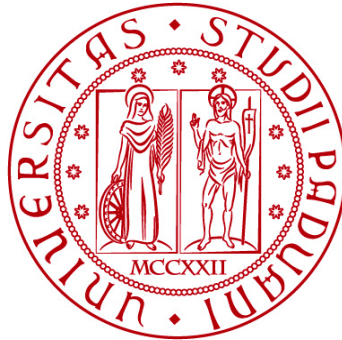


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biologia Molecolare



ELABORATO DI LAUREA

**Il ruolo delle uORFs codificanti uPeptidi e ligandi HLA
nella regolazione dell'espressione genica a livello
traduzionale in trascritti associati al cancro.**

**Tutor: Prof. Gianluca Occhi
Dipartimento di Biologia**

Laureanda: Giulia Casarin

ANNO ACCADEMICO 2023/2024

Sommario

1. STATO DELL'ARTE.....	4
1.1 STRUTTURA E FUNZIONE DELLE uORFs NEI TRASCritti UMANI CODIFICANTI PROTEINE.....	4
1.2 L'ESPRESSIONE CELLULARE DI uPEPTIDI CODIFICATI DA uORFs.....	5
1.3 GLI uPEPTIDI POSSONO FUNGERE DA ANTIGENI IMMUNOGENICI.....	6
1.4 L'IMPATTO DELLA REGOLAZIONE MEDIATA DA uORFs IN UN CONTESTO TUMORALE.....	6
2. APPROCCIO SPERIMENTALE.....	7
2.1 SELEZIONE DELLE uORF PER L'ANALISI FUNZIONALE.....	7
2.2 SAGGIO DELLA DOPPIA LUCIFERASI.....	8
2.3 RILEVAZIONE DEGLI uPEPTIDI MARCATI MEDIANTE IMMUNOBLOTTING.....	10
2.4 MICROSCOPIA A IMMUNOFLUORESCENZA.....	11
2.5 RACCOLTA DEI CAMPIONI PER L'ANALISI DELL'IMMUNOPEPTIDOMA.....	11
2.6 ISOLAMENTO DEI LIGANDI HLA.....	12
2.7 CROMATOGRAFIA LIQUIDA CON SPETTROMETRIA DI MASSA TANDEM (LC-MS-MS).....	12
2.8 VALIDAZIONE DEGLI SPETTRI.....	12
2.9 REFOLDING E PRIMING DEI LINFOCITI T CON CELLULE PRESENTANTI L'ANTIGENE.....	13
2.10 COLORAZIONE DEI TETRAMERI E DELLE CITOCHINE INTRACELLULARI (ICS).....	13
3. RISULTATI E DISCUSSIONE.....	14
3.1 REGOLAZIONE DELL'ESPRESSIONE GENICA A LIVELLO TRADUZIONALE MEDIATA DA uORFs IN TRASCritti ASSOCIATI AL CANCRO.....	14
3.2 TRADUZIONE E LOCALIZZAZIONE CELLULARE DEGLI uPEPTIDI.....	15
3.3 L'ANALISI DELL'IMMUNOPEPTIDOMA HA IDENTIFICATO LA PRESENTAZIONE DEI LIGANDI HLA.....	17
3.4 ATTIVAZIONE DEI LINFOCITI T CD8+ INDOTTA DAI LIGANDI HLA.....	18
4. CONCLUSIONI.....	19

ABSTRACT

Le upstream open reading frame (uORFs) sono brevi sequenze codificanti, situate nella regione 5'UTR dei trascritti eucariotici, coinvolte nella regolazione dell'espressione genica a livello traduzionale del trascritto a valle a cui sono associate. Dati recenti hanno dimostrato che queste sequenze possono codificare per proteine biologicamente attive e per peptidi presentati dalle molecole HLA, suggerendo una loro implicazione nello sviluppo e nella sopravvivenza delle cellule tumorali. Nello studio riportato è stato esaminato l'effetto regolatorio di 29 uORFs associate a 13 trascritti correlati al cancro osservando l'espressione cellulare e la localizzazione degli uPeptidi mediante immunoblotting e immunofluorescenza. In seguito è stata eseguita l'analisi dell'immunopeptidoma basata sulla spettrometria di massa al fine di identificare peptidi derivati dalle uORF e presentati in HLA. I risultati ottenuti hanno rivelato una frequente espressione delle uORF in uPeptidi caratterizzati da una localizzazione cellulare specifica nei tessuti umani benigni e maligni suggerendo un ruolo regolatorio diffuso delle uORF e degli uPeptidi in un contesto tumorale. Inoltre, tramite analisi di tipo comparativo, utilizzando immunopeptidomi derivati da tessuti benigni e maligni, è stato possibile delineare un insieme di ligandi HLA derivati dalle uORF associati al tumore in grado di indurre una risposta specifica da parte dei linfociti T.

1. STATO DELL'ARTE

1.1 STRUTTURA E FUNZIONE DELLE uORFs NEI TRASCritti UMANI CODIFICANTI PROTEINE

All'interno di una molecola di RNA messaggero (mRNA) maturo sono presenti, a monte e a valle della sequenza codificante la proteina principale (CDS), due regioni con funzione regolatoria denominate 5'UTR e 3'UTR. La regione 5'UTR rappresenta la sequenza leader di un mRNA maturo e svolge un ruolo fondamentale di controllo dell'espressione genica grazie alla presenza di upstream ORFs (uORFs), siti di legame per i microRNA e siti di inizio interni per i ribosomi (IRESs). Mutazioni a livello di questi elementi in cis possono andare a influire sull'espressione genica e risultano essere spesso correlati allo sviluppo di diverse patologie, incluso il cancro.

Le uORFs sono definite da un codone di inizio a livello del 5'UTR di un trascritto e un codone di terminazione che precede nella sequenza il codone di stop della CDS quindi, possono iniziare e concludersi prima della CDS oppure sovrapporsi parzialmente ad essa. Analisi recenti hanno individuato uORFs contenute in quasi la metà dei trascritti umani codificanti proteine ma in realtà la scoperta riguardante l'utilizzo di codoni di inizio non canonici o affini (aTISs), nel processo di traduzione delle uORF, ha ampliato le stime relative alla prevalenza di questi elementi.

La presenza di uORF funzionali risulta generalmente associata a una soppressione della traduzione del trascritto a valle; tuttavia la funzione esplicita da parte delle uORFs non è sempre costante e può variare in base alle condizioni della cellula.

Sono stati descritti diversi meccanismi molecolari che vedono il coinvolgimento delle uORFs nella regolazione dell'espressione genica a livello traduzionale [1]. Questi includono il riinizio della traduzione, in cui uORF e CDS sono tradotti in serie; un secondo meccanismo, definito leaky-scanning, consiste nella scansione debole da parte della subunità minore ribosomiale del codone di inizio dell'uORF nel caso in cui questo non sia supportato da un contesto adeguato che ne permetta il riconoscimento avviando la traduzione. Infine un altro meccanismo denominato ribosome-stalling consiste in una diminuzione dell'attività di traduzione di un trascritto causata dalla ritenzione del ribosoma a livello dell'uORF associata. Una pausa prolungata del ribosoma durante la fase di elongazione può infatti provocare una situazione di stallo del ribosoma in cui i ribosomi successivi si accodano dietro al ribosoma in pausa. Ciò è dovuto alla presenza del codone di terminazione dell'uORF che viene riconosciuto come prematuro e innesca il processo di nonsense-mediated mRNA decay (NMD) portando, nella maggior parte dei casi, alla degradazione dell'intero mRNA. La regolazione della traduzione da parte delle uORFs può essere legata anche alla dissociazione dei fattori di inizio della traduzione; dopo la traduzione dell'uORF, la subunità maggiore del ribosoma si distacca dall'mRNA mentre eIF3 rimane associato. Gli

altri fattori necessari per avviare la traduzione devono essere quindi reclutati nuovamente e l'efficienza di reclutamento dipende dalla distanza che intercorre tra uORF e CDS. In genere, la presenza di un lungo 5'UTR è associata a una riduzione dei livelli di traduzione della CDS a valle mentre trascritti con 5'UTR più brevi presentano livelli di espressione più elevati.

Tipicamente i codoni di inizio più frequenti in corrispondenza delle uORFs sono AUG e CUG. L'utilizzo di codoni di inizio differenti per avviare il processo di traduzione può dipendere dalle condizioni di stress della cellula e può svolgere un ruolo chiave nel processo di regolazione. Nel caso del gene ATF4 nei mammiferi, l'espressione è regolata dalla presenza di 2 uORFs nel 5'UTR. In risposta a condizioni di stress il meccanismo coinvolto porta a una regolazione positiva della traduzione della CDS a valle. La fosforilazione del fattore eIF2 induce una diminuzione dei livelli di eIF2-GTP aumentando così il tempo necessario affinché il ribosoma diventi competente per l'inizio della traduzione. In tale situazione l'uORF2, un elemento inibitorio che blocca l'espressione di ATF4, non viene riconosciuto dalla subunità ribosomiale minore. Questa quindi procede la scansione lungo l'mRNA iniziando il processo di traduzione di ATF4. In condizioni normali la stessa uORF viene invece tradotta portando alla formazione di un uPeptide e andando a inibire la traduzione della CDS a valle [2].

1.2 L'ESPRESSIONE CELLULARE DI uPEPTIDI CODIFICATI DA uORFs

I progressi nello sviluppo delle metodiche di profiling dei ribosomi hanno permesso l'identificazione di numerose uORFs tradotte in proteine nel genoma umano, delle quali circa il 20% risultano essere conservate. Queste sequenze possono quindi sia codificare per peptidi specifici definiti uPeptidi sia svolgere la loro funzione regolatoria a livello traduzionale sull'ORF principale a cui sono associate. Alcuni uPeptidi sono espressi stabilmente e rilasciati nel citosol della cellula contribuendo al micropeptidoma della stessa; anche in questo caso presentano una funzione regolatoria che esplicano individualmente o associati ad altre proteine all'interno di un complesso proteico. Ad esempio l'uPeptide uPEP2, presente nel trascritto del gene PRKCH codificante la proteina PKC- η , contiene un dominio che funge da pseudosubstrato per PKC tramite il quale si lega all'enzima inibendone l'attività chinasi. Nell'uomo esistono 15 isoforme di questo enzima suddivise in tre sottofamiglie sulla base del secondo messaggero che necessitano e uPEP2, in particolare, esplica la sua funzione inibitoria solamente nei confronti delle nuove PKC ma non ha alcun effetto sulle PKC convenzionali e atipiche. È stata dimostrata la capacità da parte di uPEP2 di sopprimere la progressione e la metastasi del tumore in diversi modelli di cancro al seno portando a una proliferazione ridotta, una maggior morte cellulare e livelli di espressione ridotti di nuove PKC, parallelamente a una diminuzione della fosforilazione dei substrati PKC [3]. Si può osservare pertanto il ruolo di alcune

uORFs nel codificare peptidi biologicamente attivi, che agiscono in cis e/o in trans prendendo parte a meccanismi di segnalazione cellulare, aprendo nuove prospettive per quanto riguarda la regolazione delle proteine eucariotiche e la biologia del cancro.

1.3 GLI uPEPTIDI POSSONO FUNGERE DA ANTIGENI IMMUNOGENICI

Gli uPeptidi costituiscono la categoria maggiormente rappresentata di peptidi criptici finora rilevati. I peptidi criptici sono epitopi presentati dalle molecole MHC di classe I derivanti da una regione non codificante; questi, se presenti nel citoplasma della cellula, possono essere poliubiquitinati e giungere, tramite le vie di degradazione nel proteasoma a livello del quale vengono processati. Il complesso proteico TAP li trasporta al reticolo endoplasmatico in cui avviene il caricamento del peptide in MHC-I. Infine il complesso MHC-peptide antigenico, mediante trasporto vescicolare, viene esposto sulla superficie cellulare affinché avvenga il riconoscimento da parte di un linfocita T CD8+ che presenta un TCR specifico per quel complesso. Mediante questo meccanismo alcuni uPeptidi, denominati uLigandi HLA, contribuiscono a quello che viene definito l'immunopeptidoma di una cellula ossia l'insieme dei peptidi presentati dalle molecole MHC sulla superficie delle cellule presentanti l'antigene (APC). Specialmente in un contesto cellulare patologico, come in una trasformazione maligna, la traduzione differenziale delle uORFs e il differente processamento degli uPeptidi potrebbero produrre uLigandi HLA specifici per il tumore in grado di indurre una risposta citotossica specifica da parte dei linfociti T CD8+. Tali cambiamenti a livello delle vie cellulari essenziali, specifici di un determinato contesto tumorale, possono quindi portare a un'alterazione nel processo di traduzione di un uORF che si riflette in un'alterata espressione dell'uPeptide corrispondente. In tal modo la composizione dell'immunopeptidoma caratteristico di quel determinato tipo cellulare viene modificata. Questi cambiamenti a livello di immunopeptidoma potrebbero quindi permettere una discriminazione tra cellule trasformate e cellule sane.

1.4 L'IMPATTO DELLA REGOLAZIONE MEDIATA DA uORFs IN UN CONTESTO TUMORALE

La sintesi proteica è un processo finemente controllato ed eventuali alterazioni nei meccanismi di regolazione coinvolti possono contribuire all'insorgere di malattie come il cancro. Le cellule tumorali sono in grado di gestire lo stress durante la tumorigenesi attivando selettivamente la traduzione di specifiche proteine cruciali per lo sviluppo e la progressione del tumore. Le uORFs in quest'ambito regolano la sintesi proteica in condizioni di stress aiutando la cellula tumorale, che necessita di sintetizzare più proteine per sostenere la sua rapida crescita e

divisione, a sopravvivere. Ad esempio, le uORFs del gene CEBPB possono indurre l'utilizzo di siti d'inizio differenti per la traduzione della CDS a valle portando alla sintesi di tre diverse isoforme del fattore di trascrizione C/EBP β . Condizioni di stress promuovono la produzione dell'isoforma piú corta, LIP, in grado di promuovere la tumorigenesi e la metastasi. È stato dimostrato come l'eliminazione del codone AUG dell'uORF associata al gene annulli l'espressione di LIP a dimostrazione del fatto che l'espressione di quest'isoforma dipenda strettamente dall'uORF in questione.

Uno studio ha identificato circa 400 mutazioni geniche a livello delle uORFs associate al cancro [4] e sono state individuate mutazioni loss of function nelle uORF di EPHB1 e MAP2K6 associate allo sviluppo di tumori. È stato dimostrato come una delezione nell'uORF di CDKN1B determini una riduzione del livello della proteina p27, un soppressore tumorale, favorendo in tal modo la progressione del tumore. Altri meccanismi che coinvolgono le uORFs nel cancro sono associati all'attività dei fattori trans-acting che possono interagire con le uORFs influenzando la loro attività regolatoria. Infine, anche il livello di magnesio cellulare può influire sulla traduzione di alcuni mRNA contenenti uORFs coinvolti nel metabolismo energetico delle cellule tumorali tramite la via AMPK/mTORC2 [5]. Queste scoperte evidenziano quindi il ruolo delle uORFs nella regolazione di oncogeni e oncosoppressori e l'impatto che possono avere eventuali deregolazioni nei meccanismi mediati da queste sequenze sullo sviluppo e progressione del tumore.

2. APPROCCIO SPERIMENTALE

2.1 SELEZIONE DELLE uORF PER L'ANALISI FUNZIONALE

Sono stati selezionati i geni contenenti le uORFs per l'analisi sperimentale in base al coinvolgimento documentato, da un punto di vista oncogeno, dei trascritti ad essi associati. Sono stati scelti i geni dotati di determinati requisiti quali un'alta conservazione della sequenza uORF o nei quali fosse presente l'evidenza di una traduzione attiva delle uORFs. Tra i geni presi in considerazione, sono quindi stati selezionati gli oncogeni presenti nei pathways tumorali canonici RTK/RAS, cell cycle, PI3K, P53, MYC e WNT nei quali la deregolazione di una singola proteina può mediare effetti oncogeni. Le uORFs dei geni sono state quindi classificate secondo uno specifico punteggio uORF, descritto in un precedente studio di McGillivray et al, in grado di fornire una stima della loro rilevanza funzionale in base a determinate caratteristiche come lunghezza, posizione e conservazione delle sequenze, considerando inoltre la loro influenza sull'espressione del gene associato a valle. Le uORFs con una lunghezza minore di 24 bp sono state eliminate non essendo rilevabili tramite immunoblot. Per ogni gene sono poi stati selezionati l'uAUG e il sito d'inizio della traduzione alternativo (aTIS) con i

punteggi uORF maggiori e il uAUG/aTIS con lo score uPeptide più alto; dopo aver controllato che gli uAUGs e aTISs selezionati fossero presenti in tutte le varianti trascrizionali dei geni in analisi è stata scelta una sequenza leader dei trascritti (TLS) rappresentativa come riferimento da utilizzare negli esperimenti successivi. Nel caso in cui fossero presenti differenti varianti trascrizionali contenenti tutte le uORFs in analisi sono state preferite TLS a bassa complessità per semplificare la gestione sperimentale. In ultima analisi, dall'insieme dei trascritti ottenuti ne sono stati selezionati 13 in base alla letteratura precedente che indicasse un qualche loro coinvolgimento da un punto di vista oncogeno o che suggerisse un'attiva regolazione dei rispettivi geni tramite uORFs.

Attraverso il processo di filtrazione appena descritto è stato quindi possibile selezionare un set di 29 uORF presenti in 13 geni associati al tumore (ASNSD1, ATF5, CDK4, CTNBB1, EGFR, ERBB2, JAK2, KRAS, MAPK1, MDM2, MET, MYC, RET) con lo scopo di analizzare il loro effetto regolatorio a livello traduzionale sulla sequenza CDS a valle.

2.2 SAGGIO DELLA DOPPIA LUCIFERASI

Le 29 uORF selezionate sono state testate per il loro impatto regolatorio sulla traduzione della CDS a valle con il saggio della doppia luciferasi. Questa tecnica, ampiamente utilizzata con il fine di valutare l'effetto di una sequenza regolatrice di DNA o RNA sull'espressione genica, si basa sull'utilizzo simultaneo di due geni reporter codificanti due enzimi distinti: la luciferasi di lucciola (da *Photinus pyralis*) e la luciferasi di Renilla (da *Renilla reniformis*), consentendo di misurare e confrontare entrambi i segnali bioluminescenti prodotti dai due enzimi reporter all'interno di uno stesso campione in analisi.

Il primo gene codificante la luciferasi Firefly, ossia il gene reporter sperimentale, è accoppiato a un promotore regolato ed è sotto controllo della sequenza regolatrice in analisi. Il secondo gene, codificante la luciferasi Renilla, funge da controllo interno ed è attivato da un promotore costitutivo; questo permette di normalizzare la luminescenza dell'altro reporter con lo scopo di compensare eventuali variazioni nell'efficienza di trasfezione o nella vitalità delle cellule, riducendo in tal modo la variabilità sperimentale. Entrambi i geni reporter vengono quindi introdotti all'interno di due plasmidi reporter e cotrasfettati nelle cellule in coltura. Dopo il periodo di incubazione, le cellule trasfettate con i due plasmidi contenenti i geni reporter vengono lisate e infine viene effettuata la misurazione del segnale luminescente prodotto dall'attività dei due enzimi in seguito all'aggiunta del loro substrato specifico.

