

TRADUZIONE DEL TESTO DEL BREVETTO EUROPEO N. 2714071

DAL TITOLO:

"VACCINI INDIVIDUALIZZATI PER IL CANCRO"

Depositata il:

*** **

DESCRIZIONE

CAMPO TECNICO DELL'INVENZIONE

La presente invenzione riguarda un vaccino individualizzato contro il cancro per uso in un metodo di trattamento di un paziente con il cancro.

SFONDO DELL'INVENZIONE

Il cancro è una causa primaria di mortalità, spiegando 1 su 4 di tutti i decessi. Il trattamento del cancro è stato tradizionalmente basato sulla legge delle medie - cosa funziona meglio per il maggior numero di pazienti. Tuttavia, a causa dell'eterogeneità molecolare nel cancro, spesso meno del 25% degli individui trae profitto dalle terapie approvate. Un medicinale individualizzato basato su un trattamento personalizzato di pazienti è considerato come potenziale soluzione alle efficacie basse ed ai costi elevati per l'innovazione nello sviluppo del farmaco.

L'immunoterapia specifica per l'antigene si propone di aumentare o di indurre risposte immunitarie specifiche nei pazienti ed è stata utilizzata con successo per controllare le malattie cancerose. I linfociti T svolgono un ruolo centrale nell'immunità mediata dalle cellule negli esseri umani e negli animali. Il riconoscimento ed il legame di un particolare antigene sono mediati dai recettori dei linfociti T (TCR) espressi sulla superficie dei linfociti T. Il recettore dei linfociti T (TCR) di un linfocita T è in grado di interagire con peptidi immunogenici (epitopi) legati a molecole del complesso maggiore di istocompatibilità (MHC) e presentati sulla superficie delle cellule bersaglio. Il legame specifico del TCR scatena una cascata di segnali all'interno del linfocita T conducendo alla proliferazione ed alla differenziazione in un linfocita T effetore maturo.

L'identificazione di un numero crescente di antigeni associati ad agenti patogeni e a tumori (TAA) ha condotto ad un'ampia raccolta di bersagli adatti per l'immunoterapia. Le cellule che presentano peptidi immunogenici (epitopi) derivati da questi antigeni possono essere specificamente mirate mediante strategie di immunizzazione attiva o passiva. L'immunizzazione attiva può tendere ad indurre e ad espandere i linfociti T specifici per un antigene nel paziente, i quali sono in grado di riconoscere e di uccidere specificamente le cellule malate. Differenti formati di antigene possono essere utilizzati per la vaccinazione contro i tumori compresi cellule cancerose intere, proteine, peptidi o vettori immunizzanti quali RNA, DNA o vettori virali che possono essere applicati direttamente in vivo o in vitro stimolando le DC in seguito al trasferimento nel paziente.

I tumori possono originarsi dall'accumulo di mutazioni genomiche e di variazioni epigenetiche, di cui una frazione può avere un ruolo causativo. Oltre agli antigeni associati a tumore, i tumori umani trasportano mediamente da 100 a 120 mutazioni non sinonime, di cui molte possono essere mirate mediante vaccini. Più del 95% delle mutazioni in un tumore è unico e specifico per un paziente (Weide et al. 2008: J. Immunother. 31, da 180 a 188). Il numero di mutazioni somatiche che cambiano le proteine, le quali possono portare ad epitopi dei linfociti T specifici per il tumore, è nell'intervallo da 30 a 400. È stato previsto in silico che vi sono da 40 a 60 epitopi ristretti per l'HLA di classe I per paziente derivati da mutazioni somatiche specifiche per il tumore (Azuma et al. 1993: Nature 366, da 76 a 79). Inoltre, epitopi immunogenici de novo ristretti per l'HLA di classe II probabilmente risultano anch'essi da mutazioni associate al tumore, tuttavia il loro numero non è ancora noto.

In particolare, alcune mutazioni non sinonime sono coinvolte in modo causale nella trasformazione neoplastica, essenziale per mantenere il fenotipo oncogeno (mutazioni scatenanti) e possono rappresentare un potenziale "tallone d'Achille" delle cellule cancerose. Poiché tali mutazioni non sinonime non sono soggette a tolleranza immunitaria centrale, esse possono essere candidati ideali per lo sviluppo di vaccini individuali contro il cancro. Le mutazioni trovate nel tumore primario possono anche essere presenti nelle metastasi. Tuttavia, diversi studi hanno dimostrato che i tumori metastatici di un paziente acquisiscono ulteriori mutazioni genetiche durante lo sviluppo del singolo tumore che sono spesso clinicamente rilevanti (Suzuki et al. 2007: Mol. Oncol. 1 (2), da 172 a 180; Campbell et al. 2010: Nature 467 (7319), da 1109 a 1113). Inoltre, anche le caratteristiche molecolari di molte metastasi deviano significativamente da quelle dei tumori primari.

Il problema tecnico alla base della presente invenzione è fornire un vaccino individualizzato contro il cancro altamente efficace.

Rammensee et al. (Genomic and Personalized Medicine; 11 novembre 2008; pagine da 573 a 589) riguarda considerazioni di base riguardanti i vaccini contro il cancro.

La presente invenzione si basa sull'identificazione di mutazioni specifiche del cancro per un paziente e sul mirare una mutazione singola "firma" del cancro del paziente. Specificamente, la presente invenzione che implica un approccio individualizzato di immunoterapia basato sul sequenziamento del genoma, preferibilmente esoma, o del trascrittoma punta a mirare immunoterapeuticamente multiple mutazioni singole nel cancro. Il sequenziamento che utilizza il sequenziamento di nuova generazione (NGS) permette un'identificazione veloce ed economica di mutazioni specifiche del cancro per un paziente.

L'identificazione di mutazioni puntiformi non sinonime che risulta in variazioni amminoacidiche che saranno presentate alle molecole del complesso maggiore di istocompatibilità (MHC) del paziente fornisce nuovi epitopi (neoepitopi) che sono specifici per il cancro del paziente ma non si trovano nelle cellule normali del paziente. La raccolta di una serie di mutazioni dalle cellule cancerose quali le cellule tumorali circolanti

(CTC) permette la fornitura di un vaccino che induce una risposta immunitaria che mira potenzialmente il tumore primario anche se contenente sottopopolazioni geneticamente distinte così come metastasi tumorali. Per la vaccinazione, tali neoepitopi identificati secondo la presente domanda vengono forniti in un paziente in forma di polipeptide comprendente i suddetti neoepitopi e in seguito ad appropriati processamento e presentazione dalle molecole dell'MHC i neoepitopi vengono mostrati al sistema immunitario del paziente per la stimolazione dei linfociti T appropriati.

Preferibilmente, tale polipeptide viene fornito nel paziente somministrando RNA codificante per il polipeptide. Una strategia in cui l'RNA trascritto in vitro (IVT-RNA) è direttamente iniettato in un paziente mediante differenti vie di immunizzazione è stata testata con successo in vari modelli animali. L'RNA può essere tradotto nelle cellule trasfettate e la proteina espressa in seguito a processamento presentata sulle molecole dell'MHC sulla superficie delle cellule per provocare una risposta immunitaria.

I vantaggi dell'utilizzo di RNA come tipo di terapia genica reversibile includono l'espressione transiente ed il carattere non trasformante. L'RNA non necessita di entrare nel nucleo allo scopo di essere espresso ed inoltre non può integrarsi nel genoma dell'ospite, eliminando quindi il rischio di oncogenesi. I tassi di trasfezione che si possono ottenere con l'RNA sono relativamente elevati. Inoltre, le quantità della proteina ottenuta corrispondono a quelle nell'espressione fisiologica.

Il razionale per l'indirizzamento a bersaglio immunoterapeutico di multiple mutazioni singole è che (i) queste mutazioni sono esclusivamente espresse, (ii) si attende che gli epitopi mutati siano ideali per l'immunoterapia a linfociti T poiché i linfociti T che li riconoscono non hanno subito selezione timica, (iii) l'elusione immunitaria del tumore può essere ridotta per esempio mirando le "mutazioni scatenanti" che sono altamente rilevanti per il fenotipo tumorale, e (iv) una risposta immunitaria multiepitopica ha una probabilità superiore di portare ad un beneficio clinico migliorato.

DESCRIZIONE DELL'INVENZIONE

RIASSUNTO DELL'INVENZIONE

Vengono qui descritti metodi efficienti per fornire vaccini ricombinanti individualizzati contro il cancro che inducono una risposta immunitaria efficiente e specifica in un paziente con il cancro e che mirano potenzialmente il tumore primario così come le metastasi tumorali. I vaccini contro il cancro per uso secondo l'invenzione quando somministrati ad un paziente forniscono una raccolta di epitopi presentati dall'MHC specifici per il tumore del paziente adatti per stimolare, per innescare e/o per espandere i linfociti T diretti contro le cellule esprimenti antigeni da cui sono derivati gli epitopi presentati dall'MHC. Così, i vaccini qui descritti sono preferibilmente in grado di indurre o di promuovere una risposta cellulare, preferibilmente l'attività dei linfociti T citotossici, nei confronti di una malattia oncologica caratterizzata dalla presentazione di uno o più antigeni

tumorali espressi con l'MHC di classe I. Poiché un vaccino per uso secondo la presente invenzione avrà come bersaglio mutazioni specifiche del cancro esso sarà specifico per il tumore del paziente.

L'ambito della presente invenzione viene definito nelle rivendicazioni allegate.

La presente invenzione riguarda un vaccino individualizzato contro il cancro per uso in un metodo di trattamento di un paziente con il cancro, il suddetto metodo comprendendo i passaggi di:

(a) fornire il suddetto vaccino individualizzato contro il cancro mediante un metodo comprendente i passaggi di:

(aa) identificare le mutazioni somatiche specifiche per il cancro in un campione tumorale del paziente con il cancro per fornire una firma delle mutazioni tumorali del paziente; e

(ab) fornire un vaccino a RNA che comprende la firma delle mutazioni tumorali ottenuta nel passaggio (aa), in cui il vaccino a RNA che comprende la firma delle mutazioni del paziente comprende RNA codificante per un polipeptide poliepitopico ricombinante comprendente neoepitopi basati su mutazioni; e

(b) somministrare il suddetto vaccino individualizzato contro il cancro al paziente.

In una forma di realizzazione, il metodo per fornire il suddetto vaccino individualizzato contro il cancro comprende i seguenti passaggi:

i) fornire un campione tumorale dal paziente con il cancro ed un campione non tumorigenico che preferibilmente è derivato dal paziente con il cancro;

ii) identificare differenze di sequenza tra il genoma, l'esoma e/o il trascrittoma del campione tumorale ed il genoma, l'esoma e/o il trascrittoma del campione non tumorigenico;

iii) progettare un polipeptide comprendente epitopi incorporanti le differenze di sequenza determinate nel passaggio (ii);

iv) fornire l'RNA codificante per il suddetto polipeptide progettato nel passaggio (iii); e

v) fornire un vaccino comprendente l'RNA fornito nel passaggio (iv).

Secondo l'invenzione un campione tumorale riguarda qualsiasi campione quale un campione corporeo derivato da un paziente contenente o che è atteso contenga cellule tumorali o cancerose. Il campione corporeo può essere qualsiasi campione di tessuto quale sangue, un campione di tessuto ottenuto dal tumore primario o da metastasi tumorali o qualsiasi altro campione contenente cellule tumorali o cancerose. Preferibilmente, un campione corporeo è il sangue e mutazioni somatiche specifiche per il cancro o differenze di sequenza vengono determinate in una o più cellule tumorali circolanti (CTC) contenute nel sangue. In un'altra forma di realizzazione, un campione tumorale riguarda una o più cellule tumorali o

cancerose isolate quali le cellule tumorali circolanti (CTC) o un campione contenente una o più cellule tumorali o cancerose isolate quali le cellule tumorali circolanti (CTC).

Un campione non tumorigenico riguarda qualsiasi campione quale un campione corporeo derivato da un paziente o da un altro individuo che preferibilmente è della stessa specie del paziente, preferibilmente un individuo sano non contenente o che è atteso non contenga cellule tumorali o cancerose. Il campione corporeo può essere qualsiasi campione di tessuto quale sangue o un campione da un tessuto non tumorigenico.

Secondo l'invenzione, il termine "firma delle mutazioni tumorali" può riferirsi a tutte le mutazioni tumorali presenti in una o più cellule cancerose di un paziente oppure può riferirsi solo ad una porzione delle mutazioni tumorali presenti in una o più cellule cancerose di un paziente. Conseguentemente, la presente invenzione può riguardare l'identificazione di tutte le mutazioni specifiche del tumore presenti in una o più cellule cancerose di un paziente oppure può implicare l'identificazione di solo una porzione delle mutazioni specifiche del cancro presenti in una o più cellule cancerose di un paziente. Generalmente, il metodo per fornire il suddetto vaccino individualizzato contro il cancro fornisce l'identificazione di un certo numero di mutazioni che fornisce un numero sufficiente di neoepitopi da includere in un vaccino. Una "mutazione tumorale" riguarda una differenza di sequenza tra l'acido nucleico contenuto in una cellula cancerosa e l'acido nucleico contenuto in una cellula normale.

Preferibilmente, le mutazioni identificate secondo la presente invenzione sono mutazioni non sinonime, preferibilmente mutazioni non sinonime di proteine espresse in una cellula tumorale o cancerosa.

In una forma di realizzazione, le mutazioni somatiche specifiche per il cancro o le differenze di sequenza vengono determinate nel genoma, preferibilmente l'intero genoma, di un campione tumorale. Così, il metodo per fornire il suddetto vaccino individualizzato contro il cancro può comprendere l'identificazione della firma delle mutazioni tumorali del genoma, preferibilmente l'intero genoma di una o più cellule cancerose. In una forma di realizzazione, il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro in un campione tumorale di un paziente con il cancro comprende l'identificazione del profilo di mutazioni tumorali ad ampio genoma.

