

genome, including untranslated regions (UTRs), frameshift, intergenic and intronic regions, and endogenous retroelements (EREs), are translated and represent a major source of aeTSAs (Chong et al., 2020; Ehx et al., 2021; Erhard et al., 2020; Laumont et al., 2018; Wei and Guo, 2020; Zhao et al., 2020). These aeTSAs are more common than mTSAs and highly shared within cancer types, representing attractive targets for DC-based vaccines (Zitvogel et al., 2021).

Oncofetal antigens represent a different class of non-mutated MAPs that represent attractive targets for cancer immunotherapy but have, thus far, not been explored by systematic immunopeptidomic analyses. The antigenic similarity between embryos and tumors has been acknowledged for more than a century. Indeed, several reports show that vaccination of mice with embryonic material induces killing of tumors from different origins and that natural pregnancy (in animals including humans) generates anti-tumor antibodies and cytotoxic lymphocytes (for a detailed history, see Brewer et al., 2009). More recently, several studies have shown that mouse immunization with irradiated embryonic stem cells (ESCs) or autologous induced pluripotent stem cells (iPSCs) mounts effective long-term protection against multiple cancer types (Kooreman et al., 2018; Yaddanapudi et al., 2012). With a "two-way immunity" experiment, Kooreman et al. demonstrated that effector T cells from iPSC-vaccinated mice recognized shared epitopes between iPSCs and cancer cells (Kooreman et al., 2018). This report indicates that antigens expressed specifically in pluripotent stem cells (PSCs) can generate MAPs and are an important source of actionable aeTSAs. Such aeTSAs would represent attractive targets for developing vaccines against a wide range of cancers, mainly for two reasons: their absence on normal cells should minimize collateral damage, and they should be more highly shared among tumors than mTSAs. Therefore, herein we used mass spectrometry (MS) to study the immunopeptidome of human iPSCs in search of pluripotency-associated MAPs (paMAPs) shared by human cancers.

RESULTS

MS-based identification of paMAPs using human iPSCs

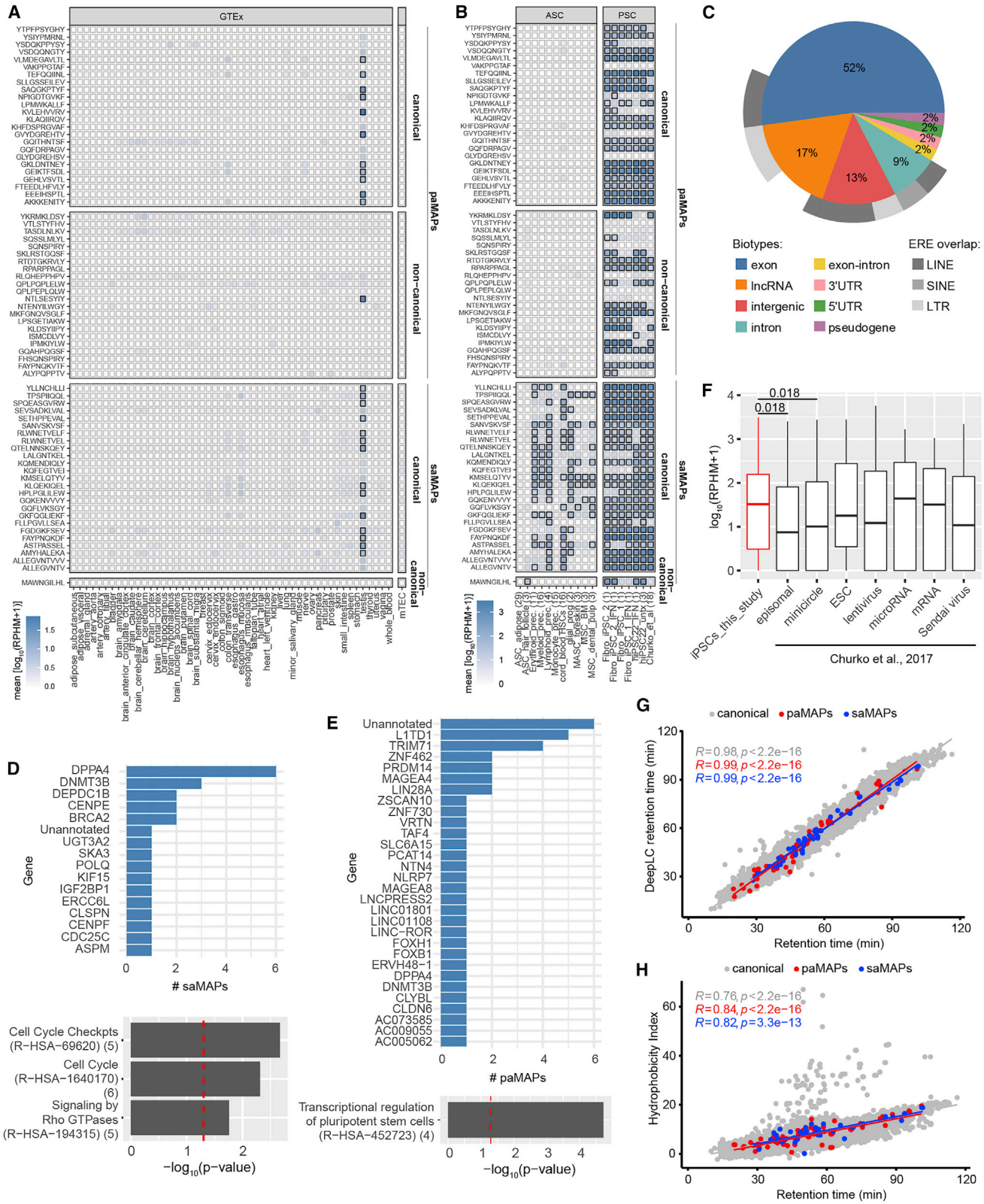
To identify paMAPs derived from all possible genomic regions, we used a proteogenomic strategy that we previously developed for TSA identification (Laumont et al., 2018). In essence, we constructed iPS-cell-line-specific MS databases by combining (1) annotated proteome-derived sequences (canonical proteome) and (2) three-frame translations of canonical and non-canonical iPSC-specific contigs depleted of subsequences expressed in human medullary thymic epithelial cells (mTECs) (Figure 1A). This method maintains an optimal database size and, due to the role of mTECs in mediating central tolerance, enables the identification of MAPs that may be immunogenic.

Because the abundance of MAPs is limiting in MS-based identifications and iPSCs express low levels of surface MHC I molecules (Suárez-Alvarez et al., 2010; Vogel and Marcotte, 2009), we added a 72-h treatment with interferon-γ (IFN-γ) prior to collection for MS analysis (Figure 1A). IFN-γ treatment induced, on average, a 34-fold increase in surface HLA-A/B/C levels for the three fibroblast-derived iPSC samples studied (see STAR Methods; Table S1), without affecting the pluripotency of the iPSCs, as evaluated using trilineage differentiation assays and the expression of canonical pluripotency markers (Stewart et al., 2006) (Figure S1). As a result, this treatment allowed the detection of 1.8- to 4.5-fold more non-redundant MAPs than for untreated iPSCs (Figure 1B; Tables S1 and S2), thus expanding our search space for paMAPs.

The immunopeptidome of iPSCs reflects their pluripotency state

The probability that a MAP will be presented at the cell surface depends mainly on two factors. First, it depends on the expression of MHC I genes, which is high in hematopoietic cells and mTECs (MHC I^{hi}) but low on non-inflamed extrathymic

CellPress
OPEN ACCESS



(legend on next page)

nonhematopoietic cells (MHC I^{lo}) (Benhammadi et al., 2020). Second, it depends on the expression of the MAP coding sequence (MCS) (Bassani-Sternberg et al., 2015; Ehx et al., 2021; Pearson et al., 2016, Ruiz Cuevas et al., 2021). Indeed, an MCS expression inferior to 8.55 RPHM (reads per hundred million) corresponds to a probability of MAP generation lower than 5% in myeloid cells (Ehx et al., 2021). We can assume that the probability would be even lower in extrathymic nonhematopoietic cells because they are MHC I^{lo}. Hence, in our search for paMAPs, we selected MAPs whose MCS were expressed at less than 8.55 RPHM in 29 different healthy tissues from the GTEx dataset (Genotype-Tissue Expression; Lonsdale et al., 2013). MCS expression in the testis was not an exclusion criterion because cells of the spermatocyte lineage do not express MHC I genes (Zhao et al., 2014) and may retain expression of some pluripotency markers (Izadyar et al., 2011; Wang et al., 2007; Zheng et al., 2009). Of the 5,424 individual MAPs identified from untreated and IFN-γ-treated iPSCs, 72 (1.33%) matched our stringent expression profile (Figures 1A and 2A; Tables S3 and S4). To distinguish MAPs associated with a stemness program as opposed to a pluripotency program, we then quantified the 72 MCSs in the RNA-seq of primary adult stem and progenitor cells (ASCs) from different origins: mesenchymal stem cells, bone marrow progenitors, hematopoietic stem cells from cord blood samples, and glial progenitors (Table S4). We found that 26 MAPs were expressed in at least one ASC dataset and were termed stemness-associated MAPs (saMAPs), whereas the remaining 46 MAPs were considered pluripotency-associated (paMAPs) (Figure 2B; Table S3). Because we lacked single nucleotide variation information for the somatic cells used for iPSC generation and could not discriminate between germline and reprogramming-associated mutations (Merkle et al., 2017), MAPs deriving from mutated DNA sequences were excluded from these analyses.

It is notable that all but one saMAP were derived from annotated protein-coding exons, whereas 48% of the paMAPs were derived from allegedly non-coding genomic regions, in particular from annotated long non-coding RNAs (lncRNAs) (17%) and intergenic (13%) and intronic (9%) sequences (Figures 2B and 2C; Table S3). Remarkably, nearly all paMAP-coding sequences from these three ostensibly non-coding regions had overlapping EREs comprising primarily long-interspersed nuclear element (LINE) and long terminal repeat sequences (Figure 2C; Table S3). These elements, most notably LINE-1 (L1) and human endogenous retrovirus subfamily H (HERV-H), are derepressed during reprogramming and essential for the maintenance of pluripotency because they enhance the specific expression of lncRNAs and neighboring genes (Fort et al., 2014; Friedli et al., 2014; Kelley and Rinn, 2012; Klawitter et al., 2016; Lu et al., 2014). The major overlap with EREs provides a mechanistic rationale for the PSC-specific transcription and translation of these allegedly non-coding regions. By contrast, ERE overlap with canonical paMAP-coding sequences was found only in paMAPs derived from LINE-1 type transposase domain containing 1 (L1TD1), a domesticated RNA-binding protein derived from an L1 retroelement and required for the self-renewal of PSCs (McLaughlin et al., 2014; Närvä et al., 2012) (Table S3).

Several features of paMAPs reinforce their association to pluripotency. Firstly, the paMAP- and saMAP-source genes were non-redundant, except for two genes, *DNMT3B* and *DPPA4*. These two genes generated iPSC-specific expression of non-canonical MAPs, whereas their exonic MAPs were also highly expressed in ASCs (Figures 2B, 2D, and 2E; Table S3). Accordingly, the only biological pathway significantly enriched in the paMAP-source genes was *transcriptional regulation of PSCs,* represented by the pluripotency-regulating genes *LIN28A*, *ZSCAN10*, *PRDM14,* and *DPPA4* (Chia et al., 2010; Hernandez et al., 2018; Wang et al., 2007; Zhang et al., 2016) (Figure 2E). By contrast, saMAP-source genes were primarily involved in cell cycle regulation (Figure 2D). Secondly, paMAPs were expressed similarly between ESCs and iPSCs generated from six different reprogramming methods (Churko et al., 2017) (Figures 2F and S2A). Besides, paMAPs were expressed similarly in the untreated and IFN-γ-treated samples studied, and most paMAPs identified from untreated iPSCs were also found in the immunopeptidome of IFN-γ-treated samples (Figures S2B and S2C). This indicates that paMAPs are processed similarly between untreated and IFN-γ-treated iPSCs. Finally, we evaluated whether annotated saMAP- and paMAP-source genes (Table S3) can infer stemness and pluripotency using single-sample gene set enrichment analysis (ssGSEA), as previously described (Barbie et al., 2009; Hänzelmann et al., 2013). To this end, we used the stemness signatures extracted by Miranda and colleagues from different cancer datasets (Miranda et al., 2019). Our saMAP- and paMAP-source gene signatures (saMAP

Figure 2. The immunopeptidome of iPSCs reflects their pluripotency state

(A) Heatmap showing the mean RNA expression [log10(RPHM+1)] of paMAPs and saMAPs in healthy tissues from the GTEx consortium (n = 5–150, depending on sample availability) and in mTECs (n = 11) (Table S4). Boxed: tissues with expression >8.55 RPHM in >25% of samples.

(B) Heatmap showing the mean RNA expression [log10(RPHM+1)] of paMAPs and saMAPs in PSCs (from this study and from Churko et al., 2017) and ASCs (healthy sorted primary adult stem cells, normal hematopoietic precursors (prec.), or cord blood samples) (Table S4). Boxed: mean expression across samples >8.55 RPHM. The number of samples in each sample group is in parentheses. MSC, mesenchymal stem cells.

(C) Pie chart displaying the percentage of paMAP-source genes corresponding to each biotype and the class of the ERE overlapping at the respective paMAP-coding region, if applicable.

(D and E) Top: Number of saMAPs (D) or paMAPs (E) derived from each source gene. Bottom: Reactome pathways significantly enriched in saMAP-source (D) or paMAP-source (E) genes. Red dotted lines, p value = 0.05.

(F) Boxplot showing the expression [log10(RPHM+1)] of paMAP-coding sequences in the iPSCs from this study and the PSCs from Churko et al. (2017), with iPSCs grouped according to the method used for reprogramming. Data are represented as the median and inter-quartile range, and whiskers extend to the largest value no further than 1.5 * IQR from the box hinges. p values from pairwise Wilcoxon rank-sum test, adjusted for multiple comparisons using the Benjamini-Hochberg method.

(G and H) Pearson correlations between observed retention times and predicted retention time (G) or hydrophobicity index (H).

See also Figures S2 and S3.

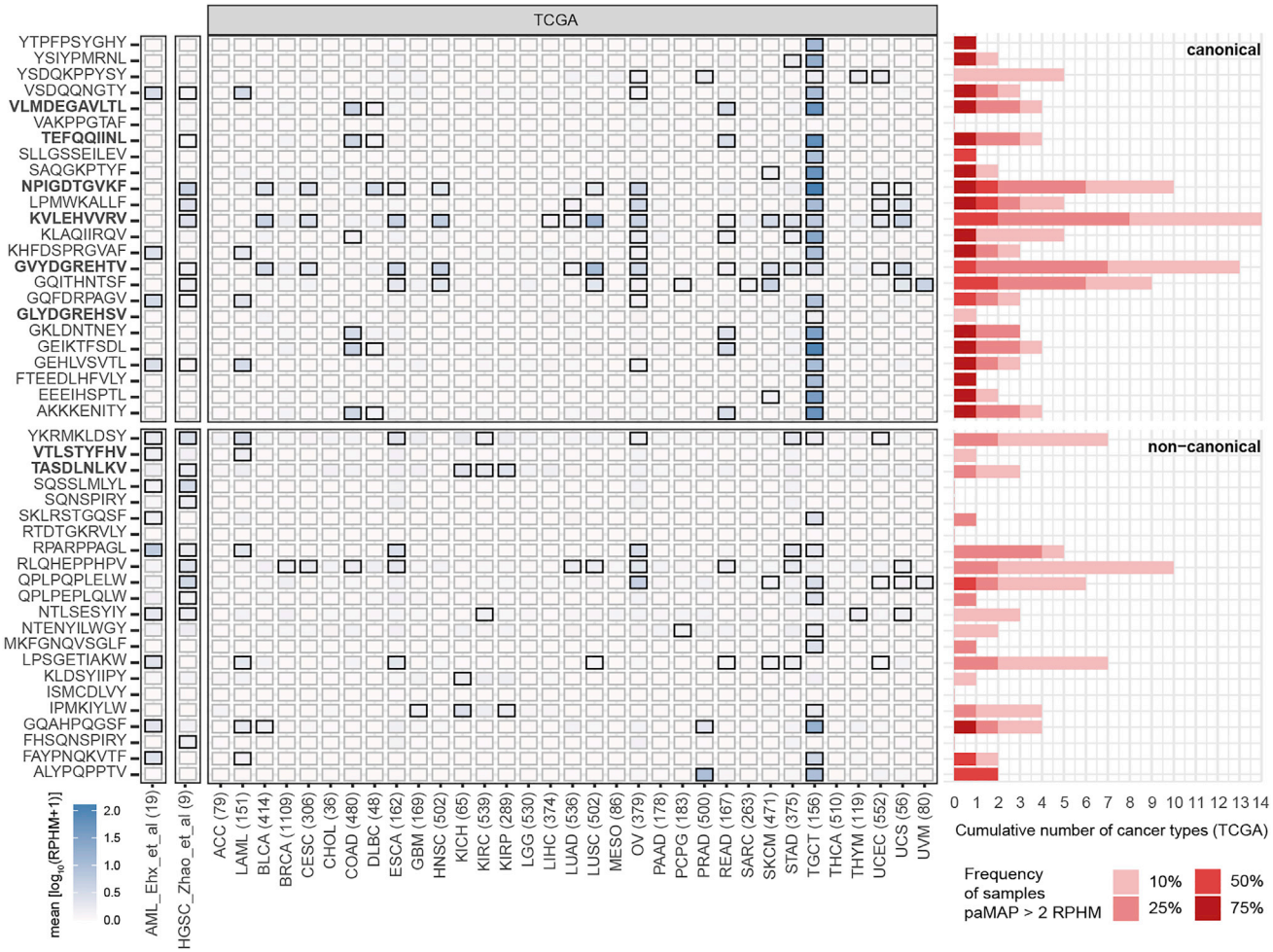


Figure 3. paMAPs derived from iPSCs are shared across cancer types

Left: Heatmap showing the mean RNA expression [log10(RPHM+1)] of paMAPs in cancer samples from our lab or TCGA, and the respective number of samples per cancer type in parentheses (Table S4). Boxed: tissues with expression >2 RPHM in >10% of samples. Right: Bar plot showing the cumulative number of TCGA cancer types expressing the paMAP-coding sequence at different levels of sharing among samples. TCGA acronyms were used as defined by TCGA (https://portal.gdc.cancer.gov/). paMAPs in bold were previously reported.

ssGSEA and paMAP ssGSEA, respectively) showed a good correlation with the stemness signatures in an array of RNA-seq data from PSCs, sorted progenitor and differentiated cells from various sources (Pearson's R > 0.5 for most gene sets, Figures S2D and S2E). Although additional analyses will be required to validate this finding in other datasets, the paMAP-source gene enrichment achieved the highest specificity to PSCs in our analyses (Figures S2F and S2G). Hence, we conclude that the immunopeptidome of iPSCs contains paMAPs derived from pluripotency-associated transcription events absent from healthy tissues and ASCs.