L'intensità della luce prodotta dalla luciferasi di lucciola, proporzionale alla sua attività enzimatica, può essere dunque rilevata tramite un lettore di micropiastre e successivamente confrontata con quella della luciferasi Renilla in modo tale da essere normalizzata ottenendo un risultato più accurato [6].

Nell'esperimento in questione, le TLS dei trascritti wild type, incluso il codone di inizio dell'ORF principale (mORF) e la base Kozak in posizione + 4, sono state sintetizzate commercialmente utilizzando GeneArt (Thermo Fisher Scientific), isolate mediante enzimi di restrizione appropriati e infine legate in un plasmide reporter di controllo traduzionale (TCRP) basato sul vettore di base pGL3. Sono state poi introdotte delle mutazioni a livello dei codoni d'inizio delle uORFs, tramite mutagenesi sito specifica (SDM), mutando il codone d'inizio in CUC o in alternativa in un codone di non inizio. Introducendo queste mutazioni è stato possibile misurare l'impatto relativo di eventuali mutazioni presenti a livello dell'uORF nella TLS sulla regolazione dell'espressione genica del trascritto associato. La correttezza delle mutazioni introdotte è stata verificata tramite sequenziamento Sanger.

Sia i TCRP contenenti l'uORF wild type che quelli contenenti quella mutata (Δ uORF) sono stati cotrasfettati nelle cellule HEK293T, una linea cellulare derivante da cellule renali embrionali umane immortalizzate, insieme al vettore di controllo luciferasi Renilla (pRL-CMV, Promega). 44 ore dopo la trasfezione, le cellule sono state lavate con PBS e lisate mediante un buffer di lisi della luciferasi contenente un cocktail di inibitori di proteasi per evitare la proteolisi delle proteine. I lisati cellulari sono stati poi centrifugati per separare i detriti cellulari, il DNA cromosomiale e altre impurità dal supernatante contenente le proteine d'interesse; infine sono state effettuate misurazioni in triplicato di ciascun lisato, utilizzando un lettore di micropiastre, per valutare le attività delle luciferasi Firefly e Renilla.

I livelli di espressione genica delle luciferasi Firefly e Renilla sono stati misurati effettuando una RT-qPCR ossia una tecnica che permette l'amplificazione e il simultaneo monitoraggio, ciclo per ciclo, della quantità di amplicone prodotto permettendo così la quantificazione del template presente in un determinato campione in analisi. Nello specifico, il DNA complementare (cDNA), sintetizzato tramite l'enzima trascrittasi inversa, funge da template per la reazione di real time PCR. L'amplificazione del template avviene all'interno di uno strumento apposito, un termociclatore, dotato di un sistema di rilevazione della fluorescenza all'interno del pozzetto di reazione. All'interno del pozzetto è presente una molecola intercalante, come il sybr green, che emette un segnale quando lega la doppia elica di DNA consentendo di rilevare, per ogni ciclo, la quantità di fluorescenza emessa man mano che il template viene amplificato. Più amplicone si genera nel pozzetto della piastra maggiore è il segnale di fluorescenza rilevato. Ciò permette di identificare tramite un software il ciclo soglia, ossia il ciclo di reazione della PCR in cui il segnale è al di sopra di quello rilevabile, analizzando la curva di amplificazione nella fase di aumento esponenziale della quantità di prodotto della PCR. All'interno di questa fase, esiste infatti una proporzionalità che rimane più o meno lineare tra la quantità di template iniziale e la quantità di amplicone prodotto.

2.3 RILEVAZIONE DEGLI uPEPTIDI MARCATI MEDIANTE IMMUNOBLOTTING

Con il fine di rilevare gli uPeptidi prodotti a partire dalle uORFs selezionate, per determinare se queste venissero o meno tradotte in vitro, come primo passaggio è stato necessario creare dei vettori di espressione per gli uPeptidi. I tag peptidici 3xHA, ossia dei peptidi biologicamente attivi utilizzati per facilitare la rilevazione delle proteine in analisi, sono stati inseriti all'interno del vettore pcDNA3.1(+) utilizzando i siti di restrizione BamHI e XbaI. In seguito mediante PCR sono state amplificate le sequenze TLS con le uORFs in esame dai vettori della luciferasi Firefly eliminando il codone di terminazione dell'uORF e includendo le sporgenze del sito di restrizione HindIII e BamHI per la ligazione. Gli ampliconi così prodotti sono stati inseriti a monte del tag 3xHA nel vettore pcDNA3.1(+)-3xHA con il codone di inizio dell'uPeptide in-frame con 3xHA in modo da permettere una traduzione continua della uORF con il tag aggiunto. Tramite questo procedimento sono stati quindi generati dei vettori di espressione per tutte le uORF presentanti il codone d'inizio AUG, per le uORF con i punteggi uPeptide più alti e per l'uORF CTNNB1 aTIS contenente un codone alternativo di inizio UUG. Infatti nel caso del gene CTNNB1 le uORFs con i punteggi uORF e uPeptide più alti, quindi con una maggior importanza da un punto di vista regolatorio, sono state mappate su varianti trascrizionali distinte del gene in questione. I vettori di espressione generati per tutte le uORFs in analisi sono stati successivamente trasfettati in cellule HEK293T e dopo 44 ore le cellule sono state trattate con MG123, un inibitore del proteasoma, per evitare la degradazione degli uPeptidi. Successivamente le cellule sono state lavate con PBS e lisate utilizzando un buffer di lisi contenente inibitori di proteasi per preservare le proteine d'interesse. Infine, dopo aver centrifugato il lisato, è stato raccolto il surnatante contenente le proteine estratte.

La concentrazione delle proteine estratte è stata calcolata mediante BCA, ossia un saggio colorimetrico per determinare la concentrazione proteica in un campione: in condizioni alcaline la proteina presente nell'estratto proteico in analisi lega lo ione rame rameico (Cu^{2+}) riducendolo a ione rameoso (Cu^+), il quale coordina due molecole di acido bicinconinico formando un complesso solubile in acqua. Dopodiché si opera la misurazione a 560 nm, ossia la lunghezza d'onda corrispondente al picco di assorbimento del complesso che si è formato.

Successivamente le proteine estratte sono state separate in base alla loro dimensione tramite SDS-PAGE, trasferite su una membrana e rilevate con anticorpi specifici mediante immunoblotting. L'immunoblotting è una tecnica immunochimica che deriva dall'elettroforesi e prevede il trasferimento delle molecole da un gel a una membrana mantenendo la separazione ottenuta con il gel; questa metodologia unisce due tecniche con risoluzione elevata, l'elettroforesi e il legame antigene-anticorpo, consentendo una capacità risolutiva finale ancora più alta.

Dopo l'iniziale separazione elettroforetica del lisato proteico mediante SDS-PAGE al 22% è stato effettuato il trasferimento delle proteine su una membrana di PVDF, mantenendo la separazione ottenuta precedentemente. La membrana prima di essere trattata con l'anticorpo primario è stata bloccata con latte scremato in polvere al 5%. Questo passaggio, chiamato blocking, è necessario affinché l'anticorpo non si leghi in modo aspecifico alla membrana, essendo questa reattiva, generando un rumore di fondo elevato che produrrebbe come risultato un segnale diffuso e aspecifico. Completata la saturazione, la parte inferiore della membrana, contenente proteine più piccole di 100 kDa inclusi gli uPeptidi, è stata incubata con un anticorpo primario contro HA, con il fine di rilevare gli uPeptidi in analisi. Successivamente, dopo i lavaggi, la membrana è stata incubata con un anticorpo secondario, coniugato all'enzima perossidasi di rafano (HRP), che interagendo con gli anticorpi primari ha consentito la visualizzazione delle proteine target presenti sulla membrana. L'utilizzo dell'enzima legato all'anticorpo secondario amplifica il segnale rilevabile che dipende da quanto substrato chemiluminescente viene fornito all'enzima.

2.4 MICROSCOPIA A IMMUNOFLUORESCENZA

Con lo scopo di osservare la localizzazione e l'espressione degli uPeptidi tramite immunofluorescenza nei diversi compartimenti all'interno della cellula, per ognuna delle uORFs è stata creata una proteina di fusione isolando le TLSs dal vettore pcDNA3.1(+)-3xHA, tramite gli enzimi di restrizione HindIII e BamHI, e inserendole nel vettore di espressione pEGFP N3 a monte e in fase con la sequenza EGFP. Questo tag consiste in una proteina fluorescente e viene utilizzato per marcare le proteine rilevando così la loro localizzazione subcellulare.

200.000 cellule HEK293T sono state quindi seminate su vetrini coprioggetto per microscopio posti sul fondo di una piastra di coltura da 12 pozzetti. Dopo 24 ore le cellule sono state trasfettate con il vettore pEGFP N3 e successivamente, dopo altre 24 ore, sono state lavate con PBS freddo, permeabilizzate mediante un trattamento con metanolo per 5 minuti a -20 °C. Le stesse poi sono state fissate in paraformaldeide al 4% e infine colorate trattando i vetrini coprioggetto con un mezzo di montaggio contenente DAPI, un colorante fluorescente che legandosi al DNA mette in risalto i nuclei delle cellule facilitandone l'identificazione al microscopio. Infine l'acquisizione, l'analisi e l'elaborazione delle immagini sono state eseguite utilizzando un microscopio FLIM Leica SP8 e il software ImageJ.

2.5 RACCOLTA DEI CAMPIONI PER L'ANALISI DELL'IMMUNOPEPTIDOMA

Con lo scopo di identificare eventuali uPeptidi espressi sulla superficie delle cellule tramite HLA, è stata effettuata l'analisi dell'immuno-peptidoma. Sono stati quindi raccolti nel complesso 45 campioni provenienti da tessuto maligno e

altrettanti campioni provenienti da tessuto benigno. Per quanto riguarda i campioni di tessuto maligno, i 15 provenienti da pazienti affetti da leucemia mieloide acuta (AML) e i 15 provenienti da pazienti con leucemia linfatica cronica (CLL) sono stati raccolti da campioni di sangue prelevando le cellule mononucleate del sangue periferico (PBMCs) o da campioni di midollo osseo prelevando le cellule mononucleate del midollo osseo (BMNCs) presso il Dipartimento di Ematologia e Oncologia dell'Ospedale Universitario di Tübingen, Germania. I 10 campioni provenienti da carcinoma ovarico (OvCa) e i 5 campioni di melanoma (Mel) sono stati selezionati da dataset immunopeptidomici aggiuntivi provenienti da altri studi relativi al carcinoma ovarico e al melanoma. Relativamente ai campioni di tessuto benigno, i 30 provenienti da cellule mononucleate del sangue periferico (PBMCs) e i 5 provenienti da cellule progenitrici ematopoietiche (HPC) CD34+ sono stati prelevati sempre presso il Dipartimento di Ematologia e Oncologia dell'Ospedale Universitario di Tübingen mentre i rimanenti 10 campioni provenivano da un dataset relativo a uno studio effettuato precedentemente riguardante il tumore benigno all'ovaio (OvN).

2.6 ISOLAMENTO DEI LIGANDI HLA

Le molecole HLA di classe I sono state isolate mediante cromatografia a immunoaffinità utilizzando l'anticorpo monoclonale W6/32 pan-HLA di classe I. Questa tecnica di laboratorio si basa sulla capacità di una proteina di legarsi a un determinato anticorpo immobilizzato su un supporto solido e consente di isolare le biomolecole da una miscela eterogenea in base alla loro funzione biologica e capacità di legame.

2.7 CROMATOGRAFIA LIQUIDA CON SPETTROMETRIA DI MASSA TANDEM (LC-MS-MS)

I ligandi HLA estratti sono stati analizzati mediante la cromatografia liquida con spettrometro di massa, una tecnica che combina la cromatografia liquida, utilizzata per separare i peptidi in analisi, con la spettrometria di massa per l'analisi e l'identificazione dei peptidi eluiti. Una volta separati, i peptidi sono stati poi analizzati tramite spettrometria di massa, una metodologia che misura con un'elevata sensibilità il rapporto massa/carica delle molecole dopo averle ionizzate mediante uno spettrometro di massa.

2.8 VALIDAZIONE DEGLI SPETTRI

Per validare l'identificazione dei peptidi, che era stata effettuata mediante la spettrometria di massa, è stato scelto un set comprendente 18 ligandi HLA associati ai tumori. Dopo la sintesi dei peptidi, utilizzando il sintetizzatore di peptidi Liberty Blue, è stata effettuata la validazione degli spettri mettendo a

confronto gli spettri dei peptidi eluiti ottenuti precedentemente con quelli di origine sintetica.

2.9 REFOLDING E PRIMING DEI LINFOCITI T CON CELLULE PRESENTANTI L'ANTIGENE

I linfociti T CD8⁺ sono stati isolati magneticamente con delle microbeads tramite centrifugazione a gradiente di densità da campioni di cellule mononucleate del sangue periferico (PBMC) ottenute da pazienti sani presso l'Istituto di Medicina TrASFusionale Clinica e Sperimentale dell'Ospedale Universitario di Tübingen. Sono stati preparati dei complessi HLA-peptide biotinilati, tetramerizzati usando la streptavidina coniugata PE con un rapporto molare 4:1, in grado di interagire con i linfociti T. Infine, il priming di linfociti T citotossici specifici per i peptidi in analisi è stato condotto utilizzando cellule presentanti l'antigene artificiali (aAPC). La stimolazione mediante aAPC e IL-12 delle cellule T CD8⁺, coltivate con i fattori di crescita IL-2 e IL-7, è stata eseguita settimanalmente per un totale di quattro volte con il fine di attivare i linfociti T.

2.10 COLORAZIONE DEI TETRAMERI E DELLE CITOCHINE INTRACELLULARI (ICS)

La frequenza e la funzionalità delle cellule T CD8⁺ peptide-specifiche sono state analizzate mediante tetramer staining e intracellular cytokine staining (ICS). La prima tecnica utilizza proteine tetrameriche per rilevare e quantificare le cellule T specifiche per un determinato antigene in un campione in analisi. I complessi HLA-peptide tetramerizzati sono stati utilizzati per identificare le cellule T con recettori che corrispondono al peptide specifico presentato dal complesso. L'intracellular cytokine staining è una tecnica basata sulla citometria a flusso che consente di misurare la produzione e l'accumulo di citochine nel reticolo endoplasmatico delle cellule T in seguito a una stimolazione antigene-specifica. Le cellule sono state stimulate con il singolo peptide e incubate con due inibitori del trasporto, Brefeldin A e GolgiStop per bloccare la secrezione delle citochine prodotte in modo tale da permetterne il rilevamento. La colorazione delle cellule è stata quindi eseguita utilizzando diversi anticorpi contro marker cellulari e citochine con il fine di identificare i linfociti T attivi. Infine, tutti i campioni sono stati analizzati tramite un citometro.

3. RISULTATI E DISCUSSIONE

3.1 REGOLAZIONE DELL'ESPRESSIONE GENICA A LIVELLO TRADUZIONALE MEDIATA DA uORFs IN TRASCRIPTI ASSOCIATI AL CANCRO

L'effetto delle 29 uORFs selezionate è stato studiato mediante il saggio della doppia luciferasi confrontando le sequenze TLS wt dei trascritti in analisi con quelle presentanti un codone d'inizio dell'uORF mutato (Δ uORF). La figura 1.b illustra in scala le sequenze TLS relative alle uORFs analizzate.

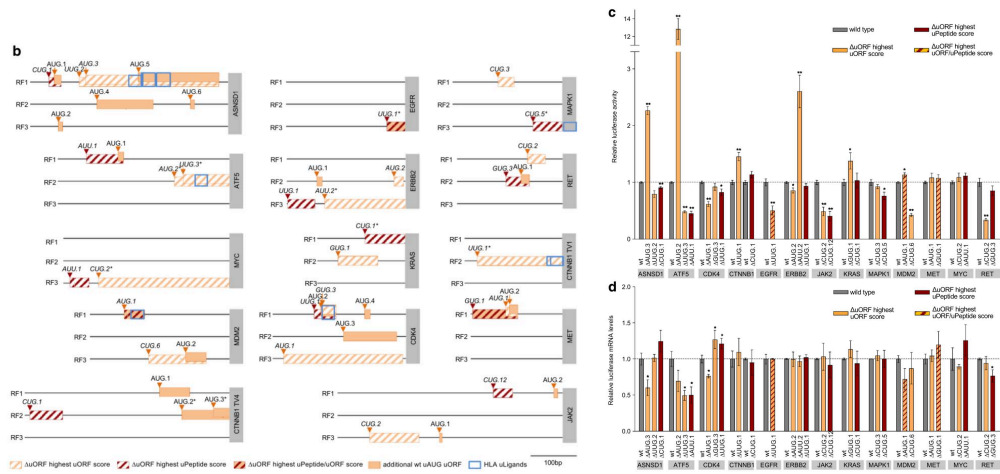


Figura 1

Come è possibile osservare nella Figura 1.c la mutazione delle uORFs, a livello dei codoni d'inizio canonici (uAUG) o alternativi (aTIS) ha causato variazioni significative nell'attività della luciferasi firefly per 18 delle 29 uORFs totali in esame. In alcuni casi, l'attività e quindi la traduzione della CDS a valle è aumentata fino a 12,83 volte, mentre in altri è diminuita fino a 0,34 volte. In 15 di questi casi inoltre, il monitoraggio dei livelli di mRNA della luciferasi Firefly tramite RT-qPCR ha escluso alterazioni significative nei livelli di trascritto, suggerendo un effetto regolatorio delle uORFs prevalentemente a livello traduzionale (Figura 1.d). Un'induzione significativa dell'attività relativa della luciferasi è stata rilevata mutando il codone d'inizio AUG.2 in CUC nella sequenza TLS del gene ATF5, mutando il codone d'inizio AUU.2 in CUC nella TLS del gene ERBB2 e infine mutando il codone d'inizio AUG.3 in CUA nella sequenza TLS della proteina ASNSD1. Una riduzione dell'attività della luciferasi è stata invece osservata mutando il codone d'inizio CUG.2 in CUC nella TLS del gene JAK2, mutando i codoni d'inizio CUG.12 e CUG.2 in CUC nella sequenza TLS del gene RET e infine mutando il codone d'inizio CUG.6 in CUC nella sequenza TLS del gene MDM2. Complessivamente è stato possibile osservare un effetto regolatorio sulla traduzione mediato da 4 uORFs su 6 presentanti un codone d'inizio uAUG e di 11 uORFs su 23 con un codone d'inizio aTIS. È stato

quindi dimostrato come l'efficienza di traduzione delle uORFs a partire sia dai codoni di inizio canonici (uAUG) che da quelli alternativi (aTIS) sia simile, implicando un coinvolgimento di entrambi i tipi di uORFs nella regolazione a livello traduzionale di geni associati al cancro. I risultati hanno confermato nella maggior parte dei casi un'attività regolatoria delle uORFs sulla traduzione della CDS a valle. Tuttavia, alcune delle uAUG e aTIS uORF analizzate non hanno provocato variazioni nell'attività del reporter della luciferasi e non sono state tradotte in uPeptidi, evidenziando una certa variabilità nei risultati ottenibili a seconda del contesto cellulare e delle condizioni sperimentali.