In una forma di realizzazione, le mutazioni somatiche specifiche per il cancro o le differenze di sequenza vengono determinate nell'esoma, preferibilmente l'intero esoma, di un campione tumorale. L'esoma è parte del genoma di un organismo formato dagli esoni, i quali sono porzioni codificanti di geni espressi. L'esoma fornisce il progetto genetico utilizzato nella sintesi di proteine e di altri prodotti genici funzionali. È la parte più funzionalmente rilevante del genoma e, quindi, è più probabile che contribuisca al fenotipo di un organismo. Si stima che l'esoma del genoma umano comprenda l'1,5% del genoma totale (Ng, PC et al., PLoS Gen., 4(8): da 1 a 15, 2008). Così, il metodo per fornire il suddetto vaccino individualizzato contro il cancro può comprendere l'identificazione della firma delle mutazioni tumorali dell'esoma, preferibilmente l'intero esoma

di una o più cellule cancerose. In una forma di realizzazione, il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro in un campione tumorale di un paziente con il cancro comprende l'identificazione del profilo di mutazioni tumorali ad ampio esoma.

In una forma di realizzazione, le mutazioni somatiche specifiche per il cancro o le differenze di sequenza vengono determinate nel trascrittoma, preferibilmente l'intero trascrittoma, di un campione tumorale. Il trascrittoma è la serie di tutte le molecole di RNA, compresi mRNA, rRNA, tRNA, ed altri RNA non codificanti prodotte in una cellula o in una popolazione di cellule. Nel contesto della presente invenzione il trascrittoma significa la serie di tutte le molecole di RNA prodotte in una cellula, in una popolazione di cellule, preferibilmente una popolazione di cellule cancerose, o in tutte le cellule di un dato individuo ad un certo istante. Così, il metodo per fornire il suddetto vaccino individualizzato contro il cancro può comprendere l'identificazione della firma delle mutazioni tumorali del trascrittoma, preferibilmente l'intero trascrittoma di una o più cellule cancerose. In una forma di realizzazione, il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro in un campione tumorale di un paziente con il cancro comprende l'identificazione del profilo di mutazioni tumorali ad ampio trascrittoma.

In una forma di realizzazione, il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro o di identificazione delle differenze di sequenza comprende il sequenziamento a singola cellula di una o più, preferibilmente 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20 od ancora più cellule cancerose. Così, il metodo per fornire il suddetto vaccino individualizzato contro il cancro può comprendere l'identificazione di una firma delle mutazioni tumorali delle suddette una o più cellule cancerose. In una forma di realizzazione, le cellule cancerose sono cellule tumorali circolanti. Le cellule cancerose quali le cellule tumorali circolanti possono essere isolate prima del sequenziamento a singola cellula.

In una forma di realizzazione, il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro o di identificazione delle differenze di sequenza implica l'utilizzo del sequenziamento di nuova generazione (NGS).

In una forma di realizzazione, il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro o di identificazione delle differenze di sequenza comprende il sequenziamento del DNA e/o dell'RNA genomici del campione tumorale.

Per rivelare le mutazioni somatiche specifiche per il cancro o le differenze di sequenza le informazioni di sequenza ottenute dal campione tumorale sono preferibilmente confrontate con un riferimento quale informazioni di sequenza ottenute dal sequenziamento di un acido nucleico quale DNA o RNA di cellule normali non cancerose quali le cellule della linea germinale che possono essere ottenute dal paziente o da un individuo differente. In una forma di realizzazione, il DNA normale genomico della linea germinale viene ottenuto dalle cellule mononucleate del sangue periferico (PBMC).

Un vaccino per uso secondo la presente invenzione riguarda un vaccino che quando somministrato ad un paziente fornisce preferibilmente una raccolta di epitopi presentati dall'MHC, quali 2 o più, 5 o più, 10 o più, 15 o più, 20 o più, 25 o più, 30 o più e preferibilmente fino a 60, fino a 55, fino a 50, fino a 45, fino a 40, fino a 35 o fino a 30 epitopi presentati dall'MHC, i quali incorporano variazioni di sequenza basate sulle mutazioni o sulle differenze di sequenza identificate. Tali epitopi presentati dall'MHC incorporanti variazioni di sequenza basate sulle mutazioni o sulle differenze di sequenza identificate vengono chiamati anche "neoepitopi" qui. La presentazione di questi epitopi da parte delle cellule di un paziente, in particolare le cellule presentanti l'antigene, risulta preferibilmente in linfociti T che hanno come bersaglio gli epitopi quando legati all'MHC e così, il tumore del paziente, preferibilmente il tumore primario così come le metastasi tumorali, esprimenti gli antigeni da cui sono derivati gli epitopi presentati dall'MHC e presentanti gli stessi epitopi sulla superficie delle cellule tumorali.

Il metodo per fornire il suddetto vaccino individualizzato contro il cancro può comprendere l'inclusione arbitraria di un numero sufficiente di neoepitopi (preferibilmente in forma di acido nucleico codificante) in un vaccino oppure può comprendere l'ulteriore passaggio di determinazione dell'utilizzabilità delle mutazioni identificate negli epitopi per la vaccinazione contro il cancro. Così ulteriori passaggi possono implicare uno o più dei seguenti: (i) valutare se le variazioni di sequenza siano localizzate in epitopi presentati dall'MHC noti o previsti, (ii) testare in vitro e/o in silico se le variazioni di sequenza siano localizzate in epitopi presentati dall'MHC, per esempio testare se le variazioni di sequenza facciano parte di sequenze peptidiche che vengono processate in e/o presentate come epitopi presentati dall'MHC, e (iii) testare in vitro se gli epitopi mutati previsti, in particolare quando presenti nel loro contesto naturale di sequenza, per esempio quando fiancheggiati da sequenze amminoacidiche fiancheggianti anche i suddetti epitopi nella proteina presente in natura, e quando espressi nelle cellule presentanti l'antigene sono in grado di stimolare i linfociti T del paziente aventi la specificità desiderata. Tali sequenze fiancheggianti possono ciascuna comprendere 3 o più, 5 o più, 10 o più, 15 o più, 20 o più e preferibilmente fino a 50, fino a 45, fino a 40, fino a 35 o fino a 30 amminoacidi e possono fiancheggiare la sequenza epitopica aminoterminalmente e/o carbossiterninalmente.

Le mutazioni o le differenze di sequenza determinate secondo l'invenzione possono essere classificate per la loro usabilità come epitopi per la vaccinazione contro il cancro. Così, in un aspetto, il metodo per fornire il suddetto vaccino individualizzato contro il cancro comprende un procedimento analitico manuale o informatizzato in cui le mutazioni identificate vengono analizzate e scelte per la loro usabilità nel rispettivo vaccino da fornire. In una forma di realizzazione preferita, il suddetto procedimento analitico è un procedimento basato su un algoritmo computazionale. Preferibilmente, il suddetto procedimento analitico comprende uno o più, preferibilmente tutti i seguenti passaggi:
identificare le mutazioni espresse modificanti le proteine, per esempio analizzando i trascritti;

identificare le mutazioni che sono potenzialmente immunogeniche, cioè confrontando i dati ottenuti con le serie di dati disponibili di epitopi immunogenici confermati, per esempio quelli contenuti in banche dati pubbliche di epitopi immunitari quali cioè IMMUNE EPITOPE DATABASE AND ANALYSIS RESOURCE a <http://www.immunoepitope.org>.

Il passaggio di identificazione delle mutazioni che sono potenzialmente immunogeniche può comprendere la determinazione e/o la classificazione degli epitopi secondo una predizione della loro capacità di legame all'MHC, preferibilmente capacità di legame all'MHC di classe I.

In un'altra forma di realizzazione dell'invenzione, gli epitopi possono essere scelti e/o classificati utilizzando ulteriori parametri quali l'impatto della proteina, l'espressione genica associata, l'unicità della sequenza, la probabilità di presentazione prevista, e l'associazione con oncogeni.

Le analisi multiple delle CTC permettono anche la selezione e la prioritizzazione delle mutazioni. Per esempio, una mutazione che si trova in una porzione maggiore di CTC può essere prioritizzata in modo superiore rispetto ad una mutazione trovata in una porzione minore di CTC.

La raccolta di neoepitopi basati su mutazioni identificati secondo l'invenzione e forniti da un vaccino per uso secondo l'invenzione è presente in forma di RNA codificante per un polipeptide comprendente i suddetti neoepitopi (polipeptide poliepitopico). Inoltre, i neoepitopi possono essere presenti nel polipeptide in forma di sequenza vaccinale, cioè presenti nel loro contesto naturale di sequenza, per esempio fiancheggiati da sequenze amminoacidiche fiancheggianti anche i suddetti epitopi nella proteina presente in natura. Tali sequenze fiancheggianti possono ciascuna comprendere 5 o più, 10 o più, 15 o più, 20 o più e preferibilmente fino a 50, fino a 45, fino a 40, fino a 35 o fino a 30 amminoacidi e possono fiancheggiare la sequenza epitopica aminoterminalmente e/o carbossiterninalmente. Così, una sequenza vaccinale può comprendere 20 o più, 25 o più, 30 o più, 35 o più, 40 o più e preferibilmente fino a 50, fino a 45, fino a 40, fino a 35 o fino a 30 amminoacidi. In una forma di realizzazione, i neoepitopi e/o le sequenze vaccinali sono allineati in modo testa-coda nel polipeptide.

In una forma di realizzazione, i neoepitopi e/o le sequenze vaccinali sono distanziati mediante gruppi di collegamento, in particolare gruppi di collegamento neutri. Il termine "gruppo di collegamento" secondo l'invenzione riguarda un peptide aggiunto tra due domini peptidici quali epitopi o sequenze vaccinali per collegare i suddetti domini peptidici. Non vi è alcuna limitazione particolare riguardante la sequenza di collegamento. Tuttavia, si preferisce che la sequenza di collegamento riduca l'ingombro sterico tra i due domini peptidici, sia ben tradotta, e supporti o permetta il processamento degli epitopi. Inoltre, il gruppo di collegamento non dovrebbe avere alcun elemento immunogenico di sequenza o dovrebbe avere solo pochi elementi immunogenici di sequenza. I gruppi di collegamento preferibilmente non dovrebbero creare neoepitopi non endogeni come quelli generati dalla sutura delle giunzioni tra neoepitopi adiacenti, il che potrebbe generare reazioni immunitarie indesiderate. Quindi, il vaccino poliepitopico dovrebbe contenere preferibilmente sequenze di collegamento che sono in grado di ridurre il numero di epitopi di giunzione

indesiderati che legano l'MHC. Hoyt et al. (EMBO J. 25(8), da 1720 a 1729, 2006) e Zhang et al. (J. Biol. Chem., 279(10), da 8635 a 8641, 2004) hanno mostrato che sequenze ricche in glicina danneggiano il processamento proteasomico e così l'uso di sequenze di collegamento ricche in glicina agisce per minimizzare il numero di peptidi contenenti gruppi di collegamento che possono essere processati dal proteasoma. Inoltre, è stato osservato che la glicina inibisce un legame forte nelle posizioni dei solchi di legame all'MHC (Abastado et al., J. Immunol. 151(7), da 3569 a 3575, 1993). Schlessinger et al. (Proteins, 61(1), da 115 a 126, 2005) avevano trovato che gli aminoacidi glicina e serina inclusi in una sequenza aminoacidica portano ad una proteina più flessibile che viene più efficientemente tradotta e processata dal proteasoma, permettendo un migliore accesso ai neoepitopi codificati. Il gruppo di collegamento può ciascuno comprendere 3 o più, 6 o più, 9 o più, 10 o più, 15 o più, 20 o più e preferibilmente fino a 50, fino a 45, fino a 40, fino a 35 o fino a 30 aminoacidi. Preferibilmente il gruppo di collegamento è arricchito in aminoacidi glicina e/o serina. Preferibilmente, almeno il 50%, almeno il 60%, almeno il 70%, almeno l'80%, almeno il 90%, o almeno il 95% degli aminoacidi del gruppo di collegamento è rappresentato da glicina e/o serina. In una forma di realizzazione preferita, un gruppo di collegamento è sostanzialmente costituito dagli aminoacidi glicina e serina. In una forma di realizzazione, il gruppo di collegamento comprende la sequenza aminoacidica $(GGS)_a(GSS)_b(GGG)_c(SSG)_d(GSG)_e$ in cui a, b, c, d ed e sono indipendentemente un numero scelto tra 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, o 20 ed in cui $a+b+c+d+e$ sono diversi da 0 e sono preferibilmente 2 o superiore, 3 o superiore, 4 o superiore o 5 o superiore. In una forma di realizzazione, il gruppo di collegamento comprende una sequenza come qui descritta comprendente le sequenze di collegamento descritte negli esempi quali la sequenza GGSGGGGSG.

In un'altra forma di realizzazione della presente invenzione la raccolta di neoepitopi basati su mutazioni identificate secondo l'invenzione e forniti da un vaccino per uso secondo l'invenzione è presente preferibilmente in forma di raccolta di RNA codificanti per i polipeptidi comprendente i suddetti neoepitopi su polipeptidi differenti, in cui i suddetti polipeptidi comprendono ciascuno uno o più neoepitopi, i quali possono anche essere sovrapposti.