To assess the robustness of paMAP and saMAP identifications, we used the two best-in-class metrics to validate MAPs identified with high-throughput MS. We thus compared the distribution of the observed MAP retention times (RT) with the distribution of the RT predicted using the DeepLC algorithm (Bouwmeester et al., 2021) and with the distribution of the hydrophobicity index assessed with SSRcalc (Krokhin, 2006). Both of these metrics had a strong correlation with the observed RTs for paMAPs and saMAPs (Figures 2G and 2H), and the RT distributions were not significantly different from the distribution of canonical proteome-derived peptides (F-test), supporting their correct identification. Additionally, 11 paMAPs and five saMAPs used in immunogenicity assays reported below were also validated by comparing their tandem mass spectra with that of synthetic peptides (see data and code availability).

paMAPs derived from iPSCs are shared across cancer types

We next evaluated whether cancer cells present aberrant expression of paMAPs. To this end, we queried the MCS of paMAPs in the RNA-seq of cancer samples from the 33 cancer types included in The Cancer Genome Atlas (TCGA) and from previous proteogenomic studies of acute myeloid leukemia (AML) (Ehx et al., 2021) and ovarian high-grade serous carcinoma (HGSC) (Zhao et al., 2020) from our laboratory. We found

that 40 of the 46 paMAPs were expressed in at least 10% of the samples in up to 14 cancer types from TCGA, and nine of these paMAPs were shared by more than 50% of the samples in one or two cancer types (Figure 3). AML and HGSC samples studied in our lab also shared expression of 13 and 19 paMAPs, respectively (Figure 3). Of all paMAPs, eight were previously reported in the context of cancer immunotherapy and are shared by many TCGA cancer types (Figure 3, bold). Of the reported paMAPs, six derive from in-frame exonic translation (Duffour et al., 1999; Huang et al., 1991; Jia et al., 2010; Schuster et al., 2017), whereas two derive from a 3′UTR and an intron, and they have been independently identified in HGSC and AML samples from our lab (Ehx et al., 2021; Zhao et al., 2020) (Table S3). Hence, commonly expressed paMAP-coding sequences have a high potential to generate shared TSAs between patients across multiple cancer types.

Human tumors with stemness properties acquire paMAP expression

Consistent with previous reports showing that stemness varies across TCGA samples (Malta et al., 2018; Miranda et al., 2019; Smith et al., 2018), we found that at the RNA level, the number of expressed paMAPs ranged from 0 to 19 per sample (excluding testis cancer, Figure 4A; Table S5). However, the inclusion of purity estimates for 21 solid cancers from Aran and colleagues (Aran et al., 2015) revealed that the number of expressed paMAPs increased with sample purity. From this correlation, we infer that paMAPs are specifically expressed in cancer cells and that a low purity may underestimate their number in some cancer samples (Figure 4B; Table S5). However, the situation was different with saMAPs, which had similar counts in high- and low-purity samples. The latter likely reflected the expression of saMAPs in healthy or pre-cancerous adult stem cells or healthy proliferating cells (Figure S3A).

Two additional observations could be made by comparing the expression of paMAPs and saMAPs: (1) saMAPs were more widely expressed in cancer samples, with 86% of TCGA samples expressing at least one saMAP, and only 60% of samples expressing one paMAP or more (Figures 4A, S3B, and S3C), and (2) paMAP expression co-occurred with saMAPs, but not all high-stemness samples were paMAP-positive, even when accounting for sample purity (Figures 4D, S3D, and S3E). This suggests that paMAP expression appears with cancer progression and further dedifferentiation from stemness to a pluripotency-associated program. Accordingly, we found a positive association between the accumulation of mutations and the degree of undifferentiation. The non-synonymous mutation load followed the following hierarchy: cancer samples with no paMAP/saMAP expression < samples with saMAP expression only < paMAP (and saMAP)-expressing samples (Figures 4D and S3D). We then performed a differential gene expression analysis (controlling for purity and tumor type) between samples with paMAPs versus those with many saMAPs but no paMAPs. This analysis revealed that tumors with paMAPs overexpressed genes involved in cancer cell migration and invasiveness, namely *CDH12* (cadherin-12) and *HIF3A* (hypoxia-inducible factor 3 alpha subunit) (Ma et al., 2016; Wang et al., 2011; Zhou et al., 2018) (Figure 4E; Table S6). High paMAP-expressing samples also showed overexpression of embryonic antigens and CGAs in addition to paMAP-source genes, including *TPTE* (transmembrane phosphatase with tensin homology) and *MAGEA3* (melanoma-associated antigen A3). Notably, RNA vaccines targeting *TPTE* and *MAGEA3* were reported to induce durable immune responses in patients with unresectable melanoma (Sahin et al., 2020).

Supporting the notion that pluripotency is associated with cancer progression and invasiveness (Ben-Porath et al., 2008), paMAPs were preferentially expressed in cancer subtypes with poor prognosis or advanced stages (Figure 4F). In breast cancer (BRCA), the basal subtype had the highest number of paMAPs, followed by the HER2 subtype and the luminal A and luminal B subtypes. Glioblastoma (GBM, G4) samples also showed a significantly higher number of paMAPs compared with low-grade gliomas (LGG, G2, and G3), while stage III and IV endometrial cancers (UCEC) expressed more paMAPs than early-stage tumors. Similarly, metastatic melanoma samples expressed more paMAPs compared with primary lesions (Figure 4F). Nonetheless, we observed that cancers from all subtypes and stages could re-express paMAPs (Figures 4F and S3F), indicating that immune targeting of paMAPs could be envisioned at any tumor stage.

Shared epigenetic and signaling events associate with paMAP and saMAP expression across cancers

To elucidate the mechanisms regulating paMAP expression, we first evaluated its potential correlation with epigenetic and focal DNA copy number aberrations. This analysis showed that, within and across cancer types, the DNA methylation status at source-gene promoter regions negatively correlated with the MCS expression for most paMAPs (Figures 5A and S4A). In contrast, only a small fraction of paMAPs was associated with an increase in DNA copy number (Figures 5B and S4B). We observed a similar trend for saMAPs (Figures S4C and S4D), indicating that, irrespective of the tissue of origin, epigenetics is an important mechanism in acquiring stemness and pluripotency features in cancer.

Next, the hallmark gene set collection from the Molecular Signature Database (MSigDB) (Liberzon et al., 2015) was used to explore other events that may drive paMAP and saMAP expression (interrogated together given their co-occurrence) and to understand their effect on global patterns. First, these data revealed that proliferation-related gene sets such as mitotic spindle assembly, G2/M checkpoint, E2F, and MYC signaling, were the most enriched and shared across all paMAP and saMAP-expressing samples (Figures 5C and S4E). In accordance with the high demands of proliferation and functionality, DNA repair, protein synthesis, and the unfolded protein response programs were also robustly upregulated. A second PSC pattern correlated with paMAP and saMAP expression within cancers: metabolic rewiring to glycolysis and downregulation of pathways active in differentiated tissues (i.e., oxidative phosphorylation, bile acid, and fatty acid metabolism) (Aran et al., 2017a; Kroemer and Pouyssegur, 2008; Zhang et al., 2012).

Notably, two signaling pathways were highly enriched in paMAP and saMAP-expressing samples: MYC signaling and the phosphoinositide 3-kinase (PI3K)/protein kinase B (AKT)/mammalian

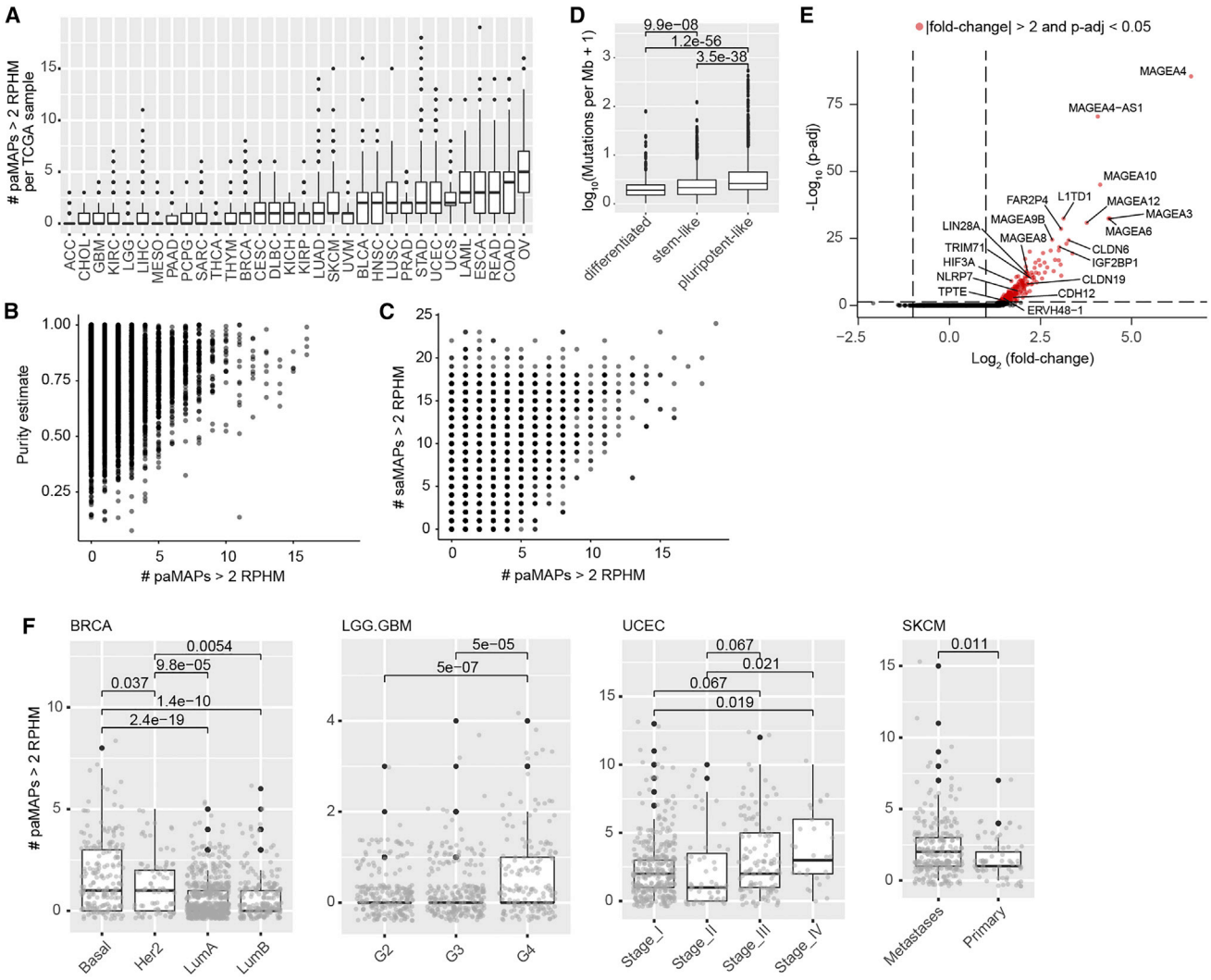


Figure 4. Human tumors with stemness properties acquire paMAP expression

(A) Boxplot showing the number of paMAPs expressed per TCGA sample within cancer types, in the increasing order of the median.

(B) Scatterplot displaying the number of paMAPs expressed across TCGA cancers (n = 21 cancer types) according to the estimated tumor purity from Aran et al. (2015).

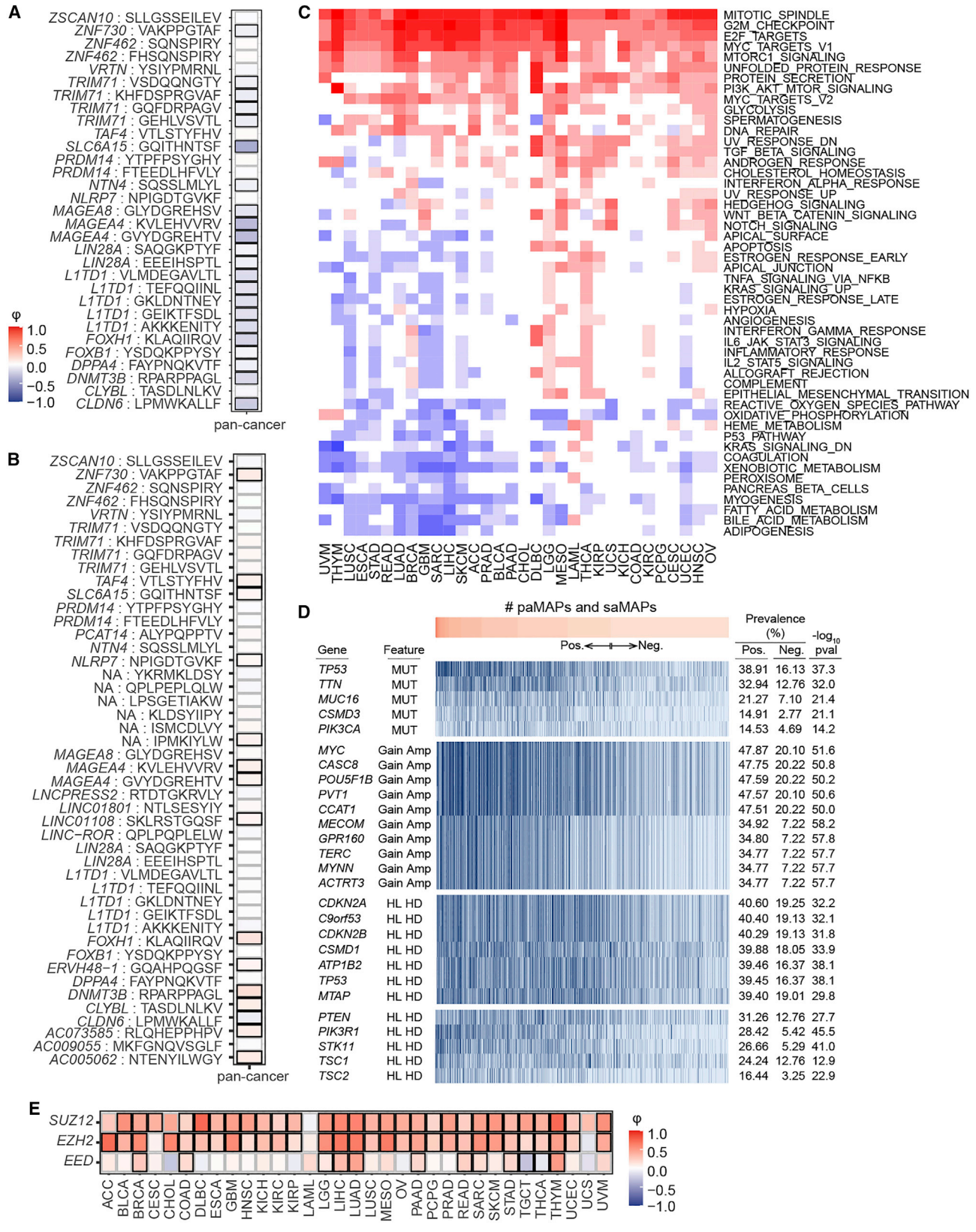
(C) Scatterplot displaying the number of paMAPs versus the number of saMAPs expressed across TCGA cancers (n = 32 cancer types, excluding TGCT).

(D) Mutation load [log10(Non-synonymous mutations per mega base pairs +1)] in TCGA samples with no paMAP/saMAP expression (differentiated), with saMAP but no paMAP expression (stem-like), or with paMAP expression (pluripotent-like). Only samples with estimated purity >0.75 (Aran et al., 2015) were included (Figure S3D).

(E) Volcano plot showing genes differentially expressed (red dots) between samples with high paMAP expression (>4 paMAPs, n = 775) and high saMAP expression (>4 saMAPs and 0 paMAPs, n = 1,270). (F) Boxplots showing the number of paMAPs at different molecular subtypes, grades, or stages of breast (BRCA), glioma (LGG.GBM), and endometrial (UCEC) cancers, respectively, and within primary and metastatic melanoma (SKCM) samples. Each gray dot represents one sample.

(A, D, and F) All boxplots show the median and interquartile range (IQR), and whiskers extend to the largest value no further than 1.5 * IQR from the box hinges. Pairwise comparisons were performed using the unpaired two-tailed Wilcoxon test, and p values were corrected for multiple comparisons using the Benjamini-Hochberg method. See also Figure S3.

target of rapamycin (mTOR) pathway. These two pathways regulate cellular growth, protein synthesis, and metabolism to promote the survival and dissemination of cancer cells (Heiden et al., 2009; Janku et al., 2018) (Figure 5C). Though activating alterations in these pathways are the most frequent among TCGA samples (Sanchez-Vega et al., 2018), their prevalence in paMAP- and saMAP-expressing samples was significantly higher than in non-expressing samples. This was particularly true for mutations in *PIK3CA*, deletions in PI3K/AKT signaling antagonists *PTEN*, *PIK3R1,* and *STK11* (also called *LKB1*), and *MYC* amplification (Figure 5D). The activating *PIK3CA*^{H1047R} mutation induces multipotency by dedifferentiation in mouse models of breast, lung, and



(legend on next page)

colorectal cancer, consistent with the ability of PI3K/AKT activation to increase the expression of pluripotency genes and self-renewal in human PSCs (Madsen, 2020). Moreover, the link between metabolism and epigenetics under the regulation of MYC and PI3K/AKT/mTOR pathways has been described in both PSCs and cancer (Dai et al., 2020; Fagnocchi and Zippo, 2017; Madsen, 2020; Zhang et al., 2012). In addition to promoting cell growth and proliferation, *MYC* overexpression induces transcriptional repression of lineage-specifying transcription factors. This is achieved via upregulation and recruitment of chromatin modifiers like the Polycomb repressive complex 2 (PRC2), which promotes epigenetic reprogramming toward a stem-like state, tumorigenesis, and self-renewal (Dardenne et al., 2016; Das et al., 2019; Fagnocchi and Zippo, 2017; Poli et al., 2018; Stine et al., 2015; Zhang et al., 2019). Accordingly, we found that the number of paMAPs and saMAPs strongly correlated with the expression of PRC2 components (*SUZ12*, *EZH2*, *EED*) within cancers (Figure 5E).

Other core signaling pathways that cross-talk to promote oncogenic dedifferentiation, namely Hedgehog, transforming growth factor (TGF)-β, WNT/β-catenin, and NOTCH signaling (Madsen, 2020; Malta et al., 2018; Pelullo et al., 2019), were also enriched in high paMAP and saMAP-expressing samples (Figure 5C), whereas tumor suppressors had a high prevalence of deletions (Figure 5D). Among them, we found that the pluripotency inhibitor *TP53* (Lin and Lin, 2017; Merkle et al., 2017) had a strong negative enrichment signature and the highest prevalence of mutations within and across cancers with paMAP and saMAP expression (Figures 5C, 5D, and S4F). Altogether, these data indicate that common genomic and signaling aberrations cooperate to induce a unifying PSC-like program across cancers.

Immunogenicity of paMAPs and saMAPs

Given that paMAPs are appealing targets for immunotherapy, we tested their immunogenicity using *in vitro* T cell assays with peripheral blood mononuclear cells (PBMCs) from healthy donors. paMAPs were prioritized based on four criteria: (1) the immunogenicity score predicted by Repitope (Table S3), a machine learning algorithm that uses public T cell receptor (TCR) databases to predict a probability of T cell response (Ogishi and Yotsuyanagi, 2019), (2) the HLA allotype presenting the paMAP (HLA-A*02:01 or HLA-B*53:01 shared by our iPSCs and PBMCs donors), (3) expression in minimum 10% of the samples in at least one TCGA cancer type (Figure 3), and (4) previously unreported MAP. We tested CD8 T cell response against 11 paMAPs using peptide-HLA tetramer staining and/or more sensitive functional expansion of specific T cells (FEST) assays (Danilova et al., 2018). In FEST assays, TCR sequencing is performed on T cells stimulated or not with synthetic paMAPs. T cells were stimulated *in vitro* with autologous T cell-depleted PBMCs pulsed with individual or pooled (n = 5 or 6) paMAPs (Table S7).