3.2 TRADUZIONE E LOCALIZZAZIONE CELLULARE DEGLI uPEPTIDI

Come è possibile osservare dalla Figura 2.a, tramite immunoblotting sono stati rilevati 5 dei 19 uPeptidi marcati con il tag HA tradotti a partire dalle uORFs wt presenti nelle sequenze leader dei trascritti di ASNSD1, ATF5, beta-catenina (CTNNB1), protein chinasi 1 attivata dal mitogeno (MAPK1) e MDM2. Mutando il codone d'inizio uAUG in CUC a livello delle sequenze TLS dei ASNSD1 e ATF5, si è osservata la scomparsa delle bande degli uPeptidi corrispondenti a riprova del fatto che questi codoni, inizialmente previsti computazionalmente, fossero effettivamente i codoni d'inizio della traduzione delle uORFs in analisi.

È stato possibile constatare come non sempre l'eliminazione di un codone d'inizio di un uORF si traduca nell'eliminazione completa dell'uPeptide corrispondente. Sono state quindi introdotte ulteriori mutazioni a livello dei codoni d'inizio delle uORFs relative ai geni CTNNB1, MAPK1, MDM2 con il fine di identificare i codoni d'inizio della traduzione degli uPeptidi. È stato riscontrato come l'eliminazione del codone d'inizio UUG.1 dell'uORF del gene CTNNB1 portasse a una riduzione dell'espressione dell'uPeptide codificato; lo stesso peptide inoltre, non risultava più rilevabile andando a mutare il codone d'inizio UUG.3 in UCG.

Nel caso di MAPK1, la delezione del codone CUG.5 previsto non ha avuto alcun effetto sull'espressione dell'uPeptide mentre una mutazione del codone CUG.1 in CGC ha fortemente ridotto l'espressione dell'uPeptide MAPK1 rispetto ai livelli rilevati a partire dalla TLS wild type. Allo stesso modo, l'eliminazione del codone di inizio CUG.6 dell'uORF nel TLS di MDM2 non ha bloccato l'espressione di uPeptide, il cui segnale è stato perso introducendo una mutazione in un codone differente ossia mutando il codone AUC.2 in ACC. Infine, la rimozione del codone AUG.2 subito a monte di AUC.2 non ha causato alcun effetto rilevabile sull'espressione dell'uPeptide MDM2.

Per identificare la localizzazione cellulare degli uPeptidi è stata eseguita una microscopia basata sull'immunofluorescenza degli uPeptidi in analisi marcandoli con EGFP. Facendo riferimento alla Figura 2.b è possibile notare come gli uPeptidi ASNSD1 e ATF5 abbiano mostrato rispettivamente una localizzazione cellulare ubiquitaria e prevalentemente citosolica. La mutazione del codone

d'inizio AUG.1 in CUC a livello del TLS di ATF5, oltre ad indurre livelli più elevati dell'uPeptide tradotto a partire dal codone d'inizio AUG.2 nelle analisi di immunoblot, è stata associata a un cambiamento nella localizzazione cellulare dell'uPeptide AUG. 2 che in seguito alla mutazione si accumula più frequentemente a livello perinucleare. Riguardo il peptide MAPK1 CUG.1 è stata osservata una localizzazione subcellulare specifica a livello del nucleo, implicando un suo potenziale ruolo nella via di segnalazione MAPK1.

I peptidi MAPK1 CUG.1 e TMEM203 AUG.1 sono stati localizzati in compartimenti intracellulari specifici, suggerendo la possibilità che questi possano avere un ruolo preciso all'interno della cellula come fattori agenti in trans.

Nonostante la localizzazione cellulare specifica, individuata per diversi uPeptidi in analisi, sia stata attribuita in modo predominante alle caratteristiche funzionali del peptide specifico, è stato preso in considerazione il fatto che il tag EGFP, di dimensioni relativamente grandi, possa aver influenzato sia il livello di espressione che la localizzazione degli uPeptidi ad esso associati.

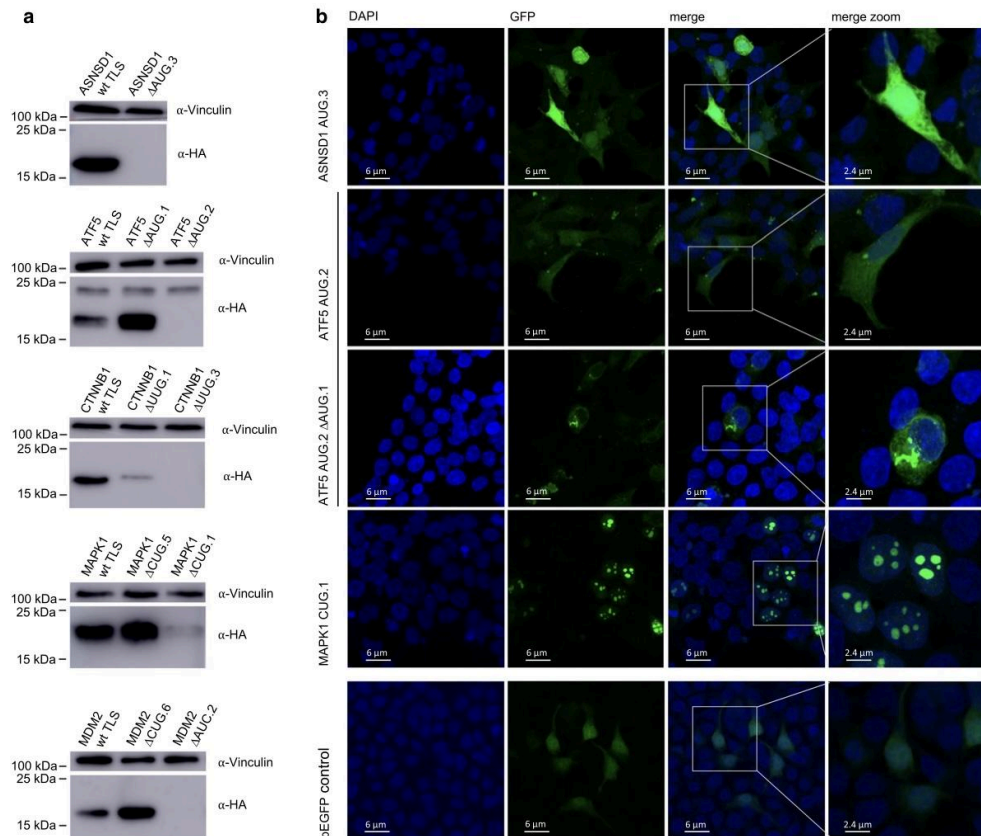


Figura 2

3.3 L'ANALISI DELL'IMMUNOPEPTIDOMA HA IDENTIFICATO LA PRESENTAZIONE DEI LIGANDI HLA

Tramite l'analisi dell'immunopeptidoma sono stati identificati gli uPeptidi presentati dalle molecole HLA di classe I, denominati uLigandi HLA. Il dataset analizzato comprendeva 49 allotipi HLA di classe I presenti in un totale di 90 campioni, 45 provenienti da tessuti cancerosi e 45 provenienti da tessuti benigni. Il 96% di questi allotipi presentava una frequenza nel dataset comparabile con quella degli allotipi all'interno della popolazione mondiale ed europea, rendendo il dataset altamente rappresentativo, come rappresentato nella Figura 3.b. Come riportato nella Figura 3.d, l'analisi ha permesso di osservare come l'82% dei campioni in analisi contenesse almeno un uLigando HLA, con una presenza maggiore nei campioni di tessuto maligno (91%) rispetto ai campioni di tessuto benigno (73%). Nel complesso sono stati identificati 125 uLigandi distinti, provenienti da 79 geni, inclusi geni di interesse come ASNSD1, ATF5, MAPK1 e TMEM203. È stata identificata la lunghezza degli uLigandi HLA, pari a 9-10 aminoacidi, paragonabile a quella degli altri peptidi presentati dalle molecole HLA di classe I non provenienti da uORFs (Figura 3.h).

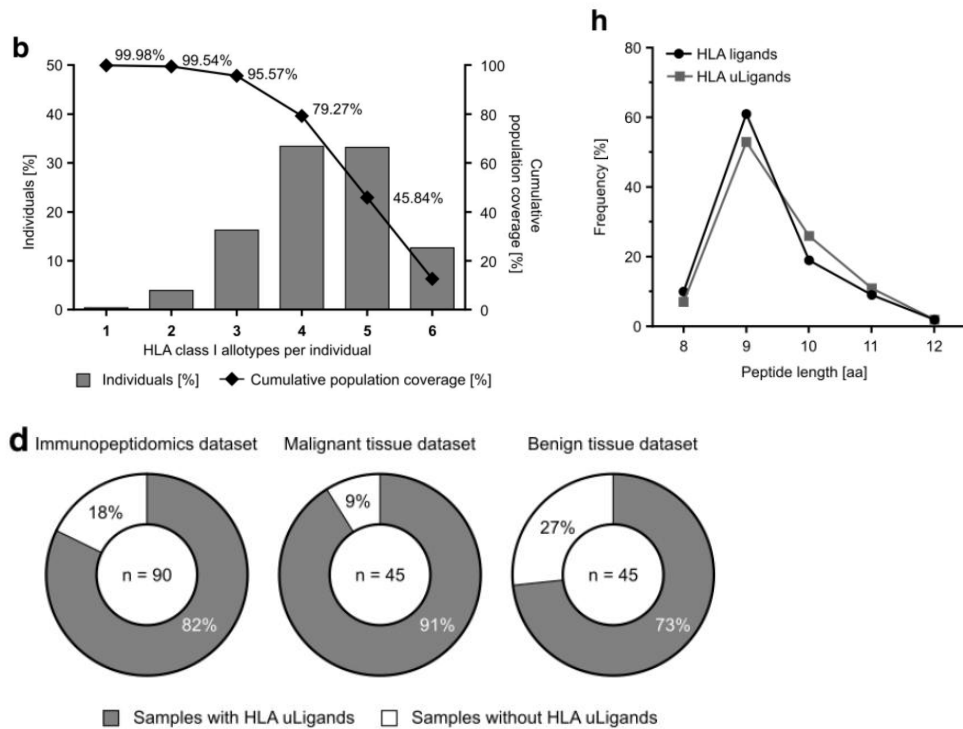


Figura 3

Confrontando i risultati ottenuti dall'analisi dell'immunopeptidoma, provenienti dai dataset dei campioni di tessuto maligno e benigno, sono stati identificati 82 uLigandi HLA su 125 presenti esclusivamente nei campioni di tessuto maligno. Questi sono stati classificati in base alla loro frequenza nei campioni di tessuto

canceroso e ciò ha portato alla scoperta del fatto che alcuni uLigandi fossero presenti in diversi campioni tumorali indipendentemente dal tipo di HLA mentre altri fossero associati a un solo tipo di tumore. Sono stati identificati 4 uLigandi HLA specifici per la leucemia mieloide acuta (AML), 1 per la leucemia linfatica cronica (CLL), 2 per il melanoma (Mel). È stato inoltre possibile identificare un gruppo di uLigandi HLA, 15 su 82, con una frequenza nei campioni maligni almeno doppia rispetto a quella relativa ai campioni benigni, suggerendo una loro associazione con un contesto tumorale. 16 uLigandi su 82 sono stati rilevati in due o più campioni di tessuto tumorale indipendentemente dall'allotipo HLA (Figure 4.a e 4.b). In particolare, l'uLigando HLA, derivato dal gene TMEM203, RSAGPRPAL, essendo stato rilevato in 8 campioni tumorali su 45, è stato associato alla frequenza di presentazione più alta nei campioni maligni. In seguito alla mutazione del codone d'inizio AUG.1 nella TLS del gene TMEM l'attività della luciferasi è aumentata di 6 volte rispetto a quella rilevata a partire dalla TLS wild type e l'uPeptide prodotto da questa variante è stato localizzato principalmente a livello nucleare e, solamente nel 20% delle cellule, a livello citoplasmatico (Figure 4.c-f).

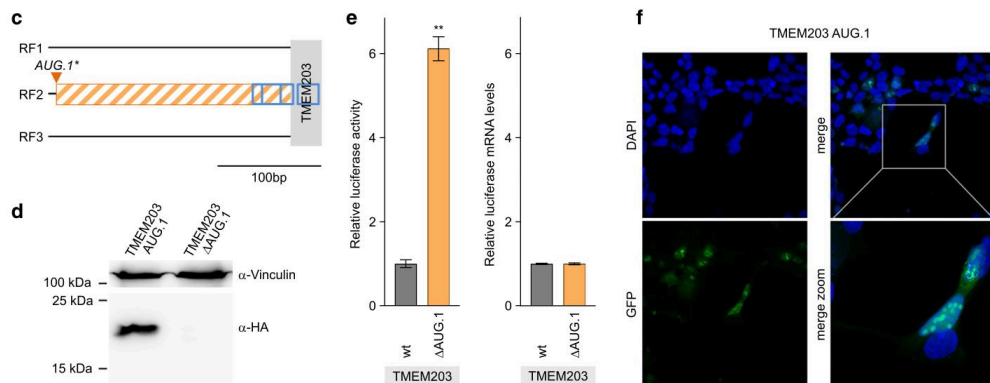


Figura 4

3.4 ATTIVAZIONE DEI LINFOCITI T CD8+ INDOTTA DAI LIGANDI HLA

Mediante l'utilizzo di peptidi sintetici marcati con isotopi, è stato convalidato il 94% dei 18 uLigandi HLA selezionati comprendenti quelli derivati dagli uPeptidi ASNSD1, ATF5, CTNNB1, MAPK1 e TMEM203 precedentemente rilevati tramite esperimenti di immunoblot. Per valutare l'immunogenicità degli uLigandi HLA, è stato selezionato un pannello di uLigandi HLA associati al tumore presentati dagli allotipi HLA-A*02, HLA-A*03 e HLA-B*07. È stato quindi eseguito il priming in vitro, basato su aAPC, dei linfociti T CD8+ naive isolati da campioni di individui sani, utilizzando i complessi HLA-uLigandi.

L'espansione dei linfociti T CD8+, specifici per ciascuno dei peptidi in esame, è stata osservata per tutti e tre gli uLigandi HLA nel 100% dei campioni provenienti dai 3 individui sani in analisi. È stata infine rilevata la presenza intracellulare delle citochine IFN- γ , di TNF e dal marcatore di degranulazione CD107a nei linfociti T attivati che ha confermato l'induzione di una risposta multifunzionale nei linfociti T attivati da parte degli epitopi derivanti dalle uORFs di trascritti associati al cancro. I risultati evidenziano un possibile ruolo degli uPeptidi come bersagli per l'immunoterapia contro il cancro. Inoltre, a differenza dei marcatori tumorali tradizionali, gli uPeptidi che sono stati identificati hanno mostrato una variabilità maggiore, suggerendo la possibilità che possano indurre una risposta immunitaria più forte. Sono quindi necessari ulteriori studi per approfondire il ruolo degli uPeptidi come potenziali bersagli nei diversi stadi di evoluzione della malattia tumorale, ad esempio diagnosi primaria rispetto a recidiva, e sotto trattamenti antitumorali sempre con il fine di comprendere il ruolo specifico di questi uPeptidi nel riconoscimento immunitario e nella progressione del tumore.

4. CONCLUSIONI

I risultati ottenuti mediante questo studio relativamente al ruolo delle uORFs nella biologia del cancro hanno evidenziato l'impatto regolatorio a livello traduzionale e potenzialmente immunologico di queste sequenze. È stato osservato come la traduzione delle sequenze uORFs, presenti nella TLS di trascritti associati ai tumori, possa mediare sia un effetto attivante, come nel caso di ATF5 Δ uAUU.1, JAK2 Δ uCUG.2 e 12, MDM2 Δ uCUG.6, che inibitorio, come osservato nel caso di ASNSD1 Δ uAUG.3, ATF5 Δ uAUG.2, ERBB2 Δ uAUU.2 sulla traduzione della CDS a valle.

Riguardo alla rilevazione degli uPeptidi, la mutazione dei codoni d'inizio uAUG, relativamente alle uORFs dei geni ASNSD1 e ATF5, ha causato la scomparsa dei corrispondenti uPeptidi, confermando i codoni d'inizio delle uORFs predetti computazionalmente come codoni d'inizio effettivi da cui aveva origine la traduzione. Tuttavia, in altri casi sono stati ottenuti risultati più variabili nei quali non sempre l'eliminazione di un codone d'inizio di un uORF portava all'eliminazione dell'uPeptide rilevato, come nel caso delle uORFs MAPK1 CUG.5 e MDM2 CUG.6.

È stata confermata la possibilità che queste uORFs possano codificare per uPeptidi espressi nelle cellule tumorali a livello di localizzazioni cellulari specifiche e presentati sulla superficie cellulare tramite i complessi MHC di classe I. Ciò implica un potenziale ruolo degli uPeptidi come fattori agenti in trans e dei uLigandi HLA, derivanti dagli uPeptidi presenti all'interno di una cellula tumorale, nel riconoscimento immunitario. Uno studio recente, ad esempio, attribuisce all'uORF presente nella TLS del gene ASNSD1 un ruolo essenziale

nella sopravvivenza delle cellule MYC-driven di medulloblastoma, evidenziando il suo potenziale come target terapeutico [7].

Relativamente all'analisi dell'immuno-peptidoma, i risultati ottenuti hanno permesso di identificare uPeptidi presentati dalle molecole HLA condivisi tra campioni e tipi tumorali differenti, fornendo una prova diretta del fatto che alcuni uPeptidi siano più abbondanti nei tessuti maligni e rari, se non del tutto assenti, nei tessuti sani. Questo potrebbe quindi suggerire un loro potenziale ruolo come marcatori tumorali comuni a tipi di tumore differenti. Inoltre è stata evidenziata la capacità di questi uLigandi HLA di attivare i linfociti T CD8+ per riconoscere e indurre una risposta immunitaria specifica contro le cellule tumorali presentanti queste molecole sulla loro superficie. Ad esempio, uno studio incentrato sul riconoscimento di nuovi antigeni specifici tumorali ha identificato un uPeptide, prodotto dalla traduzione di un uORF relativa al gene RNF10 in grado di attivare una risposta immunitaria da parte delle cellule T CD8+. È stato infatti osservato come un clone di linfociti T citotossici, derivato da un paziente con cancro pancreatico, riconoscendo l'epitopo dell'uPeptide RNF10 fosse in grado di uccidere le cellule di carcinoma pancreatico presentanti tale peptide [8].

In conclusione si sottolinea l'importanza delle uORFs presenti in circa il 49% del trascrittoma umano, inclusi i due terzi degli oncogeni e molti altri trascritti che codificano proteine coinvolte in importanti processi cellulari [1], nella regolazione dell'espressione genica e nelle risposte immunitarie. I risultati ottenuti suggeriscono la necessità di ulteriori ricerche che potrebbero portare allo sviluppo di nuovi approcci nel trattamento del cancro e di immunoterapie dirette contro i ligandi HLA presenti sulla superficie delle cellule tumorali derivanti dalle uORFs.