Un polipeptide poliepitopico secondo la presente invenzione viene somministrato ad un paziente in forma di RNA quale RNA trascritto in vitro o sintetico, il quale può essere espresso nelle cellule di un paziente quali le cellule presentanti l'antigene per produrre il polipeptide. La presente invenzione prevede la somministrazione di uno o più polipeptidi multiepitopici che per lo scopo della presente invenzione sono compresi nel termine "polipeptide poliepitopico", in forma di RNA quale RNA trascritto in vitro o sintetico, il quale può essere espresso nelle cellule di un paziente quali le cellule presentanti l'antigene per produrre l'uno o più polipeptidi. Nel caso di una somministrazione di più di un polipeptide multiepitopico i neoepitopi forniti dai differenti polipeptidi multiepitopici possono essere differenti o parzialmente sovrapposti. Una volta presente

nelle cellule di un paziente quali le cellule presentanti l'antigene il polipeptide qui descritto viene processato per produrre i neoepitopi identificati secondo l'invenzione. La somministrazione di un vaccino per uso secondo l'invenzione può fornire epitopi presentati dall'MHC di classe II che sono in grado di suscitare una risposta dei linfociti T helper CD4+ nei confronti delle cellule esprimenti antigeni da cui sono derivati gli epitopi presentati dall'MHC. In alternativa o in aggiunta, la somministrazione di un vaccino per uso secondo l'invenzione può fornire epitopi presentati dall'MHC di classe I che sono in grado di suscitare una risposta dei linfociti T CD8+ nei confronti delle cellule esprimenti antigeni da cui sono derivati gli epitopi presentati dall'MHC. Inoltre, la somministrazione di un vaccino per uso secondo l'invenzione può fornire uno o più neoepitopi (compresi neoepitopi noti e neoepitopi identificati secondo l'invenzione) così come uno o più epitopi non contenenti mutazioni somatiche specifiche per il cancro ma che sono espressi dalle cellule cancerose e che inducono preferibilmente una risposta immunitaria nei confronti delle cellule cancerose, preferibilmente una risposta immunitaria specifica per il cancro. In una forma di realizzazione, la somministrazione di un vaccino per uso secondo l'invenzione fornisce neoepitopi che sono epitopi presentati dall'MHC di classe II e/o sono in grado di suscitare una risposta dei linfociti T helper CD4+ nei confronti delle cellule esprimenti antigeni da cui sono derivati gli epitopi presentati dall'MHC così come epitopi non contenenti mutazioni somatiche specifiche del cancro che sono epitopi presentati dall'MHC di classe I e/o sono in grado di suscitare una risposta dei linfociti T CD8+ nei confronti delle cellule esprimenti antigeni da cui sono derivati gli epitopi presentati dall'MHC. In una forma di realizzazione, gli epitopi non contenenti mutazioni somatiche specifiche del cancro sono derivati da un antigene tumorale. In una forma di realizzazione, i neoepitopi e gli epitopi non contenenti mutazioni somatiche specifiche del cancro hanno un effetto sinergico nel trattamento del cancro. Preferibilmente, un vaccino per uso secondo l'invenzione è utile per la stimolazione poliepitopica delle risposte dei linfociti T citotossici e/o helper.

Viene inoltre qui descritto un vaccino che può essere ottenuto mediante il metodo qui descritto. Conseguentemente, la presente descrizione riguarda anche un vaccino comprendente un polipeptide ricombinante comprendente neoepitopi basati su mutazioni, i suddetti neoepitopi risultando da mutazioni somatiche specifiche per il cancro in un campione tumorale di un paziente con il cancro, o un acido nucleico codificante per il suddetto polipeptide. Tale polipeptide ricombinante può anche comprendere epitopi non comprendenti mutazioni somatiche specifiche per il cancro come discusso sopra. Forme di realizzazione preferite di tale vaccino sono come descritte sopra nel contesto del metodo qui descritto.

Un vaccino per uso secondo l'invenzione può comprendere un veicolante farmaceuticamente accettabile e può facoltativamente comprendere uno o più adiuvanti, stabilizzanti ecc. Il vaccino può avere la forma di un vaccino terapeutico o profilattico.

Viene inoltre qui descritto un metodo per indurre una risposta immunitaria in un paziente, comprendente la somministrazione al paziente di un

vaccino qui descritto.

Viene inoltre qui descritto un metodo di trattamento di un paziente con il cancro comprendente la somministrazione del vaccino qui descritto al paziente.

La descrizione fornisce anche i vaccini qui descritti per uso nei metodi di trattamento qui descritti, in particolare per uso nel trattamento o nella prevenzione del cancro.

I trattamenti del cancro qui descritti possono essere combinati con la resezione chirurgica e/o la radiazione e/o la chemioterapia tradizionale.

Viene inoltre qui descritto un metodo per determinare un tasso di scoperta falsa sulla base dei dati del sequenziamento di nuova generazione, il suddetto metodo comprendendo:

prendere un primo campione di materiale genetico da un animale o dall'uomo;

prendere un secondo campione di materiale genetico da un animale o dall'uomo;

prendere un primo campione di materiale genetico dalle cellule tumorali;

prendere un secondo campione di materiale genetico dalle suddette cellule tumorali;

determinare un comune confronto di copertura tumorale contando tutte le basi del genoma di riferimento che è incluso in entrambi il tumore ed almeno uno tra il suddetto primo campione di materiale genetico da un animale o dall'uomo ed il suddetto secondo campione di materiale genetico da un animale o dall'uomo;

determinare un comune confronto di copertura stesso contro stesso contando tutte le basi del genoma di riferimento che sono coperte da entrambi il suddetto primo campione di materiale genetico da un animale o dall'uomo ed il suddetto secondo campione di materiale genetico da un animale o dall'uomo;

dividere il suddetto comune confronto di copertura tumorale mediante il suddetto comune confronto di copertura stesso contro stesso per formare una normalizzazione;

determinare un tasso di scoperta falsa dividendo 1) il numero di singole variazioni nucleotidiche con un punteggio di qualità superiore a Q in un confronto del suddetto primo campione di materiale genetico da un animale o dall'uomo ed il suddetto secondo campione di materiale genetico da un animale o dall'uomo, mediante 2) il numero di singole variazioni nucleotidiche con un punteggio di qualità superiore a Q in un confronto del suddetto primo campione di materiale genetico dalle suddette cellule tumorali ed il suddetto secondo campione di materiale genetico dalle suddette cellule tumorali e 3) moltiplicando il risultato per la suddetta normalizzazione.

In una forma di realizzazione, il suddetto materiale genetico è un DNA.

In una forma di realizzazione, Q viene determinato:

stabilendo una serie di proprietà di qualità $S=(s_1, \dots, s_n)$ in cui S è preferibile a $T=(t_1, \dots, t_n)$, indicato mediante $S>T$, quando $s_i>t_i$ per tutti $i=1, \dots, n$;

definendo un tasso intermedio di scoperta falsa dividendo 1) il numero di singole variazioni nucleotidiche con un punteggio di qualità $S>T$ in un confronto del suddetto primo campione di DNA da un animale o dall'uomo ed il suddetto secondo campione di DNA da un animale o dall'uomo, per 2) il numero di singole variazioni nucleotidiche con un punteggio di qualità $S>T$ in un confronto del suddetto primo campione di DNA dalle suddette cellule tumorali ed il suddetto secondo campione di DNA dalle suddette cellule tumorali e 3) moltiplicando il risultato per la suddetta normalizzazione;

determinando l'intervallo di valori per ciascuna proprietà per m mutazioni con n proprietà di qualità ciascuna;

campionando fino a valori di p fuori dal suddetto intervallo di valori;

creando ciascuna possibile combinazione di valori campionati di qualità che risulta in p^n punti di dati;

utilizzando un campione casuale dei suddetti punti di dati come indicatore per un addestramento a foresta casuale;

utilizzando il corrispondente valore del tasso intermedio di scoperta falsa come risposta per il suddetto addestramento a foresta casuale,

in cui il risultante punteggio di regressione del suddetto addestramento a foresta casuale è Q.

In una forma di realizzazione, il suddetto secondo campione di DNA da un animale o dall'uomo è allogenic al suddetto primo campione di DNA da un animale o dall'uomo. In una forma di realizzazione, il suddetto secondo campione di DNA da un animale o dall'uomo è autologo al suddetto primo campione di DNA da un animale o dall'uomo. In una forma di realizzazione, il suddetto secondo campione di DNA da un animale o dall'uomo è xenogenic al suddetto primo campione di DNA da un animale o dall'uomo.

In una forma di realizzazione, il suddetto materiale genetico è un RNA.

In una forma di realizzazione, Q viene determinato:

stabilendo una serie di proprietà di qualità $S=(s_1, \dots, s_n)$ in cui S è preferibile a $T=(t_1, \dots, t_n)$, indicato mediante $S>T$, quando $s_i>t_i$ per tutti $i=1, \dots, n$;

definendo un tasso intermedio di scoperta falsa dividendo 1) il numero di singole variazioni nucleotidiche con un punteggio di qualità $S>T$ in un confronto del suddetto primo campione di RNA da un animale o dall'uomo ed il suddetto secondo campione di RNA da un animale o dall'uomo, per 2) il numero di singole variazioni nucleotidiche con un punteggio di qualità $S>T$ in un confronto del suddetto primo campione di RNA dalle suddette cellule tumorali ed il suddetto secondo campione di RNA dalle suddette cellule tumorali e 3) moltiplicando il risultato per la suddetta

normalizzazione;

determinando l'intervallo di valori per ciascuna proprietà per m mutazioni con n proprietà di qualità ciascuna;

campionando fino a valori di p fuori dal suddetto intervallo di valori;

creando ciascuna possibile combinazione di valori campionati di qualità che risulta in p^n punti di dati;

utilizzando un campione casuale dei suddetti punti di dati come indicatore per un addestramento a foresta casuale;

utilizzando il corrispondente valore del tasso intermedio di scoperta falsa come risposta per il suddetto addestramento a foresta casuale,

in cui il risultante punteggio di regressione del suddetto addestramento a foresta casuale è Q.

In una forma di realizzazione, il suddetto secondo campione di RNA da un animale o dall'uomo è allogenico al suddetto primo campione di RNA da un animale o dall'uomo. In una forma di realizzazione, il suddetto secondo campione di RNA da un animale o dall'uomo è autologo al suddetto primo campione di RNA da un animale o dall'uomo. In una forma di realizzazione, il suddetto secondo campione di RNA da un animale o dall'uomo è xenogenico al suddetto primo campione di RNA da un animale o dall'uomo.

In una forma di realizzazione, il suddetto tasso di scoperta falsa viene utilizzato per preparare una formulazione del vaccino. In una forma di realizzazione, il suddetto vaccino è rilasciabile per via endovenosa. In una forma di realizzazione, il suddetto vaccino è rilasciabile per via dermica.

In una forma di realizzazione, il suddetto vaccino è rilasciabile per via muscolare. In una forma di realizzazione, il suddetto vaccino è rilasciabile per via sottocutanea. In una forma di realizzazione, il suddetto vaccino viene personalizzato per un paziente specifico.

In una forma di realizzazione, uno del suddetto primo campione di materiale genetico da un animale o dall'uomo e del suddetto secondo campione di materiale genetico da un animale o dall'uomo è dal suddetto paziente specifico.

In una forma di realizzazione, il suddetto passaggio di determinazione di un confronto di copertura comune tumorale contando tutte le basi del genoma di riferimento che è incluso in entrambi il tumore ed almeno uno tra il suddetto primo campione di materiale genetico da un animale o dall'uomo ed il suddetto secondo campione di materiale genetico da un animale o dall'uomo utilizza un sistema automatizzato per contare tutte le basi.

In una forma di realizzazione, il suddetto passaggio di determinazione di un confronto di copertura comune stesso contro stesso contando tutte le basi del genoma di riferimento che sono coperte da entrambi il suddetto primo campione di materiale genetico da un animale o dall'uomo ed il suddetto secondo campione di materiale genetico da un animale o dall'uomo utilizza il suddetto sistema automatizzato.

In una forma di realizzazione, il suddetto passaggio di divisione del suddetto confronto di copertura comune tumorale per il suddetto comune

confronto di copertura stesso contro stesso per formare una normalizzazione utilizza il suddetto sistema automatizzato.

In una forma di realizzazione, il suddetto passaggio di determinazione di un tasso di scoperta falsa dividendo 1) il numero di singole variazioni nucleotidiche con un punteggio di qualità superiore a Q in un confronto del suddetto primo campione di materiale genetico da un animale o dall'uomo ed il suddetto secondo campione di materiale genetico da un animale o dall'uomo, per 2) il numero di singole variazioni nucleotidiche con un punteggio di qualità superiore a Q in un confronto del suddetto primo campione di materiale genetico dalle suddette cellule tumorali ed il suddetto secondo campione di materiale genetico dalle suddette cellule tumorali e 3) moltiplicando il risultato per la suddetta normalizzazione utilizza il suddetto sistema automatizzato.

Viene inoltre qui descritto un metodo per determinare una curva operante stimata del ricevente (ROC), il suddetto metodo comprendendo:

ricevere una serie di dati di mutazioni, ciascuna mutazione associata ad un tasso di scoperta falsa (FDR); e

per ciascuna mutazione:

determinare un tasso di veri positivi (TPR) sottraendo il suddetto FDR da uno; e

determinare un tasso di falsi positivi (FPR) impostando il suddetto FPR uguale al suddetto FDR; e

formando una ROC stimata riportando in grafico, per ciascuna mutazione, un punto ai valori cumulativi di TPR e di FPR fino alla suddetta mutazione, divisi per la somma di tutti i valori di TPR e di FPR.

Altre caratteristiche e vantaggi della presente invenzione saranno evidenti dalle seguenti descrizione dettagliata e rivendicazioni.