Significant T cell responses against canonical paMAPs SLLGSSEILEV and KLAQIIRQV were detected by tetramer staining in one out of four donors (D13 and D14, respectively) (Figure 6A). The FEST assay also revealed the immunogenicity of four paMAPs in D12, for which no specific T cell expansion was detected by tetramer staining (Figure 6B). Following stimulation with a single peptide, we identified a specific expansion of two to four TCRβ clonotypes against canonical paMAPs SLLGSSEILEV and LPMWKALLF, and the non-canonical paMAPs VTLSTYFHV and ALYPQPPTV (Figure 6B; Table S7). Two additional TCRβ clonotypes were expanded following stimulation with a pool of HLA-A*02:01-binding paMAP (Figure 6B; Table S7). Additionally, we also evaluated the immunogenicity of five saMAPs. We found that, despite its expression in lymphoid precursor cells (Figure 2B), the canonical saMAP FLLPGVLLSEA, deriving from the UDP glycosyltransferase family 3 member A2 (UGT3A2) gene, was immunogenic in one donor by tetramer staining (Figures 6A and 6B).

The stochasticity of paMAP and saMAP detection can be explained by low frequencies of antigen-specific (i.e., tetramer$^+$) CD8$^+$ T cells in donor PBMCs before *in vitro* stimulation, with a median of $\leq$0.75 paMAP-specific cells per 10^6 CD8 T cells (Figure 6C). Indeed, positive control peptides with high specific T cell frequencies (MelanA$_{27}$ and NS3$_{1073}$) were consistently immunogenic by tetramer staining post-stimulation. In contrast, the positive control epitope Gag$_{77}$ (derived from the human immunodeficiency virus), which had specific T cell frequencies similar to our paMAPs and saMAPs, was not immunogenic in any of the three PBMC donors tested (Figure 6C). The low frequencies of antigen-specific T cells detected before *in vitro* priming suggest that they were in the naive (rather than the memory) T cell compartment.

The specificity and functional quality of the paMAP-directed T cell response was also investigated using cytotoxicity assays. We chose to study the T cell reactivity against the AML-associated paMAP VTLSTYFHV (VTLS) (Figure 3), as the HLA-A*02:01-positive AML cell lines THP-1 and OCI-AML3 were

Figure 5. Shared epigenetic and signaling events associate with paMAP and saMAP expression across cancers

(A) Heatmap showing the Spearman correlation between the paMAP expression (RPHM) and the methylation β-value at the promoter region of the respective source gene across cancers. All available data for the 450K methylation dataset were included. Boxed: p-adj < 10^{-4} (Benjamini-Hochberg).

(B) Heatmap showing the Spearman correlation between the paMAP expression (RPHM) and the focal DNA copy number. Source gene symbols are added for reference; NA, no annotated source gene; all available data were included. Boxed: p-adj < 10^{-4} (Benjamini-Hochberg).

(C) Within-cancer Spearman correlation between the number of paMAPs and saMAPs expressed per sample (>2 RPHM) and the ssGSEA score for hallmark gene sets from MSigDB; only significant correlations are presented (p-adj < 0.05, Benjamini-Hochberg), otherwise the cell is white.

(D) Prevalence of the indicated genomic feature in cancer samples that express paMAPs and saMAPs (>2 RPHM) versus those with no expression. The top three blocks were selected based on the highest prevalence in paMAP- and saMAP-positive samples or lowest p values. In contrast, features in the last block are PI3K/AKT signaling antagonists. p value calculated based on the difference in prevalence in the two groups of samples using the chi-square test. Features MUT, somatic mutation; Gain, single-copy gain; Amp, amplification; HL, heterozygous loss; HD, homozygous deletion.

(E) Heatmap showing the Spearman correlation between the number of paMAPs and saMAPs expressed and the expression of PRC2 components within cancer types. Boxed: correlations with p-adj < 0.05 (Benjamini-Hochberg).

See also Figure S4.

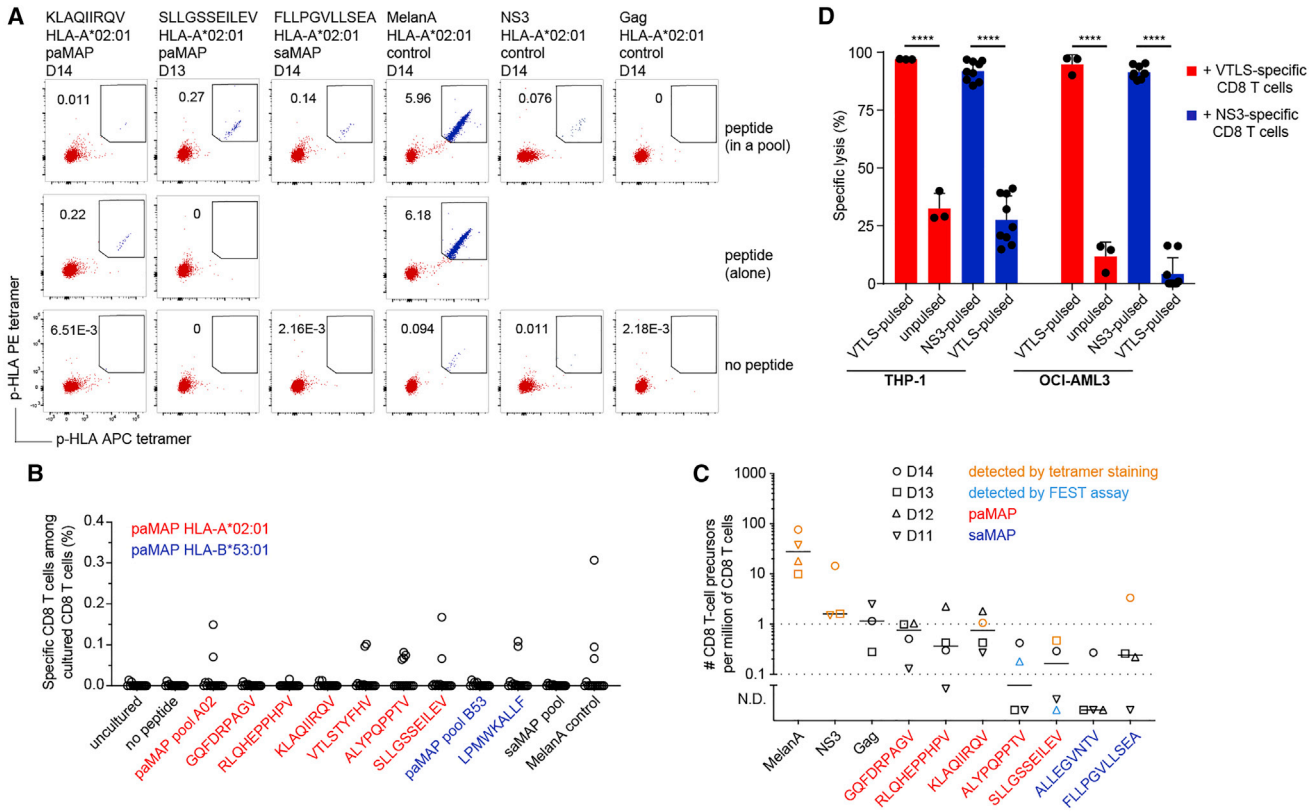


Figure 6. Immunogenicity of paMAPs and saMAPs

(A) Flow cytometry plots of peptide-HLA tetramer staining of specific CD8$^+$ T cells following *in vitro* stimulation, with numbers indicating the frequency of total CD8$^+$ T cells.

(B) FEST assay showing the expansion of specific T cell clonotypes following *in vitro* stimulation with the indicated peptides alone or in a pool compared with the control without peptides (Table S7).

(C) Number of specific cells per million of CD8$^+$ T cells in the pre-immune repertoire for each donor (D11-14), quantified using tetramer staining after immunomagnetic pre-enrichment. N.D., not detected.

(D) Specific lysis (%) of peptide-pulsed (VTLS or NS3) or unpulsed (DMSO) AML cells after overnight co-incubation with VTLS-specific or NS3-specific CD8$^+$ T cells (from D25). Data points pooled from two independent experiments per cell line; data are shown as mean ± SD; ****, $p < 0.0001$ (two-sided unpaired Student's t test). No significant differences were found between specific lysis of target cells that were unpulsed versus pulsed with an irrelevant peptide.

See also Figure S5.

available in our lab for use as target cell models. Thus, the specific lysis of VTLS-pulsed or unpulsed AML cells was quantified after overnight incubation with VTLS-specific CD8$^+$ T cells isolated and expanded from a healthy HLA-A*02:01 donor (D25). We found that VTLS-specific CD8$^+$ T cells efficiently killed AML cells pulsed with VTLS. However, the lysis of unpulsed AML cells was similar to the irrelevant NS3-specific T cell control (Figure 6D). Hence, the potent cytotoxic activity of these activated CD8$^+$ T cells was exquisitely VTLS specific. The low VTLS-specific CD8$^+$ T cell number available prevented us from determining their functional avidity using peptide titration.

In summary, five previously unreported paMAPs (3/4 canonical and 2/7 non-canonical) and 1/5 canonical saMAPs were immunogenic in at least one T cell assay (Figure S5A). Their MCS was significantly more expressed in cancer samples than the corresponding normal tissues (Figures 2A, 3, S5B, and S5C). These paMAPs had different origins: (1) ZSCAN10, FOXH1, and TAF4, which are transcription factors involved in pluripotency maintenance and embryonic development and are known to promote self-renewal in cancer (Kazantseva et al., 2016; Loizou et al., 2019; Wang et al., 2007, 2019; Yu et al., 2009), (2) the oncofetal antigen CLDN6 (Reinhard et al., 2020), and (3) the prostate-cancer associated, ''exonized'' transposable element, PCAT14 (Babarinde et al., 2020; Prensner et al., 2011) (Figure S5A). In addition, two of our paMAPs derived from MAGEA4 (GVYDGREHTV and KVLEHVVRV) and overexpressed in cancer samples (Figure S5B) were immunogenic in previous studies (Duffour et al., 1999; Jia et al., 2010), altogether reinforcing the therapeutic potential of paMAPs.

paMAP and saMAP expression correlates with immune evasion

Having determined that paMAPs and saMAPs could be immunogenic, we evaluated the effect of their HLA presentation on patient survival for the TCGA patient cohorts using a Cox regression analysis. We inferred that a paMAP or an saMAP was

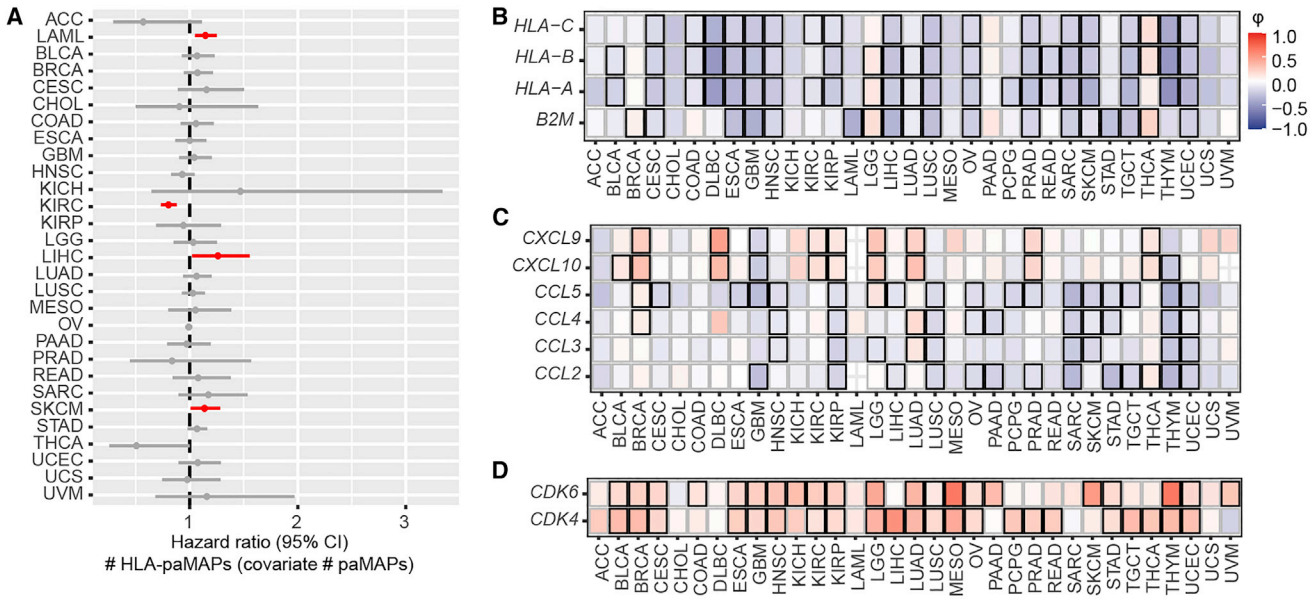


Figure 7. paMAP and saMAP expression correlates with immune evasion

(A) Hazard ratio (risk of death) (±95% CI) for the association between the risk of death and the number of paMAPs with predicted presentation (# HLA-paMAPs), taking the number of paMAPs expressed (>0 RPHM) as a covariate. Red dots and whiskers, p value < 0.05 (Cox proportional hazards model). Patients with more than one sample were excluded from the analysis.

(B–D) Spearman correlation between the number of paMAPs and saMAPs expressed and the expression of MHC I-related genes (B), immune recruitment chemokine-encoding genes (C), or *CDK4/6* genes (D) within cancer types. Boxed: correlations with p-adj < 0.05 (Benjamini-Hochberg).

See also Figure S6.

presented in a given sample if two conditions were met: expression of the MCS and presence of an HLA allotype that can bind and present the MAP according to the NetMHCpan-4.0 software (Jurtz et al., 2017) (Table S8). Hence, each MAP was assumed to be presented only in a fraction of the samples (bearing a relevant HLA allotype) in which its MCS was expressed. The mere expression of saMAPs correlated in many cancer types with a shortened survival (Figure S6B). However, the presence of a relevant HLA allotype had a positive effect in KIRC and thyroid carcinoma (THCA), a negative impact in AML (LAML), and no effect in all other cancer types (Figure S6B). The same analysis performed using paMAPs showed similar results. Presentation of paMAPs had an HLA-dependent positive impact on survival in renal clear cell carcinoma (KIRC), but either no or a minimally negative impact in other cancer types (Figure 7A). The negative association between the sole expression of paMAPs and patient survival was less pronounced than for saMAPs, likely due to the presence of high-stemness cancers with poor survival in the low pluripotency group (Figures 4C, S6A, and S6B). Therefore, considering that inter-group differences were minimal, we concluded that the presentation of paMAPs and saMAPs did not confer a clear survival advantage in patients from the TCGA cohorts, which prompted us to investigate possible immune evasion mechanisms associated with their expression.

Consistent with other studies (Malta et al., 2018; Miranda et al., 2019), we found that, within most cancer types, the immune infiltration signature, derived herein using xCell (Aran et al., 2017b), was decreased in samples with high paMAP and saMAP counts or their source gene enrichment (Figure S6C). In considering cell-autonomous mechanisms that could mediate escape from T cell recognition, we first evaluated the expression of MHC I molecules, whose downregulation leads to evasion from immune detection (Agudo et al., 2018; Castro et al., 2019). We found a significant negative correlation between the number of paMAPs and saMAPs expressed per sample and the expression of genes involved in surface HLA expression (Figure 7B). In addition, we found a negative association with the expression of genes encoding chemokines that recruit immune cells, including the $BATF3^{+}$ DCs, which are essential for cross-presenting tumor antigens (Spranger et al., 2017) (Figure 7C). Furthermore, pathways strongly associated with paMAP and saMAP expression (Figures 5C and 5D), namely the activation of MYC and WNT-β-catenin signaling, the loss of function of P53, and the loss of PI3K pathway inhibitors, are known to inhibit T cell activation and infiltration (Spranger and Gajewski, 2018). Accordingly, the number of paMAPs and saMAPs showed a strong positive correlation with the expression of *CDK4* and *CDK6* in nearly all TCGA cancer types (Figure 7D). Importantly, WNT-β-catenin and TGF-β signaling, and CDK4/6, regulate cancer cell programs that promote T cell exclusion and immune evasion in BRCA and melanoma (Bagati et al., 2021; Goel et al., 2017; Jerby-Arnon et al., 2018; Spranger et al., 2015). Lastly, our stemness signature showed a positive pan-cancer correlation with the immunosuppressive genes *PVR* (*CD155*) and *CD276* (*B7-H3*) (Figure S6D). Collectively, these data suggest that cancers with high numbers of paMAPs and saMAPs employ multiple resistance mechanisms as a shield from immune detection and destruction.

In conclusion, our data show that an increased immune evasion and repression program may hinder immune recognition of paMAPs and saMAPs, which are found in poorly differentiated advanced cancers. Furthermore, they suggest that WNT-β-catenin, TGF-β, and CDK4/6 inhibitors, currently used to treat several types of cancer (Álvarez-Fernández and Malumbres, 2020; Hinze et al., 2020; Huang et al., 2021), could potentially enhance immune recognition of tumor cells expressing paMAPs and saMAPs.

DISCUSSION

Despite convincing evidence that some oncofetal CD8 T cell epitopes are shared by tumor cells and embryonic or induced PSCs (Kooreman et al., 2018), the molecular nature of these epitopes has remained largely elusive. In the present study, we performed in-depth MS analyses to investigate the global repertoire of MAPs present on iPSCs. We identified a set of pluripotency- or stemness-associated antigens that are non-mutated and shared between patients across multiple cancer types. Indeed, by studying the immunopeptidome of iPSCs, we found 72 MAPs whose RNA was minimally or not detected in normal differentiated tissues but was expressed in normal adult stem and progenitor cells (saMAPs) or expressed in human PSCs from different sources (paMAPs). In TCGA tumors, paMAPs and saMAPs were expressed in up to 14 and 32 cancer types, respectively. In the present study, we inferred the presentation of paMAPs on cancer cells from transcriptomic data. However, the independent identification of eight of our paMAPs on cancer cells using peptide-based approaches (Duffour et al., 1999; Huang et al., 1991; Jia et al., 2010; Schuster et al., 2017; Zhao et al., 2020) supports the validity of our inferences. Using PSC lines to identify paMAPs is advantageous since low purity, sample heterogeneity, and decreased MHC expression on cancer cells acquiring pluripotency features may limit the detection of such MAPs from primary tumor samples using shotgun MS. Nevertheless, sensitive targeted MS analyses of primary tumor samples that express paMAPs at the RNA level are warranted to confirm their pan-cancer presentation.

Growing evidence shows that a high stemness index correlates with tumor aggressiveness and resistance to therapy (Ben-Porath et al., 2008; Malta et al., 2018; Miranda et al., 2019; Van Rhenen et al., 2005; Smith et al., 2018). It is postulated that the immune pressure selects tumor clones capable of immune escape, a characteristic shared with normal adult and PSCs (Åkesson et al., 2009; Li et al., 2004; Miranda et al., 2019; Di Nicola et al., 2002). Our data highlighted that the presence of paMAPs and saMAPs is associated with a pan-cancer rewiring of the epigenome, metabolome, and signaling networks that are instrumental in immune escape. Salient and prevalent changes across cancers were an enrichment in the MYC and PI3K/AKT signaling pathways and a strong downregulation in the P53 signature. The oncogenic functions of these regulators include cellular growth, cell cycle progression, and suppression of apoptosis. Thus, alterations in MYC, PI3K, and P53 oncogenic pathways (Figures 5C, 5D, S4E, and S4F) may directly explain the expression of paMAP- and saMAP-coding transcripts in cancer. Indeed, downstream metabolic changes impinge on the epigenome dynamics and determine cell fate (Dai et al., 2020), and MYC is directly involved in modulating core pluripotency and lineage-differentiation factors (Fagnocchi and Zippo, 2017). Furthermore, the expression of paMAPs and saMAPs correlated with hypomethylation of their source genes and the overexpression of PRC2 components in most TCGA cancer types.