Bibliografia

- [1] Cristina Barbosa, Isabel Peixeiro, Luisa Romão. Gene Expression Regulation by Upstream Open Reading Frames and Human Disease. *PloS genetics*. 2013 August 8;9(8):e1003529. doi: 10.1371/journal.pgen.1003529.PMID: 23950723.
- [2] Sara K Young, Ronald C Wek. Upstream Open Reading Frames Differentially Regulate Gene-specific Translation in the Integrated Stress Response. *J Biol Chem*. 2016 Jun 29;291(33):16927–16935. doi: 10.1074/jbc.R116.733899. PMID: 27358398. PMID: PMC5016099.
- [3] Divya Ram Jayaram, Sigal Frost, Chanan Argov, Vijayasteltar Belsamma Liju, Nikhil Ponnor Anto, Amitha Muraleedharan, Assaf Ben-Ari, Rose Sinay, Ilan Smoly, Ofra Novoplansky, Noah Isakov, Debra Toiber, Chen Keasar, Moshe Elkabets, Esti Yeger-Lotem, Etta Livneh. Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. *Proc Natl Acad Sci USA*. 2021 Oct 5;118(40):e2018899118. doi: 10.1073/pnas.2018899118. PMID: 34593629. PMID: PMC8501901.
- [4] Julia Schulz, Nancy Mah, Martin Neuenschwander, Tabea Kischka, Richard Ratei, Peter M Schlag, Esmeralda Castaños-Vélez, Iduna Fichtner, Per-Ulf Tunn, Carsten Denkert, Oliver Klaas, Wolfgang E Berdel, Jens P von Kries, Wojciech Makalowski, Miguel A Andrade-Navarro, Achim Leutz, Klaus Wethmar. Loss-of-function uORF mutations in human malignancies. *Sci Rep*. 2018 Feb 5;8:2395. doi: 10.1038/s41598-018-19201-8. PMID: 29402903. PMID: PMC5799362.
- [5] Hung-Hsi Chen, Woan-Yuh Tarn. uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. *RNA Biol*. 2019 Jun 24;16(10):1327–1338. doi: 10.1080/15476286.2019.1632634. PMID: 31234713. PMID: PMC6779392.
- [6] B. Sherf, Shauna L. Navarro, R. Hannah, K. Wood. Dual-Luciferase TM Reporter Assay: An Advanced Co-Reporter Technology Integrating Firefly and Renilla Luciferase Assays. *Promega Notes Magazine Number 57*, 1996, p.02.
- [7] Damon A. Hofman, Jorge Ruiz-Orera, Ian Yannuzzi, Rakesh Murugesan, Adam Brown, Karl R. Clauser, Alexandra L. Condurat, Jip T. van Dinter, Sem A.G. Engels, Amy Goodale, Jasper van der Lugt, Tanaz Abid, Li Wang, Kevin N. Zhou, Jayne Vogelzang, Keith L. Ligon, Timothy N. Phoenix, Jennifer A. Roth, David E. Root, Norbert Hubner, Todd R. Golub, Pratiti Bandopadhyay, Sebastiaan van Heesch, John R. Prensner. Translation of non-canonical open reading frames as a cancer cell survival mechanism in childhood

medulloblastoma. *Molecular Cell*. January 2024;84(2):261-276. doi: 10.1016/j.molcel.2023.12.003. PMID: 38176414. PMCID: PMC10872554.

[8] Lili Zeng, Wei Zheng, Jiahui Zhang, Jiawen Wang, Qing Ji, Xinglong Wu, Yaming Meng, Xiaofeng Zhu. An epitope encoded by uORF of RNF10 elicits a therapeutic anti-tumor immune response. *Mol Ther Oncolytics*. 2023 Oct 19;31:100737. doi: 10.1016/j.omto.2023.100737. PMCID: PMC10654591. PMID: 38020063.



Upstream open reading frames regulate translation of cancer-associated transcripts and encode HLA-presented immunogenic tumor antigens

Annika Nelde^{1,2,3} · Lea Flötotto⁴ · Lara Jürgens⁴ · Laura Szymik⁴ · Elvira Hubert⁴ · Jens Bauer^{1,2,3} · Christoph Schliemann⁴ · Torsten Kessler⁴ · Georg Lenz⁴ · Hans-Georg Rammensee^{2,3,5} · Juliane S. Walz^{1,2,3,6} · Klaus Wethmar⁴

Received: 7 October 2021 / Revised: 21 December 2021 / Accepted: 10 January 2022 / Published online: 3 March 2022
© The Author(s) 2022

Abstract

Background Upstream open reading frames (uORFs) represent translational control elements within eukaryotic transcript leader sequences. Recent data showed that uORFs can encode for biologically active proteins and human leukocyte antigen (HLA)-presented peptides in malignant and benign cells suggesting their potential role in cancer cell development and survival. However, the role of uORFs in translational regulation of cancer-associated transcripts as well as in cancer immune surveillance is still incompletely understood.

Methods We examined the translational regulatory effect of 29 uORFs in 13 cancer-associated genes by dual-luciferase assays. Cellular expression and localization of uORF-encoded peptides (uPeptides) were investigated by immunoblotting and immunofluorescence-based microscopy. Furthermore, we utilized mass spectrometry-based immunopeptidome analyses in an extensive dataset of primary malignant and benign tissue samples for the identification of naturally presented uORF-derived HLA-presented peptides screening for more than 2000 uORFs.

Results We provide experimental evidence for similarly effective translational regulation of cancer-associated transcripts through uORFs initiated by either canonical AUG codons or by alternative translation initiation sites (aTISs). We further demonstrate frequent cellular expression and reveal occasional specific cellular localization of uORF-derived peptides, suggesting uPeptide-specific biological implications. Immunopeptidome analyses delineated a set of 125 naturally presented uORF-derived HLA-presented peptides. Comparative immunopeptidome profiling of malignant and benign tissue-derived immunopeptidomes identified several tumor-associated uORF-derived HLA ligands capable to induce multifunctional T cell responses.

Conclusion Our data provide direct evidence for the frequent expression of uPeptides in benign and malignant human tissues, suggesting a potentially widespread function of uPeptides in cancer biology. These findings may inspire novel approaches in direct molecular as well as immunotherapeutic targeting of cancer-associated uORFs and uPeptides.

Keywords Cancer · Immunopeptidomics · Mass spectrometry · Translational control · uORF

Abbreviations

Aa	Amino acid
aAPC	Artificial antigen-presenting cell
AML	Acute myeloid leukemia
aTIS	Alternative translation initiation site
BMNCs	Bone marrow mononuclear cells
CDS	Protein-coding sequence
CID	Collision-induced dissociation
CLL	Chronic lymphocytic leukemia
CTNNB1	Beta-catenin
FDR	False discovery rate
HLA	Human leukocyte antigen

Annika Nelde, Lea Flötotto contributed equally to this work.

Juliane S. Walz, Klaus Wethmar contributed equally to this work.

✉ Juliane S. Walz
juliane.walz@med.uni-tuebingen.de

✉ Klaus Wethmar
klaus.wethmar@ukmuenster.de

Extended author information available on the last page of the article

HLA uEpitope	Upstream ORF-derived T cell epitope
HLA uLigand	Upstream ORF-derived HLA ligand
HPCs	Hematopoietic progenitor cells
HVs	Healthy volunteers
ICS	Intracellular cytokine staining
IDs	Identifications
LC-MS/MS	Liquid chromatography-coupled tandem mass spectrometry
mAb	Monoclonal antibody
MAPK1	Mitogen-activated protein kinase 1
MDM2	Murine double minute 2 homolog
Mel	Melanoma
mORF	Main open reading frame
OvCa	Ovarian carcinoma
OvN	Benign ovaries
PBMCs	Peripheral blood mononuclear cells
RF	Reading frame
RT-qPCR	Real time PCR
SDM	Site-directed mutagenesis
TCRP	Translational control reporter plasmid
TLS	Transcript leader sequence
TMEM203	Transmembrane protein 203
uAUG	Upstream AUG
uORF	Upstream open reading frame
uPeptide	Upstream ORF-encoded peptide
wt	Wild type
Δ uORF	Upstream ORF with mutated start codon

Background

The development and advances of ribosome profiling [1, 2] has uncovered numerous sites of active translation at upstream open reading frames (uORFs) preceding the main protein-coding sequences (CDS) of eukaryotic transcripts [3, 4]. While approximately 55% of human transcript leader sequences (TLSs) contain canonical upstream AUG (uAUG) initiation codons [5], virtually all human transcripts carry near-cognate alternative translational initiation sites (aTISs), differing in one base from the canonical AUG sequence [6, 7]. Computational and experimental studies demonstrated compelling evidence for an important regulatory role of uORF-mediated translational control in (patho-)physiology [3, 8–10], and several uORF-associated genetic variants have been linked to the development of disease [8, 10–15]. A recent study also demonstrated that virus-derived uORFs are translated during infection and contribute to virulence [16].

Upstream ORFs represent important relays of gene expression regulation, as translation of the downstream CDS from uORF-bearing transcripts requires leaky scanning across the uORF start site or reinitiation of ribosomes after translating the uORF [9, 17, 18]. Upstream ORF-mediated translational regulation has been observed in

multiple transcripts across eukaryotic species [8, 19–21]. Specific arrangements of multiple uORFs have been shown to mediate the paradoxical induction of downstream protein translation under conditions of cellular stress, as studied in detail for the transcription factors GCNA4 in yeast, and for ATF4 and ATF5 in mammals [22–25]. Furthermore, several sequencing studies demonstrated frequent genetic variability of uORFs in human cancer [11, 15, 26] and additional individual reports on CDKN1B and CDKN2A directly linked defective uORF-mediated translational control to tumorigenesis [12, 13]. However, only a few reports provided individual experimental evidence for the regulatory impact of uAUG and aTIS uORFs in human proto-oncogenes [11–13, 15, 26, 27].

Very recent studies combining ribosome profiling, proteomics and immunopeptidomics [28–32] confirmed the widespread translation of cryptic peptides from non-coding regions, including 5'-TLSs and 3'-UTRs, non-coding RNAs, intronic, intergenic, and off-frame regions, and provided first insights into their presentation on human leukocyte antigen (HLA) class I molecules. Thereby, uORF-derived peptides encoded in the TLSs of protein-coding transcripts represent the largest category of detected cryptic peptides [30]. Upstream ORF-derived peptides may form direct *cis*- and *trans*-regulatory ways, and may sense the cellular levels of small molecules or metabolites to serve as pepto-switches regulating downstream translation [33–35]. For example, a uPeptide in the TLS of PKC was recently shown to suppress tumor progression, proliferation, invasion and metastasis in different models of breast cancer [36]. Especially in the context of pathologically altered cellular processes such as malignant transformation, the differential translation of uORFs and differential uPeptide processing could produce tumor-specific uORF-derived HLA ligands (HLA uLigands) that may serve as rejection antigens [12, 13, 37]. However, previous immunopeptidomic studies were mainly limited to cell lines and only applied sample-specific proteogenomic approaches using personalized reference databases [29, 32]. Furthermore, cancer-associated HLA presentation was retrospectively determined based on RNA sequencing data [29, 32] due to the lack of complete tissue immunopeptidomics reference libraries from healthy tissues, calling for the direct immunopeptidome analysis of primary malignant and benign tissue samples to further delineate the role of HLA uLigands as tumor-specific targets and their role in anti-tumor immunity.

Here we experimentally characterized the translational regulatory role of selected uORFs in cancer-associated transcripts and present evidence for frequent translation and cellular expression of the related uORF-derived uPeptides. Mass spectrometry-based immunopeptidome analyses using a broadly applicable non-personalized uORF

database in an extensive dataset of primary malignant and benign tissue samples further delineated several tumor-associated HLA uLigands capable to induce multifunctional peptide-specific T cell responses.

Materials and methods

Selection of uORFs for functional analysis

The selection of uORF containing genes for functional analysis was based on documented oncogenic functions of the associated transcripts [38, 39], high conservation of uORF sequences (PhyloCFS score > 50), or evidence for active uORF translation (TEscore ≥ 5) [4]. This selection yielded a total of 536 genes. Oncogenes were filtered for presence in canonical cancer pathways (RTK/RAS, cell cycle, PI3K, P53, MYC and WNT pathway) where deregulation of a single protein may be sufficient to mediate oncogenic effects on downstream signalling [40]. Then, uORFs of the selected genes were ranked according to the highest uORF score (top 10%) as described in McGillivray *et al.* [3]. This score predicts functional uORF relevance based on specific features, including uORF length, position, and conservation as well as expression of the associated downstream main protein. For each gene, we selected the uAUG with the highest uORF score, the aTIS with the highest uORF score, and the uAUG/aTIS with the highest uPeptide score for experimental analysis. Upstream ORFs < 24 bp were excluded as they were considered to be too small for immunoblot detection. We then determined the presence of the selected uAUGs and aTISs in all RefSeq transcript variants of the respective gene according to the genomic position given in McGillivray *et al.* [3] (hg19) and picked one representative TLS to be used for further experiments. If there were multiple transcript variants including all uORFs under investigation we preferred low complexity TLS as defined by short length, low number of additional uORFs and low number of exons to ease experimental handling. From the remaining set, we finally selected 13 transcripts based on the abundance of previous literature indicating functional oncogenic importance or suggesting active uORF regulation in the respective genes.

Cell culture

HEK293T cells (obtained from ATCC) were cultivated at 37 °C, 5% (v/v) CO₂ in humidified and DMEM culture medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin.

Dual-luciferase assay

Complete wt TLSs including the endogenous main ORF (mORF) initiation codon and the Kozak base at position +4 were synthesized by GeneArt (Thermo Fisher Scientific) (Supplementary Table 1). TLSs were isolated from GeneArt vectors using the appropriate restriction enzymes and were ligated into a translational control reporter plasmid (TCRP) based on the pGL3 basic vector as previously described [5] (Supplementary Figure 1a). Individual uORF initiation codons were mutated to CUC (Δ uORF) wherever possible by site-directed mutagenesis (SDM) or to alternative non-initiation codons (Supplementary Tables 2 and 3). Correctness of all insertions and SDMs was verified by Sanger sequencing. The wt and Δ uORF TCRPs were co-transfected together with a Renilla luciferase control vector (pRL-CMV, Promega) into HEK293T cells using METAFECTENE® transfection reagent (Biontex) according to manufactures instructions. 44 h later cells were washed, lysed, and Firefly and Renilla luciferase activities were measured using a multilabel plate reader (Victor™ X3, PerkinElmer) as described before [41]. In detail, 50,000 HEK293T cells were seeded in 24-well plates, grown for one day and subsequently transfected using TLS-specific amounts of translational reporter plasmid, 75 ng/well Renilla luciferase vector and 3 μ l METAFECTENE® mixed in 100 μ l Opti-MEM® (Thermo Fisher Scientific). The use of TLS-specific amounts of TCRP (range 1–364.5 ng/well) was required to adjust luminescent signals to the linear range of detection of the plate reader, as individual wt TLS caused largely diverging global inhibitory effects on luciferase expression (Supplementary Figure 2). After 15 min of incubation at room temperature, 50 μ l of transfection mix was added drop-wise to the cells in duplicates. 44 h later cells were washed with 500 μ l PBS and then lysed using Luciferase Lysis Buffer (90 mM K₂HPO₄, 9 mM KH₂PO₄, 0.2% Triton X-100) containing 40 μ l Proteinase-Inhibitor Cocktail Complete (Sigma Aldrich). For complete lysis cells were shaken on ice for 30 min. Cell lysates were transferred to a 1.5 ml tube and centrifuged at 21,000g at 4 °C for 10 min. The supernatant was transferred in a new 1.5 ml tube and triplicate measurements of each lysate were performed in a Nunc™ F96 MicroWell™ polystyrol plate (Thermo Fisher Scientific).

Real-time quantitative PCR (RT-qPCR)

Whole RNA was isolated from washed and pelleted cells using the Nucleo Spin® RNA Kit (Macherey–Nagel) including a first DNaseI digestion step according to the manufacturer's instructions. Afterwards, a second DNaseI treatment was performed with 1 μ g of isolated RNA. cDNA was synthesised from 200 ng RNA according to the protocol of the RevertAid Strand cDNA Synthesis Kit (Thermo

Fisher Scientific) and the final cDNA concentration was adjusted to 100 ng/μl. Relative real-time quantitative PCR (RT-qPCR) was performed in a MicroAmp® Fast 96-well Reaction Plate (0.1 ml) (Thermo Fisher Scientific) using the Luna® Universal qPCR Mastermix (NEB) and following primers: Firefly_for ATCCATCTTGCTCCAACACC, Firefly_rev TCGCGGTTGTTACTTGACTG, Renilla_for GGA ATTATAATGCTTATCTACGTGC, Renilla_rev CTTGCG AAAAATGAAGACCTTTTAC. RT-qPCR was performed in a StepOnePlus Real-Time PCR System (Applied Biosystems). To exclude relevant plasmid DNA contamination in RNA extracts, we always included RNA control samples without reverse transcription in RT-qPCR experiments.

Detection of HA-tagged uPeptides by immunoblotting

Oligonucleotides including the 3xHA-Tag sequence were annealed and ligated into the pcDNA3.1(+) vector using BamHI and XbaI restriction sites. Next, complete TLSs sequences including the uORFs under investigation were amplified by PCR from Firefly luciferase vectors deleting the uORF's termination codon and including HindIII and BamHI restriction site-overhangs for ligation. Amplicons were ligated upstream of the 3xHA-tag into the pcDNA3.1(+)-3xHA vector with the uPeptide initiation codon being in-frame with the 3xHA-tag (Supplementary Figure 1b). We generated expression vectors for all AUG uORFs, all uORFs with the highest uPeptides scores, and additionally included the CTNNB1 aTIS uORF, as this gene lacked an AUG uORF but contained a UUG.1 aTIS uORF with the highest uORF score on a distinct transcript variant. Correctness of insertions was verified by Sanger sequencing. 500,000 HEK293T cells were seeded in a 6-well plate and grown for 24 h. 3 μg of expression vector was transfected using 5 μl METAFECTENE® (Biontexas). After 44 h cells were treated with 2 μl of 10 μM MG132 (Enzo life sciences) and 8 h later cell lysates were subjected to immunoblotting following standardized protocols using Vinculin (7F9) (sc-73614) and HA (F-7) (sc-7392) antibodies (Santa Cruz). In detail, cells were washed with 1 ml PBS and lysed with 150 μl Immunoblot Lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 0.1% Triton X-100, 1 mM EDTA) containing 6 μl Proteinase-Inhibitor Cocktail Complete (Sigma-Aldrich) and 1.5 μl DTT by shaking for 30 min at 4 °C. Lysates were centrifuged for 20 min by 21,000×g at 4 °C and supernatants were transferred to a new 1.5 ml Eppendorf tube. Total protein amounts were determined using the BCA assay. 5 μl of NuPage® LDS Sample Buffer (4x) (Invitrogen) was added to 20 μl protein lysate containing 50 μg of total protein. The mixture was incubated at 9 °C for 5 min and subsequently applied on a 22% SDS-Page gel. After electrophoretic separation proteins were transferred

on a PVDF membrane using the Mini Trans-Blot® cell at 100 V for 1.5 h. After blocking the membrane in 5% skim milk for 1 h, the upper part (> 100 kDa) was incubated with Vinculin antibody (1:5,000 in 5% skim milk) and the lower part (< 100 kDa) with the HA antibody (1:1,000 in 5% skim milk) at 4 °C over night. Membrane was washed with 1×TBST three times for 5 min and then incubated with goat anti-mouse antibody (1:5,000 in 5% skim milk, Jackson Immuno Research AB_2338461) for 1 h. The membrane was again washed three times in 1×TBST for 5 min and the Super Signal™ West Pico PLUS Chemiluminescent Substrate (Thermo Fisher Scientific) was added to the membrane according to the manufacturer's protocol. Immunoblots were developed using the Amersham Imager 600 (GE Healthcare). Exposure times were one minute for the immunoblots showing ASNSD1, ATF5, MAPK1, and MDM2 uPeptides, 4 min for the immunoblot of the CTNNB1 uPeptide and 3.5 min for the immunoblot of the TMEM203 uPeptide.