DESCRIZIONE DETTAGLIATA DELL'INVENZIONE

Sebbene la presente invenzione venga descritta in dettaglio sotto, si deve intendere che questa invenzione non è limitata alle metodologie, ai protocolli e ai reagenti specifici qui descritti poiché questi possono variare. Si deve anche intendere che la terminologia qui utilizzata è allo scopo di descrivere solo forme di realizzazione particolari, e non è intesa a limitare l'ambito della presente invenzione che sarà limitata solo dalle rivendicazioni allegate. Se non definiti diversamente, tutti i termini tecnici e scientifici qui utilizzati hanno gli stessi significati come comunemente intesi da una persona esperta nell'arte.

In seguito, gli elementi della presente invenzione saranno descritti. Questi elementi vengono elencati con forme di realizzazione specifiche, tuttavia, dovrebbe essere inteso che essi possono essere combinati in qualsiasi modo ed in qualsiasi numero per creare ulteriori forme di realizzazione. Gli esempi variamente descritti e le forme di realizzazione preferite non dovrebbero essere intesi a limitare la presente invenzione a solo le forme di realizzazione descritte esplicitamente. Dovrebbe essere inteso che questa descrizione supporti e comprenda forme di realizzazione che combinano le

forme di realizzazione descritte esplicitamente con qualsiasi numero degli elementi descritti e/o preferiti. Inoltre, qualsiasi permutazione e combinazione di tutti gli elementi descritti in questa domanda dovrebbe essere considerata descritta dalla descrizione della presente domanda a meno che il contesto non indichi diversamente. Per esempio, se in una forma di realizzazione preferita l'RNA comprende una coda di poli(A) composta da 120 nucleotidi ed in un'altra forma di realizzazione preferita la molecola di RNA comprende un analogo del cappuccio al 5', allora in una forma di realizzazione preferita, l'RNA comprende la coda di poli(A) composta da 120 nucleotidi e l'analogo del cappuccio al 5'.

Preferibilmente, i termini qui utilizzati vengono definiti come descritto in "A multilingual glossary of biotechnological terms: (IUPAC Recommendations)", H.G.W. Leuenberger, B. Nagel, e H. Kölbl, a cura di, (1995) Helvetica Chimica Acta, CH-4010 Basilea, Svizzera.

La pratica della presente invenzione impiegherà, se non indicato diversamente, metodi convenzionali di biochimica, di biologia cellulare, di immunologia, e tecniche del DNA ricombinante che sono spiegate nella letteratura nel campo (si confronti, per esempio, Molecular Cloning: A Laboratory Manual, seconda edizione, J. Sambrook et al. a cura di, Cold Spring Harbor Laboratory Press, Cold Spring Harbor 1989).

In ogni punto di questa descrizione e nelle rivendicazioni che seguono, a meno che il contesto non richieda diversamente, il termine "comprendere", e variazioni quali "comprende" e "comprendente", sarà inteso implicare l'inclusione di uno specificato membro, numero intero o passaggio o gruppo di membri, di numeri interi o di passaggi ma non l'esclusione di alcun altro membro, numero intero o passaggio o gruppo di membri, di numeri interi o di passaggi sebbene in alcune forme di realizzazione tale altro membro, numero intero o passaggio o gruppo di membri, di numeri interi o di passaggi possa essere escluso, cioè l'argomento consiste nell'inclusione di uno specificato membro, numero intero o passaggio o gruppo di membri, di numeri interi o di passaggi. I termini "un" e "uno/a" e "il/lo/la" e simile riferimento utilizzati nel contesto di descrizione dell'invenzione (specialmente nel contesto delle rivendicazioni) devono essere intesi a coprire sia il singolare che il plurale, se non diversamente qui indicato o chiaramente contraddetto dal contesto. L'elenco di intervalli di valori qui è solamente inteso a servire come metodo abbreviato per riferirsi singolarmente a ciascun valore separato che ricade nell'intervallo. Se non diversamente qui indicato, ciascun singolo valore è incorporato nella descrizione come se fosse stato elencato singolarmente qui.

Tutti i metodi qui descritti possono essere eseguiti in qualsiasi ordine adatto se non diversamente qui indicato o altrimenti chiaramente contraddetto dal contesto. L'uso di qualsiasi e di tutti gli esempi, o linguaggio degli esempi (per esempio "quale"), qui fornito è inteso solamente ad illustrare meglio l'invenzione e non pone una limitazione all'ambito dell'invenzione altrimenti rivendicato. Nessun linguaggio nella descrizione dovrebbe essere inteso come indicante qualsiasi elemento essenziale non rivendicato alla pratica dell'invenzione.

Diversi documenti sono citati in ogni parte del testo di questa descrizione. Nulla qui deve essere inteso come ammissione che l'invenzione non è

intitolata ad anticipare tale descrizione in virtù della precedente invenzione.

Il vaccino per uso secondo l'invenzione è un vaccino ricombinante.

Il termine "ricombinante" nel contesto della presente invenzione significa "prodotto attraverso l'ingegneria genetica". Preferibilmente, un'"entità ricombinante" quale un polipeptide ricombinante nel contesto della presente invenzione non è presente in natura, ed è preferibilmente un risultato di una combinazione di entità quali sequenze amminoacidiche o nucleotidiche che non sono combinate in natura. Per esempio, un polipeptide ricombinante nel contesto della presente invenzione può contenere diverse sequenze amminoacidiche quali neoepitopi o sequenze vaccinali derivate da proteine differenti o porzioni differenti della stessa proteina fuse insieme, per esempio, mediante legami peptidici o gruppi di collegamento appropriati.

Il termine "comunemente presente in natura" come qui utilizzato si riferisce al fatto che un oggetto si può trovare in natura. Per esempio, un peptide oppure un acido nucleico che è presente in un organismo (compresi virus) e può essere isolato da una fonte in natura e che non è stato intenzionalmente modificato dall'uomo in laboratorio è presente in natura.

Secondo l'invenzione, il termine "vaccino" riguarda una preparazione farmaceutica (composizione farmaceutica) od un prodotto che dopo la somministrazione induce una risposta immunitaria, in particolare una risposta immunitaria cellulare, la quale riconosce ed attacca un patogeno o una cellula malata quale una cellula cancerosa. Un vaccino può essere utilizzato per la prevenzione o per il trattamento di una malattia. Il termine "vaccino individualizzato contro il cancro" riguarda un particolare paziente con il cancro e significa che un vaccino contro il cancro è adatto alle necessità o alle speciali circostanze di un singolo paziente con il cancro.

Il termine "risposta immunitaria" si riferisce ad una risposta corporea integrata ad un antigene e si riferisce preferibilmente ad una risposta immunitaria cellulare o ad una risposta immunitaria cellulare nonché umorale. La risposta immunitaria può essere protettiva/preventiva/profilattica e/o terapeutica.

"Indurre una risposta immunitaria" può significare che non vi è stata alcuna risposta immunitaria nei confronti di un particolare antigene prima dell'induzione, ma può anche significare che vi è stato un certo livello di risposta immunitaria nei confronti di un particolare antigene prima dell'induzione e dopo l'induzione la suddetta risposta immunitaria è aumentata. Così, "indurre una risposta immunitaria" comprende anche "aumentare una risposta immunitaria". Preferibilmente, dopo l'induzione di una risposta immunitaria in un soggetto, il suddetto soggetto è protetto dallo sviluppo di una malattia quale una malattia oncologica o la condizione patologica è migliorata inducendo una risposta immunitaria. Per esempio, una risposta immunitaria nei confronti di un antigene tumorale espresso può essere indotta in un paziente avente una malattia oncologica o

in un soggetto che è a rischio di sviluppare una malattia oncologica. Indurre una risposta immunitaria in questo caso può significare che la condizione patologica del soggetto è migliorata, che il soggetto non sviluppa metastasi, o che il soggetto che è a rischio di sviluppare una malattia oncologica non sviluppa una malattia oncologica.

Una "risposta immunitaria cellulare", una "risposta cellulare", una "risposta cellulare nei confronti di un antigene" o un termine simile è inteso ad includere una risposta cellulare diretta alle cellule caratterizzata dalla presentazione di un antigene con l'MHC di classe I o di classe II. La risposta cellulare riguarda cellule chiamate linfociti T o linfociti T che agiscono come "helper" o "killer". I linfociti T helper (chiamati anche linfociti T CD4+) svolgono un ruolo centrale regolando la risposta immunitaria e le cellule killer (chiamate anche linfociti T citotossici, linfociti T citolitici, linfociti T CD8+ o CTL) uccidono le cellule malate quali le cellule cancerose, prevenendo la produzione di più cellule malate. In forme di realizzazione preferite, la presente invenzione implica la stimolazione di una risposta antitumorale dei CTL nei confronti delle cellule tumorali esprimenti uno o più antigeni tumorali espressi e preferibilmente presentanti tali antigeni tumorali espressi con l'MHC di classe I.

Un "antigene" secondo l'invenzione copre qualsiasi sostanza che provocherà una risposta immunitaria. In particolare, un "antigene" riguarda qualsiasi sostanza, preferibilmente un peptide o una proteina, la quale reagisce specificamente con anticorpi o con linfociti T. Secondo la presente invenzione, il termine "antigene" comprende qualsiasi molecola che comprende almeno un epitopo. Preferibilmente, un antigene nel contesto della presente invenzione è una molecola che, facoltativamente dopo il processamento, induce una reazione immunitaria, la quale è preferibilmente specifica per l'antigene (comprese cellule esprimenti l'antigene). Secondo la presente invenzione, può essere utilizzato qualsiasi antigene adatto, il quale è un candidato per una reazione immunitaria, in cui la reazione immunitaria è preferibilmente una reazione immunitaria cellulare. Nel contesto delle forme di realizzazione della presente invenzione, l'antigene è preferibilmente presentato da una cellula, preferibilmente da una cellula presentante l'antigene che comprende una cellula malata, in particolare una cellula cancerosa, in associazione a molecole dell'MHC, il che porta ad una reazione immunitaria contro l'antigene. Un antigene è preferibilmente un prodotto che corrisponde a o è derivato da un antigene presente in natura. Tali antigeni comunemente presenti in natura includono gli antigeni tumorali.

In una forma di realizzazione preferita, l'antigene è un antigene tumorale, cioè, una parte di una cellula tumorale quale una proteina o un peptide espresso in una cellula tumorale che può essere derivato dal citoplasma, dalla superficie cellulare o dal nucleo cellulare, in particolare quelli che principalmente si trovano a livello intracellulare o come antigeni di superficie delle cellule tumorali. Per esempio, gli antigeni tumorali includono l'antigene carcinoembrionale, la fetoproteina $\alpha 1$, l'isoferritina, e la solfoglucoproteina fetale, l' $\alpha 2$ -H-ferroproteina e la γ -fetoproteina. Secondo la presente invenzione, un antigene tumorale comprende preferibilmente qualsiasi antigene che è espresso in e facoltativamente caratteristico di cancri

o tumori così come di cellule tumorali o cancerose in relazione al tipo e/o al livello di espressione. In una forma di realizzazione, il termine “antigene tumorale” o “antigene associato al tumore” riguarda proteine che in condizioni normali sono specificamente espresse in un numero limitato di tessuti e/o di organi o in specifici stadi dello sviluppo, per esempio, l’antigene tumorale può essere in condizioni normali specificamente espresso nel tessuto dello stomaco, preferibilmente nella mucosa gastrica, negli organi riproduttivi, per esempio, nei testicoli, nel tessuto trofoblastico, per esempio, nella placenta, o nelle cellule della linea germinale, e vengono espresse o espresse in modo aberrante in uno o più tessuti tumorali o cancerosi. In questo contesto, “un numero limitato” significa preferibilmente non superiore a 3, più preferibilmente non superiore a 2. Gli antigeni tumorali nel contesto della presente invenzione includono, per esempio, antigeni di differenziazione, preferibilmente antigeni di differenziazione specifici per il tipo cellulare, cioè, proteine che in condizioni normali sono specificamente espresse in un certo tipo cellulare ad un certo stadio di differenziazione, antigeni del cancro/testicolo, cioè, proteine che in condizioni normali sono specificamente espresse nei testicoli e qualche volta nella placenta, e antigeni specifici della linea germinale. Preferibilmente, l’antigene tumorale o l’espressione aberrante dell’antigene tumorale identifica le cellule cancerose. Nel contesto della presente invenzione, l’antigene tumorale che viene espresso da una cellula cancerosa in un soggetto, per esempio, un paziente che soffre di una malattia oncologica, è preferibilmente un’autoproteina nel suddetto soggetto. In forme di realizzazione preferite, l’antigene tumorale nel contesto della presente invenzione viene espresso in condizioni normali specificamente in un tessuto o in un organo che è non essenziale, cioè, tessuti od organi che quando danneggiati dal sistema immunitario non conducono alla morte del soggetto, o in organi o in strutture del corpo che non sono accessibili o sono solo difficilmente accessibili da parte del sistema immunitario.

Secondo l’invenzione, i termini “antigene tumorale”, “antigene tumorale espresso”, “antigene canceroso” e “antigene canceroso espresso” sono equivalenti e vengono qui utilizzati in modo intercambiabile. Il termine “immunogenicità” riguarda l’efficacia relativa di un antigene per indurre una reazione immunitaria.