The immunopeptidome projects at the cell surface a representation of biochemical networks and metabolic events regulated at multiple levels inside the cell (Caron et al., 2011). Accordingly, paMAPs derived from genes and annotated lncRNAs with essential roles in PSC identity. Of note, 48% of the paMAPs derived from supposedly non-coding genomic regions, primarily from the transcription and translation of lncRNAs, intergenic regions, and introns overlapping EREs, which are primary players in reprogramming to pluripotency (Klawitter et al., 2016; Lu et al., 2014). This suggests that the onco-exaptation of EREs in tumorigenesis (Ishak and De Carvalho, 2020) may generate actionable targets for immune targeting of cancer stem cells.

Two key findings support a model where paMAP expression occurs as a co-opted program during cancer progression. This program is marked by a gradual transition of cancer cells from a differentiated state to a stem cell and eventually a pluripotent cell state: (1) paMAP-expressing samples were a subset of saMAP-expressing samples, and (2) dedifferentiation was associated with an increase in tumor mutational burden. Nevertheless, further analyses will be required to confirm this model and to evaluate the extent to which bona fide cancer stem cells, a subpopulation of cells responsible for sustaining tumor growth (Batlle and Clevers, 2017), may contribute to the paMAP and saMAP expression across cancers.

Notably, the expression of paMAPs and saMAPs correlated with activation of the most potent immunosuppressive pathways in cancer: CDK4/6, TGF-β, and WNT-β-catenin. Collectively these pathways transform the tumor environment in an immune desert. They repress MHC expression, inhibit the recruitment of $BATF3^{+}$ DCs (required for cross-presentation) and CD8 T cells, and favor the accumulation of immune suppressive cell subsets (Bagati et al., 2021; Goel et al., 2017; Spranger et al., 2015). This strongly suggests that, in the absence of therapeutic intervention, paMAPs are shielded from the immune system. This would explain that, despite their potential immunogenicity, the presentation of paMAPs did not spontaneously trigger protective immune responses in the TCGA cohorts. An interesting corollary is that paMAP-based vaccines may induce de novo anti-tumor T cell responses that could be potentiated by selective inhibitors of CDK4/6, TGF-β, and WNT-β-catenin. Other immune checkpoints inhibitors and drugs inducing immunogenic cancer cell death could also synergize with paMAP-targeted immunotherapies (Anderson et al., 2017; Humeau et al., 2020; Petroni et al., 2021). In addition, inhibitors of MYC and PI3K, whose activation correlates with the expression of paMAPs and saMAPs, could further synergize with paMAP-targeted immunotherapy.

In conclusion, the present study provides a molecular framework for understanding several reports suggesting commonalities between the immunopeptidome of PSCs and cancer cells and provides further insights into the molecular underpinnings of stemness in cancer. Furthermore, at least some of our

paMAPs were immunogenic and shared among patients across multiple cancer types, making them attractive targets for immunotherapy of poorly differentiated cancers.

Limitations of study

The set of paMAPs and saMAPs identified herein is limited to those presented by the HLA allotypes of our iPSCs. Hence, immunopeptidomic analyses of additional iPSCs derived from somatic cells with different HLA allotypes are warranted to obtain a global overview of the paMAP and saMAP repertoire. Also, considering that (1) our transcriptomic analyses of TCGA samples were based on bulk RNA-seq data and that (2) stem/pluripotent cells may represent a small fraction of tumor cells, we have probably significantly underestimated the frequency of tumors expressing paMAPs and saMAPs.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Human iPSC culture
  - OCI-AML3 and THP-1 cell culture
  - B lymphoblastoid cell lines culture
- METHOD DETAILS
  - Flow cytometry analysis of iPSCs
  - Trilineage differentiation assay
  - RNA extraction and sequencing
  - Database generation for shotgun mass spectrometry analyses
  - Isolation of MHC-associated peptides
  - Mass spectrometry analyses
  - paMAP and saMAP validation with synthetic peptides
  - Bioinformatic analyses
  - TCGA analyses
  - Immunogenicity assays
- QUANTIFICATION AND STATISTICAL ANALYSIS

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.celrep.2022.111241.

ACKNOWLEDGMENTS

We wish to thank Qingchuan Zhao, Jean-David Larouche, Assya Trofimov, Maria Virginia Ruiz Cuevas, Eralda Kina, and Krystel Vincent for valuable discussions and suggestions and help with preliminary exploratory analyses. We are grateful to Raphaëlle Lambert, Jennifer Huber, Sarah Boissel, and the IRIC genomics core facility staff for technical assistance with RNA-seq and Patrick Gendron from the IRIC bioinformatics platform for help with RNA-seq and TCGA data analyses. We also thank Dr. Hannah Carter (UC San Diego) for providing the HLA typing information for TCGA patients and the NIH Tetramer Core Facility for providing the tetramers used in this study.

Finally, we thank the TCGA and the GTEx Consortium for access to data that enabled this study. This study was supported by grants from the Canadian Cancer Society (705604) (to C.P. and P.T.), the Canadian Institutes of Health Research (FDN 148400) (to C.P.), The Oncopole (EMC^2 Grant) (to C.P. and P.T.), and the Fonds Vaccins Thérapeutiques Contre le Cancer (to C.P.). A.A. is supported by doctoral studentships from the IRIC, The Cole Foundation, and the Fonds de Recherche du Québec - Santé. For the Bhatia program, this work was supported by the Canada Chair Program (Tier 1) for Human Stem Cell Biology and DeGroote Chair in Stem Cell Biology to M.B. and Canadian Institutes of Health Research (CIHR) foundation grant FRN 159925 to M.B. approved by the Stem Cell Oversight Committee in Canada.

AUTHOR CONTRIBUTIONS

A.A., M.-P.H., and C.P. designed the study. A.A. performed the cell culture experiments, cytotoxicity assays, main bioinformatic analyses, and data interpretation and wrote the first manuscript draft. C.D. and J.L. performed mass spectrometry experiments. L.H. and C.T. performed *in vitro* immunogenicity assays. B.B. performed trilineage differentiation assays. M.-P.H. and D.K.C. assisted with preliminary iPSC experiments. N.N. provided assistance with AML cell culture. M.-P.H., G.E., J.-P.L., and M.C. contributed to bioinformatic analyses. M.-P.H., L.H., G.E., B.B., N.N., S.L., C.B., M.B., P.T., and C.P. contributed to the analysis and interpretation of data. All authors edited and approved the final manuscript.

DECLARATION OF INTERESTS

A.A., M.-P.H., P.T., and C.P. are named inventors on a patent application filed by Université de Montréal and covering antigens described in this article.

Received: August 2, 2021
Revised: June 20, 2022
Accepted: July 27, 2022
Published: August 16, 2022

REFERENCES

Agudo, J., Park, E.S., Rose, S.A., Alibo, E., Sweeney, R., Dhainaut, M., Kobayashi, K.S., Sachidanandam, R., Baccarini, A., Merad, M., and Brown, B.D. (2018). Quiescent tissue stem cells evade immune surveillance. Immunity *48*, 271–285.e5. https://doi.org/10.1016/j.immuni.2018.02.001.

Åkesson, E., Wolmer-Solberg, N., Cederarv, M., Falci, S., and Odeberg, J. (2009). Human neural stem cells and astrocytes, but not neurons, suppress an allogeneic lymphocyte response. Stem Cell Res. *2*, 56–67. https://doi.org/10.1016/j.scr.2008.06.002.

Álvarez-Fernández, M., and Malumbres, M. (2020). Mechanisms of sensitivity and resistance to CDK4/6 inhibition. Cancer Cell *37*, 514–529. https://doi.org/10.1016/j.ccell.2020.03.010.

Anderson, K.G., Stromnes, I.M., and Greenberg, P.D. (2017). Obstacles posed by the tumor microenvironment to T cell activity: a case for synergistic therapies. Cancer Cell *31*, 311–325. https://doi.org/10.1016/j.ccell.2017.02.008.

Apavaloaei, A., Hardy, M.P., Thibault, P., and Perreault, C. (2020). The origin and immune recognition of tumor-specific antigens. Cancers *12*, 2607. https://doi.org/10.3390/cancers12092607.

Aran, D., Sirota, M., and Butte, A.J. (2015). Systematic pan-cancer analysis of tumour purity. Nat. Commun. *6*, 8971–9012. https://doi.org/10.1038/ncomms9971.

Aran, D., Camarda, R., Odegaard, J., Paik, H., Oskotsky, B., Krings, G., Goga, A., Sirota, M., and Butte, A.J. (2017a). Comprehensive analysis of normal adjacent to tumor transcriptomes. Nat. Commun. *8*, 1077–1113. https://doi.org/10.1038/s41467-017-01027-z.

Aran, D., Hu, Z., and Butte, A.J. (2017b). xCell: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. *18*, 220–314. https://doi.org/10.1186/s13059-017-1349-1.

Babarinde, I.A., Ma, G., Li, Y., Deng, B., Luo, Z., Liu, H., Abdul, M.M., Ward, C., Chen, M., Fu, X., et al. (2020). Transposable element-gene splicing modulates the transcriptional landscape of human pluripotent stem cells. Preprint at bioRxiv. https://doi.org/10.1101/2020.07.26.220608.

Bagati, A., Kumar, S., Jiang, P., Pyrdol, J., Zou, A.E., Godicelj, A., Mathewson, N.D., Cartwright, A.N.R., Cejas, P., Brown, M., et al. (2021). Integrin αvβ6–TGFβ–SOX4 pathway drives immune evasion in triple-negative breast cancer. Cancer Cell *39*, 54–67.e9. https://doi.org/10.1016/j.ccell.2020.12.001.

Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. Nature *462*, 108–112. https://doi.org/10.1038/nature08460.

Bassani-Sternberg, M., Pletscher-Frankild, S., Jensen, L.J., and Mann, M. (2015). 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. https://doi.org/10.1074/mcp.M114.042812.

Batlle, E., and Clevers, H. (2017). Cancer stem cells revisited. Nat. Med. *23*, 1124–1134. https://doi.org/10.1038/nm.4409.

Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat. Genet. *40*, 499–507. https://doi.org/10.1038/ng.127.

Benhammadi, M., Mathé, J., Dumont-Lagacé, M., Kobayashi, K.S., Gaboury, L., Brochu, S., and Perreault, C. (2020). IFN-Λ enhances constitutive expression of MHC class I molecules on thymic epithelial cells. J. Immunol. *205*, 1268–1280. https://doi.org/10.4049/jimmunol.2000225.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics *30*, 2114–2120. https://doi.org/10.1093/bioinformatics/btu170.

Bouwmeester, R., Gabriels, R., Hulstaert, N., Martens, L., and Degroeve, S. (2021). DeepLC can predict retention times for peptides that carry as-yet unseen modifications. Nat. Methods *18*, 1363–1369. https://doi.org/10.1038/s41592-021-01301-5.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

Brewer, B.G., Mitchell, R.A., Harandi, A., and Eaton, J.W. (2009). Embryonic vaccines against cancer: an early history. Exp. Mol. Pathol. *86*, 192–197. https://doi.org/10.1016/j.yexmp.2008.12.002.

Caron, E., Vincent, K., Fortier, M.-H., Laverdure, J.-P., Bramoullé, A., Hardy, M.-P., Voisin, G., Roux, P.P., Lemieux, S., Thibault, P., and Perreault, C. (2011). The MHC I immunopeptidome conveys to the cell surface an integrative view of cellular regulation. Mol. Syst. Biol. *7*, 533–615. https://doi.org/10.1038/msb.2011.68.

Castro, A., Ozturk, K., Pyke, R.M., Xian, S., Zanetti, M., and Carter, H. (2019). Elevated neoantigen levels in tumors with somatic mutations in the HLA-A, HLA-B, HLA-C and B2M genes. BMC Med. Genom. *12*, 107–113. https://doi.org/10.1186/s12920-019-0544-1.

Chia, N.Y., Chan, Y.S., Feng, B., Lu, X., Orlov, Y.L., Moreau, D., Kumar, P., Yang, L., Jiang, J., Lau, M.S., et al. (2010). A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. Nature *468*, 316–320. https://doi.org/10.1038/nature09531.

Chong, C., Müller, M., Pak, H., Harnett, D., Huber, F., Grun, D., Leleu, M., Auger, A., Arnaud, M., Stevenson, B.J., et al. (2020). Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. Nat. Commun. *11*, 1293. https://doi.org/10.1038/s41467-020-14968-9.

Churko, J.M., Lee, J., Ameen, M., Gu, M., Venkatasubramanian, M., Diecke, S., Sallam, K., Im, H., Wang, G., Gold, J.D., et al. (2017). Transcriptomic and epigenomic differences in human induced pluripotent stem cells generated from six reprogramming methods. Nat. Biomed. Eng. *1*, 826–837. https://doi.org/10.1038/s41551-017-0141-6.

Colaprico, A., Silva, T.C., Olsen, C., Garofano, L., Cava, C., Garolini, D., Sabedot, T.S., Malta, T.M., Pagnotta, S.M., Castiglioni, I., et al. (2016). TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. Nucleic Acids Res. *44*, e71. https://doi.org/10.1093/nar/gkv1507.

Coulie, P.G., Van Den Eynde, B.J., Van Der Bruggen, P., and Boon, T. (2014). Tumour antigens recognized by T lymphocytes: at the core of cancer immunotherapy. Nat. Rev. Cancer *14*, 135–146. https://doi.org/10.1038/nrc3670.

Courcelles, M., Durette, C., Daouda, T., Laverdure, J.-P., Vincent, K., Lemieux, S., Perreault, C., and Thibault, P. (2020). MAPDP: a cloud-based computational platform for immunopeptidomics analyses. J. Proteome Res. *19*, 1873–1881. https://doi.org/10.1021/acs.jproteome.9b00859.

Dai, Z., Ramesh, V., and Locasale, J.W. (2020). The evolving metabolic landscape of chromatin biology and epigenetics. Nat. Rev. Genet. *21*, 737–753. https://doi.org/10.1038/s41576-020-0270-8.

Danilova, L., Anagnostou, V., Caushi, J.X., Sidhom, J.W., Guo, H., Chan, H.Y., Suri, P., Tam, A., Zhang, J., Asmar, M.E., et al. (2018). The mutation-associated neoantigen functional expansion of specific T cells (MANAFEST) assay: a sensitive platform for monitoring antitumor immunity. Cancer Immunol. Res. *6*, 888–899. https://doi.org/10.1158/2326-6066.CIR-18-0129.

Daouda, T., Perreault, C., and Lemieux, S. (2016). pyGeno: a Python package for precision medicine and proteogenomics. F1000Res. *5*, 381. https://doi.org/10.12688/F1000RESEARCH.8251.2.

Dardenne, E., Beltran, H., Benelli, M., Gayvert, K., Berger, A., Puca, L., Cyrta, J., Sboner, A., Noorzad, Z., MacDonald, T., et al. (2016). N-myc induces an EZH2-mediated transcriptional program driving neuroendocrine prostate cancer. Cancer Cell *30*, 563–577. https://doi.org/10.1016/j.ccell.2016.09.005.

Das, B., Pal, B., Bhuyan, R., Li, H., Sarma, A., Gayan, S., Talukdar, J., Sandhya, S., Bhuyan, S., Gogoi, G., et al. (2019). MYC regulates the HIF2a stemness pathway via nanog and Sox2 to maintain self-renewal in cancer stem cells versus non-stem cancer cells. Cancer Res. *79*, 4015–4025. https://doi.org/10.1158/0008-5472.CAN-18-2847.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics *29*, 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Duffour, M.T., Chaux, P., Lurquin, C., Cornelis, G., Boon, T., and Van Der Bruggen, P. (1999). A MAGE-A4 peptide presented by HLA-A2 is recognized by cytolytic T lymphocytes. Eur. J. Immunol. *29*, 3329–3337.

Ehx, G., and Perreault, C. (2019). Discovery and characterization of actionable tumor antigens. Genome Med. *11*, 29–32. https://doi.org/10.1186/s13073-019-0642-x.

Ehx, G., Larouche, J.-D., Durette, C., Laverdure, J.-P., Hesnard, L., Vincent, K., Hardy, M.-P., Thériault, C., Rulleau, C., Lanoix, J., et al. (2021). Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I- associated epitopes Article Atypical acute myeloid leukemia-specific transcripts generate shared and immunogenic MHC class-I-associated epitopes. Immunity *54*, 737–752.e10. https://doi.org/10.1016/j.immuni.2021.03.001.

Ellrott, K., Bailey, M.H., Saksena, G., Covington, K.R., Kandoth, C., Stewart, C., Hess, J., Ma, S., Chiotti, K.E., McLellan, M., et al. (2018). Scalable open science approach for mutation calling of tumor exomes using multiple genomic pipelines. Cell Syst. *6*, 271–281.e7. https://doi.org/10.1016/j.cels.2018.03.002.

Erhard, F., Dölken, L., Schilling, B., and Schlosser, A. (2020). Identification of the cryptic HLA-I immunopeptidome. Cancer Immunol. Res. *8*, 1018–1026. https://doi.org/10.1158/2326-6066.CIR-19-0886.

Fagnocchi, L., and Zippo, A. (2017). Multiple roles of MYC in integrating regulatory networks of pluripotent stem cells. Front. Cell Dev. Biol. *5*, 7–19. https://doi.org/10.3389/fcell.2017.00007.

Fort, A., Hashimoto, K., Yamada, D., Salimullah, M., Keya, C.A., Saxena, A., Bonetti, A., Voineagu, I., Bertin, N., Kratz, A., et al. (2014). Deep transcriptome profiling of mammalian stem cells supports a regulatory role for

retrotransposons in pluripotency maintenance. Nat. Genet. *46*, 558–566. https://doi.org/10.1038/ng.2965.

Friedli, M., Turelli, P., Kapopoulou, A., Rauwel, B., Castro-Díaz, N., Rowe, H.M., Ecco, G., Unzu, C., Planet, E., Lombardo, A., et al. (2014). Loss of transcriptional control over endogenous retroelements during reprogramming to pluripotency. Genome Res. *24*, 1251–1259. https://doi.org/10.1101/gr.172809.114.

Frisch, B.J., Hoffman, C.M., Latchney, S.E., LaMere, M.W., Myers, J., Ashton, J., Li, A.J., Saunders, J., 2nd, Palis, J., Perkins, A.S., et al. (2019). Aged marrow macrophages expand platelet-biased hematopoietic stem cells via Interleukin1B. JCI Insight *5*, e124213. https://doi.org/10.1172/jci.insight.124213.

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. Preprint at ArXiv. https://doi.org/10.48550/arXiv.1207.3907.

Goel, S., Decristo, M.J., Watt, A.C., Brinjones, H., Sceneay, J., Li, B.B., Khan, N., Ubellacker, J.M., Xie, S., Metzger-Filho, O., et al. (2017). CDK4/6 inhibition triggers anti-tumour immunity. Nature *548*, 471–475. https://doi.org/10.1038/nature23465.

Goldman, M.J., Craft, B., Hastie, M., Repečka, K., McDade, F., Kamath, A., Banerjee, A., Luo, Y., Rogers, D., Brooks, A.N., et al. (2020). Visualizing and interpreting cancer genomics data via the Xena platform. Nat. Biotechnol. *38*, 675–678. https://doi.org/10.1038/s41587-020-0546-8.

Granados, D.P., Sriranganadane, D., Daouda, T., Zieger, A., Laumont, C.M., Caron-Lizotte, O., Boucher, G., Hardy, M.-P., Gendron, P., Côté, C., et al. (2014). Impact of genomic polymorphisms on the repertoire of human MHC class I-associated peptides. Nat. Commun. *5*, 3600. https://doi.org/10.1038/ncomms4600.