Immunofluorescence-based microscopy

For each uORF individual TLSs including the sequence from the 5'-cap to the disrupted termination codon of the uORF under investigation were isolated from the pcDNA3.1(+)-3xHA vector using the restriction enzymes HindIII and BamHI and were ligated upstream and in-frame to the EGFP coding sequence (with deleted EGFP-initiation codon) into the pEGFP N3 vector (Addgene) and correct insertion was verified by Sanger sequencing (Supplementary Figure 1c). 200,000 HEK293T cells were seeded on a microscope cover-glass in a 12-well plate. After 24 h cells were transfected using METAFECTENE® as described above. After another 24 h the cells were washed with cold PBS, permeabilized by a 5 min treatment with 100 μl Methanol (−20 °C) and fixed in 4% paraformaldehyde for 10 min, and washed again with cold PBS. Cover-glasses were then treated with 30 μl DAPI-containing mounting medium (Thermo Fisher Scientific) and placed in the middle of a microscopy object slide. Image acquisition, analysis, and processing were carried out using a Leica SP8 FLIM Microscope and ImageJ software [42]. Here, the captured z-stack images are presented as merged z-stack images.

Sample collection for immunopeptidomic analysis

For immunopeptidome analysis, peripheral blood mononuclear cells (PBMCs) or bone marrow mononuclear cells (BMNCs) from acute myeloid leukemia (AML) and chronic lymphocytic leukemia (CLL) patients were collected at the Department of Hematology and Oncology at the University Hospital Tübingen, Germany. Samples of CLL patients ($n = 15$) were collected at the time of first therapy indication according to iwCLL guidelines

[43]. Samples of AML patients ($n = 15$) were collected at the time of diagnosis ($n = 13$), under palliative therapy ($n = 1$) or at relapse ($n = 1$). PBMCs from healthy volunteers (HVs) and CD34⁺ magnetically enriched hematopoietic progenitor cells (HPCs, CD34 MicroBead Kit, human, Miltenyi Biotec) from hematopoietic stem cell aphereses from G-CSF mobilized blood donations of HVs and patients with non-hematological malignancies (*e.g.* germ cell tumors) were collected at the University Hospital Tübingen, Germany. Cells were isolated by density gradient centrifugation and stored at $-80\text{ }^{\circ}\text{C}$ until further use. Informed consent was obtained in accordance with the Declaration of Helsinki protocol. The study was performed according to the guidelines of the local ethics committees (373/2011B02, 454/2016B02, 406/2019B02). HLA typing was carried out by the Department of Hematology and Oncology, Tübingen, Germany. Furthermore, we used two publically available immunopeptidomic datasets comprising samples of ovarian carcinoma (OvCa) and benign ovaries (OvN) [44] as well as melanoma (Mel) [45]. Sample characteristics of malignant and benign tissue samples are provided in Supplementary Tables 4 and 5, respectively.

Isolation of HLA ligands

HLA class I molecules were isolated by standard immunoaffinity purification as described before [47] using the pan-HLA class I-specific W6/32 monoclonal antibody (produced in-house).

Analysis of HLA ligands by liquid chromatography-coupled tandem mass spectrometry (LC-MS/MS)

HLA ligand extracts were analyzed as described previously [47, 48]. Peptides were separated by nanoflow high-performance liquid chromatography (RSLCnano, Thermo Fisher Scientific) using a $50\text{ }\mu\text{m} \times 25\text{ cm}$ PepMap rapid separation liquid chromatography column (Thermo Fisher Scientific) and a gradient ranging from 2.4% to 32.0% acetonitrile over the course of 90 min. Eluted peptides were analyzed in an online-coupled LTQ Orbitrap XL or LTQ Orbitrap Fusion Lumos mass spectrometer (Thermo Fisher Scientific) equipped with nano-electrospray ion sources using a data-dependent acquisition mode employing a top five or a top speed collision-induced dissociation (CID) fragmentation method (normalized collision energy 35%), respectively. The mass range was set to 400–650 m/z with charge states 2+ and 3+ selected for fragmentation.

Data processing, uORF database structure, and HLA annotation

For data processing, the software Proteome Discoverer (v1.4.0, Thermo Fisher) was used to integrate the search results of the SEQUEST HT search engine (University of Washington) [49] against the human proteome as comprised in the Swiss-Prot database (20,367 reviewed protein sequences, January 7th 2020) supplemented with two datasets of uORF sequences. The datasets of uORF sequences contained 1062 uORFs (877 different amino acid sequences) with the highest scores predicting uORF functionality (McGillivray set [3]) and 1236 uORFs (1235 different amino acid sequences with five sequences contained in both sets) selected based on indications of functional relevance from previous experimental data, genetic context or sequence analysis (in-house set). No enzymatic restriction was applied. Precursor mass tolerance was set to 5 ppm, and fragment mass tolerance to 0.5 Da for ion trap spectra and 0.02 Da for orbitrap spectra, respectively. Oxidized methionine was allowed as a dynamic modification. The false discovery rate (FDR) was estimated using the Percolator algorithm (v2.04) [50] and limited to 5%. Peptide lengths were limited to 8–12 amino acids. Protein inference was disabled, allowing for multiple protein annotations of peptides. HLA class I annotation was performed using NetMHCpan 4.0 [51–53] and SYFPEITHI 1.0 [54] annotating peptides with percentile rank below 2% and $\geq 60\%$ of the maximal score, respectively. We screened the immunopeptidomes for uORF-derived peptide sequences, which are uniquely mapped on uORF sequences and not on any other non-uORF human protein sequence (except for ASDURF_HUMAN, a reviewed uORF).

Peptide synthesis

Peptides were produced by the peptide synthesizer Liberty Blue (CEM) using the 9-fluorenylmethyl-oxycarbonyl/tert-butyl strategy [55].

Spectrum validation

Spectrum validation of the experimentally eluted peptides was performed by computing the similarity of the spectra with corresponding isotope-labeled synthetic peptides measured in a complex matrix. The spectral correlation was calculated between eluted peptide spectra and synthetic peptide spectra using the intensities of annotated b- and y-ion peaks. For synthetic peptide-based validation of mass spectrometry-based peptide identifications a panel

of 18 tumor-associated, tumor-enriched, high frequent, and reviewed HLA uLigands was selected.

Blood samples for T cell-based assays

PBMCs from whole blood samples of HVs were isolated by standard density gradient centrifugation and CD8⁺ T cells were magnetically isolated (CD8 MicroBeads, human, Miltenyi Biotec). Blood samples were kindly provided by the Institute for Clinical and Experimental Transfusion Medicine at the University Hospital Tübingen after obtaining written informed consent.

Refolding

Biotinylated HLA-peptide complexes were manufactured as described previously [56] and tetramerized using PE-conjugated streptavidin (Invitrogen Life Technologies) at a 4:1 molar ratio.

Induction of peptide-specific CD8⁺ T cells with artificial antigen-presenting cells (aAPC)

Priming of peptide-specific cytotoxic T lymphocytes was conducted using artificial antigen-presenting cells (aAPCs) as described previously [57]. In detail, 800,000 streptavidin-coated microspheres were loaded with 200 ng biotinylated HLA:peptide monomer and 600 ng biotinylated anti-human CD28 monoclonal antibody (mAb, clone 9.3, in-house production). CD8⁺ T cells were cultured with 4.8 U/μl IL-2 (R + D) and 1.25 ng/ml IL-7 (PromoKine). Weekly stimulation with aAPCs (200,000 aAPCs per 1 × 10⁶ CD8⁺ T cells) and 5 ng/ml IL-12 (PromoKine) was performed four times.

Cytokine and tetramer staining

The frequency and functionality of peptide-specific CD8⁺ T cells was analyzed by tetramer [58] and intracellular cytokine staining (ICS) [59, 60], respectively, as described previously. For ICS, cells were pulsed with 10 μg/ml of individual peptide and incubated with 10 μg/ml Brefeldin A (Sigma–Aldrich) and 10 μg/ml GolgiStop (BD) for 12–16 h. Staining was performed using Cytofix/Cytoperm (BD), PerCP anti-human CD8, PacificBlue anti-human TNF, FITC anti-human CD107a (BioLegend), and PE antihuman IFN-γ antibodies (BD). PMA and ionomycin (Sigma–Aldrich) served as a positive control. The peptides YLLPAIVHI (HLA-A*02, DDX5_HUMAN₁₄₈₋₁₅₆), RLR-PGGKKK (HLA-A*03, GAG_HV1BR₂₀₋₂₈), and TPGPG-VRYPL (HLA-B*07, NEF_HV1BR₁₂₈₋₁₃₇) served as negative control peptides. The frequency of peptide-specific CD8⁺ T cells after aAPC-based priming was determined by tetramer staining using PerCP anti-human CD8 antibody

and HLA:peptide tetramer-PE. For negative control, tetramers of the same HLA allotype containing irrelevant control peptides were used. The priming was considered successful if the frequency of peptide-specific CD8⁺ T cells was > 0.1% of CD8⁺ T cells within the viable single-cell population and at least three-fold higher than the frequency of peptide-specific CD8⁺ T cells in the negative control. The same evaluation criteria were applied for the ICS results. All samples were analyzed on a FACS Canto II cytometer (BD).

Software and statistical analysis

Overlap analysis was performed using BioVenn [61]. The population coverage of HLA allotypes was calculated by the IEDB population coverage tool (www.iedb.org) [62, 63]. Fisher's exact test was used for the analysis of HLA allotype distribution between the immunopeptidome dataset ($n=90$), the world population ($n=90,046$) and the European population ($n=32,856$) [64] as well as between the malignant ($n=45$) and benign ($n=45$) tissue dataset. Flow cytometric data were analyzed using FlowJo 10.0.8 (Treestar). All figures were generated using GraphPad Prism 9.0.2 (GraphPad Software).

Results

Upstream ORF-mediated translational regulation of CDS expression in cancer-associated transcripts

Aiming to characterize the functional impact of uORF-mediated translational regulation and the prevalence of uPeptide expression in cancer-related genes, we first selected a set of 29 uAUG and aTIS uORFs from 13 cancer-associated genes (Fig. 1a, Supplementary Table 3). Upstream ORFs were selected based on literature research [4, 38, 39] and categorized according to the type of initiation codon (uAUG vs. aTIS) and the computationally defined uORF and uPeptide scores [3] predicting functional relevance. All candidate uORFs were tested for their translational regulatory impact on downstream CDS translation in dual-luciferase reporter assays [41] using wild type (wt) TLSs and TLSs carrying a functionally deleted uORF initiation codon (Δ uORF) (Supplementary Figure 1a).

Structurally, individual TLSs showed high variability with respect to TLS length, as well as uORF number, length and position (Fig. 1b), resulting in variable levels of baseline relative luciferase activity for individual wt TLSs (Supplementary Figure 2). The functional ablation of uAUG and aTIS initiation codons caused significant changes of luciferase activity for 18 of 29 uORFs, ranging from 12.83-fold induction to 0.34-fold repression of luciferase signals as compared to wt TLSs (Fig. 1c). In 15 of these cases,

concomitant monitoring of luciferase mRNA levels excluded major contributions of alterations in luciferase transcript levels, suggesting a mostly translational regulatory effect of the Δ uORF variants (Fig. 1d). Six of these uORF ablations induced a ≥ 2 -fold increase or ≤ 0.5 -fold decrease of luciferase activity, respectively. Most prominent induction of relative luciferase activity was detected for the AUG.2 > CUC Δ uORF TLS of the activating transcription factor 5 (ATF5, 12.83 ± 1.64 SEM, $p \leq 0.01$), the AUU.2 > CUC Δ uORF TLS of receptor tyrosine kinase Erb-B2 (ERBB2, 2.60 ± 0.63 SEM, $p \leq 0.01$), and the AUG.3 > CUA Δ uORF TLS of Asparagin synthetase domain-containing protein 1 (ASNSD1, 2.27 ± 0.10 SEM, $p \leq 0.01$). Major reduction of relative luciferase activity compared to wt TLS levels was observed for the CUG.2 > CUC Δ uORF TLS of the receptor tyrosine kinase Ret (RET, 0.34 ± 0.03 SEM, $p \leq 0.01$), the CUG.12 > CUC and the CUG.2 > CUC Δ uORF TLS of janus kinase 2 (JAK2, 0.41 ± 0.11 SEM, $p \leq 0.01$ and 0.49 ± 0.12 , $p \leq 0.01$, respectively), and the CUG.6 > CUC Δ uORF-TLS of the murine double minute 2 homolog (MDM2, 0.53 ± 0.09 SEM, $p \leq 0.01$). Overall, a translational regulatory effect was observed for 4 of 6 uAUG and 11 of 23 aTIS uORFs, suggesting that both, canonical uAUG and aTIS initiation codons may be similar functionally relevant for the regulation of cancer-associated gene expression and may impact cancer onset and progression.

Translation and cellular localization of uPeptides

Focussing on AUG uORFs and uORFs with the highest uPeptide scores, we analyzed whether uORFs encoded by cancer-associated transcripts were translated into uPeptides in vitro. Five of 19 HA-tagged uPeptides, translated from the TLSs of ASNSD1, ATF5, beta-Catenin (CTNNB1), mitogen-activated protein kinase 1 (MAPK1), and MDM2, were detected by immunoblotting (Fig. 2a). Introduction of start codon ablating AUG.3 > CUC and AUG.2 > CUC mutations into the ASNSD1 and ATF5 TLSs resulted in complete losses of the uPeptide bands (Fig. 2a), validating translational initiation at the computationally predicted uORF start codons. To identify the origin of the larger 23 kDa band in the ATF5 immunoblots, we introduced several regional and codon-specific mutations to the ATF5 TLS (Supplementary Figure 3). The data indicated that this band represented an extended ATF5 uPeptide initiated by an obscure start site differing from classical uAUG and aTIS codons. As the functional deletions of the predicted uORF initiation codons did not always result in complete ablation of the HA-tagged uPeptides (Fig. 2a), we next inserted a number of additional uStart deleting mutations into the CTNNB1, MAPK1 and MDM2 TLSs (Supplementary Table 6). For CTNNB1, expression of the uPeptide was markedly reduced upon

deletion of the predicted UUG.1 uORF start site and was undetectable upon insertion of a UUG.3 > UCG mutation. In the case of MAPK1, the deletion of the predicted CUG.5 codon had no effect on uPeptide expression in immunoblot analysis, while an alternative CUG.1 > CGC mutation strongly reduced MAPK1 uPeptide expression as compared to wt levels (Fig. 2a). Similarly, deletion of the predicted CUG.6 uORF start codon in the TLS of MDM2 did not abolish uPeptide expression, but the uPeptide signal was lost upon insertion of an AUC.2 > ACC mutation (Fig. 2a). Of note, mutational ablation of an AUG.2 codon immediately upstream of the AUC.2 codon had no detectable effect on the MDM2 uPeptide expression (Supplementary Figure 4).

Aiming to validate uPeptide expression by an independent experimental approach, we performed immunofluorescence-based microscopy of EGFP-labeled uPeptides. The ASNSD1 and ATF5 uPeptides showed ubiquitous and predominantly cytosolic cellular localization, respectively (Fig. 2b, Supplementary Figure 5). Interestingly, an AUG.1 > CUC deletion of the ATF5 AUG.1 uORF not only induced higher ATF5 AUG.2 uPeptide levels in immunoblot analyses (Fig. 2a), but was also associated with a marked change in cellular localization of the AUG.2 uPeptide, now frequently accumulating in perinuclear focal structures (Fig. 2b, Supplementary Figure 5). The MAPK1 CUG.1 peptide also showed specific focal localization, exclusively mapping to the nucleus and implying a potential functional relevance for MAPK1 signalling (Fig. 2b, Supplementary Figure 5). Together, these data confirmed the cellular expression of uPeptides from cancer-associated transcripts and demonstrated uPeptide-specific cellular localizations, suggesting distinct functions of individual uPeptides as *trans*-acting factors.

Mass spectrometry-based immunopeptidome analysis identified naturally presented HLA uLigands

To evaluate if uORFs encode HLA-presented peptides that might serve as antigenic targets for cancer immune surveillance and immunotherapeutic approaches, mass spectrometry-based immunopeptidome profiling of primary malignant [$n=45$, AML ($n=15$), CLL ($n=15$), OvCa [44] ($n=10$), and Mel [45] ($n=5$)] as well as benign tissue samples [$n=45$, PBMCs ($n=30$), CD34-enriched HPC ($n=5$), and OvN ($n=10$)] was applied (Fig. 3a). This immunopeptidomic dataset covers 49 different HLA class I allotypes including 14 different HLA-A, 22 HLA-B, and 13 HLA-C allotypes. HLA allotype frequencies are comparable to the world and European population with 96% of the allotypes showing no significant differences in the frequency between the immunopeptidome dataset compared to the