Un “peptide antigenico” secondo l’invenzione riguarda preferibilmente una porzione o un frammento di un antigene che è in grado di stimolare una risposta immunitaria, preferibilmente una risposta cellulare nei confronti dell’antigene o delle cellule caratterizzate dall’espressione dell’antigene e preferibilmente mediante la presentazione dell’antigene quali le cellule malate, in particolare le cellule cancerose. Preferibilmente, un peptide antigenico è in grado di stimolare una risposta cellulare nei confronti di una cellula caratterizzata dalla presentazione di un antigene con l’MHC di classe I e preferibilmente è in grado di stimolare un linfocita T citotossico (CTL) sensibile all’antigene. Preferibilmente, i peptidi antigenici sono peptidi presentati dall’MHC di classe I e/o di classe II o possono essere processati per produrre peptidi presentati dall’MHC di classe I e/o di classe II. Preferibilmente, i peptidi antigenici comprendono una sequenza amminoacidica sostanzialmente corrispondente alla sequenza amminoacidica di

un frammento di un antigene. Preferibilmente, il suddetto frammento di un antigene è un peptide presentato dall'MHC di classe I e/o di classe II. Preferibilmente, un peptide antigenico comprende una sequenza amminoacidica sostanzialmente corrispondente alla sequenza amminoacidica di tale frammento e viene trattato per produrre tale frammento, cioè, un peptide presentato dall'MHC di classe I e/o di classe II derivato da un antigene.

Se un peptide deve essere presentato direttamente, cioè, senza processamento, in particolare senza scissione, esso ha una lunghezza che è adatta per il legame ad una molecola di MHC, in particolare una molecola dell'MHC di classe I, e preferibilmente ha una lunghezza da 7 a 20 amminoacidi, più preferibilmente una lunghezza da 7 a 12 amminoacidi, più preferibilmente una lunghezza da 8 ad 11 amminoacidi, in particolare una lunghezza di 9 o 10 amminoacidi.

Se un peptide fa parte di un'entità più grande comprendente ulteriori sequenze, per esempio di una sequenza o un polipeptide vaccinale, e deve essere presentato in seguito a processamento, in particolare in seguito a scissione, il peptide prodotto mediante il processamento ha una lunghezza che è adatta per il legame ad una molecola di MHC, in particolare una molecola dell'MHC di classe I, e preferibilmente ha una lunghezza da 7 a 20 amminoacidi, più preferibilmente una lunghezza da 7 a 12 amminoacidi, più preferibilmente una lunghezza da 8 ad 11 amminoacidi, in particolare una lunghezza di 9 o 10 amminoacidi. Preferibilmente, la sequenza del peptide che deve essere presentato in seguito a processamento è derivata dalla sequenza amminoacidica di un antigene, cioè, la sua sequenza corrisponde sostanzialmente ed è preferibilmente completamente identica ad un frammento di un antigene. Così, un peptide antigenico o una sequenza vaccinale in una forma di realizzazione comprende una sequenza da 7 a 20 amminoacidi di lunghezza, più preferibilmente una lunghezza da 7 a 12 amminoacidi, più preferibilmente una lunghezza da 8 ad 11 amminoacidi, in particolare una lunghezza di 9 o 10 amminoacidi che corrisponde sostanzialmente ed è preferibilmente completamente identica ad un frammento di un antigene ed in seguito a processamento del peptide antigenico o della sequenza vaccinale costituisce il peptide presentato. Tale peptide prodotto mediante processamento comprende la variazione di sequenza identificata.

Un peptide o un epitopo antigenico può essere presente in un vaccino come parte di un'entità più grande quale una sequenza e/o un polipeptide vaccinale comprendente più di un peptide o di un epitopo antigenico. Il peptide o l'epitopo antigenico presentato viene prodotto in seguito a processamento adatto.

Peptidi aventi sequenze amminoacidiche sostanzialmente corrispondenti ad una sequenza di un peptide che è presentato dall'MHC di classe I possono differire ad uno o più residui che non sono essenziali per il riconoscimento da parte del TCR del peptide come presentato dall'MHC di classe I, o per il legame del peptide all'MHC. Tali peptidi sostanzialmente corrispondenti sono anche in grado di stimolare un CTL sensibile

all'antigene e possono essere considerati immunologicamente equivalenti. Peptidi aventi sequenze amminoacidiche differenti da un peptide presentato a residui che non influenzano il riconoscimento da parte del TCR ma migliorano la stabilità di legame all'MHC possono migliorare l'immunogenicità del peptide antigenico, e possono essere qui indicati come "peptide ottimizzato". Utilizzando la conoscenza esistente circa quali di questi residui possono più probabilmente influenzare il legame all'MHC o al TCR, può essere impiegato un approccio razionale alla progettazione di peptidi sostanzialmente corrispondenti. I peptidi derivanti che sono funzionali sono contemplati come peptidi antigenici.

Un peptide antigenico quando presentato dall'MHC dovrebbe essere riconoscibile da parte di un recettore dei linfociti T. Preferibilmente, il peptide antigenico se riconosciuto da un recettore dei linfociti T è in grado di indurre, in presenza di appropriati segnali costimolatori, l'espansione clonale del linfocita T recante il recettore dei linfociti T che riconosce specificamente il peptide antigenico. Preferibilmente, i peptidi antigenici, in particolare se presentati in associazione a molecole dell'MHC, sono in grado di stimolare una risposta immunitaria, preferibilmente una risposta cellulare nei confronti dell'antigene da cui essi sono derivati o di cellule caratterizzate dall'espressione dell'antigene e preferibilmente caratterizzate dalla presentazione dell'antigene. Preferibilmente, un peptide antigenico è in grado di stimolare una risposta cellulare nei confronti di una cellula caratterizzata dalla presentazione dell'antigene con l'MHC di classe I e preferibilmente è in grado di stimolare un CTL sensibile all'antigene. Tale cellula è preferibilmente una cellula bersaglio.

"Processamento dell'antigene" o "processamento" si riferisce alla degradazione di un polipeptide o di un antigene in prodotti di processamento, i quali sono frammenti del suddetto polipeptide o antigene (per esempio la degradazione di un polipeptide in peptidi) e l'associazione di uno o più di questi frammenti (per esempio attraverso il legame) con molecole dell'MHC per la presentazione da parte delle cellule, preferibilmente cellule presentanti l'antigene, a linfociti T specifici.

"Cellule presentanti l'antigene" (APC) sono cellule che presentano frammenti peptidici di antigeni proteici in associazione a molecole dell'MHC sulla loro superficie cellulare. Alcune APC possono attivare i linfociti T specifici per un antigene.

Le cellule presentanti l'antigene professionali sono molto efficienti nell'internalizzare l'antigene, mediante la fagocitosi o mediante l'endocitosi mediata da recettore, e quindi nel mostrare un frammento dell'antigene, legato ad una molecola dell'MHC di classe II, sulla loro membrana. Il linfocita T riconosce ed interagisce con il complesso antigene-molecola dell'MHC di classe II sulla membrana della cellula presentante l'antigene. Un ulteriore segnale costimolatorio è quindi prodotto dalla cellula presentante l'antigene, conducendo all'attivazione del linfocita T. L'espressione di molecole costimolatorie è una caratteristica che definisce le cellule presentanti l'antigene professionali.

I principali tipi di cellule presentanti l'antigene professionali sono le cellule dendritiche, le quali hanno l'intervallo più ampio di presentazione

dell'antigene, e sono probabilmente le più importanti cellule presentanti l'antigene, i macrofagi, i linfociti B, e certe cellule epiteliali attivate.

Le cellule dendritiche (DC) sono popolazioni leucocitarie che presentano antigeni catturati nei tessuti periferici ai linfociti T attraverso entrambe le vie metaboliche di presentazione dell'antigene degli MHC di classe II e I. È ben noto che le cellule dendritiche sono potenti induttori delle risposte immunitarie e l'attivazione di queste cellule è un passaggio critico per l'induzione dell'immunità antitumorale.

Le cellule dendritiche sono convenientemente classificate come cellule "immature" e "mature", le quali possono essere utilizzate come modo semplice per distinguere tra due fenotipi ben caratterizzati. Tuttavia, questa nomenclatura non dovrebbe essere intesa ad escludere tutti i possibili stadi intermedi di differenziazione.

Le cellule dendritiche immature sono caratterizzate come cellule presentanti l'antigene con una capacità elevata di assorbimento e processamento dell'antigene, la quale si correla con l'espressione elevata del recettore per l'Fcγ e del recettore per il mannosio. Il fenotipo maturo è tipicamente caratterizzato da un'espressione inferiore di questi marcatori, ma da un'espressione elevata di molecole della superficie cellulare responsabili dell'attivazione dei linfociti T quali MHC di classe I e di classe II, molecole di adesione (per esempio CD54 e CD11) e molecole costimolatorie (per esempio CD40, CD80, CD86 e 4-1 BB).

La maturazione delle cellule dendritiche viene indicata come lo stato di attivazione delle cellule dendritiche a cui tali cellule dendritiche presentanti l'antigene conducono alla stimolazione dei linfociti T, mentre la presentazione da parte delle cellule dendritiche immature risulta in tolleranza. La maturazione delle cellule dendritiche è principalmente causata da biomolecole con caratteristiche microbiche rilevate mediante recettori innati (DNA batterico, RNA virale, endotossina, ecc.), citochine proinfiammatorie (TNF, IL-1, IFN), ligazione di CD40 sulla superficie delle cellule dendritiche da parte di CD40L, e sostanze rilasciate da cellule che vanno incontro a morte cellulare stressante. Le cellule dendritiche possono essere derivate coltivando cellule di midollo osseo in vitro con citochine, quali il fattore stimolante le colonie di granulociti e di macrofagi (GM-CSF) ed il fattore di necrosi tumorale alfa.

Le cellule presentanti l'antigene non professionali non esprimono costitutivamente le proteine dell'MHC di classe II richieste per l'interazione con i linfociti T vergini; queste vengono espresse solo a seguito della stimolazione delle cellule presentanti l'antigene non professionali mediante certe citochine quali IFN-γ.

Le "cellule presentanti l'antigene" possono essere caricate con peptidi presentati dall'MHC di classe I mediante trasduzione delle cellule con acido nucleico, preferibilmente RNA, codificante per un peptide o per un polipeptide comprendente il peptide da presentare, per esempio un acido nucleico codificante per l'antigene.

In alcune forme di realizzazione, una composizione farmaceutica comprendente un veicolo di rilascio del gene che ha come bersaglio una cellula dendritica o un'altra cellula presentante l'antigene può essere somministrata ad un paziente, risultando in una trasfezione che avviene in vivo. La trasfezione in vivo di cellule dendritiche, per esempio, in generale può essere eseguita utilizzando qualsiasi dei metodi noti nell'arte, quali quelli descritti in WO 97/24447, o l'approccio della pistola genica descritto da Mahvi et al., Immunology e cell Biology 75: da 456 a 460, 1997.

Secondo l'invenzione, il termine "cellula presentante l'antigene" comprende anche le cellule bersaglio.

"Cellula bersaglio" deve significare una cellula che è un bersaglio per una risposta immunitaria quale una risposta immunitaria cellulare. Le cellule bersaglio includono cellule che presentano un antigene o un epitopo antigenico, cioè un frammento peptidico derivato da un antigene, ed includono qualsiasi cellula indesiderata quale una cellula cancerosa. In forme di realizzazione preferite, la cellula bersaglio è una cellula esprimente un antigene come qui descritto e preferibilmente che presenta il suddetto antigene con l'MHC di classe I.

Il termine "epitopo" si riferisce ad un determinante antigenico in una molecola quale un antigene, cioè, ad una parte o ad un frammento della molecola che è riconosciuto dal sistema immunitario, per esempio, che è riconosciuto da un linfocita T, in particolare quando presentato in associazione a molecole dell'MHC. Un epitopo di una proteina quale un antigene tumorale comprende preferibilmente una porzione continua o discontinua della suddetta proteina ed ha preferibilmente una lunghezza da 5 a 100, preferibilmente da 5 a 50, più preferibilmente da 8 a 30, il più preferibilmente da 10 a 25 amminoacidi, per esempio, l'epitopo può avere preferibilmente una lunghezza di 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, o 25 amminoacidi. Si preferisce particolarmente che l'epitopo nel contesto della presente invenzione sia un epitopo dei linfociti T.

Secondo l'invenzione un epitopo può legarsi a molecole dell'MHC quali molecole dell'MHC sulla superficie di una cellula e così, può essere un "peptide legante l'MHC" o "peptide antigenico". Il termine "peptide legante l'MHC" riguarda un peptide che si lega ad una molecola dell'MHC di classe I e/o dell'MHC di classe II. Nel caso di complessi MHC di classe I/peptide, i peptidi di legame hanno tipicamente una lunghezza da 8 a 10 amminoacidi sebbene peptidi più lunghi o più corti possano essere efficaci. Nel caso di complessi MHC di classe II/peptide, i peptidi di legame hanno tipicamente una lunghezza da 10 a 25 amminoacidi ed hanno in particolare una lunghezza da 13 a 18 amminoacidi, mentre peptidi più lunghi e più corti possono essere efficaci.

I termini "epitopo", "peptide antigenico", "epitopo antigenico", "peptide immunogenico" e "peptide legante l'MHC" vengono qui utilizzati in modo intercambiabile e riguardano preferibilmente una rappresentazione incompleta di un antigene che è preferibilmente in grado di stimolare una risposta immunitaria nei confronti dell'antigene o di una cellula esprimente o comprendente e preferibilmente presentante l'antigene.

Preferibilmente, i termini riguardano una porzione immunogenica di un antigene. Preferibilmente, è una porzione di un antigene che è riconosciuta (cioè, specificamente legata) da un recettore dei linfociti T, in particolare se presentata in associazione a molecole dell'MHC. Tali porzioni immunogeniche preferite si legano ad una molecola dell'MHC di classe I o di classe II. Come qui utilizzata, una porzione immunogenica è detta "legarsi" ad una molecola dell'MHC di classe I o di classe II se tale legame è rilevabile utilizzando qualsiasi saggio noto nell'arte.