Haen, S.P., Löffler, M.W., Rammensee, H.G., and Brossart, P. (2020). Towards new horizons: characterization, classification and implications of the tumour antigenic repertoire. Nat. Rev. Clin. Oncol. *17*, 595–610. https://doi.org/10.1038/s41571-020-0387-x.

Hänzelmann, S., Castelo, R., and Guinney, J. (2013). GSVA: gene set variation analysis for microarray and RNA-Seq data. BMC Bioinf. *14*, 7. https://doi.org/10.1186/1471-2105-14-7.

Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the warburg effect: the metabolic Requirements of cell proliferation. Science *324*, 1029–1033. https://doi.org/10.1126/science.1160809.

Hernandez, C., Wang, Z., Ramazanov, B., Tang, Y., Mehta, S., Dambrot, C., Lee, Y.W., Tessema, K., Kumar, I., Astudillo, M., et al. (2018). Dppa2/4 facilitate epigenetic Remodeling during reprogramming to pluripotency. Cell Stem Cell *23*, 396–411.e8. https://doi.org/10.1016/j.stem.2018.08.001.

Hinze, L., Labrosse, R., Degar, J., Han, T., Schatoff, E.M., Schreek, S., Karim, S., McGuckin, C., Sacher, J.R., Wagner, F., et al. (2020). Exploiting the therapeutic interaction of WNT pathway activation and asparaginase for colorectal cancer therapy. Cancer Discov. *10*, 1690–1705. https://doi.org/10.1158/2159-8290.CD-19-1472.

Hong, S.-H., Lee, J.-H., Lee, J.B., Ji, J., and Bhatia, M. (2011). ID1 and ID3 represent conserved negative regulators of human embryonic and induced pluripotent stem cell hematopoiesis. J. Cell Sci. *124*, 1445–1452. https://doi.org/10.1242/jcs.077511.

Huang, C.-Y., Chung, C.-L., Hu, T.-H., Chen, J.-J., Liu, P.-F., and Chen, C.-L. (2021). Recent progress in TGF-β inhibitors for cancer therapy. Biomed. Pharmacother. *134*, 111046. https://doi.org/10.1016/j.biopha.2020.111046.

Huang, L.-Q., Brasseur, F., Serrano, A., De Plaen, E., Bruggen, P. van der, Boon, T., and Pel, A.V. (1991). Cytolytic T lymphocytes recognize an antigen encoded by MAGE-A10 on a human melanoma. J. Immunol. *162*, 6849–6854. https://doi.org/10.1126/science.1840703.

Humeau, J., Sauvat, A., Cerrato, G., Xie, W., Loos, F., Iannantuoni, F., Bezu, L., Lévesque, S., Paillet, J., Pol, J., et al. (2020). Inhibition of transcription by dactinomycin reveals a new characteristic of immunogenic cell stress. EMBO Mol. Med. *12*, e11622. https://doi.org/10.15252/emmm.201911622.

Ishak, C.A., and De Carvalho, D.D. (2020). Reactivation of endogenous retroelements in cancer development and therapy. Annu. Rev. Cell Biol. *4*, 159–176. https://doi.org/10.1146/annurev-cancerbio-030419-033525.

Izadyar, F., Wong, J., Maki, C., Pacchiarotti, J., Ramos, T., Howerton, K., Yuen, C., Greilach, S., Zhao, H.H., Chow, M., et al. (2011). Identification and characterization of repopulating spermatogonial stem cells from the adult human testis. Hum. Reprod. *26*, 1296–1306. https://doi.org/10.1093/humrep/der026.

Janku, F., Yap, T.A., and Meric-Bernstam, F. (2018). Targeting the PI3K pathway in cancer: are we making headway? Nat. Rev. Clin. Oncol. *15*, 273–291. https://doi.org/10.1038/nrclinonc.2018.28.

Jerby-Arnon, L., Shah, P., Cuoco, M.S., Rodman, C., Su, M.J., Melms, J.C., Leeson, R., Kanodia, A., Mei, S., Lin, J.R., et al. (2018). A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. Cell *175*, 984–997.e24. https://doi.org/10.1016/j.cell.2018.09.006.

Jia, Z.C., Ni, B., Huang, Z.-M., Tian, Y., Tang, J., Wang, J.-X., Fu, X.-L., and Wu, Y.-Z. (2010). Identification of two novel HLA-A *0201-restricted CTL epitopes derived from MAGE-A4. Clin. Dev. Immunol. *2010*, 567594. https://doi.org/10.1155/2010/567594.

Jurtz, V., Paul, S., Andreatta, M., Marcatili, P., Peters, B., and Nielsen, M. (2017). NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. J. Immunol. *199*, 3360–3368. https://doi.org/10.4049/jimmunol.1700893.

Kazantseva, J., Sadam, H., Neuman, T., and Palm, K. (2016). Targeted alternative splicing of TAF4: a new strategy for cell reprogramming. Sci. Rep. *6*, 30852–30911. https://doi.org/10.1038/srep30852.

Kelley, D., and Rinn, J. (2012). Transposable elements reveal a stem cell-specific class of long noncoding RNAs. Genome Biol. *13*, R107. https://doi.org/10.1186/gb-2012-13-11-r107.

Kim, J.Y., Ohn, J., Yoon, J.-S., Kang, B.M., Park, M., Kim, S., Lee, W., Hwang, S., Kim, J.-I., Kim, K.H., and Kwon, O. (2019). Priming mobilization of hair follicle stem cells triggers permanent loss of regeneration after alkylating chemotherapy. Nat. Commun. *10*, 3694. https://doi.org/10.1038/s41467-019-11665-0.

Klawitter, S., Fuchs, N.V., Upton, K.R., Muñoz-Lopez, M., Shukla, R., Wang, J., Garcia-Cañadas, M., Lopez-Ruiz, C., Gerhardt, D.J., Sebe, A., et al. (2016). Reprogramming triggers endogenous L1 and Alu retrotransposition in human induced pluripotent stem cells. Nat. Commun. *7*, 10286. https://doi.org/10.1038/ncomms10286.

Kooreman, N.G., Kim, Y., de Almeida, P.E., Termglinchan, V., Diecke, S., Shao, N.Y., Wei, T.T., Yi, H., Dey, D., Nelakanti, R., et al. (2018). Autologous iPSC-based vaccines elicit anti-tumor responses in vivo. Cell Stem Cell *22*, 501–513.e7. https://doi.org/10.1016/j.stem.2018.01.016.

Kroemer, G., and Pouyssegur, J. (2008). Tumor cell metabolism: cancer's achilles' heel. Cancer Cell *13*, 472–482. https://doi.org/10.1016/j.ccr.2008.05.005.

Krokhin, O.V. (2006). Sequence-specific retention calculator. Algorithm for peptide retention prediction in ion-pair RP-HPLC: application to 300- and 100-A pore size C18 sorbents. Anal. Chem. *78*, 7785–7795. https://doi.org/10.1021/ac060777w.

Lamoliatte, F., McManus, F.P., Maarifi, G., Chelbi-Alix, M.K., and Thibault, P. (2017). Uncovering the SUMOylation and ubiquitylation crosstalk in human cells using sequential peptide immunopurification. Nat. Commun. *8*, 14109. https://doi.org/10.1038/ncomms14109.

Larouche, J.D., Trofimov, A., Hesnard, L., Ehx, G., Zhao, Q., Vincent, K., Durette, C., Gendron, P., Laverdure, J.P., Bonneil, É., et al. (2020). Widespread and tissue-specific expression of endogenous retroelements in human somatic tissues. Genome Med. *12*, 40. https://doi.org/10.1186/s13073-020-00740-7.

Laumont, C.M., Vincent, K., Hesnard, L., Audemard, É., Bonneil, É., Laverdure, J.-P., Gendron, P., Courcelles, M., Hardy, M.-P., Côté, C., et al. (2018). Noncoding regions are the main source of targetable tumor-specific antigens. Sci. Transl. Med. *10*, eaau5516. https://doi.org/10.1126/scitranslmed.aau5516.

Li, L., Baroja, M.L., Majumdar, A., Chadwick, K., Rouleau, A., Gallacher, L., Ferber, I., Lebkowski, J., Martin, T., Madrenas, J., and Bhatia, M. (2004). Human embryonic stem cells possess immune-privileged properties. Stem Cell. *22*, 448–456. https://doi.org/10.1634/stemcells.22-4-448.

Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J.P., and Tamayo, P. (2015). The molecular signatures database hallmark gene set collection. Cell Syst. *1*, 417–425. https://doi.org/10.1016/j.cels.2015.12.004.

Lin, T., and Lin, Y. (2017). P53 switches off pluripotency on differentiation. Stem Cell Res. Ther. *8*, 44. https://doi.org/10.1186/s13287-017-0498-1.

Liu, J., Lichtenberg, T., Hoadley, K.A., Poisson, L.M., Lazar, A.J., Cherniack, A.D., Kovatich, A.J., Benz, C.C., Levine, D.A., Lee, A.V., et al. (2018). An integrated TCGA pan-cancer clinical data resource to drive high-quality survival outcome analytics. Cell *173*, 400–416.e11. https://doi.org/10.1016/j.cell.2018.02.052.

Löffler, M.W., Mohr, C., Bichmann, L., Freudenmann, L.K., Walzer, M., Schroeder, C.M., Trautwein, N., Hilke, F.J., Zinser, R.S., Mühlenbruch, L., et al. (2019). Multi-omics discovery of exome-derived neoantigens in hepatocellular carcinoma. Genome Med. *11*, 28–36. https://doi.org/10.1186/s13073-019-0636-8.

Loizou, E., Banito, A., Livshits, G., Ho, Y.J., Koche, R.P., Sánchez-Rivera, F.J., Mayle, A., Chen, C.C., Kinalis, S., Bagger, F.O., et al. (2019). A gain-of-function p53-mutant oncogene promotes cell fate plasticity and myeloid leukemia through the pluripotency factor foxh1. Cancer Discov. *9*, 962–979. https://doi.org/10.1158/2159-8290.CD-18-1391.

GTEx Consortium; Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., Hasz, R., Walters, G., Garcia, F., Young, N., et al. (2013). The genotype-tissue expression (GTEx) project. Nat. Genet. *45*, 580–585. https://doi.org/10.1038/ng.2653.

Lu, X., Sachs, F., Ramsay, L., Jacques, P.É., Göke, J., Bourque, G., and Ng, H.H. (2014). The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. Nat. Struct. Mol. Biol. *21*, 423–425. https://doi.org/10.1038/nsmb.2799.

Ma, J., Zhao, J., Lu, J., Wang, P., Feng, H., Zong, Y., Ou, B., Zheng, M., and Lu, A. (2016). Cadherin-12 enhances proliferation in colorectal cancer cells and increases progression by promoting EMT. Tumour Biol. *37*, 9077–9088. https://doi.org/10.1007/s13277-015-4555-z.

Macrae, T., Sargeant, T., Lemieux, S., Hébert, J., Deneault, E., and Sauvageau, G. (2013). RNA-Seq reveals spliceosome and proteasome genes as most consistent transcripts in human cancer cells. PLoS One *8*, e72884. https://doi.org/10.1371/journal.pone.0072884.

Madsen, R.R. (2020). PI3K in stemness regulation: from development to cancer. Biochem. Soc. Trans. *48*, 301–315. https://doi.org/10.1042/BST20190778.

Maiga, A., Lemieux, S., Pabst, C., Lavallée, V.P., Bouvier, M., Sauvageau, G., and Hébert, J. (2016). Transcriptome analysis of G protein-coupled receptors in distinct genetic subgroups of acute myeloid leukemia: identification of potential disease-specific targets. Blood Cancer J. *6*, e431. https://doi.org/10.1038/bcj.2016.36.

Malta, T.M., Sokolov, A., Gentles, A.J., Burzykowski, T., Poisson, L., Weinstein, J.N., Kamińska, B., Huelsken, J., Omberg, L., Gevaert, O., et al. (2018). Machine learning identifies stemness features associated with oncogenic dedifferentiation. Cell *173*, 338–354.e15. https://doi.org/10.1016/j.cell.2018.03.034.

Marçais, G., and Kingsford, C. (2011). A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics *27*, 764–770. https://doi.org/10.1093/bioinformatics/btr011.

McLaughlin, R.N., Young, J.M., Yang, L., Neme, R., Wichman, H.A., and Malik, H.S. (2014). Positive selection and multiple losses of the LINE-1-derived L1TD1 gene in mammals suggest a dual role in genome defense and pluripotency. PLoS Genet. *10*, e1004531. https://doi.org/10.1371/journal.pgen.1004531.

Merkle, F.T., Ghosh, S., Kamitaki, N., Mitchell, J., Avior, Y., Mello, C., Kashin, S., Mekhoubad, S., Ilic, D., Charlton, M., et al. (2017). Human pluripotent stem cells recurrently acquire and expand dominant negative P53 mutations. Nature *545*, 229–233. https://doi.org/10.1038/nature22312.

Mi, H., Ebert, D., Muruganujan, A., Mills, C., Albou, L.-P., Mushayamaha, T., and Thomas, P.D. (2021). PANTHER version 16: a revised family classification, tree-based classification tool, enhancer regions and extensive API. Nucleic Acids Res. *49*, D394–D403. https://doi.org/10.1093/nar/gkaa1106.

Miranda, A., Hamilton, P.T., Zhang, A.W., Pattnaik, S., Becht, E., Mezheyeuski, A., Bruun, J., Micke, P., de Reynies, A., and Nelson, B.H. (2019). Cancer stemness, intratumoral heterogeneity, and immune response across cancers. Proc. Natl. Acad. Sci. USA *116*, 9020–9029. https://doi.org/10.1073/pnas.1818210116.

Närvä, E., Rahkonen, N., Emani, M.R., Lund, R., Pursiheimo, J.P., Nästi, J., Autio, R., Rasool, O., Denessiouk, K., Lähdesmäki, H., et al. (2012). RNA-binding protein L1TD1 interacts with LIN28 via RNA and is required for human embryonic stem cell self-renewal and cancer cell proliferation. Stem Cell. *30*, 452–460. https://doi.org/10.1002/stem.1013.

Di Nicola, M., Carlo-Stella, C., Magni, M., Milanesi, M., Longoni, P.D., Matteucci, P., Grisanti, S., and Gianni, A.M. (2002). Human bone marrow stromal cells suppress T-lymphocyte proliferation induced by cellular or nonspecific mitogenic stimuli. Blood *99*, 3838–3843. https://doi.org/10.1182/blood.V99.10.3838.

Noronha, N., Ehx, G., Meunier, M.-C., Laverdure, J.-P., Thériault, C., and Perreault, C. (2020). Major multilevel molecular divergence between THP-1 cells from different biorepositories. Int. J. Cancer *147*, 2000–2006. https://doi.org/10.1002/ijc.32967.

Ochsenbein, A.F. (2005). Immunological ignorance of solid tumors. Springer Semin. Immunopathol. *27*, 19–35. https://doi.org/10.1007/s00281-004-0192-0.

Ogishi, M., and Yotsuyanagi, H. (2019). Quantitative prediction of the landscape of T cell epitope immunogenicity in sequence space. Front. Immunol. *10*, 827. https://doi.org/10.3389/fimmu.2019.00827.

Durette, C., Daouda, T., Granados, D.P., Pearson, H., Bonneil, E., Courcelles, M., Rodenbrock, A., Laverdure, J.P., Côté, C., Mader, S., et al. (2016). MHC class I–associated peptides derive from selective regions of the human genome. J. Clin. Invest. *126*, 4690–4701. https://doi.org/10.1172/JCI88590.defective.

Pellagatti, A., Armstrong, R.N., Steeples, V., Sharma, E., Repapi, E., Singh, S., Sanchi, A., Radujkovic, A., Horn, P., Dolatshad, H., et al. (2018). Impact of spliceosome mutations on RNA splicing in myelodysplasia: dysregulated genes/pathways and clinical associations. Blood *132*, 1225–1240. https://doi.org/10.1182/blood-2018-04-843771.

Pelullo, M., Zema, S., Nardozza, F., Checquolo, S., Screpanti, I., and Bellavia, D. (2019). Wnt, Notch, and TGF-β pathways impinge on hedgehog signaling complexity: an open window on cancer. Front. Genet. *10*, 711–716. https://doi.org/10.3389/fgene.2019.00711.

Petroni, G., Buqué, A., Zitvogel, L., Kroemer, G., and Galluzzi, L. (2021). Immunomodulation by targeted anticancer agents. Cancer Cell *39*, 310–345. https://doi.org/10.1016/j.ccell.2020.11.009.

Poli, V., Fagnocchi, L., Fasciani, A., Cherubini, A., Mazzoleni, S., Ferrillo, S., Miluzio, A., Gaudioso, G., Vaira, V., Turdo, A., et al. (2018). MYC-driven epigenetic reprogramming favors the onset of tumorigenesis by inducing a stem cell-like state. Nat. Commun. *9*, 1024. https://doi.org/10.1038/s41467-018-03264-2.

Prensner, J.R., Iyer, M.K., Balbin, O.A., Dhanasekaran, S.M., Cao, Q., Brenner, J.C., Laxman, B., Asangani, I.A., Grasso, C.S., Kominsky, H.D., et al. (2011). Transcriptome sequencing across a prostate cancer cohort identifies PCAT-1, an unannotated lincRNA implicated in disease progression. Nat. Biotechnol. *29*, 742–749. https://doi.org/10.1038/nbt.1914.

Reinhard, K., Rengstl, B., Oehm, P., Michel, K., Billmeier, A., Hayduk, N., Klein, O., Kuna, K., Ouchan, Y., Wöll, S., et al. (2020). An RNA vaccine drives expansion and efficacy of claudin-CAR-T cells against solid tumors. Science *367*, 446–453. https://doi.org/10.1126/science.aay5967.

Van Rhenen, A., Feller, N., Kelder, A., Westra, A.H., Rombouts, E., Zweegman, S., Van Der Pol, M.A., Waisfisz, Q., Ossenkoppele, G.J., and Schuurhuis, G.J. (2005). High stem cell frequency in acute myeloid leukemia at diagnosis predicts high minimal residual disease and poor survival. Clin. Cancer Res. *11*, 6520–6527. https://doi.org/10.1158/1078-0432.CCR-05-0468.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. *29*, 24–26. https://doi.org/10.1038/nbt.1754.

Ruiz Cuevas, M.V., Hardy, M.P., Hollý, J., Bonneil, É., Durette, C., Courcelles, M., Lanoix, J., Côté, C., Staudt, L.M., Lemieux, S., et al. (2021). Most non-canonical proteins uniquely populate the proteome or immunopeptidome. Cell Rep. *34*, 108815. https://doi.org/10.1016/j.celrep.2021.108815.

Sahin, U., Oehm, P., Derhovanessian, E., Jabulowsky, R.A., Vormehr, M., Gold, M., Maurus, D., Schwarck-Kokarakis, D., Kuhn, A.N., Omokoko, T., et al. (2020). An RNA vaccine drives immunity in checkpoint-inhibitor-treated melanoma. Nature *585*, 107–112. https://doi.org/10.1038/s41586-020-2537-9.

Salmon, H., Idoyaga, J., Rahman, A., Leboeuf, M., Remark, R., Jordan, S., Casanova-Acebes, M., Khudoynazarova, M., Agudo, J., Tung, N., et al. (2016). Expansion and activation of CD103+ dendritic cell progenitors at the tumor site enhances tumor responses to therapeutic PD-L1 and BRAF inhibition. Immunity *44*, 924–938. https://doi.org/10.1016/j.immuni.2016.03.012.