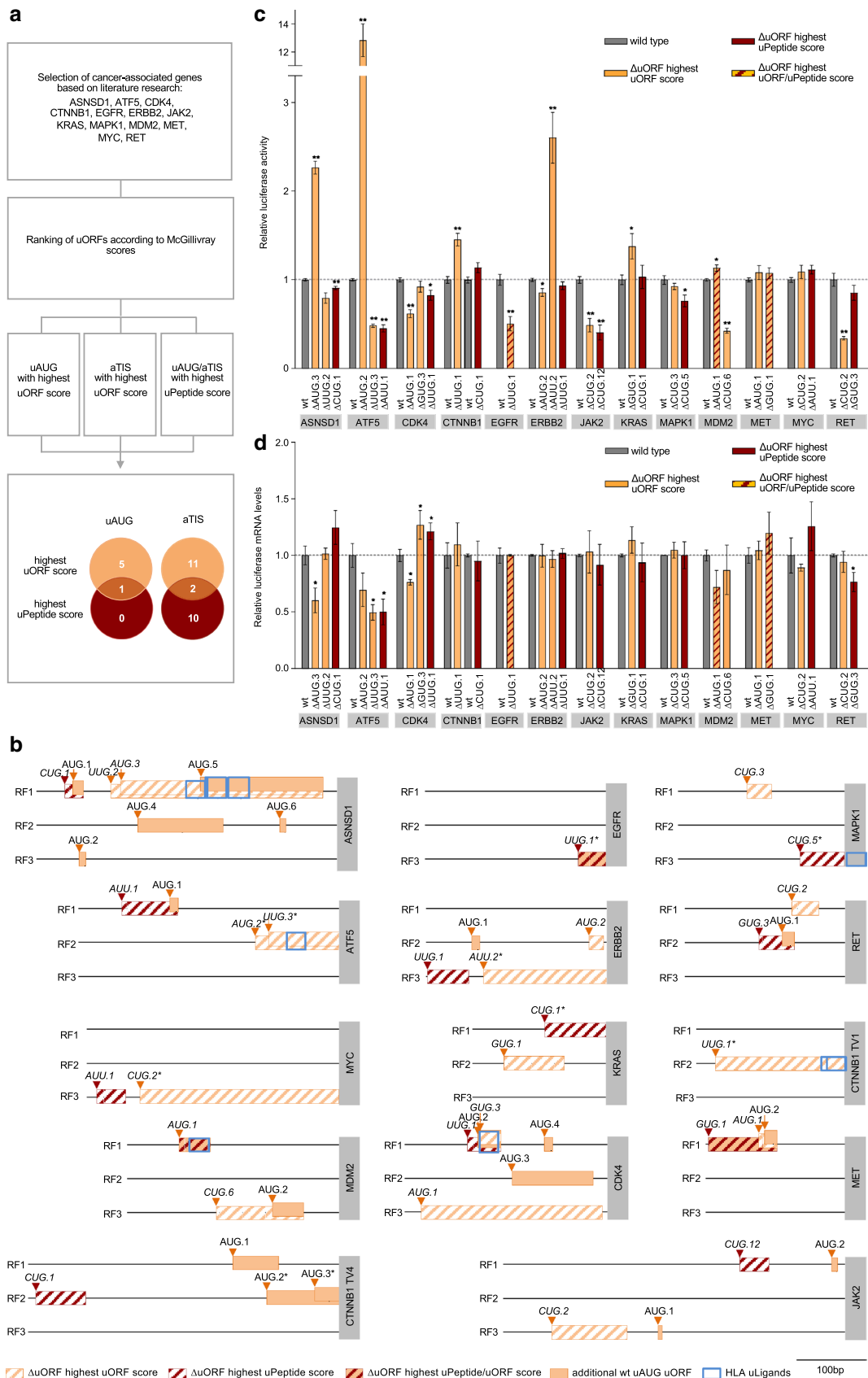


Fig. 1 Translational regulation of CDS expression through uORFs in cancer-associated transcripts. **a** Flowchart illustrating the selection process for uORFs initiated at canonical (uAUG) and at alternative translational initiation sites (aTIS) for functional analysis. **b** Schematic illustrations of indicated in-scale TLSs displaying the analyzed uORFs (stripped boxes) and additional wild type AUG uORFs (wt, filled orange boxes) on reading frames (RF) 1–3 (black lines). Upstream ORFs overlapping into the CDS are marked with an asterisk at the respective start codon. Grey boxes on the right contain gene symbols and indicate the start of the CDS. Note that for CTNNB1 the uORFs with the highest uORF- and uPeptid-scores mapped to distinct transcript variants. Accordingly, both were selected for experimental analysis. **c** Bar graphs showing relative Firefly luciferase activities detected for indicated wild type (wt) and Δ uORF TLSs normalized to Renilla luciferase internal controls. Results are combined from three independent experiments and error bars indicate SEM. Levels of significance are $p \leq 0.05$ (*) and $p \leq 0.01$ (**) as determined by two-tailed nonparametric Mann–Whitney *U*-tests. **d** Bar graphs indicating relative Firefly luciferase mRNA levels for indicated wt and Δ uORF TLSs normalized to Renilla luciferase internal control. Results are combined from three independent experiments and error bars indicate SEM. Levels of significance are $p \leq 0.05$ (*) and $p \leq 0.01$ (**) as determined by two-tailed nonparametric Mann–Whitney *U*-tests

world and the European population (Supplementary Figure 6a). Between the malignant and benign tissue datasets comparable HLA allotype frequencies are observed for 96% of the allotypes (Supplementary Figure 6b). 99.98% of the world population carries at least one of the HLA class I allotypes included in the dataset (Fig. 3b). We identified a total of 127,766 unique HLA class I ligands (peptides assigned to their HLA allotype, range 684–25,249, mean 4368 per sample, Supplementary Tables 4 and 5) with a FDR of 5% from 15,336 different source proteins, obtaining 98% (97% and 94% for malignant and benign samples, respectively) of the estimated maximum attainable coverage in HLA ligand source proteins (Fig. 3c and Supplementary Figure 7). For the identification of naturally presented HLA uLigands we screened the immunopeptidomes for HLA ligands derived from 1062 uORFs with highest scores predicting uORF functionality [3] and 1236 uORFs manually selected due to previous experimental data [9], genetic context or cancer association (database available at PRIDE PXD025716). Strikingly, HLA uLigands were identified in 82% (74/90) of the samples (91% (41/45) of malignant and 73% (33/45) of benign tissue samples, Fig. 3d). A total of 125 unique HLA uLigands derived from 120 different uORFs of 79 different genes including ASNSD1, ATF5, MAPK1, and transmembrane protein 203 (TMEM203) were identified (Supplementary Data 1). The frequency of HLA uLigands within the total immunopeptidome varies from 0.00 to 0.35% (median 0.06%) with no significant differences between malignant (range 0.00–0.32%, median 0.05%) and benign (range 0.00–0.35%, median 0.07%) tissue samples (Fig. 3e, Supplementary Data 1). The number of identified HLA uLigands correlates significantly with the size of the individual immunopeptidomes (Fig. 3f). HLA uLigands are presented

by 30 different HLA class I allotypes (7 HLA-A, 17 HLA-B, 6 HLA-C) with 14/125 HLA uLigands presented on more than one allotype resulting in 140 unique HLA uLigand-allotype combinations (Supplementary Data 1). HLA uLigands showed different ligand- and sample-specific intensity ranks covering the whole range of immunopeptidome peptide abundance (Fig. 3g and Supplementary Figure 8). The peptide length distribution is similar between uORF-derived and non-uORF-derived HLA ligands with 79% and 80% of the HLA ligands being 9/10mers, respectively, showing the characteristic length distribution of HLA class I-presented peptides (Fig. 3h).

Comparative immunopeptidome profiling delineates tumor-associated and high abundant HLA uLigand presentation in malignancy

For the identification of tumor-associated HLA uLigands, we performed comparative immunopeptidome profiling of the malignant and benign tissue datasets. Overlap analysis of all identified HLA uLigands revealed 66% (82/125) tumor-exclusive HLA uLigands that were never detected on benign tissue samples (Fig. 4a, Supplementary Data 1). For the identification of high frequent tumor-associated uORF antigens, tumor-exclusive HLA uLigands were ranked according to their frequency within the malignant tissue dataset (Fig. 4b). We identified 16/82 (20%) HLA uLigands with representation in two or more malignant tissue samples independent of the HLA allotype. The allotype-specific frequencies within the malignant dataset rose up to 80% for HLA-A*68-, 29% for HLA-A*03-, and 21% for HLA-B*07-restricted HLA uLigands in HLA-matched samples. The 16 tumor-associated uORF antigens could be further divided into tumor entity-specific subgroups with 4/16 AML-specific, 1/16 CLL-specific, and 2/16 Mel-specific HLA uLigands as well as 9/16 HLA uLigands presented by multiple entities (Fig. 4b, Supplementary Data 1). Furthermore, 15/82 (18%) tumor-enriched HLA uLigands defined by at least two-fold higher frequency in the malignant tissue dataset compared to the benign tissue dataset were identified (Fig. 4b, Supplementary Data 1). As a third interesting group of HLA uLigands, high frequent (identified in ≥ 5 samples in the total immunopeptidomics dataset) HLA uLigands (9/125, 7%) presented on both, malignant and benign tissue samples, were distinguished with allotype-specific representation frequencies up to 88% in allotype-matched samples (Fig. 4b, Supplementary Data 1). Since the HLA uLigand P_{TMEM203_B*07/C*16} (RSAGPRPAL) showed the highest frequency of tumor-specific presentation (8/45, 17.8%), we performed additional functional testing on the TMEM203 TLS (Fig. 4c). We detected a strong signal of HA-tagged TMEM203 AUG.1 peptide ectopically expressed in HEK293T cells that was lost upon introduction of an

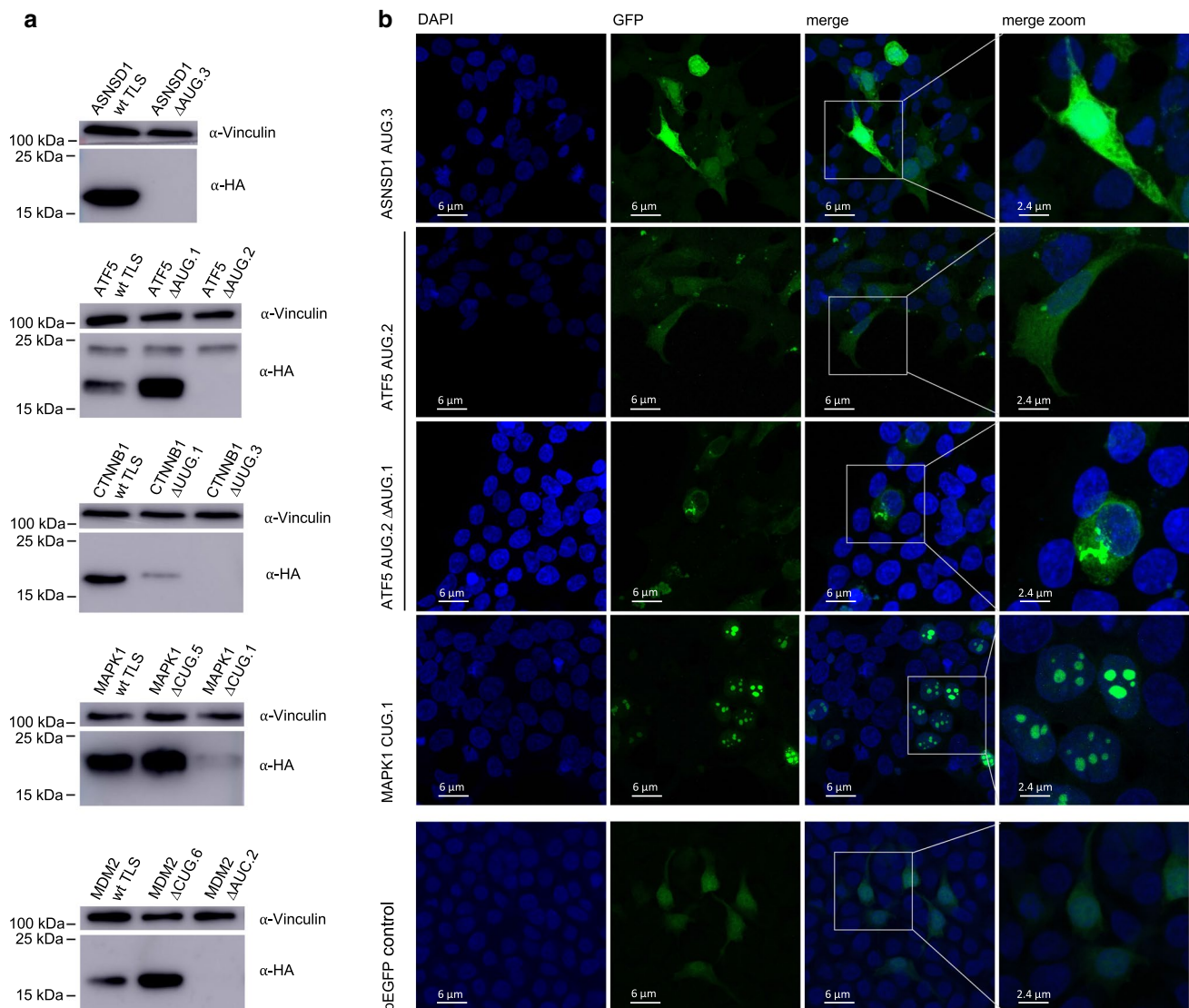


Fig. 2 Evidence for translation and specific cellular localization of uPeptides in cancer-associated transcripts. **a** Representative immunoblots of ≥ 3 independent experiments using HEK293T cell lysates prepared 52 h after transfection of expression vectors containing indicated triple HA-tagged wt and Δ uORF TLS variants. Eight hours prior to lysis cells were exposed to proteasome inhibitor MG132. **b** Representative pictures of ≥ 3 independent experiments showing the

expression and intracellular localization of indicated EGFP-tagged uPeptides as detected 24 h after transfection of HEK293T cells. Upstream peptides were expressed from TLS-EGFP-expression vectors containing the complete 5'-upstream sequence of indicated wt TLSs and an EGFP-tag replacing the uStop codon of the investigated uORF. The pictures shown here are presented as merged z-stack images. Additional pictures are presented in Supplementary Fig. 5

AUG.1 > ACC mutation to the TMEM203 TLS (Fig. 4d). In dual-luciferase reporter assays, we observed a 6.12-fold (± 0.32 SEM, $p \leq 0.01$) increase of luciferase activity for an AUG.1 > ACC Δ uORF TLS variant compared to wt TLS signals (Fig. 4e). Furthermore, in immunofluorescence experiments the TMEM203 AUG.1 uPeptide localized to the nucleus in the majority of cells, resembling the focal enrichment observed for the MAPK1 CUG.1 uPeptide before. Additionally, in approximately 20% of analyzed EGFP⁺ cells, a focal localization was also observed within

the cytoplasm, indicating two potential sites of functional implication (Fig. 4f, Supplementary Figure 5).

Naturally presented HLA uLigands induce multifunctional peptide-specific T cells

Using isotope-labeled synthetic peptides, we could validate 94% (17/18) of a selected panel of experimental HLA uLigands identifications (Fig. 5a–c, Supplementary Figure 9, Supplementary Table 7), including those derived from the ASNSD1, ATF5, CTNMB1, MAPK1,

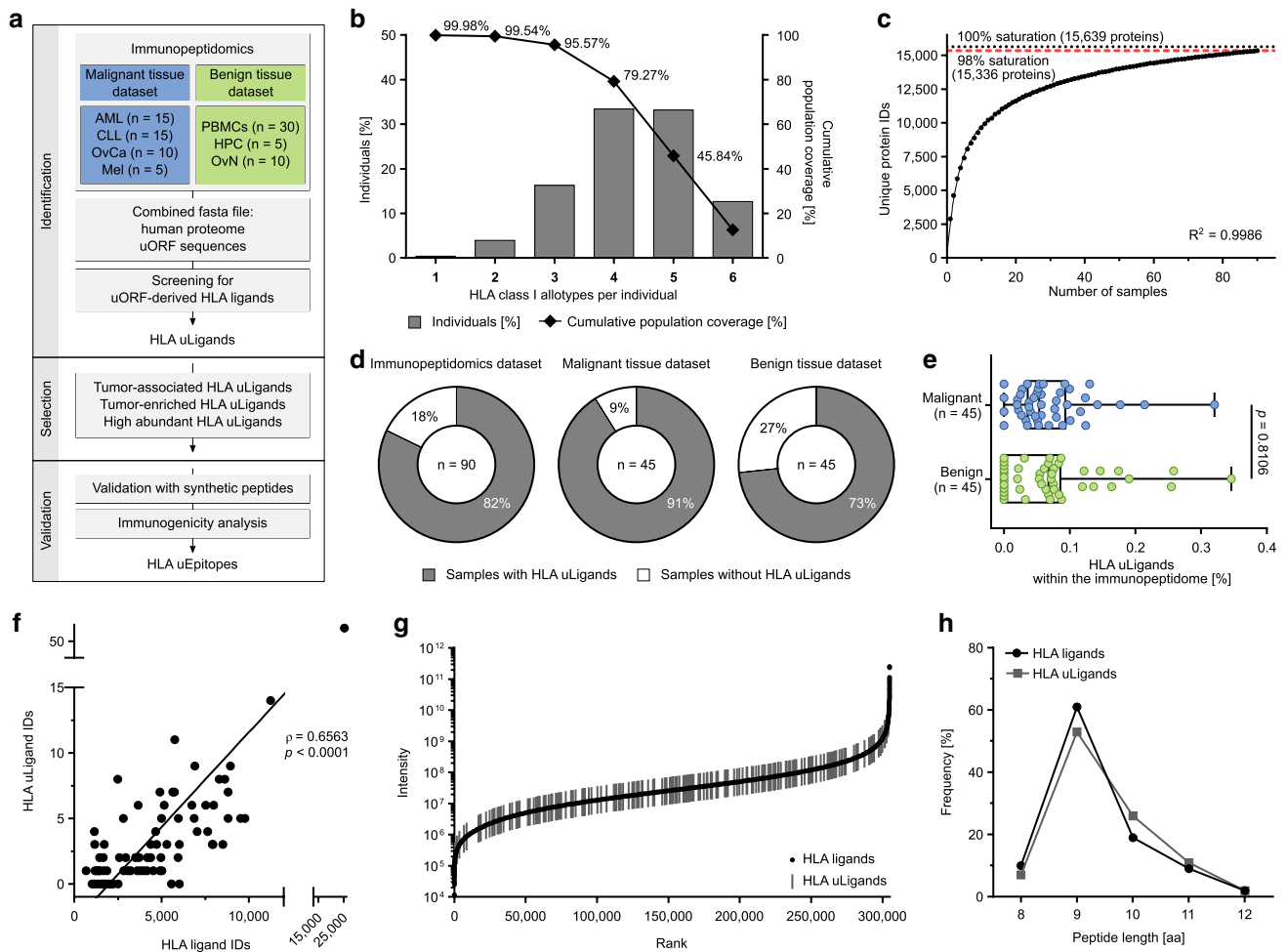


Fig. 3 Mass spectrometry-based identification of uORF-derived HLA ligands. **a** Workflow of immunopeptidomics-based identification of uORF-derived HLA ligands (HLA uLigands) and T cell epitopes (HLA uEpitopes). **b** HLA class I allotype population coverage within the immunopeptidomics dataset ($n=90$) compared to the world population (calculated by the IEDB population coverage tool, www.iedb.org). The frequencies of individuals within the world population carrying up to six HLA allotypes (x -axis) included in the immunopeptidomics dataset are indicated as grey bars on the left y -axis. The cumulative percentage of population coverage is depicted as black dots on the right y -axis. **c** Saturation analysis of HLA ligand source proteins of the immunopeptidomics dataset ($n=90$). Number of unique HLA ligand source protein identifications shown as a function of cumulative immunopeptidome analysis. Exponential regression allowed for the robust calculation ($R^2=0.9986$) of the maximum attainable number of different source protein identifications (100% saturation, dotted line). The dashed red line depicts the source proteome coverage achieved in the immunopeptidome dataset. **d** Pie charts depicting the percentage of samples with identified HLA uLigands within the total

immunopeptidomics dataset comprising malignant and benign tissue samples ($n=90$, left panel) as well as within the malignant ($n=45$, middle panel) and benign ($n=45$, right panel) tissue datasets separately. **e** Percentage of HLA uLigands within the immunopeptidome of malignant and benign tissue samples (boxes represent median and 25th–75th percentiles, whiskers are minimum to maximum, two-sided Mann–Whitney U -test). **f** Correlation of total HLA ligand identifications with HLA uLigand identifications in the immunopeptidome dataset ($n=90$). Dots represent individual samples. Spearman's rho (ρ) and p -value. **g** Ranked intensity values of mass spectrometry-acquired data derived from the combined immunopeptidomes of all samples ($n=90$). Positions of HLA uLigands are projected on the curve. **h** Peptide length distribution of HLA uLigands and all identified HLA ligands. *AML* acute myeloid leukemia, *CLL* chronic lymphocytic leukemia, *OvCa* ovarian carcinoma, *Mel* melanoma, *PBMCs* peripheral blood mononuclear cells, *HPC* CD34-enriched hematopoietic progenitor cells, *OvN* benign ovaries, *HLA uLigands* uORF-derived HLA ligands, *IDs* identifications, *aa* amino acid

and TMEM203 uPeptides previously detected by immunoblot experiments (Figs. 2, 4d). To assess the immunogenicity of HLA uLigands, we selected a panel of tumor-associated HLA uLigands presented on the common HLA allotypes HLA-A*02, -A*03, and -B*07. We

performed in vitro aAPC-based priming of naive CD8⁺ T cells from HVs using HLA:peptide monomers of the HLA-A*02-, -A*03-, and -B*07-restricted HLA uLigands P_{ATF5_A*02} (SILQSLVPA), P_{MAPK1_A*03} (ALHQPLVHR), and P_{TMEM203_B*07/C*16} (RSAGPRPAL). De novo priming