Come qui utilizzato, il termine "neoepitopo" si riferisce ad un epitopo che non è presente in un riferimento quale una cellula normale non cancerosa o una cellula della linea germinale ma si trova nelle cellule cancerose. Questo include, in particolare, situazioni in cui un corrispondente epitopo si trova in una cellula normale non cancerosa o in una cellula della linea germinale, tuttavia, a causa di una o più mutazioni in una cellula cancerosa la sequenza dell'epitopo è cambiata in modo tale da risultare nel neoepitopo.

Il termine "porzione" si riferisce ad una frazione. Con rispetto ad una struttura particolare quale una sequenza amminoacidica o una proteina il termine "porzione" di questa può designare una frazione continua o discontinua della suddetta struttura. Preferibilmente, una porzione di una sequenza amminoacidica comprende almeno l'1%, almeno il 5%, almeno il 10%, almeno il 20%, almeno il 30%, preferibilmente almeno il 40%, preferibilmente almeno il 50%, più preferibilmente almeno il 60%, più preferibilmente almeno il 70%, ancora più preferibilmente almeno l'80%, ed il più preferibilmente almeno il 90% degli amminoacidi della suddetta sequenza amminoacidica. Preferibilmente, se la porzione è una frazione discontinua la suddetta frazione discontinua è costituita da 2, 3, 4, 5, 6, 7, 8, o più parti di una struttura, ciascuna parte essendo un elemento continuo della struttura. Per esempio, una frazione discontinua di una sequenza amminoacidica può essere costituita da 2, 3, 4, 5, 6, 7, 8, o più, preferibilmente non più di 4 parti della suddetta sequenza amminoacidica, in cui ciascuna parte comprende preferibilmente almeno 5 amminoacidi continui, almeno 10 amminoacidi continui, preferibilmente almeno 20 amminoacidi continui, preferibilmente almeno 30 amminoacidi continui della sequenza amminoacidica.

I termini "parte" e "frammento" vengono qui utilizzati in modo intercambiabile e si riferiscono ad un elemento continuo. Per esempio, una parte di una struttura quale una sequenza amminoacidica o una proteina si riferisce ad un elemento continuo della suddetta struttura. Una porzione, una parte o un frammento di una struttura comprende preferibilmente una o più proprietà funzionali della suddetta struttura. Per esempio, una porzione, una parte o un frammento di un epitopo, di un peptide o di una proteina è preferibilmente immunologicamente equivalente all'epitopo, al peptide o alla proteina da cui è derivato. Nel contesto della presente invenzione, una "parte" di una struttura quale una sequenza amminoacidica comprende preferibilmente, preferibilmente consiste di almeno il 10%, almeno il 20%, almeno il 30%, almeno il 40%, almeno il 50%, almeno il 60%, almeno il 70%, almeno l'80%, almeno l'85%, almeno il 90%, almeno il 92%, almeno il 94%, almeno il 96%, almeno il 98%, almeno il 99% della struttura o

della sequenza amminoacidica intera.

Il termine “cellula immunoreattiva” nel contesto della presente invenzione riguarda una cellula che esercita funzioni effettrici durante una reazione immunitaria. Una “cellula immunoreattiva” è preferibilmente in grado di legarsi ad un antigene o a una cellula caratterizzata dalla presentazione di un antigene o di un peptide antigenico derivato da un antigene e che media una risposta immunitaria. Per esempio, tali cellule secernono citochine e/o chemochine, secernono anticorpi, riconoscono cellule cancerose, e facoltativamente eliminano tali cellule. Per esempio, le cellule immunoreattive comprendono linfociti T (linfociti T citotossici, linfociti T helper, linfociti T infiltranti il tumore), linfociti B, cellule natural killer, neutrofili, macrofagi, e cellule dendritiche. Preferibilmente, nel contesto della presente invenzione, le “cellule immunoreattive” sono i linfociti T, preferibilmente linfociti T CD4+ e/o CD8+.

Preferibilmente, una “cellula immunoreattiva” riconosce un antigene o un peptide antigenico derivato da un antigene con qualche grado di specificità, in particolare se presentato in associazione a molecole dell’MHC quale sulla superficie di cellule presentanti l’antigene o di cellule malate quali le cellule cancerose. Preferibilmente, il suddetto riconoscimento permette alla cellula che riconosce un antigene o un peptide antigenico derivato dal suddetto antigene di essere sensibile o reattiva. Se la cellula è un linfocita T helper (linfocita T CD4+) recante recettori che riconoscono un antigene o un peptide antigenico derivato da un antigene in associazione a molecole dell’MHC di classe II tale sensibilità o reattività può implicare il rilascio di citochine e/o l’attivazione di linfociti CD8+ (CTL) e/o di linfociti B. Se la cellula è un CTL tale sensibilità o reattività può implicare l’eliminazione di cellule presentate in associazione a molecole dell’MHC di classe I, cioè, cellule caratterizzate dalla presentazione di un antigene con l’MHC di classe I, per esempio, attraverso l’apoptosi o la lisi cellulare mediata da perforina. Secondo l’invenzione, la sensibilità dei CTL può includere flusso prolungato di calcio, divisione cellulare, produzione di citochine quali IFN- γ e TNF- α , sovraregolazione di marcatori di attivazione quali CD44 e CD69, e uccisione citolitica specifica delle cellule bersaglio esprimenti l’antigene. La sensibilità dei CTL può anche essere determinata utilizzando un reporter artificiale che indica accuratamente la sensibilità dei CTL. Tali CTL che riconoscono un antigene o un peptide antigenico derivato da un antigene e sono sensibili o reattivi vengono chiamati anche “CTL sensibili all’antigene” qui. Se la cellula è un linfocita B tale sensibilità può implicare il rilascio di immunoglobuline.

I termini “cellula T” e “linfocita T” vengono qui utilizzati in modo intercambiabile ed includono i linfociti T helper (linfociti T CD4+) ed i linfociti T citotossici (CTL, linfociti T CD8+) che comprendono i linfociti T citolitici.

I linfociti T appartengono ad un gruppo di globuli bianchi noti come linfociti, e svolgono un ruolo centrale nell’immunità mediata dalle cellule. Possono essere distinti da altri tipi di linfociti, quali i linfociti B e le cellule natural killer dalla presenza di uno speciale recettore sulla loro

superficie cellulare chiamato recettore dei linfociti T (TCR). Il timo è l'organo principale responsabile della maturazione dei linfociti T. Sono stati scoperti diversi sottoinsiemi differenti di linfociti T, ciascuno con una funzione distinta.

I linfociti T helper aiutano altri globuli bianchi nei processi immunologici, compresi la maturazione dei linfociti B in cellule del plasma e l'attivazione dei linfociti T citotossici e dei macrofagi, tra le altre funzioni. Queste cellule sono note anche come linfociti T CD4+ poiché esprimono la proteina CD4 sulla loro superficie. I linfociti T helper diventano attivati quando vengono presentati con antigeni peptidici dalle molecole dell'MHC di classe II che vengono espresse sulla superficie di cellule presentanti l'antigene (APC). Una volta attivati, essi si dividono rapidamente e secernono piccole proteine chiamate citochine che regolano o aiutano nella risposta immunitaria attiva.

I linfociti T citotossici distruggono cellule infettate dai virus e cellule tumorali, e sono anche implicati nel rigetto dei trapianti. Queste cellule sono note anche come linfociti T CD8+ poiché esprimono la glicoproteina CD8 sulla loro superficie. Queste cellule riconoscono i loro bersagli legando un antigene associato all'MHC di classe I, il quale è presente sulla superficie di quasi ogni cellula del corpo.

Una maggioranza di linfociti T ha un recettore dei linfociti T (TCR) esistente come complesso di diverse proteine. Il reale recettore dei linfociti T è costituito da due catene peptidiche separate, le quali vengono prodotte dai geni indipendenti alfa e beta del recettore dei linfociti T (TCR α e TCR β) e vengono chiamate catene del TCR α e β . I linfociti T $\gamma\delta$ (linfociti T gamma delta) rappresentano un piccolo sottoinsieme di linfociti T che possiedono un recettore distinto dei linfociti T (TCR) sulla loro superficie. Tuttavia, nei linfociti T $\gamma\delta$, il TCR è costituito da una catena γ e da una catena δ . Questo gruppo di linfociti T è molto meno comune (2% dei linfociti T totali) rispetto ai linfociti T $\alpha\beta$.

Il primo segnale nell'attivazione dei linfociti T viene fornito dal legame del recettore dei linfociti T ad un peptide corto presentato dal complesso maggiore di istocompatibilità (MHC) su un'altra cellula. Questo assicura che solo un linfocita T con un TCR specifico per quel peptide venga attivato. La cellula partner è solitamente una cellula presentante l'antigene professionale (APC), solitamente una cellula dendritica nel caso di risposte non vergini, sebbene i linfociti B ed i macrofagi possano essere importanti APC. I peptidi presentati ai linfociti T CD8+ dalle molecole dell'MHC di classe I sono lunghi tipicamente da 8 a 10 amminoacidi; i peptidi presentati ai linfociti T CD4+ dalle molecole dell'MHC di classe II sono tipicamente più lunghi, poiché le estremità della fessura di legame della molecola dell'MHC di classe II sono aperte.

Secondo la presente invenzione, un recettore dei linfociti T è in grado di legarsi ad un bersaglio predeterminato se esso ha una significativa affinità per il suddetto bersaglio predeterminato e si lega al suddetto bersaglio predeterminato nei saggi convenzionali. L'"affinità" o "affinità di legame" è spesso misurata mediante la costante di dissociazione all'equilibrio (K_D). Un recettore dei linfociti T non è (sostanzialmente) in grado di legarsi ad un bersaglio se esso non ha significativa affinità per il suddetto bersaglio e non si lega significativamente al suddetto bersaglio nei saggi

convenzionali.

Un recettore dei linfociti T è preferibilmente in grado di legarsi specificamente ad un bersaglio predeterminato. Un recettore dei linfociti T è specifico per un bersaglio predeterminato se è in grado di legarsi al suddetto bersaglio predeterminato mentre non è (sostanzialmente) in grado di legarsi ad altri bersagli, cioè non ha significativa affinità per altri bersagli e non si lega significativamente ad altri bersagli nei saggi convenzionali.

I linfociti T citotossici possono essere generati in vivo mediante l'incorporazione di un antigene o di un peptide antigenico nelle cellule presentanti l'antigene in vivo. L'antigene o il peptide antigenico può essere rappresentato come proteina, come DNA (per esempio all'interno di un vettore) o come RNA. L'antigene può essere processato per produrre un peptide partner per la molecola dell'MHC, mentre un suo frammento può essere presentato senza la necessità di ulteriore processamento. L'ultimo è in particolare il caso, se questi possono legarsi a molecole dell'MHC. In generale, è possibile la somministrazione ad un paziente mediante iniezione intradermica. Tuttavia, l'iniezione può anche essere realizzata per via intranodale in un linfonodo (Maloy et al. (2001), Proc. Natl. Acad. Sci. USA 98: da 3299 a 3303). Le cellule risultanti presentano il complesso d'interesse e vengono riconosciute dai linfociti T citotossici autologhi che quindi si propagano.

L'attivazione specifica dei linfociti T CD4+ o CD8+ può essere rilevata in una varietà di modi. I metodi per rilevare l'attivazione specifica dei linfociti T includono la rilevazione della proliferazione dei linfociti T, la produzione di citochine (per esempio linfocine), o la generazione di attività citolitica. Per i linfociti T CD4+, un metodo preferito per rilevare l'attivazione specifica dei linfociti T è il rilevamento della proliferazione dei linfociti T. Per i linfociti T CD8+, un metodo preferito per rilevare l'attivazione specifica dei linfociti T è il rilevamento della generazione di attività citolitica.

Il termine "complesso maggiore di istocompatibilità" e l'abbreviazione "MHC" includono le molecole dell'MHC di classe I e dell'MHC di classe II e riguardano un complesso di geni che sono presenti in tutti i vertebrati. Le proteine o le molecole dell'MHC sono importanti per la trasduzione del segnale tra i linfociti e le cellule presentanti l'antigene o le cellule malate in reazioni immunitarie, in cui le proteine o le molecole dell'MHC legano peptidi e li presentano per il riconoscimento da parte del recettore dei linfociti T. Le proteine codificate dall'MHC vengono espresse sulla superficie delle cellule, e mostrano sia autoantigeni (frammenti peptidici dalla cellula stessa) che antigeni non propri (per esempio frammenti di microrganismi invasori) ad un linfocita T.

La regione dell'MHC è divisa in tre sottogruppi, classe I, classe II, e classe III. Le proteine dell'MHC di classe I contengono una catena α e β 2-microglobulina (non parte dell'MHC codificato dal cromosoma 15). Esse presentano frammenti antigenici ai linfociti T citotossici. Sulla maggior parte delle cellule del sistema immunitario, specificamente sulle cellule presentanti l'antigene, le proteine dell'MHC di classe II contengono catene

α e β e presentano frammenti antigenici per i linfociti T helper. La regione dell'MHC di classe III codifica per altri componenti immunologici, quali componenti del complemento ed alcuni che codificano citochine.

Nell'uomo, i geni nella regione dell'MHC che codificano per proteine presentanti l'antigene sulla superficie cellulare vengono indicati come geni dell'antigene leucocitario umano (HLA). Tuttavia, l'abbreviazione MHC è utilizzata spesso per riferirsi ai prodotti dei geni HLA. I geni HLA includono i nove geni cosiddetti classici dell'MHC: HLA-A, HLA-B, HLA-C, HLA-DPA1, HLA-DPB1, HLA-DQA1, HLA-DQB1, HLA-DRA, e HLA-DRB1.