Sanchez-Vega, F., Mina, M., Armenia, J., Chatila, W.K., Luna, A., La, K.C., Dimitriadoy, S., Liu, D.L., Kantheti, H.S., Saghafinia, S., et al. (2018). Oncogenic signaling pathways in the cancer genome Atlas. Cell *173*, 321–337.e10. https://doi.org/10.1016/j.cell.2018.03.035.

Schonberg, D.L., Miller, T.E., Wu, Q., Flavahan, W.A., Das, N.K., Hale, J.S., Hubert, C.G., Mack, S.C., Jarrar, A.M., Karl, R.T., et al. (2015). Preferential iron trafficking characterizes glioblastoma stem-like cells. Cancer Cell *28*, 441–455. https://doi.org/10.1016/j.ccell.2015.09.002.

Schumacher, T.N., Scheper, W., and Kvistborg, P. (2019). Cancer neoantigens. Annu. Rev. Immunol. *37*, 173–200. https://doi.org/10.1146/annurev-immunol-042617-053402.

Schuster, H., Peper, J.K., Bösmüller, H.C., Röhle, K., Backert, L., Bilich, T., Ney, B., Löffler, M.W., Kowalewski, D.J., Trautwein, N., et al. (2017). The immunopeptidomic landscape of ovarian carcinomas. Proc. Natl. Acad. Sci. USA *114*, E9942–E9951. https://doi.org/10.1073/pnas.1707658114.

Shan, X., Roberts, C., Kim, E.J., Brenner, A., Grant, G., and Percec, I. (2017). Transcriptional and cell cycle alterations mark aging of primary human adipose-derived stem cells. Stem Cell. *35*, 1392–1401. https://doi.org/10.1002/stem.2592.

Shen, W.-C., Lai, Y.-C., Li, L.-H., Liao, K., Lai, H.-C., Kao, S.-Y., Wang, J., Chuong, C.-M., and Hung, S.-C. (2019). Methylation and PTEN activation in dental pulp mesenchymal stem cells promotes osteogenesis and reduces oncogenesis. Nat. Commun. *10*, 2226. https://doi.org/10.1038/s41467-019-10197-x.

Smith, B.A., Balanis, N.G., Nanjundiah, A., Sheu, K.M., Tsai, B.L., Zhang, Q., Park, J.W., Thompson, M., Huang, J., Witte, O.N., and Graeber, T.G. (2018). A human adult stem cell signature marks aggressive variants across epithelial cancers. Cell Rep. *24*, 3353–3366.e5. https://doi.org/10.1016/j.celrep.2018.08.062.

Spranger, S., and Gajewski, T.F. (2018). Impact of oncogenic pathways on evasion of antitumour immune responses. Nat. Rev. Cancer *18*, 139–147. https://doi.org/10.1038/nrc.2017.117.

Spranger, S., Bao, R., and Gajewski, T.F. (2015). Melanoma-intrinsic β-catenin signalling prevents anti-tumour immunity. Nature *523*, 231–235. https://doi.org/10.1038/nature14404.

Spranger, S., Dai, D., Horton, B., and Gajewski, T.F. (2017). Tumor-residing Batf3 dendritic cells are required for effector T cell trafficking and adoptive T cell therapy. Cancer Cell *31*, 711–723.e4. https://doi.org/10.1016/j.ccell.2017.04.003.

Stewart, M.H., Bossé, M., Chadwick, K., Menendez, P., Bendall, S.C., and Bhatia, M. (2006). Clonal isolation of hESCs reveals heterogeneity within the pluripotent stem cell compartment. Nat. Methods *3*, 807–815. https://doi.org/10.1038/nmeth939.

Stine, Z.E., Walton, Z.E., Altman, B.J., Hsieh, A.L., and Dang, C.V. (2015). MYC, metabolism, and cancer. Cancer Discov. *5*, 1024–1039. https://doi.org/10.1158/2159-8290.CD-15-0507.

Suárez-Alvarez, B., Rodriguez, R.M., Calvanese, V., Blanco-Gelaz, M.A., Suhr, S.T., Ortega, F., Otero, J., Cibelli, J.B., Moore, H., Fraga, M.F., and López-Larrea, C. (2010). Epigenetic mechanisms regulate MHC and antigen processing molecules in human embryonic and induced pluripotent stem cells. PLoS One *5*, e10192. https://doi.org/10.1371/journal.pone.0010192.

Szolek, A., Schubert, B., Mohr, C., Sturm, M., Feldhahn, M., and Kohlbacher, O. (2014). OptiType: precision HLA typing from next-generation sequencing data. Bioinformatics *30*, 3310–3316. https://doi.org/10.1093/bioinformatics/btu548.

Thorsson, V., Gibbs, D.L., Brown, S.D., Wolf, D., Bortone, D.S., Ou Yang, T.H., Porta-Pardo, E., Gao, G.F., Plaisier, C.L., Eddy, J.A., et al. (2018). The immune landscape of cancer. Immunity *48*, 812–830.e14. https://doi.org/10.1016/j.immuni.2018.03.023.

Vasileiou, S., Lulla, P.D., Tzannou, I., Watanabe, A., Kuvalekar, M., Callejas, W.L., Bilgi, M., Wang, T., Wu, M.J., Kamble, R., et al. (2021). T-cell therapy for lymphoma using nonengineered multiantigen-targeted T cells is safe and produces durable clinical effects. J. Clin. Oncol. *39*, 1415–1425. https://doi.org/10.1200/jco.20.02224.

Vogel, C., and Marcotte, E.M. (2009). Absolute abundance for the masses. Nat. Biotechnol. *27*, 825–826. https://doi.org/10.1038/nbt0909-825.

Wang, J.F., She, L., Su, B.H., Ding, L.C., Zheng, F.F., Zheng, D.L., and Lu, Y.G. (2011). CDH12 promotes the invasion of salivary adenoid cystic carcinoma. Oncol. Rep. *26*, 101–108. https://doi.org/10.3892/or.2011.1286.

Wang, L., Su, Y., Huang, C., Yin, Y., Zhu, J., Knupp, A., Chu, A., and Tang, Y. (2019). FOXH1 is regulated by NANOG and LIN28 for early-stage reprogramming. Sci. Rep. *9*, 16443–16448. https://doi.org/10.1038/s41598-019-52861-8.

Wang, Z.-X., Kueh, J.L.L., Teh, C.H.-L., Rossbach, M., Lim, L., Li, P., Wong, K.-Y., Lufkin, T., Robson, P., and Stanton, L.W. (2007). Zfp206 is a transcription factor that controls pluripotency of embryonic stem cells. Stem Cell. *25*, 2173–2182. https://doi.org/10.1634/stemcells.2007-0085.

Bartolo, D., and Guo, J.U. (2020). Coding functions of ''noncoding'' RNAs. Science *367*, 1075–1076. https://doi.org/10.1126/science.aba5319.

Wu, T.D., and Nacu, S. (2010). Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics *26*, 873–881. https://doi.org/10.1093/bioinformatics/btq057.

Yaddanapudi, K., Mitchell, R.A., Putty, K., Willer, S., Sharma, R.K., Yan, J., Bodduluri, H., and Eaton, J.W. (2012). Vaccination with embryonic stem cells protects against lung cancer: is a broad-spectrum prophylactic vaccine against cancer possible? PLoS One *7*, e42289. https://doi.org/10.1371/journal.pone.0042289.

Yewdell, J.W. (2010). Designing CD8+ T cell vaccines: it's not rocket science (yet). Curr. Opin. Immunol. *22*, 402–410. https://doi.org/10.1016/j.coi.2010.04.002.

Yu, H.B., Kunarso, G., Hong, F.H., and Stanton, L.W. (2009). Zfp206, Oct4, and Sox2 are integrated components of a transcriptional regulatory network in embryonic stem cells. J. Biol. Chem. *284*, 31327–31335. https://doi.org/10.1074/jbc.M109.016162.

Zhang, H.L., Wang, P., Lu, M.Z., Zhang, S.D., and Zheng, L. (2019). c-Myc maintains the self-renewal and chemoresistance properties of colon cancer stem cells. Oncol. Lett. *17*, 4487–4493. https://doi.org/10.3892/ol.2019.10081.

Zhang, J., Nuebel, E., Daley, G.Q., Koehler, C.M., and Teitell, M.A. (2012). Metabolic regulation in pluripotent stem cells during reprogramming and self-renewal. Cell Stem Cell *11*, 589–595. https://doi.org/10.1016/j.stem.2012.10.005.

Zhang, J., Ratanasirintrawoot, S., Chandrasekaran, S., Wu, Z., Ficarro, S.B., Yu, C., Ross, C.A., Cacchiarelli, D., Xia, Q., Seligson, M., et al. (2016). LIN28 regulates stem cell metabolism and conversion to primed pluripotency. Cell Stem Cell *19*, 66–80. https://doi.org/10.1016/j.stem.2016.05.009.

Zhao, Q., Laverdure, J.-P., Lanoix, J., Durette, C., Côté, C., Bonneil, É., Laumont, C.M., Gendron, P., Vincent, K., Courcelles, M., et al. (2020). Proteogenomics uncovers a vast repertoire of shared tumor-specific antigens in ovarian

cancer. Cancer Immunol. Res. *8*, 544–555. https://doi.org/10.1158/2326-6066.cir-19-0541.

Zhao, S., Zhu, W., Xue, S., and Han, D. (2014). Testicular defense systems: immune privilege and innate immunity. Cell. Mol. Immunol. *11*, 428–437. https://doi.org/10.1038/cmi.2014.38.

Zheng, K., Wu, X., Kaestner, K.H., and Wang, P.J. (2009). The pluripotency factor LIN28 marks undifferentiated spermatogonia in mouse. BMC Dev. Biol. *9*, 38–41. https://doi.org/10.1186/1471-213X-9-38.

Zhou, X., Guo, X., Chen, M., Xie, C., and Jiang, J. (2018). HIF-3α promotes metastatic phenotypes in pancreatic cancer by transcriptional regulation of the RhoC-ROCK1 signaling pathway. Mol. Cancer Res. *16*, 124–134. https://doi.org/10.1158/1541-7786.MCR-17-0256.

Zitvogel, L., Perreault, C., Finn, O.J., and Kroemer, G. (2021). Beneficial autoimmunity improves cancer prognosis. Nat. Rev. Clin. Oncol. *18*, 591–602. https://doi.org/10.1038/s41571-021-00508-x.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
Human and Mouse Pluripotent Stem Cell Analysis Kit (containing PerCP-Cy5.5 Mouse anti-Oct3/4, PE Mouse anti-SSEA-1, Alexa Fluor 647 Mouse anti-SSEA-4 antibodies and isotypes, and intracellular staining buffers)	BD Biosciences	Cat# 560477; RRID:AB_2869350
APC/Cyanine7 anti-human/mouse SSEA-3	BioLegend	Cat# 330317; RRID:AB_2616816
APC-Cy™7 Rat IgM, κ Isotype Control	BD Biosciences	Cat# 560571; RRID:AB_1645640
Purified Mouse IgG2a κ Isotype Control	BD Biosciences	Cat# 553454; RRID:AB_479656
InVivoMAb anti-human MHC Class I (W6/32)	BioXCell	Cat# BE0079; RRID:AB_1107730
BD Pharmingen APC-H7 Mouse Anti-human CD3	BD Biosciences	Cat# 560176; RRID:AB_1645475
BD Horizon™ BV510 Mouse Anti-Human CD4	BD Biosciences	Cat# 562970; RRID:AB_2744424
BD Horizon BB515 Mouse Anti-human CD8	BD Biosciences	Cat# 564526; RRID:AB_2744458
PerCP-Cy5.5 Mouse Anti-human CD14	BD Biosciences	Cat# 561116; RRID:AB_2033939
PerCP-Cy5.5 Mouse Anti-human CD16	BD Biosciences	Cat# 560717; RRID:AB_1727434
PerCP-Cy5.5 Mouse Anti-human CD19	BD Biosciences	Cat# 561295; RRID:AB_10644017
Rabbit anti-Sox2	Life Technologies	Cat# 48-1400; RRID:AB_2533841
Mouse anti-Pax6	BioLegend	Cat# 862001; RRID:AB_2801237
Rabbit anti-CXCR4	MyBiosource	Cat# MBS7005441; RRID:AB_2895129
Mouse anti-CD56	BioLegend	Cat# 362502; RRID:AB_2563558
Mouse anti-Sox17	BioLegend	Cat# 698501; RRID:AB_2687317
Leukoagglutinin	Invitrogen	Cat# 00-4977-93
Biological samples		
Healthy human peripheral blood	BioIVT	https://bioivt.com/
Chemicals, peptides, and recombinant proteins		
Human IFN gamma Recombinant Protein	Gibco	Cat# PHC4031
DPBS, no calcium, no magnesium	Gibco	Cat# 14190144
TrypLE™ Express Enzyme (1X), no phenol red	Gibco	Cat# 12604013
Corning Matrigel Basement Membrane Matrix	CORNING	Cat# 354234
DMEM/F-12	Gibco	Cat# 11320033
mTeSR1 medium	STEMCELL	Cat# 85850
mTeSR Plus medium	STEMCELL	Cat# 100-0276
Gentle Cell Dissociation Reagent	STEMCELL	Cat# 07174
Cellartis DEF-CS™ 500 Basal Medium with Additives	Takara Bio	Cat# Y30017
Cellartis DEF-CS™ 500 COAT-1	Takara Bio	Cat# Y30012
TRIzol	Invitrogen	Cat# 15596018
AIM-V medium	Gibco	Cat# 12055-91
IL-2	Peprotech	Cat# 200-02
IL-7	Peprotech	Cat# 200-07
IL-15	Peprotech	Cat# 200-15
Penicillin-Streptomycin (10,000 U/mL)	Gibco	Cat# 15140122
RPMI 1640 Medium	Gibco	Cat# 11875093
Fetal Bovine Serum	Gibco	Cat# 12483020
MEM α, nucleosides	Gibco	Cat# 12571063

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
7-AAD	BD Biosciences	Cat# 559925
L-glutamine	Gibco	Cat# 25030081
Critical commercial assays		
QIFIKIT (Quantitative Analysis Kit)	Agilent Dako	Cat# K0078
StemMACS Trilineage Differentiation Kit	Miltenyi Biotec	Cat# 130-115-660
RNeasy Micro Kit	QIAGEN	Cat# 74004
KAPA HyperPrep RNAseq stranded kit	KAPA	Cat# KK8421
KAPA library quantification kit	KAPA	Cat# KK4973
QiaAmp DNA blood mini kit (50)	QIAGEN	Cat# 51104
PureProteome protein A magnetic beads	Millipore	Cat# LSKMAGA10
$CD8^+$ T Cell Isolation Kit, human	Miltenyi Biotec	Cat# 130-096-495
Pan T cell Isolation Kit, Human	Miltenyi Biotec	Cat# 130-096-535
Anti-APC MicroBeads	Miltenyi Biotec	Cat# 130-090-855
Anti-PE MicroBeads	Miltenyi Biotec	Cat# 130-048-801
CellTrace™ CFSE Cell Proliferation Kit, for flow cytometry	Invitrogen	Cat# C34554
Deposited data		
MS (immunopeptidomic) data from hiPSCs	This paper	PRIDE: PXD026702
RNA-seq from hiPSCs	This paper	GEO: GSE171226
RNA-seq from human mTECs	Laumont et al., 2018	BioProject: PRJNA525590
RNA-seq from human mTECs	Larouche et al., 2020	GEO: GSE127826
RNA-seq from human PSCs	Churko et al., 2017	GEO: GSE69626
RNA-seq from HGSC samples	Zhao et al., 2020	GEO: GSE131880
RNA-seq from AML samples	Ehx et al., 2021	GEO: GSE147524
RNA-seq from healthy tissues in the GTEx database	GTEx project	https://gtexportal.org/
RNA-seq, clinical, mutation data from TCGA samples	TCGA project	https://portal.gdc.cancer.gov/
RNA-seq from umbilical cord blood cells	Macrae et al., 2013	GEO: GSE48846
RNA-seq from adipose stem cells	Shan et al., 2017	GEO: GSE86244
RNA-seq from bone marrow- and dental pulp-derived mesenchymal stem cells	Shen et al., 2019	GEO: GSE105145
RNA-seq from hair follicle stem cells	Kim et al., 2019	GEO: GSE130054
RNA-seq from multipotent adult stem cell from human foreskin	Paul A. Lucas Lab, New York Medical College	GEO: GSE140119
RNA-seq from glial progenitors	Schonberg et al., 2015	GEO: GSM1857484
RNA-seq from bone marrow progenitors	Frisch et al., 2019	GEO: GSE100905
RNA-seq from bone marrow progenitors	Pellagatti et al., 2018	GEO: GSE114922
RNA-seq from bone marrow progenitors	Maiga et al., 2016	GEO: GSE98310
Experimental models: Cell lines		
hiPSC22	Takara Bio	Cat# Y00325
Fibro-iPSC.1 and Fibro-iPSC.2	Mick Bhatia lab, McMaster University	https://www.bhatiaprogram.com/
THP-1	DSMZ	ACC 16
OCI-AML3	DSMZ	ACC 582
B lymphoblastoid cell lines (B-LCLs)	Generated as in Granados et al., 2014	In-house identifiers M(2.1), R(2.2), MA(2.3)
Software and algorithms		
R v3.6.5	The R Project	https://www.r-project.org/
Python v3.6.7 or v3.9.5	Python Software Foundation	https://www.python.org/

(*Continued on next page*)

Continued

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Adobe Illustrator CS	Adobe	https://www.adobe.com/ca/products/illustrator.html
FlowJo v10	BD Biosciences	https://www.flowjo.com/
PANTHER	Mi et al., 2021	http://www.pantherdb.org/
SSRCalc	Krokhin, 2006	http://hs2.proteome.ca/SSRCalc/SSRCalcQ.html
DeepLC v0.1.16	Bouwmeester et al., 2021	https://github.com/compomics/DeepLC
xCell	Aran et al., 2017a, 2017b	https://github.com/dviraran/xCell
Trimmomatic v0.35	Bolger et al., 2014	http://www.usadellab.org/cms/?page=trimmomatic
STAR v2.5.1b	Dobin et al., 2013	https://github.com/alexdobin/STAR
freeBayes v1.0.2-16-gd466dde	Garrison and Marth, 2012	https://github.com/ekg/freebayes
Kallisto v0.43.0	Bray et al., 2016	https://pachterlab.github.io/kallisto/
pyGeno	Daouda et al., 2016	https://github.com/tariqdaouda/pyGeno
JELLYFISH v2.2.3	Marçais and Kingsford, 2011	https://www.cbcb.umd.edu/software/jellyfish/
NEKTAR	In-house software	https://github.com/iric-soft/nektar
GSNAP	Wu and Nacu, 2010	http://research-pub.gene.com/gmap
FASTX-Toolkit v0.0.14		https://github.com/agordon/fastx_toolkit
PEAKS v10.5	Bioinformatics Solutions Inc.	https://www.bioinfor.com/peaksdb/
OptiType	Szolek et al., 2014	https://github.com/FRED-2/OptiType
NetMHCpan-4.0	Jurtz et al., 2017	http://www.cbs.dtu.dk/services/NetMHCpan-4.0/
MAPDP	Courcelles et al., 2020	https://gitlab.com/iric-proteo/mapdp
Repitope	Ogishi and Yotsuyanagi, 2019	https://github.com/masato-ogishi/Repitope
immunoSEQ	Adaptive Biotechnologies	https://www.immunoseq.com/
FEST data analysis tool	Danilova et al., 2018	www.stat-apps.onc.jhmi.edu/FEST

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Claude Perreault (claude.perreault@umontreal.ca).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- RNA-sequencing data have been deposited at the NCBI Sequence Read Archive and Gene Expression Omnibus (GEO). MS raw data and associated databases, and mirror plots for 11 paMAPs and 5 saMAPs, have been deposited at the ProteomeXchange Consortium via the PRIDE partner repository. All data are publicly available as of the date of publication. Accession numbers are listed in the key resources table. Details regarding all other data used in this study are listed in Table S4.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human iPSC culture

hiPSC22 cells derived from male adult human skin fibroblasts using defective polycistronic retroviruses expressing OCT4, SOX2, KLF4, and c-MYC were obtained from Takara Bio (Cellartis human iPS cell line 22). hiPSC22 cells were cultured in the Cellartis® DEF-CS™ 500 Basal Medium with Additives (Takara Bio) on coated (Cellartis DEF-CS 500 COAT-1, Takara Bio) cell culture vessels according to the manufacturer's instructions. Fibro-iPSC.1 and Fibro-iPSC.2 cells are biological replicates of the same iPS cell line reprogrammed from female adult human dermal fibroblasts using lentiviruses expressing OCT4, SOX2, NANOG, and LIN28, as per

(Hong et al., 2011), and were provided by Dr. Mick Bathia (McMaster University, Ontario, Canada). Fibro-iPSC.1 and Fibro-iPSC.2 were cultured on Matrigel (Corning, diluted in DMEM/F-12 from Gibco)-coated cell culture vessels in mTeSR1 medium (STEMCELL), according to the manufacturer's instructions. Cells were grown in a humidified 37°C, 5% CO_2 incubator. All iPSCs were passaged using the Gentle Cell Dissociation Reagent (STEMCELL) or were dissociated to single cells using TrypLE Express (Gibco) and washed with DPBS (Gibco) for downstream analyses. After removing 3-5 x 10^6 iPSCs for RNA-seq and 5 × 10^6 cells for flow cytometry, iPSCs were pelleted and stored at −80 degrees C until MS analysis. For IFN-γ-treated samples, iPSCs were treated with a final concentration of 40 ng/mL recombinant human IFN-γ (Gibco) for 72 h before collection. MS analyses were performed on two fractions per iPS cell line as following, for each fraction: 250 × 10^6 cells for untreated Fibro-iPSC.1 and Fibro-iPSC.2, 375 × 10^6 cells for untreated hiPSC22, and 100-125 × 10^6 cells for all IFN-γ-treated iPSC samples.