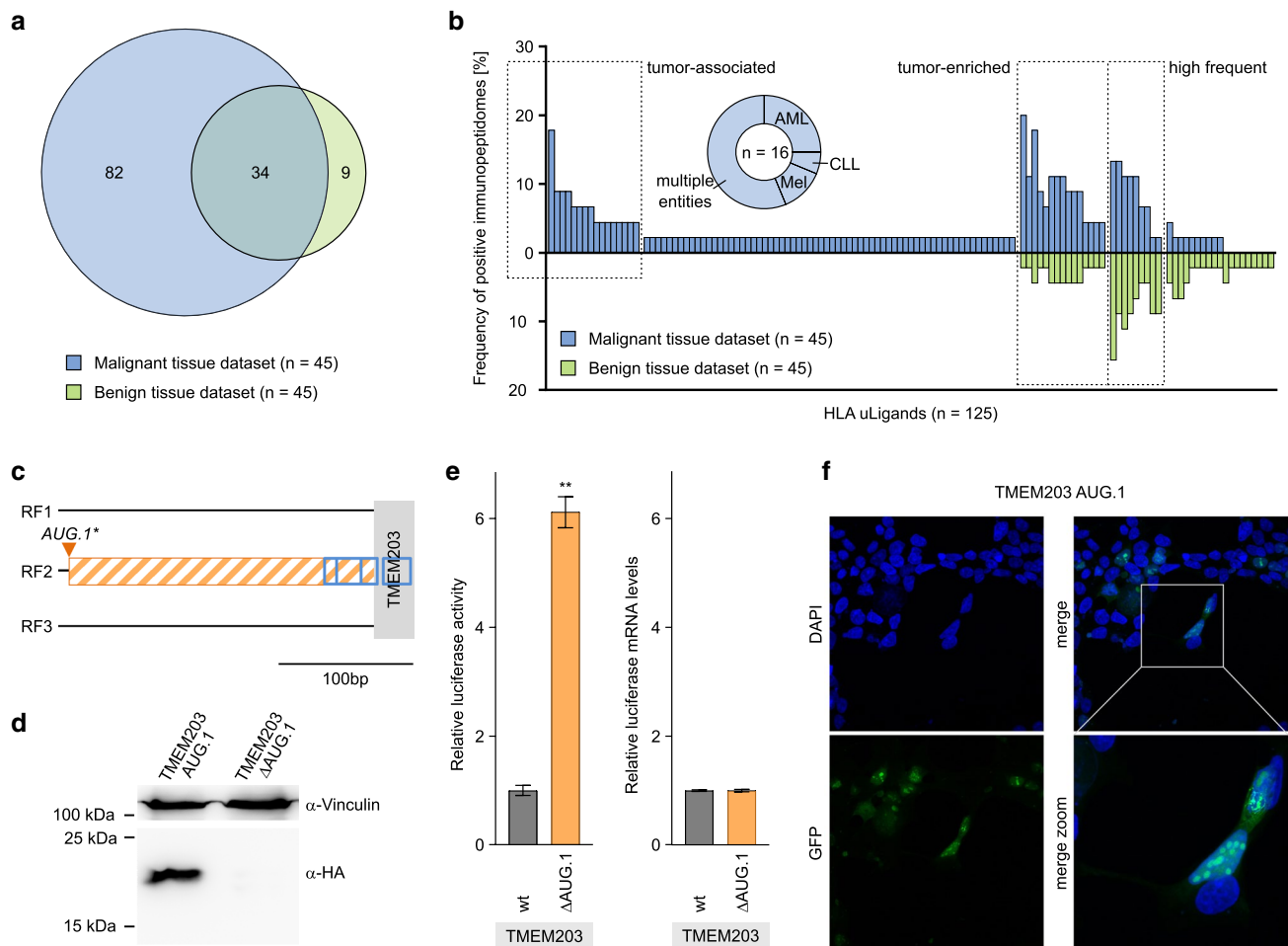


Fig. 4 Comparative immunopeptidome profiling identified uORF-derived tumor antigens. **a** Overlap analysis of uORF-derived HLA ligand identifications of primary malignant ($n=45$) and benign ($n=45$) tissue samples. **b** Comparative profiling of HLA uLigands ($n=125$) based on HLA-restricted presentation frequency in malignant and benign immunopeptidomes. Frequencies of positive immunopeptidomes for the respective HLA uLigands (x -axis) are indicated on the y -axis. The left box highlights tumor-associated antigens ($n=16$) showing malignant-exclusive frequent presentation. The donut chart displays the entity specificity of tumor-associated HLA uLigands. The middle box marks tumor-enriched antigens displaying two-fold representation frequency on malignant compared to benign tissue samples. The right box highlights high abundant antigens ($n=9$) showing a frequent presentation in ≥ 5 samples. **c** Schematic illustration of the TMEM203 TLS displaying the only canonical uORF (hatched orange box) on reading frame 2 (black line). The grey box indicates the start of the CDS. Blue open boxes mark the

identified HLA uLigands. **d** Representative immunoblot of ≥ 3 independent experiments using HEK293T cell lysates prepared 52 h after transfection of expression vectors containing 3xHA-tagged TMEM203 AUG.1 wt and Δ uORF-TLSs. Eight hours prior to lysis the cells were exposed to proteasome inhibitor MG132. Exposure time was 3.5 min. **e** Bar graph showing relative Firefly luciferase activities and mRNA levels detected for indicated wt and Δ uORF TMEM203 TLSs normalized to Renilla luciferase internal controls. Results are combined from three independent experiments. Error bars indicate SEM. Level of significance $p \leq 0.01$ (**) as determined by two-tailed nonparametric Mann–Whitney U -tests. **f** Representative pictures of ≥ 3 independent experiments showing the intracellular localization of the EGFP-tagged TMEM203 AUG.1 uPeptide as detected 24 h after transfection of HEK293T cells. AML acute myeloid leukemia, CLL chronic lymphocytic leukemia, Mel melanoma, RF reading frame, wt wild type

and expansion of antigen-specific T cells was observed for all three HLA uLigands in 100% of analyzed HVs ($n=3$) with frequencies of peptide-specific T cells ranging from 0.11–0.83% (mean 0.26%) within the viable CD8⁺ T cell population (Fig. 5d–f). Furthermore, multifunctionality of the induced P_{ATF5_A*02}⁺, P_{MAPK1_A*03}⁺,

and P_{TMEM203_B*07/C*16}-specific T cells was shown using intracellular cytokine staining (ICS) for IFN- γ and TNF as well as degranulation marker staining for CD107a (Fig. 5g–i) validating these tumor-associated HLA uLigands as uORF-derived T cell epitopes (HLA uEpitopes).

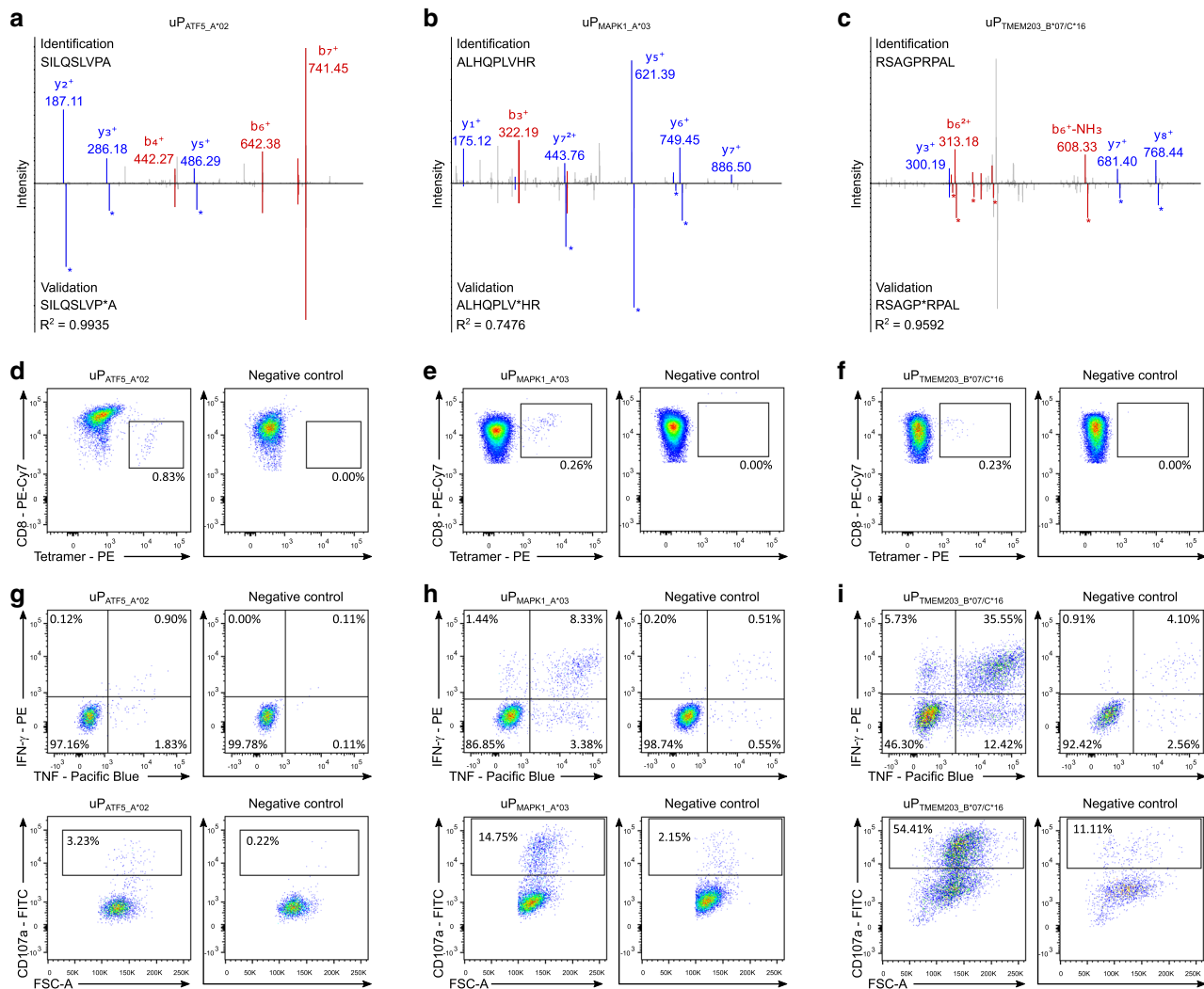


Fig. 5 Spectral validation and immunogenicity analyses of uORF-derived tumor antigens. **a–c** Validation of the experimentally eluted peptides **a** uP_{ATF5_A*02} , **b** uP_{MAPK1_A*03} , and **c** $uP_{TM2EM203_B*07/C*16}$. Comparison of fragment spectra (m/z on x -axis) of HLA uLigands eluted from primary samples (identification) to their corresponding isotope-labeled synthetic peptides (validation, mirrored on the x -axis) with the calculated spectral correlation coefficient (R^2). Identified b- and y-ions are marked in red and blue, respectively. Ions containing isotope-labeled amino acids are marked with asterisks. **d–f** Naïve CD8⁺ T cells were primed in vitro using HLA uLigand-loaded aAPCs with the HLA-A*02-, -A*03-, and -B*07-restricted peptides **d** uP_{ATF5_A*02} , **e** uP_{MAPK1_A*03} , and **f** $uP_{TM2EM203_B*07/C*16}$, respectively. Graphs show single, viable cells stained for CD8 and PE-

conjugated multimers of indicated specificity. The left panels show HLA uLigand tetramer staining, the right panels (negative control) depict tetramer staining of T cells from the same donor primed with an HLA-matched control peptide. **g–i** Functional characterization of HLA uLigand-specific CD8⁺ T cells by intracellular cytokine staining. Representative examples of IFN- γ and TNF production (upper panels) as well as CD107a expression (lower panels) after stimulation of **g** uP_{ATF5_A*02} -, **h** uP_{MAPK1_A*03} -, and **i** $uP_{TM2EM203_B*07/C*16}$ -specific CD8⁺ T cells with the HLA-A*02-, -A*03-, and -B*07-restricted peptides uP_{ATF5_A*02} , uP_{MAPK1_A*03} , and $uP_{TM2EM203_B*07/C*16}$, respectively (left panels) compared to a negative HLA-matched control peptide (right panels)

Discussion

The present work uncovers novel regulatory and immunological functions of uORFs in cancer. We provide evidence for uORF-associated translational regulation of cancer-associated genes as well as uPeptide translation and HLA-restricted presentation as cancer-associated T cell epitopes.

The translation of mRNAs into proteins is a key event in the regulation of gene expression. This is especially true in the cancer setting, as many oncogenes are regulated at this level. Upstream ORFs can impact gene expression of the downstream CDS by triggering mRNA decay or by regulating translation [10, 19–21]. Especially in the context of malignancies defective uORF-mediated regulation may have profound physiological, immunogenic and pathogenic

consequences [11, 33, 35]. The mode of action appears to be highly uORF- and uPeptide-specific and may entail co-factor-induced ribosome stalling [37, 65, 66].

Our data revealed uORF-mediated translational regulation in the majority of analyzed cancer-associated transcripts. While previous global analyses demonstrated that uORF carrying transcripts show an overall reduction of translation of the associated downstream main proteins [4, 8, 10, 15, 67–69], data from this and previous projects [15] showed that depending on the individual TLS context, the translation of distinct uAUG and aTIS uORFs may have either activating (*e.g.* ATF5 Δ uAUU.1, JAK2 Δ uCUG.2 and 12, MDM2 Δ uCUG.6) or repressive effects (*e.g.* ASNSD1 Δ uAUG.3, ATF5 Δ uAUG.2, ERBB2 Δ uAUU.2) on CDS translation.

Furthermore, our data provide direct evidence for uPeptide translation both, *in vitro* and in primary human cells, and demonstrate that the strength of translational regulation and the capability to initiate uPeptide translation can be similar for uAUG and aTIS codons. These data extend and are in line with recent genome-wide ribosome profiling studies confirming the widespread presence and translation of AUG and aTIS uORFs [1, 2, 6, 70, 71]. Dual-luciferase reporter studies of computationally predicted functional uORFs often revealed translational regulatory activity from the respective TLSs [3, 5, 8]. However, some of the high-ranking uAUG or aTIS uORFs analyzed in this study did neither result in changes of luciferase reporter activity nor did they initiate translation of uPeptides under the specific experimental conditions applied here. This highlights the need for individual experimental testing of uORF-mediated translational control and uPeptide functions, as results may vary depending on the cellular context and the global translational and environmental conditions.

In several cases, the actual uPeptide initiating codon was distinct from the one predicted by the highest uORF or uPeptide scores. Furthermore, for some TLSs several alternative uPeptides or incomplete uPeptide ablation after deletion of the initiation codon could be observed. This demonstrates that multiple upstream initiation codons may contribute to uPeptide translation, each conferring individual levels of translational regulation to the transcript. Future studies may systematically search for uPeptide interacting protein-cofactors, metabolites, or small molecule interactors, capable to specifically induce ribosome stalling and to ablate translation of harmful downstream oncogenic proteins.

Genome-scale ribosome profiling studies have allowed for the identification of large populations of uORFs known to undergo translation [1–3, 71]. However, the detectability of the translation products by standard mass spectrometry-based proteomics approaches using tryptic digestion is limited [28, 32, 72–74] due to challenges in detecting trypsin-digested fragments from these short uPeptides, which are

presumably characterized by high turnover rates [72]. We here provide direct evidence for the frequent translation and cellular expression of cancer-associated uPeptides by immunoblotting and by immunofluorescence-based microscopy. Some of the ectopically expressed uPeptides, including the MAPK1 CUG.1 and TMEM203 AUG.1 uPeptides, showed highly specific intracellular localizations, suggesting individual functional implications for the respective uPeptides. As the function of an individual uPeptide is not necessarily related to the function of the associated main protein, as exemplified for the ASNSD1 uPeptide [75], it is too early to further speculate on the potential functional implications of the uPeptides detected here. Of note, the rather large EGFP-tag may have influenced both, expression level and localization, but the differences observed across individual uPeptides argue for a predominant impact of the uPeptide causing the specific staining patterns observed. Future work is required to validate and extend on these observations in additional cell types and under various global translational conditions, for example by applying uPeptide-specific antibodies or split-GFP-based techniques [76, 77].

Furthermore, mass spectrometry-based immunopeptidome analysis in primary tumor and healthy tissues identified uORF-derived HLA-presented antigens, validating the observations of uPeptide expression upon ectopic expression *in vitro*. This demonstrates in accordance with recent immunopeptidomics studies [28, 29, 31], which, however, were mainly limited to cell lines, that uPeptides enter the HLA class I presentation pathway and contribute to the antigen repertoire also *in vivo*. In contrast to the recently published individualized proteogenomic approaches [29, 31], we here applied an approach using a generic uORF database comprising preselected sequences. Using this strategy, we were able to identify HLA uLigands shared between several samples. We further provide unprecedented evidence for tumor-associated presentation of HLA uLigands in this comprehensive cohort including various different hematological and solid tumor entities as well as different benign tissue samples. The inclusion of benign tissue-derived immunopeptidomes enabled the direct identification of tumor-associated and tumor-enriched HLA uLigands that were never or only rarely presented on benign tissues. This represents a major advantage compared to retrospective approaches using RNA sequencing data [45] facing the drawback that the immunopeptidome is an independent complex layer formed by the antigen presentation machinery and therefore does not necessarily mirror the transcriptome nor the proteome [32]. The direct comparison of benign and malignant tissue-derived immunopeptidome data further is of central importance for the definition of tumor-associated HLA uLigands as it was recently shown that the presentation of HLA uLigands and other cryptic peptides is not restricted to tumor tissues [78]. Tumor-associated cryptic peptides from non-coding regions, including

5'-TLS and 3'-UTR, non-coding RNAs, intronic, intergenic and off-frame regions, represent highly promising targets for anti-tumor immune surveillance as well as the development of immunotherapeutic approaches [30]. In contrast to classical neoepitopes, derived from tumor-specific missense point mutations affecting only one amino acid, these cryptic peptides differ by several amino acids from their respective wild type sequence and thus are even more likely to induce tumor-specific immune responses [79]. Furthermore, their shared presentation across multiple donors and even tumor entities, as so far only described for unmutated tumor-associated self-peptides derived from canonical proteins [44, 47, 80–84], enables a broader applicability compared to private neoantigens. In the future, large cohort studies are needed to analyze HLA uLigand presentation in the evolution of malignant disease (*e.g.* primary diagnosis versus relapse), in different tumor stages, and under anti-cancer treatments.

Focusing on a restricted set of uORFs, this work provides evidence that uORF-derived peptides can be processed into tumor-associated HLA-presented peptides detectable on primary human samples, even without the need for whole-exome and RNA sequencing and the assembly of sample-specific, personalized databases. These data may encourage further studies to screen all of the approximately 190 thousand uAUG and 2.5 million aTIS codons within the human genome [15] and to unravel the whole uORF-derived immunopeptidome landscape in cancer.