In una forma di realizzazione preferita di tutti gli aspetti dell'invenzione una molecola dell'MHC è una molecola HLA.

Con "cellula caratterizzata dalla presentazione di un antigene" o "cellula presentante un antigene" o espressioni simili si intende una cellula quale una cellula malata, per esempio una cellula cancerosa, o una cellula presentante l'antigene che presenta l'antigene che esprime o un frammento derivato dal suddetto antigene, per esempio mediante processamento dell'antigene, in associazione a molecole dell'MHC, in particolare molecole dell'MHC di classe I. In modo simile, i termini "malattia caratterizzata dalla presentazione di un antigene" indica una malattia riguardante cellule caratterizzate dalla presentazione di un antigene, in particolare con l'MHC di classe I. La presentazione di un antigene da parte di una cellula può essere effettuata trasferendo la cellula con un acido nucleico quale RNA codificante per l'antigene.

Con "frammento di un antigene che è presentato" o espressioni simili si intende che il frammento può essere presentato dall'MHC di classe I o di classe II, preferibilmente l'MHC di classe I, per esempio quando aggiunto direttamente a cellule presentanti l'antigene. In una forma di realizzazione, il frammento è un frammento che è naturalmente presentato dalle cellule esprimenti un antigene.

Il termine "immunologicamente equivalente" significa che la molecola immunologicamente equivalente quale la sequenza amminoacidica immunologicamente equivalente mostra le stesse o essenzialmente le stesse proprietà immunologiche e/o esercita gli stessi o essenzialmente gli stessi effetti immunologici, per esempio, in relazione al tipo dell'effetto immunologico quale l'induzione di una risposta immunitaria umorale e/o cellulare, la forza e/o la durata della reazione immunitaria indotta, o la specificità della reazione immunitaria indotta. Nel contesto della presente invenzione, il termine "immunologicamente equivalente" viene utilizzato preferibilmente in relazione agli effetti immunologici o alle proprietà di un peptide utilizzato per l'immunizzazione. Per esempio, una sequenza amminoacidica è immunologicamente equivalente ad una sequenza amminoacidica di riferimento se la suddetta sequenza amminoacidica quando esposta al sistema immunitario di un soggetto induce una reazione immunitaria che ha una specificità di reazione con la sequenza amminoacidica di riferimento.

Il termine "funzioni immunitarie effettrici" nel contesto della presente invenzione comprende qualsiasi funzione mediata da componenti del sistema

immunitario che risultano, per esempio, nell'uccisione delle cellule tumorali, o nell'inibizione della crescita tumorale e/o nell'inibizione dello sviluppo del tumore, compresi l'inibizione della diffusione del tumore e delle metastasi. Preferibilmente, le funzioni immunitarie effettrici nel contesto della presente invenzione sono funzioni effettrici mediate da linfociti T. Tali funzioni comprendono nel caso di un linfocita T helper (linfocita T CD4+) il riconoscimento di un antigene o di un peptide antigenico derivato da un antigene in associazione a molecole dell'MHC di classe II da parte di un recettore dei linfociti T, il rilascio di citochine e/o l'attivazione di linfociti CD8+ (CTL) e/o di linfociti B, e nel caso dei CTL il riconoscimento di un antigene o di un peptide antigenico derivato da un antigene in associazione a molecole dell'MHC di classe I da parte di un recettore dei linfociti T, l'eliminazione di cellule presentate in associazione a molecole dell'MHC di classe I, cioè, cellule caratterizzate dalla presentazione di un antigene con l'MHC di classe I, per esempio, attraverso l'apoptosi o la lisi cellulare mediata da perforina, la produzione di citochine quali IFN- γ e TNF- α , e l'uccisione citolitica specifica delle cellule bersaglio esprimenti l'antigene.

Il termine "genoma" riguarda la quantità totale di informazioni genetiche nei cromosomi di un organismo o di una cellula. Il termine "esoma" si riferisce alle regioni codificanti di un genoma. Il termine "trascrittoma" riguarda la serie di tutte le molecole di RNA.

Un "acido nucleico" è secondo l'invenzione preferibilmente acido deossiribonucleico (DNA) oppure acido ribonucleico (RNA), più preferibilmente RNA, il più preferibilmente RNA trascritto in vitro (IVT RNA) o RNA sintetico. Gli acidi nucleici includono secondo l'invenzione DNA, cDNA, mRNA genomico, molecole prodotte per via ricombinante e sintetizzate chimicamente. Secondo l'invenzione, un acido nucleico può essere presente come filamento singolo o filamento doppio e molecola lineare o chiusa covalentemente in modo circolare. Un acido nucleico può, secondo l'invenzione, essere isolato. Il termine "acido nucleico isolato" significa, secondo l'invenzione, che l'acido nucleico (i) è stato amplificato in vitro, per esempio attraverso la reazione a catena della polimerasi (PCR), (ii) è stato prodotto per via ricombinante mediante clonazione, (iii) è stato purificato, per esempio, mediante scissione e separazione mediante elettroforesi su gel, o (iv) è stato sintetizzato, per esempio, mediante sintesi chimica. Un nucleico può essere impiegato per l'introduzione nelle cellule, cioè la trasfezione delle cellule, in particolare, in forma di RNA che può essere preparato mediante trascrizione in vitro da uno stampo di DNA. L'RNA può inoltre essere modificato prima dell'applicazione stabilizzando le sequenze, l'incapsulazione, e la poliadenilazione.

Il termine "materiale genetico" si riferisce ad un acido nucleico isolato, DNA o RNA, ad una sezione di una doppia elica, ad una sezione di un cromosoma, oppure all'intero genoma di un organismo o di una cellula, in particolare il suo esoma o trascrittoma.

Il termine "mutazione" si riferisce ad un cambiamento della sequenza nucleotidica o ad una differenza di questa (sostituzione, aggiunta o delezione nucleotidica) in confronto ad un riferimento. Una "mutazione somatica" può avvenire in qualsiasi delle cellule del corpo eccetto le cellule germinali

(sperma ed uovo) e quindi non viene trasmessa ai bambini. Queste alterazioni possono (ma non sempre) causare il cancro o altre malattie. Preferibilmente una mutazione è una mutazione non sinonima. Il termine “mutazione non sinonima” si riferisce ad una mutazione, preferibilmente una sostituzione nucleotidica, che risulta in una modificazione amminoacidica quale una sostituzione amminoacidica nel prodotto di traduzione.

Secondo l’invenzione, il termine “mutazione” comprende mutazioni puntiformi, indel, fusioni, cromotripsi e modifiche dell’RNA.

Secondo l’invenzione, il termine “indel” descrive una speciale classe di mutazioni, definita come mutazione che risulta in un inserimento e in una delezione colocalizzati ed in un guadagno o una perdita netti in nucleotidi. In regioni codificanti del genoma, a meno che la lunghezza di un indel sia un multiplo di 3, producono una mutazione per spostamento della fase di lettura. Gli indel possono essere opposti ad una mutazione puntiforme; dove un indel inserisce e cancella nucleotidi da una sequenza, una mutazione puntiforme è una forma di sostituzione che sostituisce uno dei nucleotidi.

Le fusioni possono generare geni ibridi formati da due geni precedentemente distinti. Può avvenire come risultato di una traslocazione, di una delezione interstiziale, o di un’inversione cromosomiale. Spesso, i geni di fusione sono oncogeni. I geni oncogenici di fusione possono condurre ad un prodotto genico con una funzione nuova o differente dai due partner di fusione. In alternativa, un protooncogene è fuso ad un promotore forte, e quindi la funzione oncogenica è regolata per funzionare mediante una sovraregolazione causata dal promotore forte del partner di fusione a monte. I trascritti oncogenici di fusione possono anche essere causati da eventi di trans-splicing o di continuazione della trascrizione oltre il codone di terminazione.

Secondo l’invenzione, il termine “cromotripsi” si riferisce ad un fenomeno genetico mediante il quale regioni specifiche del genoma vengono distrutte e quindi incollate insieme attraverso un singolo evento devastante.

Secondo l’invenzione, il termine “modifica dell’RNA” si riferisce a procedimenti molecolari in cui il contenuto di informazioni in una molecola di RNA è alterato attraverso una variazione chimica nella composizione delle basi. La modifica dell’RNA comprende modifiche dei nucleosidi quali deamminazioni da citidina (C) ad uridina (U) e da adenosina (A) ad inosina (I), così come aggiunte ed inserzioni nucleotidiche non presenti nello stampo. La modifica dell’RNA negli mRNA altera efficacemente la sequenza amminoacidica della proteina codificata così che si differenzi da quella prevista dalla sequenza di DNA genomico.

Il termine “firma delle mutazioni tumorali” si riferisce ad una serie di mutazioni che sono presenti nelle cellule cancerose in confronto alle cellule non cancerose di riferimento.

Secondo l’invenzione, un “riferimento” può essere utilizzato per mettere in correlazione e confrontare i risultati ottenuti nei metodi dell’invenzione

da un campione tumorale. Tipicamente il “riferimento” può essere ottenuto sulla base di uno o più campioni normali, in campioni particolari che non sono affetti da una malattia oncologica, ottenuto da un paziente o da uno o più individui differenti, preferibilmente individui sani, in particolare individui della stessa specie. Un “riferimento” può essere determinato empiricamente testando un numero sufficientemente grande di campioni normali.

Qualsiasi metodo adatto di sequenziamento può essere utilizzato secondo l’invenzione, essendo le tecnologie di sequenziamento di nuova generazione (NGS) preferite. I metodi di sequenziamento di terza generazione potrebbero sostituire la tecnologia dell’NGS nel futuro per accelerare il passaggio di sequenziamento del metodo. Per scopi di chiarificazione: i termini “sequenziamento di nuova generazione” o “NGS” nel contesto della presente invenzione significano tutte le nuove tecnologie di sequenziamento a volumi elevati che, in contrasto con la metodologia “convenzionale” di sequenziamento nota come chimica secondo Sanger, legge gli stampi di acido nucleico casualmente in parallelo lungo l’intero genoma mediante rottura dell’intero genoma in piccoli pezzi. Tali tecnologie dell’NGS (note anche come tecnologie di sequenziamento massivamente parallelo) sono in grado di rilasciare informazioni di sequenze nucleotidiche di un intero genoma, esoma, trascrittoma (tutte le sequenze trascritte di un genoma) o metiloma (tutte le sequenze metilate di un genoma) in periodi di tempo molto brevi, per esempio entro da 1 a 2 settimane, preferibilmente entro da 1 a 7 giorni o il più preferibilmente entro meno di 24 ore e permettono, in principio, approcci di sequenziamento a cellula singola. Multiple piattaforme di NGS che sono disponibili in commercio o che vengono menzionate nella letteratura possono essere utilizzate nel contesto della presente invenzione per esempio quelle descritte in dettaglio in Zhang et al. 2011: The impact of next-generation sequencing on genomics. *J. Genet Genomics* 38 (3), da 95 a 109; o in Voelkerding et al. 2009: Next generation sequencing: From basic research to diagnostics. *Clinical chemistry* 55, da 641 a 658. Esempi non limitanti di tali tecnologie/piattaforme di NGS sono

1) la tecnologia di sequenziamento mediante sintesi nota come pirosequenziamento attuata per esempio nel GS-FLX 454 Genome Sequencer™ dell’azienda 454 Life Sciences associata a Roche (Branford, Connecticut), prima descritta in Ronaghi et al. 1998: “A sequencing method based on real-time pyrophosphate”. *Science* 281 (5375), da 363 a 365. Questa tecnologia utilizza una PCR in emulsione in cui sfere leganti il DNA a singolo filamento sono incapsulate mediante vigorosa agitazione rotante in micelle acquose contenenti reagenti di PCR circondate da olio per l’amplificazione mediante PCR in emulsione. Durante il procedimento di pirosequenziamento, la luce emessa dalle molecole di fosfato durante l’incorporazione dei nucleotidi è registrata come la polimerasi sintetizza il filamento di DNA.

2) Gli approcci di sequenziamento mediante sintesi sviluppato da Solexa (ora parte di Illumina Inc., San Diego, California) che si basa su terminatori reversibili con coloranti ed è migliorato per esempio nell’Illumina/Solexa Genome Analyzer™ e nell’Illumina HiSeq 2000 Genome

Analyzer™. In questa tecnologia, tutti e quattro i nucleotidi vengono aggiunti simultaneamente in frammenti a gruppi oligoattivati nei canali di flusso cellulare insieme alla DNA polimerasi. L'amplificazione a ponte estende i filamenti dei gruppi con tutti e quattro i nucleotidi marcati in modo fluorescente per il sequenziamento.

3) Gli approcci di sequenziamento mediante ligazione, per esempio attuati nella piattaforma SOLid™ di Applied Biosystems (ora Life Technologies Corporation, Carlsbad, California). In questa tecnologia, un gruppo di tutti i possibili oligonucleotidi di una lunghezza fissa viene marcato secondo la posizione sequenziata. Gli oligonucleotidi vengono appaiati e ligati; la ligazione preferenziale mediante DNA ligasi per appaiamento delle sequenze risulta in un segnale informativo del nucleotide in quella posizione. Prima del sequenziamento, il DNA viene amplificato mediante PCR in emulsione. La sfera risultante, ciascuna contenente solo copie della stessa molecola di DNA, sono depositate su un vetrino. Come secondo esempio, la piattaforma Polonator™ G.007 di Dover Systems (Salem, New Hampshire) impiega anch'essa un approccio di sequenziamento mediante ligazione utilizzando una PCR in emulsione, a base di sfere, disposta casualmente per amplificare frammenti di DNA per il sequenziamento parallelo.