OCI-AML3 and THP-1 cell culture

THP-1 (RRID: CVCL_0006) and OCI-AML3 (RRID:CVCL_1844) cell lines were previously purchased from the Deutsche Sammlung von Mikroorganismen und Zellkulturen (DSMZ; ACC 16 and ACC 582, respectively) (Noronha et al., 2020). Cells were thawed and seeded at 0.2 million/mL in T75 cell culture flasks (Thermo Fisher Scientific) and incubated at 37°C in 5% CO_2. Cells were maintained with the following protocol: seeding at 0.2 million/mL; 1:2 dilution with fresh medium after 48 hours; 80 and 85% media replacement, respectively, after 48 hours, re-seeding at 0.2 million/mL after 72 hours. THP-1 cells were maintained in RPMI 1640 (Gibco) containing l-glutamine and supplemented with 10% heat-inactivated fetal bovine serum (FBS, Gibco) and 1% penicillin–streptomycin (10,000 U/mL, Gibco). OCI-AML3 cells were maintained in MEM Alpha (Gibco) supplemented with 10% heat-inactivated FBS and 1% penicillin–streptomycin (10,000 U/mL). After thawing, cells were allowed to recover from cryopreservation for 7 days before use in cytotoxicity assays.

B lymphoblastoid cell lines culture

B lymphoblastoid cell lines (B-LCLs) previously generated from PBMCs (Granados et al., 2014) were thawed and maintained between 0.2 and 1 million/mL in RPMI 1640 (Gibco) containing l-glutamine and supplemented with 10% heat-inactivated FBS (Gibco) and 1% penicillin–streptomycin (10,000 U/mL, Gibco).

METHOD DETAILS

Flow cytometry analysis of iPSCs

Single-cell suspensions were stained with PerCP-Cy5.5 Mouse anti-Oct3/4, PE Mouse anti-SSEA-1, Alexa Fluor 647 Mouse anti-SSEA-4 antibodies or the respective isotypes (Human and Mouse Pluripotent Stem Cell Analysis Kit, BD Biosciences), APC/Cyanine7 anti-human/mouse SSEA-3 (BioLegend) or the APC-Cy7 Rat IgM, κ Isotype Control (BD Biosciences) according to the manufacturers' instructions. Surface HLA-A,B,C molecules were quantified using a QIFIKIT (FITC conjugate, Agilent Dako) as per the manufacturer's instructions. Flow cytometry experiments were performed on a ZE5 (Bio-Rad), and data were analyzed using the FlowJo software.

Trilineage differentiation assay

iPSCs were dissociated into single cell suspensions using TrypLE and seeded in defined cell numbers into 24-well plates in mTeSR Plus medium (STEMCELL) supplemented with 10 μM ROCK inhibitor. The StemMAC Trilineage Differentiation Kit (Miltenyi Biotec) was used to differentiate iPSCs into the three germ layers. For ectoderm, cells were differentiated in StemMACS Trilineage EctoDiff Medium with daily media changes. For mesoderm, cells are first cultured for 1 day in PSC cultivation medium before induction towards mesoderm using the StemMACS Trilineage MesoDiff Medium I. Media changes were not required on day 2 and 3. From day 4 to day 6 media changes were performed daily using StemMACS Trilineage MesoDiff Medium II. For endoderm, cells were first cultured in PSC cultivation medium for 2 days. Afterwards, endoderm differentiation was induced by exchanging the PSC medium with StemMACS Trilineage EndoDiff Medium followed by daily media changes. On day 7, the differentiated cells were fixed with 4% folmaldehyde-PBS, permeabilized with 3% Triton-PBS and stained with specific antibodies for immunofluorescence.

RNA extraction and sequencing

Total RNA extraction was done using TRIzol (Invitrogen) and further purification with the RNeasy Micro Kit (QIAGEN) from 3 × 10^6 Fibro-iPSC.2 and Fibro-iPSC.2_IFN cells, and from 5 × 10^6 cells for all other samples. The RNA quantification was performed using a QuBIT (Life Technologies), and the RNA quality was assessed using a Bioanalyzer Nano (Agilent), and all samples had an RNA integrity number of 10. cDNA library preparation was done using 1000 ng RNA for hiPSC22_IFN and 4000 ng RNA for all other samples, using the KAPA Hyperprep RNAseq stranded kit (KAPA) with polyA capture. 9 and 7 PCR cycles for hiPSC22_IFN and all other iPSC samples, respectively, were used for library amplification. Libraries were quantified by QuBit, and average library length was evaluated with the BioAnalyzer DNA1000. All libraries were diluted to 10 nM and normalized by qPCR using the KAPA library quantification kit (KAPA). Libraries were pooled to equimolar concentration. Sequencing was performed with the Illumina Nextseq500 using the Nextseq High Output 150 cycles (2 × 80bp for hiPSC22 and hiPSC22_IFN, and 2 × 75bp for all other iPSCs) using 2 pM of the pooled libraries. Around 180 × 10^6 paired-end reads were generated per hiPSC22 sample (in three technical replicates pooled for MS

database generation), 360 × 10^6 paired-end reads for hiPSC22_IFN, and 230 × 10^6 paired-end reads for all other iPSC samples. Library preparation and sequencing were done at the Institute for Research in Immunology and Cancer (IRIC) Genomics Platform.

Database generation for shotgun mass spectrometry analyses

Generation of personalized canonical proteomes

This was conducted as previously described (Laumont et al., 2018). Briefly, RNA-seq reads were trimmed using Trimmomatic v0.35 (Bolger et al., 2014) and aligned to GRCh38.88 using STAR v2.5.1b (Dobin et al., 2013) running with default parameters except for –alignSJoverhangMin, –alignMatesGapMax, –alignIntronMax, and –alignSJstitchMismatchNmax parameters for which default values were replaced by 10, 200,000, 200,000 and "5 -1 5 5", respectively, to generate bam files. Single-base mutations with a minimum alternate count setting of 5 were identified using freeBayes v1.0.2-16-gd466dde (Garrison and Marth, 2012). Transcript expression was quantified in transcripts per million (tpm) with kallisto v0.43.0 with default parameters. Finally, we used pyGeno (Daouda et al., 2016) to insert high-quality sample-specific single-base mutations (freeBayes quality >20) in the reference exome and export sample-specific sequences of known proteins generated by expressed transcripts (tpm >0) to generate fasta files of personalized canonical proteomes.

Generation of iPSC and mTEC k-mer databases

This was conducted as previously described (Laumont et al., 2018), with the following exceptions: 8 mTEC samples (GEO accessions GSE127825, GSE127826; Table S4) were used instead of 6 mTECs, and the k-mer occurrence allowed in mTECs was 1 instead of 0 (see hereafter, Figure 1A for schematic). Briefly, R1 and R2 fastq files of each sample were trimmed as reported above, and the reverse mapping reads (R1 for hiPSC22, and R2 for Fibro-iPSC.1 and Fibro-iPSC.2, with or without IFN-γ) were reverse complemented using the fastx_reverse_complement function of the FASTX-Toolkit v0.0.14. K-mer databases (24 or 33-long) were generated using Jellyfish v2.2.3 (Marçais and Kingsford, 2011). A single k-mer database was generated for each iPSC sample, while the eight mTEC samples were combined in a unique database by concatenating their fastq files. Because the duration of k-mer assembly increases exponentially above 30 million k-mers, each iPSC 33-nucleotide-long k-mer database was filtered based on a sample-specific threshold on occurrence (the number of times that a given k-mer is present in the database) in order to reach a maximum of 30 million k-mers for the assembly step (see Table S1 for thresholds applied). After this filtering, k-mers present more than once in the mTECs k-mer database were removed from each sample database, and remaining k-mers were assembled into contigs with NEKTAR, an in-house developed software. Briefly, one of the submitted 33-nucleotide-long k-mer is randomly selected as a seed that is extended from both ends with consecutive k-mers overlapping by 32 nucleotides on the same strand (-r option disabled, as we were working with stranded sets of k-mers). The assembly process stops when either no k-mers can be assembled or when more than one k-mer fits (–a 1 option for linear assembly). Then a new seed is selected, and the assembly process resumes until all k-mers from the submitted list have been used once. Finally, we 3-frame translated the contigs using an in-house python script, split amino acid sequences at internal stop codons and concatenated the resulting subsequences with the respective personalized canonical proteome for each sample.

Isolation of MHC-associated peptides

The W6/32 antibodies (BioXcell) were incubated in PBS for 60 minutes at room temperature with PureProteome protein A magnetic beads (Millipore) at a ratio of 1 mg of antibody per mL of slurry. Antibodies were covalently cross-linked to magnetic beads using dimethylpimelidate as described (Lamoliatte et al., 2017). The beads were stored at 4°C in PBS pH 7.2. Frozen hiPSC22 pellets were thawed and resuspended in PBS pH 7.2 up to 1 mL and solubilized by adding 1 mL of detergent buffer containing PBS pH 7.2, 1% (w/v) CHAPS (Sigma) supplemented with Protease inhibitor cocktail (Sigma). Frozen Fibro-iPSC.1 and Fibro-iPSC.2 pellets were thawed and resuspended in PBS pH 7.2 up to 1 mL and solubilized by adding 1 mL of detergent buffer containing 0.5% (w/v) sodium deoxycholate (Thermo Fisher)/0.4 mM iodoacetamide (Sigma)/2% (w/v) Octyl β-D-glucopyranoside (Sigma)/2 mM EDTA (Promega) supplemented with Protease inhibitor cocktail (Sigma). Solubilized cell pellets were incubated for 60 minutes with tumbling at 4°C and then spun at 16600xg for 20 minutes at 4°C. Supernatants were transferred into new tubes containing 1 mg of W6/32 antibody covalently-cross-linked protein A magnetic beads per sample and incubated with tumbling for 180 minutes at 4°C. Samples were placed on a magnet to recover bound MHC I complexes to magnetic beads. Magnetic beads were first washed with 8× 1 mL PBS, then with 1 × 1 mL of 0.1X PBS, and finally with 1 × 1 mL of water. MHC I complexes were eluted from the magnetic beads by acidic treatment using 0.2% formic acid (FA). To remove any residual magnetic beads, eluates were transferred into 2.0 mL Costar mL Spin-X centrifuge tube filters (0.45 μm, Corning) and spun for 2 minutes at 855xg. Filtrates containing peptides were separated from MHC I subunits (HLA molecules and β-2 microglobulin) using homemade stage tips packed with two 1 mm diameter octadecyl (C-18) solid-phase extraction disks (EMPORE). Stage tips were pre-washed first with methanol, then with 80% acetonitrile (ACN) in 0.2% trifluoroacetic acid (TFA), and finally with 0.2% FA. Samples were loaded onto the stage tips and washed with 0.2% FA. Peptides were eluted with 30% ACN in 0.1%TFA, dried using vacuum centrifugation, and then stored at −20°C until MS analysis.

Mass spectrometry analyses

Dried peptide extracts were resuspended in 4% formic acid and loaded on a homemade C18 analytical column (15 cm × 150 μm i.d. packed with C18 Jupiter Phenomenex) with a 56-min gradient (hiPSC22, hiPSC22_IFN) or 106-minute gradient (all other samples) from 0% to 30% acetonitrile (0.2% formic acid) and a 600 nL/min flow rate on an EasynLC II system. Samples were analyzed

with a Q-Exactive HF mass spectrometer (Thermo Fisher Scientific) in positive ion mode with Nanospray 2 source at 1.6 kV. Each full MS spectrum, acquired with a 60,000 resolution was followed by 20 MS/MS spectra, where the most abundant multiply charged ions were selected for MS/MS sequencing with a resolution of 30,000, an automatic gain control target of 2×10^4, an injection time of 100 ms (hiPSC22_IFN) or 800 ms (all other samples) and collisional energy of 25%.

paMAP and saMAP validation with synthetic peptides

Validation of 11 paMAP and 5 saMAP identifications was performed by comparing the endogenous peptide spectra with synthetic peptide spectra (Figure S3). Synthetic peptides (GenScript) were dissolved in DMSO at 1 nmol/μL and diluted at 0.1 picomol/μL in 4% formic acid. The Orbitrap Fusion mass spectrometer (Thermo Fisher Scientific) was operated at a resolution of 120,000 in MS1 scan. An inclusion list was used for MS/MS sequencing with a resolution of 30,000, an automatic gain control target of 5×10^4, an injection time of 800 ms, and collision energy of 35%.

Spectra correlations were computed using a script written in Python v3.9.5 using the following steps. The first step reads the list of peptides and computes expected peptide fragments (b/y ions, singly and doubly charged, water and ammonia losses) with pyteomics v4.4.2 (https://github.com/levitsky/pyteomics). The next step searches reproducibly detected peptide fragments using multiple synthetic peptide scans. The list of peptide scans is provided by PEAKS v10.5 search result file, and MS raw files are read using Thermo MSFileReader Python bindings (https://github.com/frallain/pymsfilereader). Fragment tolerance was set to 0.02 Da for higher-energy collisional dissociation scans. After that, root scaled intensities of reproducible fragments were correlated between all corresponding endogenous and synthetic peptide scan pairs. The Pearson correlation coefficient, p value, and confidence interval were computed using SciPy v1.6.2 (https://www.scipy.org/). Finally, the scan pair of each peptide with the lowest p value was retained to generate a mirror plot using spectrum_utils v0.3.5 (https://github.com/bittremieux/spectrum_utils).

Bioinformatic analyses

All analyses were conducted on trimmed data, and all alignments were made with STAR on the GRCh38.88 genome version as described in previous sections unless otherwise mentioned.

Identification of MAPs

All liquid chromatography (LC)-MS/MS (LC-MS/MS) data were searched against the relevant database using PEAKS 10.5 (Bioinformatics Solution Inc.). For peptide identification, tolerance was set at 10 ppm and 0.01 Da for precursor and fragment ions, respectively. The occurrences of oxidation (M) and deamidation (NQ) were set as variable modifications. Following peptide identification, we used the modified target-decoy approach built in PEAKS to apply a sample-specific threshold on the PEAKS scores to ensure a false discovery rate (FDR) of 1%, calculated as the ratio between the number of decoy hits and the number of target hits above the score threshold. PEAKS scores corresponding to a 1% FDR for each sample were as following: 14 (hiPSC22), 15 (hiPSC22_IFN), 15 (Fibro-iPSC.1), 14 (Fibro-iPSC.1_IFN), 16 (Fibro-iPSC.2), 14 (Fibro-iPSC.2_IFN). Peptides that passed the threshold were further filtered to match the following criteria: peptide length between 8 and 11 amino acids, binding affinity rank to the sample's HLA alleles <2% based on NetMHCpan-4.0 (Jurtz et al., 2017) (Figure 1A). These filtering steps were done with the use of MAPDP (Courcelles et al., 2020). The HLA types for each sample were determined from the RNA-seq data using OptiType (Szolek et al., 2014).

Identification of paMAPs

To identify paMAP candidates, each MAP and its coding sequence were queried in the relevant iPSC and mTEC canonical proteomes or the iPSC and mTEC 24-nucleotide-long k-mer databases, respectively, as previously described (Laumont et al., 2018). MAPs were retained as paMAP candidates if MAPs were not found in the mTEC canonical proteome, or if all possible MAP-coding sequences (MCS) for a given MAP i) were expressed below 2 KPHM (minimum occurrence of the MCS's 24-nucleotide-long k-mer set per hundred million reads) in mTECs, and ii) had a KPHM fold change superior or equal to 10 in iPSCs compared to mTECs.

Since leucine and isoleucine variants are not distinguishable by standard MS approaches, paMAP candidates for which an existing variant was flagged as a non-paMAP candidate were discarded unless they had a higher RNA expression than the variant. The genomic location of paMAP candidates was assigned by mapping reads containing their coding sequences on the reference genome using IGV (Robinson et al., 2011) and BLAT (tool from the UCSC genome browser). RepeatMasker (in the UCSC genome browser) was used to verify the overlap with EREs.

The RNA expression of paMAP candidates was evaluated in the RNA-seq of GTEx, mTECs, and adult stem cell (ASC) samples (Figure 1A and Table S4; see details in section *RNA expression of MAPs* below) as previously described (Ehx et al., 2021). paMAP candidates containing nucleotide variants in the MCS that did not correspond to known germline polymorphisms (dbSNP149) were classified as mutated MAPs and discarded from the analysis. All MAPs for which at least one MCS was successfully aligned to the reference genome were retained. paMAP candidates that passed the RNA expression filters in GTEx samples and ASCs (see *MAP annotation* in Figure 1A) were considered paMAPs. paMAP candidates that passed the RNA expression filters in GTEx and mTEC samples but not in ASCs were considered saMAPs.