At present the pathophysiological role of uORF-derived tumor-associated antigens in cancer immune surveillance is unsettled as spontaneous immune recognition in cancer patients was limited [29]. This might at least in part be due to immunopeptidomics analysis of patient-derived cell lines showing a different pattern of HLA uLigand presentation compared to primary samples. Moreover, high turnover rates of unfunctional uPeptides may limit the uptake by professional antigen-presenting cells, preventing effective priming of naïve T cells. This suggests that further target antigen selection should be based on the knowledge of the functional role of the respective uPeptides. For the understanding of the functional role of uPeptides the investigation of their cellular localization is of particular importance. Chen *et al.* [28] and our data suggest specific and distinct cellular localizations for individual uPeptides highlighting the variety of cellular roles and functions that uORFs might fulfill beyond translational regulation [28, 37, 75, 85].

Conclusion

The data presented in this work demonstrate the translational regulatory effect of uAUG and aTIS uORFs in cancer-associated transcripts and provide direct evidence for

the cellular expression and the HLA-restricted presentation of uORF-derived peptides on primary tissue samples. The data suggest a widespread but largely unexplored regulatory and immunological role of uORFs and uORF-derived peptides in cancer biology. These observations may inspire the development of novel anti-cancer therapies, comprising direct molecular targeting of uORFs or the respective uPeptides as well as immunotherapeutic targeting of tumor-associated HLA uLigands.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00018-022-04145-0>.

Acknowledgements We thank Karin Busch (University of Muenster, Germany, Institute of molecular cell biology) for sharing equipment and expertise for immunofluorescence microscopy. We thank Oliver Klaas for help during the establishment of experimental setups and Ulrike Schmidt, Ulrich Wulle and Claudia Falkenburger for technical support. We thank the Department of Hematology and Oncology, Tübingen, Germany for providing tumor samples analyzed in this work.

Author contributions A.N., L.F., L.J., H.-G.R., J.S.W., and K.W. designed the study; A.N. and J.B. performed immunopeptidome experiments; A.N. conducted in vitro T cell experiments; L.F., L.J., L.S., and E.H. performed Luciferase assays, immunoblot analysis and immunofluorescence-based microscopy. A.N., L.F., L.J., J.S.W., and K.W. analyzed data; A.N., L.F., L.J., J.S.W., and K.W. wrote the manuscript; all authors revised the manuscript; H.-G.R., C.S., T.K., G.L., J.S.W., and K.W. supervised the study.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by the Deutsche Krebshilfe e.V., Bonn, Germany, grant 70113632 to K.W., the ‘Clinician Scientist Program’ and the ‘MedK Program’ of the Deanery of the medical Faculty of the university of Muenster, Germany, to K.W. and L.F., the Eurostars-2 programme (Grant E!11969 compare) to C.S., the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, Grant WA 4608/1–2) to J.S.W., the Deutsche Forschungsgemeinschaft under Germany’s Excellence Strategy (Grant EXC2180 390900677) to H.-G.R. and J.S.W., the German Cancer Consortium (DKTK) to H.-G.R., the Wilhelm Sander Stiftung (Grant 2016.177.2 and 2016.177.3) to J.S.W., the José Carreras Leukämie-Stiftung (Grant DJCLS 05 R/2017) to J.S.W. and the Fortüne Program of the University of Tübingen (Fortüne number 2451–0–0) to J.S.W.

Availability of data and materials The mass spectrometry data as well as the FASTA have been deposited to the ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) via the PRIDE partner repository [86] with the dataset identifier PXD025716. Raw files of ovarian carcinoma, benign ovaries [44] and melanoma [45] samples have been downloaded from the PRIDE partner repository (ProteomeXchange Consortium (<http://proteomecentral.proteomexchange.org>) with the dataset identifier PXD007635 and PXD004894.

Declarations

Conflict of interest H.-G.R. is shareholder of Immatics Biotechnologies GmbH, Synimmune GmbH, and Curevac AG. The other authors declare no competing interests.

Ethics approval and consent to participate Informed consent was obtained in accordance with the Declaration of Helsinki protocol. The

study was performed according to the guidelines of the local ethics committees (373/2011B02, 454/2016B02, 406/2019B02).

Consent for publication Not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References




- Ingolia NT, Lareau LF, Weissman JS (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147:789–802
- Ingolia NT, Ghaemmaghami S, Newman JRS, Weissman JS (2009) Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science* 324:218–223
- McGillivray P et al (2018) A comprehensive catalog of predicted functional upstream open reading frames in humans. *Nucleic Acids Res* 46:3326–3338
- Johnstone TG, Bazzini AA, Giraldez AJ (2016) Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J* 35:706–723
- Wethmar K et al (2016) Comprehensive translational control of tyrosine kinase expression by upstream open reading frames. *Oncogene* 35:1736–1742
- Lee S et al (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. In: *Proceedings of the National Academy of Sciences of the United States of America*, vol 109
- Young SK, Wek RC (2016) Upstream open reading frames differentially regulate genespecific translation in the integrated stress response. *J Biol Chem* 291:16927–16935
- Calvo SE, Pagliarini DJ, Mootha VK (2009) Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. *PNAS* 106:7507–7512
- Wethmar K, Barbosa-Silva A, Andrade-Navarro MA, Leutz A (2014) UORFdb—a comprehensive literature database on eukaryotic uORF biology. *Nucleic Acids Res* 42:D60–D67
- Barbosa C, Peixeiro I, Romão L (2013) Gene expression regulation by upstream open reading frames and human disease. *PLoS Genet* 9:e1003529
- Whiffin N et al (2020) Characterising the loss-of-function impact of 5' untranslated region variants in 15,708 individuals. *Nat Commun* 11(1):2523
- Occhi G et al (2013) A novel mutation in the upstream open reading frame of the CDKN1B gene causes a MEN4 phenotype. *PLoS Genetics* 9:e1003350
- Liu L et al (1999) Mutation of the CDKN2A 5'UTR/5'UTR creates an aberrant initiation codon and predisposes to melanoma. *Nat Genet* 21:128–132 <http://genetics.nature.com>
- Zou Q et al (2019) Survey of the translation shifts in hepatocellular carcinoma with ribosome profiling. *Theranostics* 9:4141–4155
- Jürgens L et al (2021) Somatic functional deletions of upstream open reading frame-associated initiation and termination codons in human cancer. *Biomedicines* 9:618
- Ho JSY et al (2020) Hybrid gene origination creates human-virus chimeric proteins during infection. *Cell* 181:1502–1517.e23
- He F, Jacobson A (2015) Nonsense-mediated mRNA decay: degradation of defective transcripts is only part of the story. *Annu Rev Genet* 49:339–366
- Wethmar K (2014) The regulatory potential of upstream open reading frames in eukaryotic gene expression. *Wiley Interdiscip Rev RNA* 5:765–778
- Crappé J et al (2013) Combining in silico prediction and ribosome profiling in a genome-wide search for novel putatively coding sORFs. *BMC Genom* 14:648
- Fritsch C et al (2012) Genome-wide search for novel human uORFs and N-terminal protein extensions using ribosomal footprinting. *Genome Res* 22:2208–2218
- Dever TE, Kinzy TG, Pavitt GD (2016) Mechanism and regulation of protein synthesis in *Saccharomyces cerevisiae*. *Genetics* 203:65–107
- Vattem KM, Wek RC (2004) Reinitiation involving upstream ORFs regulates ATF4 mRNA translation in mammalian cells. *Proc Natl Acad Sci USA* 101:11269–11274. <https://doi.org/10.1073/pnas.0400541101>
- Lu PD, Harding HP, Ron D (2004) Translation reinitiation at alternative open reading frames regulates gene expression in an integrated stress response. *J Cell Biol* 167:27–33
- Hinnebusch AG (2005) Translational regulation of GCN4 and the general amino acid control of yeast. *Annu Rev Microbiol* 59:407–450
- Harding HP, Zhang Y, Zeng H, Novoa I, Lu PD, Calton M, Sadri N, Yun C, Popko B, Paules R et al (2003) An integrated stress response regulates amino acid metabolism and resistance to oxidative stress. *Mol Cell* 11:619–633
- Schulz J et al (2018) Loss-of-function uORF mutations in human malignancies. *Sci Rep* 8:2395
- Schuster SL, Hsieh AC (2019) The untranslated regions of mRNAs in cancer. *Trends Cancer* 5:245–262
- Chen J et al (2020) Pervasive functional translation of non-canonical human open reading frames. *Science* 367:1140–1146
- Chong, C et al (2020) Integrated proteogenomic deep sequencing and analytics accurately identify non-canonical peptides in tumor immunopeptidomes. *Nat Commun* 11:1293
- Erhard F, Dölken L, Schilling B, Schlosser A (2020) Identification of the cryptic HLA-I immunopeptidome. *Cancer Immunol Res* 8:1018–1026
- Laumont CM et al (2018) CANCER noncoding regions are the main source of targetable tumor-specific antigens. *Sci Transl Med* 10:470. <http://stm.sciencemag.org/>
- Ouspenskaia T et al (2020) Thousands of novel unannotated proteins expand the MHC I immunopeptidome in cancer. doi:<https://doi.org/10.1101/2020.02.12.945840>.
- Dever TE, Ivanov IP, Sachs MS (2020) Conserved upstream open reading frame nascent peptides that control translation. *Annu Rev Genet* 12:8
- Pendleton LC, Goodwin BL, Solomonson LP, Eichler DC (2005) Regulation of endothelial argininosuccinate synthase expression and NO production by an upstream open reading frame. *J Biol Chem* 280:24252–24260
- Andrews SJ, Rothnagel JA (2014) Emerging evidence for functional peptides encoded by short open reading frames. *Nat Rev Genet* 15:193–204
- Jayaram DR, Frost S (2021) Unraveling the hidden role of a uORF-encoded peptide as a kinase inhibitor of PKCs. *Proc Natl Acad Sci USA* 118:e201889118

37. Chen HH, Tarn WY (2019) uORF-mediated translational control: recently elucidated mechanisms and implications in cancer. *RNA Biol* 16:1327–1338
38. Vogelstein B et al (2013) Cancer genome landscapes. *Science* 339:1546–1558
39. Bailey MH et al (2018) Comprehensive characterization of cancer driver genes and mutations. *Cell* 173:371–385.e18
40. Sanchez-Vega F et al (2018) Oncogenic signaling pathways in the Cancer Genome Atlas. *Cell* 173:321–337.e10
41. Hampf M, Gossen M (2006) A protocol for combined photinus and renilla luciferase quantification compatible with protein assays. *Anal Biochem* 356:94–99
42. Schneider CA, Rasband WS, Eliceiri KW (2012) NIH Image to ImageJ: 25 years of image analysis. *Nat Methods* 9:671–675
43. Hallek M et al (2018) Special report iwCLL guidelines for diagnosis, indications for treatment, response assessment, and supportive management of CLL. <http://ashpublications.org/blood/article-pdf/131/25/2745/1465960/blood806398.pdf>
44. Schuster H et al (2017) The immunopeptidomic landscape of ovarian carcinomas. *Proc Natl Acad Sci USA* 114:E9942–E9951
45. Bassani-Sternberg M et al (2016) Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* 7:13404
46. Nelde A, Kowalewski DJ, Stevanović S (2019) Purification and identification of naturally presented MHC class I and II ligands. In: van Ender P (eds) *Antigen processing*. *Methods molecular biology* vol 1988, pp 123–136
47. Kowalewski DJ et al (2015) HLA ligandome analysis identifies the underlying specificities of spontaneous antileukemia immune responses in chronic lymphocytic leukemia (CLL). *Proc Natl Acad Sci USA* 112:E116–E175
48. Nelde A et al (2018) HLA ligandome analysis of primary chronic lymphocytic leukemia (CLL) cells under lenalidomide treatment confirms the suitability of lenalidomide for combination with T-cell-based immunotherapy. *OncoImmunology* 7:e1316438
49. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989
50. Käll L, Canterbury JD, Weston J, Noble WS, MacCoss MJ (2007) Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4:923–925
51. Hoof I et al (2009) NetMHCpan, a method for MHC class I binding prediction beyond humans. *Immunogenetics* 61:1–13
52. Pedersen SR et al (2016) Immunogenicity of HLA class I and II double restricted influenza a-derived peptides. *PLoS One* 11(1):e0145629
53. Jurtz V et al (2017) NetMHCpan-4.0: improved peptide–MHC class I interaction predictions integrating eluted ligand and peptide binding affinity data. *J Immunol* 199:3360–3368
54. Schuler MM, Nastke MD, Stevanović S (2007) SYFPEITHI: database for searching and T-cell epitope prediction. *Methods Mol Biol (Clifton, NJ)* 409:75–93
55. Sturm T et al (2013) Mouse urinary peptides provide a molecular basis for genotype discrimination by nasal sensory neurons. *Nat Commun* 4:1616
56. Altman JD et al (1996) Phenotypic analysis of antigen-specific T lymphocytes. *Science* 274:94–96
57. Peper JK et al (2016) HLA ligandomics identifies histone deacetylase 1 as target for ovarian cancer immunotherapy. *OncoImmunology* 5(5):e1065369
58. Rudolf D et al (2008) Potent costimulation of human CD8 T cells by anti-4-1BB and anti-CD28 on synthetic artificial antigen presenting cells. *Cancer Immunol Immunother* 57:175–183
59. Widenmeyer M et al (2012) Promiscuous survivin peptide induces robust CD4 + T-cell responses in the majority of vaccinated cancer patients. *Int J Cancer* 131:140–149
60. Neumann A et al (2013) Identification of HLA ligands and T-cell epitopes for immunotherapy of lung cancer. *Cancer Immunol Immunother* 62:1485–1497
61. Hulsen T, de Vlieg J, Alkema W (2008) BioVenn—a web application for the comparison and visualization of biological lists using area-proportional Venn diagrams. *BMC Genom* 9:488
62. Bui HH et al (2006) Predicting population coverage of T-cell epitope-based diagnostics and vaccines. *BMC Bioinform* 7:153
63. Vita R et al (2015) The immune epitope database (IEDB) 3.0. *Nucleic Acids Res* 43:D405–D412
64. Gonzalez-Galarza FF et al (2020) Allele frequency net database (AFND) 2020 update: gold-standard data classification, open access genotype data and new query tools. *Nucleic Acids Res* 48:D783–D788
65. Hardy S et al (2019) Magnesium-sensitive upstream ORF controls PRL phosphatase expression to mediate energy metabolism. *Proc Natl Acad Sci USA* 116:2925–2934
66. Nikonorova IA, Nv K, Dmitriev SE, Vassilenko KS, Ryazanov AG (2014) Identification of a Mg²⁺-sensitive ORF in the 5-leader of TRPM7 magnesium channel mRNA. *Nucleic Acids Res* 42:12779–12788
67. Somers J, Pöyry T, Willis AE (2013) A perspective on mammalian upstream open reading frame function. *Int J Biochem Cell Biol* 45:1690–1700
68. Morris DR, Geballe AP (2000) Upstream open reading frames as regulators of mRNA translation. *Mol Cell Biol* 20:8635–8642
69. Meijer HA, Thomas AAM (2002) Control of eukaryotic protein synthesis by upstream open reading frames in the 5 h-untranslated region of an mRNA. *Biochem J* 367:1–11
70. Spealman P et al (2018) Conserved non-AUG uORFs revealed by a novel regression analysis of ribosome profiling data. *Genome Res* 28:214–222
71. Brar GA et al (2012) High-resolution view of the yeast meiotic program revealed by ribosome profiling. *Science* 335:552–557
72. Oyama M et al (2004) Analysis of small human proteins reveals the translation of upstream open reading frames of mRNAs. *Genome Res* 14:2048–2052
73. Orr MW, Mao Y, Storz G, Qian SB (2021) Alternative ORFs and small ORFs: shedding light on the dark proteome. *Nucleic Acids Res* 48:1029–1042
74. Slavoff SA et al (2013) Peptidomic discovery of short open reading frame-encoded peptides in human cells. *Nat Chem Biol* 9:59–64
75. Cloutier P et al (2020) Upstream ORF-encoded ASDURF is a novel prefoldin-like subunit of the PAQosome. *J Proteome Res* 19:18–27
76. Feng S et al (2017) Improved split fluorescent proteins for endogenous protein labeling. *Nat Commun* 8(1):370
77. Kamiyama D et al (2016) Versatile protein tagging in cells with split fluorescent protein. *Nat Commun* 7:11046
78. Marcu A et al (2021) HLA Ligand Atlas: a benign reference of HLA-presented peptides to improve T-cell-based cancer immunotherapy. *J Immunol Ther Cancer* 9:e002071
79. Laumont CM, Perreault C (2018) Exploiting non-canonical translation to identify new targets for T cell-based cancer immunotherapy. *Cell Mol Life Sci* 75:607–621
80. Bilich T et al (2019) The HLA ligandome landscape of chronic myeloid leukemia delineates novel T-cell epitopes for immunotherapy. *Blood* 133:550–565
81. Walz S et al (2015) The antigenic landscape of multiple myeloma: Mass spectrometry (re)defines targets for T-cell-based immunotherapy. *Blood* 126:1203–1213

82. Berlin C et al (2015) Mapping the HLA ligandome landscape of acute myeloid leukemia: a targeted approach toward peptide-based immunotherapy. *Leukemia* 29:647–659
83. Neidert MC et al (2018) The natural HLA ligandome of glioblastoma stem-like cells: antigen discovery for T cell-based immunotherapy. *Acta Neuropathol* 135:923–938
84. Heidenreich F (2017) Mass spectrometry-based identification of a naturally presented receptor tyrosine kinase-like orphan receptor 1-derived epitope recognized by CD8+ cytotoxic T cells. *Haematologica* 102:e460–e464
85. Prensner JR et al (2021) Noncanonical open reading frames encode functional proteins essential for cancer cell survival. *Nat Biotechnol*. <https://doi.org/10.1038/s41587-020-00806-2>
86. Perez-Riverol Y et al (2019) The PRIDE database and related tools and resources in 2019: improving support for quantification data. *Nucleic Acids Res* 47:D442–D450

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Authors and Affiliations

Annika Nelde^{1,2,3}  · Lea Flötotto⁴ · Lara Jürgens⁴ · Laura Szymik⁴ · Elvira Hubert⁴ · Jens Bauer^{1,2,3}  · Christoph Schliemann⁴ · Torsten Kessler⁴ · Georg Lenz⁴ · Hans-Georg Rammensee^{2,3,5} · Juliane S. Walz^{1,2,3,6}  · Klaus Wethmar⁴

¹ Clinical Collaboration Unit Translational Immunology, Department of Internal Medicine, German Cancer Consortium (DKTK), University Hospital Tübingen, Otfried-Müller-Str. 10, 72076 Tübingen, Germany

² Department of Immunology, Institute for Cell Biology, University of Tübingen, 72076 Tübingen, Germany

³ Cluster of Excellence iFIT (EXC2180) “Image-Guided and Functionally Instructed Tumor Therapies”, University of Tübingen, 72076 Tübingen, Germany

⁴ Department of Medicine A, Hematology, Oncology, Hemostaseology and Pneumology, University Hospital

Münster, Albert-Schweitzer-Campus 1A, 48149 Münster, Germany

⁵ German Cancer Consortium (DKTK) and German Cancer Research Center (DKFZ), Partner Site Tübingen, 72076 Tübingen, Germany

⁶ Dr. Margarete Fischer-Bosch Institute of Clinical Pharmacology, Robert Bosch Center for Tumor Diseases (RBCT), 70376 Stuttgart, Germany