RNA expression of MAPs

The RNA expression of paMAP candidates was evaluated in RNA-seq samples (GTEx (Lonsdale et al., 2013), mTECs (Larouche et al., 2020; Laumont et al., 2018), PSCs (Churko et al., 2017), ASCs (Frisch et al., 2019; Kim et al., 2019; Macrae et al., 2013; Maiga et al., 2016; Pellagatti et al., 2018; Schonberg et al., 2015; Shan et al., 2017; Shen et al., 2019), TCGA, AML (Ehx et al., 2021) and HGSC (Zhao et al., 2020) samples; Figures 1A, 2A, 2B, 3, and Table S4) as previously described (Ehx et al., 2021). Briefly, all MAP amino acid

sequences were reverse translated into all possible nucleotide sequences with an in-house python script (deposited to Zenodo at DOI: 3739257). Next, all these possible sequences were mapped to the genome with GSNAP (Wu and Nacu, 2010), with -n 1000000 option, to locate all genomic regions capable of coding for a given MAP. To confidently capture MAP coded by sequences overlapping splice sites, we also mapped the possible MCS's to the transcriptome (cDNA & non-coding RNA) to extract (samtools faidx with –length 80 option) large portions (80 nucleotides) of reference transcriptomic sequences that we then mapped on the reference genome (GSNAP, with –use-splicing and –novelsplicing = 1 options). For all paMAP candidates, we also performed the genomic alignment of all reads containing their coding sequence. We filtered the outputs of GSNAP to only keep perfect matches between the sequences and the reference to generate a bed file containing all possible genomic regions susceptible to code for a given MAP. By using samtools view (-F256 option), grep and wc (-l option), we counted the number of reads containing the MAP coding sequences at their respective genomic location in each desired RNA-seq sample aligned to the reference genome with STAR (bam file). The BAM Slicing function from the GDC Data Portal (https://docs.gdc.cancer.gov/API/Users_Guide/BAM_Slicing/) was used to count the number of reads at each genomic location in the GRCh38 alignment files for TCGA samples. Finally, all read counts (from different regions and coding sequences) for a given MAP were summed and normalized to the total number of reads sequenced in each assessed sample to obtain a reads-per-hundred-million (RPHM) count.

Prediction of MAP retention time and hydrophobicity index

DeepLC 0.1.16 (Bouwmeester et al., 2021) was used to predict MAP retention times within MAPDP (Courcelles et al., 2020) (Table S2). SSRcalc (Krokhin, 2006) (http://hs2.proteome.ca/SSRCalc/SSRCalcQ.html) was used to calculate hydrophobicity indices based on peptide sequences.

Pathway enrichment analysis

paMAP- or saMAP-source genes (when annotated) were submitted to the "Statistical over-representation test" using Reactome pathways (version 65) as the annotation set in PANTHER v20200728 (Mi et al., 2021). The whole list of *Homo sapiens* genes was used as a reference. The statistical significance of each pathway's enrichment was assessed using Fisher's exact test, with the Bonferroni correction for multiple testing. Only pathways with a positive enrichment and an adjusted p value < 0.05 were kept.

Single-sample gene set enrichment analyses (ssGSEA)

ssGSEA for paMAP- and saMAP-source genes, or for the stemness gene sets compiled by (Miranda et al., 2019), were performed using the GSVA package in R, without normalization, using TPM values quantified using kallisto (Bray et al., 2016) as described in previous sections. The resulting values were subsequently normalized by the absolute difference between the minimum and the maximum (min-max normalization) across gene sets and samples.

Sample clustering

Transcript expression quantifications performed with kallisto (Bray et al., 2016) with default parameters were converted into gene-level counts using the R package *tximport*. The *edgeR* package was then used to filter out lowly expressed genes and perform TMM normalization across the samples of interest. Normalized count per million (cpm) values were used to perform sample clustering based on the expression of the ESC-associated genes from Set 1 in (Ben-Porath et al., 2008). The *heatmap.2* function was used to generate the expression heatmap and sample clustering using the default *hclust* function.

TCGA analyses

All tumor samples from TCGA were included unless otherwise specified (see sample details in Table S4. Testicular germ cell tumor (TGCT) samples were excluded from analyses performed across cancer types due to the presence of canonical paMAPs in the normal testis from GTEx. Mutation rate data were retrieved from Firebrowse (http://firebrowse.org/) as the number of nonsynonymous mutations per base (*rate_non* column). Purity estimates for solid tumors were obtained from (Aran et al., 2015). Molecular subtype and tumor grade information were obtained using the *TCGAbiolinks* (Colaprico et al., 2016) package in R, while the curated clinical-stage data from (Liu et al., 2018).

Predicted paMAP and saMAP presentation

The HLA alleles of each TCGA patient obtained using Polysolver (Castro et al., 2019) were kindly provided by Dr. Hannah Carter (UC San Diego). Promiscuous binders for a given MAP (all HLA alleles capable of presenting the MAP) were obtained using NetMHCpan-4.0 (Jurtz et al., 2017), and were those HLA alleles for which the given MAP had a binding affinity rank <2%. A given MAP (paMAP or saMAP) was considered as presented in a sample if it had an expression >0 RPHM and at least one of the patient's HLA allotypes was a potential binder. If the patient expressed more than one HLA allele capable of presenting a MAP, the MAP was counted as presented once.

Survival analysis

Pan-cancer curated clinical data for TCGA patients were obtained from (Liu et al., 2018). The cancer types for which the overall survival data were not recommended for use by (Liu et al., 2018) were excluded from the analysis. Only samples from primary solid tumors were kept, except for melanoma (SKCM) and AML, for which all samples with data available were used. The hazard ratio for the association between overall survival and the number of paMAPs (or saMAPs) expressed or with predicted presentation (see above) was conducted using the Cox proportional hazards model with the *coxph* function from the R package *survival*. For analyses using the number of paMAPs or saMAPs with predicted presentation (HLA-MAP), the Cox model controlled either for the number of paMAPs or saMAPs expressed per sample, since these two metrics are correlated, and patients expressing more paMAPs and saMAPs are expected to have a worse prognosis.

Correlations and gene expression analyses

All analyses were performed in R. RNA-seq gene expression data for hg38 were retrieved as upper quartile-normalized fragments per kilobase of transcript per million mapped reads (FPKM-UQ) using the *TCGAbiolinks* package for each cancer type. The expression data were then merged across cancers. For genes with duplicate entries, we selected the one with the highest average expression across cancers. Merged FPKM-UQ values were then used to calculate ssGSEA scores for the hallmark gene sets from MSigDB (Liberzon et al., 2015) (http://www.gsea-msigdb.org/gsea/index.jsp), as described in the ssGSEA section above. Non-normalized ssGSEA scores were then used to perform Spearman's correlations with the number of paMAPs and saMAPs expressed per sample within cancer types, using the *rcorr* function. The Spearman correlations using the estimated purity from (Aran et al., 2015) as a covariate were performed using the *pcor*.test function from the *ppcor* package. p-values were adjusted using the Benjamini-Hochberg method.

For the correlation between the number of paMAPs and saMAPs per sample and gene expression, the RNA-seq expression data were retrieved as HTSeq-Counts using the *TCGAbiolinks* package for each cancer type. For genes with duplicate entries, we selected the one with the highest average expression across cancers. The *edgeR* package was then used to normalize counts using the TMM normalization after removing lowly expressed genes using the *filterByExpr* function (min.count of 10). Spearman correlations between the resulting normalized count per million (cpm) values and the number of paMAPs and saMAPs expressed per sample were performed using the *rcorr* function. Finally, the resulting p values were corrected for multiple testing using the Benjamini-Hochberg method with the *p.adjust* function. Only samples from primary solid tumors were kept, except for melanoma (SKCM) and AML, for which all samples with data available were used. The list of immune inhibitory genes was obtained from (Miranda et al., 2019; Thorsson et al., 2018).

HTSeq counts obtained as above were merged across cancer types for the differential gene expression analysis across cancers. For genes with duplicate entries, we selected the one with the highest average expression across cancers. The *edgeR* package was used to remove lowly expressed genes (genes with >1 cpm in >50 samples were kept) and perform TMM normalization. The *limma* package with the *voom* method was then used to assess differential gene expression between samples with high paMAP vs. high saMAP numbers, controlling for tumor purity and cancer types. Only samples with purity estimates from (Aran et al., 2015) were included. TGCT was also excluded. Genes with absolute fold change >2 and adjusted p value < 0.05 were considered differentially expressed.

Methylation and focal CNV analyses

Processed level 3 methylation data (HM27 for TCGA-OV and HM450 for all other cancer types) for TCGA samples were retrieved using the *TCGAbiolinks* package. Only probes within 2 kb of the transcription start site of a given paMAP- or saMAP-source gene were kept. We then performed Spearman correlations between the RPHM expression of each MAP of interest with the beta values for the respective gene within cancers. The mean beta value was used for genes associated with multiple probes. The correlation results for TCGA-OV using HM27 were merged with the HM450 results for all other cancers for plotting. The p values were adjusted for multiple testing using the Benjamini-Hochberg method. Only HM450 beta values were used for correlations across cancer types without TGCT.

For focal CNV correlations, processed hg38 gene-level copy number scores were retrieved using the *TCGAbiolinks* package. DNA copy-number changes within paMAP or saMAP coding regions were used to perform Spearman correlations with the expression (RPHM) of each MAP of interest within cancers. Mean copy-number values were used for multiple segments associated with a MAP-coding region. TGCT samples were excluded from correlation analyses across cancers. p values were adjusted for multiple testing using the Benjamini-Hochberg method.

Genomic correlations with paMAP and saMAP counts

TCGA Unified Ensemble "MC3" somatic mutation (SNP and INDEL) calls (Ellrott et al., 2018) were downloaded from the UCSC Xena Functional Genomics Explorer (Goldman et al., 2020) (https://xenabrowser.net/), where only gene-level non-silent mutation calls with filter = PASS were kept and converted to binary values (1, non-silent mutation; 0, WT). TCGA pan-cancer gene-level copy number variation (CNV) estimated using the GISTIC2 threshold method were downloaded from UCSC Xena, where estimated values were threshold converted to −2, −1, 0, 1, 2, representing homozygous deletion, single copy deletion, diploid normal copy, low-level copy number amplification, or high-level copy number amplification, respectively. Patients with more than one sample, those with TCGT, and those that did not have both somatic mutation and CNV data, were excluded from the analysis. We then used the Chi-squared test to compare the number of patients expressing >0 paMAPs and saMAPs (>2 RPHM) vs. the others, among patients with WT or mutant variants of each gene. The same analysis was repeated for gene-level amplifications (1 and 2), and deletions (−2 and −1). Features were further selected for plotting based on their prevalence in paMAP- and saMAP-expressing samples, high statistical significance or association with signaling pathways of interest.

For Figure S5D, the "MC3" somatic mutation (SNP and INDEL) calls downloaded from UCSC Xena were used. Patients with more than one sample were excluded from the analysis. We then used Fisher's exact test to compare the number of patients expressing ≥ median numbers of paMAPs and saMAPs (>2 RPHM) vs. the others among patients with WT or mutant variants of each gene. Comparisons with a p value < 0.05 were kept, and the top three genes with the most prevalent mutations in cancer samples expressing paMAPs and saMAPs above the median number per cancer type were plotted. Genes fulfilling these criteria in at least one cancer type were plotted in all cancer types if they had p value < 0.05 to emphasize common genomic events correlated with paMAP and saMAP expression across cancer types.

Immune infiltration analysis

xCell enrichment scores were calculated in R using the *rawEnrichmentAnalysis* function, which omits adjusting the raw scores (Aran et al., 2017b). Spearman correlations were performed between the raw cell type enrichment scores and the paMAP and saMAP counts per sample or the ssGSEA enrichment score for paMAP- and saMAP-source genes (using FPKM-UQ values as above), followed by p value adjustment (Benjamini-Hochberg method) with the *p.adjust* function in R. Only primary solid tumor samples were used for correlations, except for SKCM and LAML.

Immunogenicity assays

Immunogenicity predictions

Immunogenicity predictions of paMAPs and saMAPs were performed using Repitope (Ogishi and Yotsuyanagi, 2019). Feature computation was performed with the predefined MHCI_Human_MinimumFeatureSet variable and the FeatureDF_MHCI and FragmentLibrary files provided on the Mendeley repository of the package (version July 13, 2019; https://doi.org/10.17632/sydw5xnxpt.1).

Peptide-specific T cell functional expansion

Peptide-specific $CD8^+$ T cells from 4 healthy donors were expanded *in vitro* (D11, D12, D13, and D14). The expanded cells from D12 were used for FEST and tetramer staining assays, whereas the expanded cells from D11, D13, and D14 were used only for tetramer staining assays.

T cells were cultured as previously described, with minor modifications (Danilova et al., 2018). Briefly, on day 0, thawed PBMCs from each healthy donor (BioIVT) were T cell-enriched using the Human Pan T cell isolation kit (Miltenyi Biotec). T cells were resuspended at 2×10^6/mL in AIM V media supplemented with 50 μg/mL gentamicin (ThermoFisher Scientific) and 1% HEPES. The T cell-negative fraction was irradiated at 30 Gγ, washed, and resuspended at 2×10^6/mL in AIM V media supplemented with 50 μg/mL gentamicin and 1% HEPES. 2.5 mL per well of both T cells and irradiated T cell-depleted cells were added to a 6-well plate, along with either a peptide (GenScript) alone, a peptide pool (up to 6 MAPs per pool, 1 μg/mL final concentration for each MAP) or without peptide. Cells were cultured for 10 days at 37°C, 5% CO_2. On day 3 and 7, half the culture media was replaced with fresh culture media containing 100 IU/mL IL-2, 50 ng/mL IL-7, and 50 ng/mL IL-15 (day 3) and 200 IU/mL IL-2, 50 ng/mL IL-7, and 50 ng/mL IL-15 (day 7). On day 10, thawed PBMCs from the same donor were used to generate a new batch of T cell-depleted cells. These cells were irradiated at 30 Gγ and added to cultures at a 1:1 T cell:non-T cell ratio, along with 1 μg/mL of relevant peptide(s) or without peptide. On day 13 and 17, at least half the culture media was replaced with fresh culture media (final concentrations: 100 IU/mL IL-2, 25 ng/mL IL-7, and 25 ng/mL IL-15). On day 20, cells were harvested to perform tetramer staining and/or FEST assays.

FEST assays

For FEST assays, $CD8^+$ cells were further isolated using the Human $CD8^+$ T cell Isolation Kit (Miltenyi Biotec). As a negative control, $CD8^+$ T cells were also isolated from freshly thawed uncultured PBMCs of the same healthy donor. DNA was extracted from $CD8^+$ T cells using a QIAGEN DNA blood mini kit (QIAGEN). TCR Vβ CDR3 sequencing was performed using the survey resolution of the immunoSEQ platform (Adaptive Biotechnologies). Raw data exported from the immunoSEQ portal were processed with the FEST web tool (www.stat-apps.onc.jhmi.edu/FEST) with the parameters described in Table S7.

Tetramer staining

Following 20 days coculture using peptide-loaded T cell-depleted cells and cytokines, 1×10^6 cells were stained for 30 min at 4°C with custom-made peptide-HLA tetramers (NIH) and then stained for 30 min at 4°C with a CD8 monoclonal antibody (BD Biosciences). Cells were washed with PBS (containing 2% FBS) before acquisition with a Celesta cytometer (BD Biosciences). Data were analyzed using the FlowJo v10 Software (BD Biosciences).

Ex vivo peptide-specific T cell quantification

Frequencies of peptide-MHC-specific CD8 T cells without *in vitro* expansion were also determined for the four healthy donors (D11, D12, D13, and D14). 50×10^6 to 180×10^6 of thawed PBMCs were stained with 1 μg/mL of PE- and 5 μg/mL APC-labeled peptide-MHC tetramers (NIH Tetramer Core Facility) for 30 minutes at 4°C. After washing with ice-cold sorting buffer (PBS, 2 mM EDTA, 0.5% BSA), cells were resuspended in 450 μL ice-cold sorting buffer, and 50 μL of anti-PE and anti-APC antibody conjugated magnetic microbeads (Miltenyi Biotec), then incubated for 20 minutes at 4°C. Cells were then washed, and $tetramer^+$ cells were magnetically enriched with LS columns (Miltenyi Biotec), following the manufacturer's instructions. The resulting $tetramer^+$-enriched fractions were stained with APC-H7-conjugated anti-CD3, BB515-conjugated anti-CD8, BV510-conjugated anti-CD4, PerCP-Cy5.5-conjugated anti-CD14, CD16, CD19 antibodies (BD Biosciences) for 30 min at 4°C and washed. The entire stained sample was then analyzed with 7-Aminoactinomycin D (7-AAD, BD Biosciences) on a FACS Celesta cytometer (BD Biosciences), and fluorescent counting beads (Thermo Fisher Scientific) were used to normalize the results. As a control, we also enriched the antigen-specific CD8 T-cell repertoires targeting 3 HLA-A*02:01 restricted immunodominant epitopes: $MelanA_{27}$ (a melanoma-derived Ag, ELAGIGILTV), $NS3_{1073}$ (derived from hepatitis C virus, CINGVCWTV), and Gag_{77} (derived from human immunodeficiency virus, SLYNTVATL).

In vitro peptide-specific CD8 T cell sorting and expansion

$NS3_{1073}$- and VTLSTYFHV-specific CD8 T-cell lines derived from PBMCs (from healthy donor D25) were obtained after magnetic enrichment (Miltenyi Biotech) and sorting of tetramer positive CD8+ T cells with a FACS Aria III sorter (BD Biosciences). Tetramers were produced by UV-exchange technology according to the manufacturer's instructions (Flex-T™ from BioLegend for

VTLSTYFHV-HLA-A*02:01-PE and APC-tetramers) or from the NIH Tetramer Core Facility for CINGVCWTV-HLA-A*02:01-PE and APC-tetramers. Cells were then expanded *in vitro* under non-specific conditions using IL-2 (250 UI/mL, Peprotech), leukoagglutinin (1 μg/mL, Invitrogen), irradiated PBMCs and B lymphoblastoid cells lines (B-LCL). T-cell lines were maintained for at least 2 weeks without restimulation in RPMI 1640 medium (Gibco) supplemented with 10% heat-inactivated fetal bovin serum (Gibco), 1 mM L-glutamine (Gibco), 1% penicillin-streptomycin (Gibco) and 250 IU/mL IL-2 (Peprotech) before analysis and 3 weeks before each restimulation. The specificity and purity of these two peptide-specific CD8 T-cell lines were superior to 80%.

Cytotoxicity assay

THP-1 and OCI-AML3 cells were labeled with CFSE (CellTrace^{TM} CFSE Cell Proliferation Kit, Invitrogen) according to the manufacturer's instructions, and then were pulsed with 10 ug/mL peptide (VTLSTYFHV or $NS3_{1073}$) or DMSO for 2.5-4 hours. Thereafter, AML cells were plated at 5,000 cells per well in a 96-well U-bottom plate to use as targets in cytotoxicity assays. Effector cells were added at the indicated effector-to-target (E:T) ratios (range 6.5–10.5) in a final volume of 200 μL per well. After 24–25 hours of co-incubation at 37°C, cells were stained with 7-AAD and analyzed using a ZE5 flow cytometer. Viable target cells remaining were gated based on CFSE+/7-AAD- staining, and the specific lysis was calculated using the following formula. Negative specific lysis (%) values were converted to 0.

$$\textit{Specific lysis}\ (\%) = \frac{\#\ \textit{viable target cells in target only condition} - \#\ \textit{viable target cells in effector + target condition}}{\#\ \textit{viable target cells in target only condition}} \times 100$$

QUANTIFICATION AND STATISTICAL ANALYSIS

Unless mentioned in the figure legends, all statistical tests comparing two conditions were performed with the unpaired two-tailed Wilcoxon test in R. For multiple pairwise comparisons, the p values were adjusted using the Benjamini-Hochberg method using the *compare_means* function in R. All boxplots show the median and interquartile range (IQR), and whiskers extend to the largest value no further than 1.5 * IQR from the box hinges. Unless mentioned, all correlations were performed using Spearman's correlation coefficient. Plots and statistical analyses were performed in R v3.6.5 or Python v3.6.7.