4) Tecnologie di sequenziamento a singola molecola come per esempio attuate nel sistema PacBio RS di Pacific Biosciences (Menlo Park, California) o nella piattaforma HeliScope™ di Helicos Biosciences (Cambridge, Massachusetts). La caratteristica distinta di questa tecnologia è la sua capacità di sequenziare singole molecole di DNA o di RNA senza amplificazione, definita come sequenziamento del DNA a singola molecola in tempo reale (SMRT). Per esempio, HeliScope utilizza un sistema di rilevamento di fluorescenza altamente sensibile per rilevare direttamente ciascun nucleotide quando è sintetizzato. Un approccio simile basato sul trasferimento di energia per risonanza di fluorescenza (FRET) è stato sviluppato da Visigen Biotechnology (Houston, Texas). Altre tecniche a singola molecola basate sulla fluorescenza sono da US Genomics (GeneEngine™) e da Genovox (AnyGene™).

5) Nanotecnologie per il sequenziamento a singola molecola in cui vengono utilizzate varie nanostrutture che sono per esempio disposte su un supporto per monitorare il movimento di una molecola di polimerasi su un singolo filamento durante la replicazione. Esempi non limitanti per approcci basati su nanotecnologie sono la piattaforma GridON™ di Oxford Nanopore Technologies (Oxford, Gran Bretagna), le piattaforme di sequenziamento mediante nanopori assistito da ibridazione (HANS™) sviluppate da Nabsys (Providence, Rhode Island), e la piattaforma di sequenziamento del DNA basata sulla ligasi proprietaria con la tecnologia del DNA nanoball (DNB) chiamata ligazione combinatoriale sonda-ancora (cPAL™).

6) Tecnologie basate sulla microscopia elettronica per il sequenziamento a singola molecola, per esempio quelle sviluppate da LightSpeed

Genomics (Sunnyvale, California) e da Halcyon Molecular (Redwood City, California).

7) Il sequenziamento per semiconduzione ionica che è basato sul rilevamento di ioni idrogeno che vengono rilasciati durante la polimerizzazione del DNA. Per esempio, Ion Torrent Systems (San Francisco, California) utilizza una matrice ad elevata densità di pozzetti microlavorati a macchina per eseguire questo procedimento biochimico in un modo massivamente parallelo. Ciascun pozzetto mantiene un modello differente di DNA. Sotto i pozzetti vi è uno strato sensibile agli ioni e sotto quello un sensore ionico proprietario.

Preferibilmente, le preparazioni di DNA e di RNA servono come materiale di partenza per l'NGS. Tali acidi nucleici possono essere ottenuti facilmente da campioni quali materiale biologico, per esempio da tessuti tumorali freschi, congelati molto velocemente o inclusi in paraffina fissati in formalina (FFPE) o dalle cellule isolate di fresco o dalle CTC che sono presenti nel sangue periferico dei pazienti. Il DNA o l'RNA genomico normale non mutato può essere estratto da tessuto normale, somatico, tuttavia le cellule della linea germinale sono preferite nel contesto della presente invenzione. Il DNA o l'RNA della linea germinale viene estratto dalle cellule mononucleate del sangue periferico (PBMC) in pazienti con tumori ematologici non maligni. Sebbene gli acidi nucleici estratti dai tessuti FFPE o dalle cellule singole isolate di fresco siano altamente frammentati, essi sono adatti per le applicazioni dell'NGS.

Diversi metodi di NGS mirati per il sequenziamento dell'esoma vengono descritti nella letteratura (per una rassegna si veda per esempio Teer e Mullikin 2010: Human Mol Genet 19 (2), da R145 a R151), i quali possono tutti essere utilizzati in combinazione con la presente invenzione. Molti di questi metodi (descritti per esempio come cattura genomica, partizione genomica, arricchimento genomico ecc.) utilizzano tecniche di ibridazione ed includono approcci di ibridazione a base di matrici (per esempio Hodges et al. 2007: Nat. Genet. 39, da 1522 a 1527) ed a base di liquidi (per esempio Choi et al. 2009: Proc. Natl. Acad. Sci USA 106, da 19096 a 19101). Kit commerciali per la preparazione di campioni di DNA e per la successiva cattura dell'esoma sono anch'essi disponibili: per esempio, Illumina Inc. (San Diego, California) offre il kit di preparazione di campioni di DNA TruSeq™ ed il kit di arricchimento esomico TruSeq™.

Allo scopo di ridurre il numero di scoperte false positive nel rilevare mutazioni somatiche specifiche per il cancro o differenze di sequenza confrontando per esempio la sequenza di un campione tumorale con la sequenza di un campione di riferimento quale la sequenza di un campione della linea germinale si preferisce determinare la sequenza in repliche di uno o di entrambi di questi tipi di campioni. Così, si preferisce che la sequenza di un campione di riferimento quale la sequenza di un campione della linea germinale venga determinata due volte, tre volte o più. In alternativa o in aggiunta, la sequenza di un campione tumorale viene determinata due volte, tre volte o più. Può anche essere possibile determinare la sequenza di un campione di riferimento quale la sequenza di un campione della linea germinale e/o la sequenza di un campione tumorale più di

una volta determinando almeno una volta la sequenza in DNA genomico e determinando almeno una volta la sequenza in RNA del suddetto campione di riferimento e/o del suddetto campione tumorale. Per esempio, determinando le variazioni tra repliche di un campione di riferimento quale un campione della linea germinale il tasso atteso di mutazioni somatiche false positive (FDR) come quantità statistica può essere stimato. Le ripetizioni tecniche di un campione dovrebbero generare risultati identici e qualsiasi mutazione rilevata in questo “confronto stesso contro stesso” è un falso positivo. In particolare, per determinare il tasso di scoperta falsa per rilevamento di mutazione somatica in un campione tumorale rispetto ad un campione di riferimento, una ripetizione tecnica del campione di riferimento può essere utilizzata come riferimento per stimare il numero di falsi positivi. Inoltre, varie metriche correlate alla qualità (per esempio qualità di copertura o di SNP) possono essere combinate in un singolo punteggio di qualità utilizzando un approccio di apprendimento con macchine. Per una data variazione somatica possono essere contate tutte le altre variazioni con un punteggio di qualità eccedente, il quale permette una classificazione di tutte le variazioni in una serie di dati.

Secondo l'invenzione, può essere applicato un metodo di genotipizzazione a singola cellula ad ampio genoma a volumi elevati.

In una forma di realizzazione della genotipizzazione a singola cellula ad ampio genoma a volumi elevati la piattaforma Fluidigm può essere utilizzata. Tale approccio può comprendere i seguenti passaggi:

1. tessuto/cellule tumorali del campione e tessuto sano da un dato paziente.
2. Il materiale genetico viene estratto da cellule cancerose e sane e quindi il suo esoma (DNA) è sequenziato utilizzando i protocolli standard di sequenziamento di nuova generazione (NGS). La copertura dell'NGS è tale che possano essere rilevati gli alleli eterozigoti con una frequenza di almeno il 5%. Il trascrittoma (RNA) è anch'esso estratto dalle cellule cancerose, convertito in cDNA e sequenziato per determinare quali geni vengono espressi dalle cellule cancerose.
3. Variazioni di nucleotidi singoli (SNV) espresse non sinonime vengono identificate come qui descritto. I siti che sono SNP nel tessuto sano sono esclusi.
4. Vengono scelte N=96 mutazioni da (3) che coprono frequenze differenti. I saggi di genotipizzazione degli SNP basati sul rilevamento di fluorescenza vengono progettati e sintetizzati per queste mutazioni (esempi di tali saggi includono: saggi per SNP a base di TaqMan da Life Technologies o saggi SNPtype da Fluidigm). I saggi includeranno inneschi specifici di amplificazione del bersaglio (STA) per amplificare gli ampliconi contenenti i dati di SNV (questo è standard nei saggi TaqMan e SNPtype).
5. Le cellule singole saranno isolate dal tumore e da tessuto sano mediante microdissezione laser (LMD) o mediante disaggregazione in sospensioni di singole cellule cui segue la separazione come descritto precedentemente (Dalerba P. et al. (2011) Nature Biotechnology 29: da 1120 a 1127). Le

cellule possono essere scelte senza preselezione (cioè, in modo imparziale), o, in alternativa, le cellule cancerose possono essere arricchite. I metodi di arricchimento includono: marcatura specifica, separazione mediante dimensione cellulare, ispezione istologica durante LMD, e così via.

6. Le cellule singole saranno isolate in provette per PCR contenenti una miscela principale con gli inneschi STA e gli ampliconi contenenti le SNV saranno amplificati. In alternativa il genoma della cellula singola sarà amplificato attraverso l'amplificazione dell'intero genoma (WGA) come descritto precedentemente (Frumkin D. et al. (2008) Cancer Research 68: 5924). La lisi cellulare sarà ottenuta attraverso il passaggio di riscaldamento a 95°C o attraverso un tampone di lisi dedicato.

7. I campioni amplificati con gli STA sono diluiti e caricati sulla matrice di genotipizzazione Fluidigm.

8. I campioni dal tessuto sano saranno utilizzati come controlli positivi per determinare i gruppi di alleli omozigoti (senza mutazione). Poiché i dati di NGS indicano che le mutazioni omozigoti sono estremamente rare, tipicamente solo due gruppi sono attesi: XX e XY, con X=sani.

9. Il numero delle matrici che possono essere eseguite non è limitato, permettendo in pratica fino a circa 1000 cellule singole di essere sottoposte al saggio (circa 10 matrici). Se eseguita in 384 piastre la preparazione dei campioni può essere ridotta ad alcuni giorni.

10. Sono quindi determinate le SNV per ciascuna cellula.

In un'altra forma di realizzazione della genotipizzazione a singola cellula ad ampio genoma a volumi elevati la piattaforma di NGS può essere utilizzata. Tale approccio può comprendere i seguenti passaggi:

1. i passaggi da 1 a 6 sopra sono identici, tranne per il fatto che N (numero di SNV sottoposte a saggio) può essere molto superiore a 96. Nel caso di WGA, diversi cicli di STA saranno eseguiti dopo. Gli inneschi STA conterranno due sequenze segnale universali su ciascun innesco.

2. Dopo l'STA, gli inneschi con codice identificativo saranno amplificati mediante PCR negli ampliconi. Gli inneschi con codice identificativo contengono sequenze uniche di codici identificativi e le suddette sequenze segnale universali. Ciascuna cellula conterrà così un codice identificativo unico.

3. Gli ampliconi da tutte le cellule saranno miscelati e sequenziati attraverso l'NGS. La limitazione pratica sul numero di cellule che possono essere multiplate è il numero di piastre che può essere preparato. Poiché i campioni possono essere preparati in 384 piastre, un limite pratico sarebbe circa 5000 cellule.

4. Sulla base dei dati di sequenze vengono rilevate le SNV (o altre anomalie strutturali) delle cellule singole.

Per prioritizzare gli antigeni, una ricostruzione filogenetica del tumore sulla base della genotipizzazione a singola cellula ("prioritizzazione filogenetica degli antigeni") può essere utilizzata secondo l'invenzione. Oltre alla prioritizzazione degli antigeni sulla base di criteri quali

l'espressione, il tipo di mutazione (non sinonime contro altre), le caratteristiche di legame all'MHC e così via, un'ulteriore dimensione per la prioritizzazione progettata per far fronte all'eterogeneità intratumorale ed intertumorale ed una polarizzazione della biopsia possono essere utilizzate come descritto per l'esempio sotto.

1. Identificare gli antigeni più abbondanti

La frequenza di ciascuna SNV può essere accuratamente stimata sulla base del saggio su singole cellule descritto sopra in relazione al metodo di genotipizzazione a singola cellula ad ampio genoma a volumi elevati e le SNV più abbondanti presenti possono essere scelte per fornire vaccini individualizzati contro il cancro (IVAC).

2. Identificare antigeni basali primari sulla base dell'analisi ad albero radicato

I dati di NGS dai tumori suggeriscono che le mutazioni omozigoti (risultate in entrambi gli alleli) sono eventi rari. Quindi non vi è alcuna necessità di aplotipizzazione ed un albero filogenetico delle mutazioni somatiche del tumore può essere generato dalla serie di dati delle SNV delle singole cellule. La sequenza della linea germinale sarà utilizzata per radicare l'albero. Utilizzando algoritmi per riprodurre sequenze ancestrali le sequenze dei nodi vicino alla radice dell'albero saranno riprodotte. Queste sequenze contengono le prime mutazioni previste esistere nel tumore primario (qui definite come le/gli mutazioni/antigeni basali primarie/i). A causa della bassa probabilità che due mutazioni avvengano sugli stessi alleli nella stessa posizione sul genoma, è previsto che le mutazioni nelle sequenze ancestrali siano fissate nel tumore.

La prioritizzazione degli antigeni basali primari non è equivalente alla prioritizzazione delle mutazioni più frequenti nella biopsia (sebbene si attenda che le mutazioni basali primarie siano tra le più frequenti nella biopsia). La ragione è la seguente: diciamo che due SNV sembrano essere presenti in tutte le cellule derivate da una biopsia (e così hanno la stessa frequenza - 100%), ma una mutazione è basale e l'altra non lo è, allora la mutazione basale dovrebbe essere scelta per l'IVAC. Questo avviene poiché la mutazione basale si presenta probabilmente in tutte le regioni del tumore, mentre quest'ultima mutazione può essere una mutazione più recente che per caso è stata fissata nella regione in cui la biopsia è stata presa. Inoltre, gli antigeni basali esistono probabilmente in tumori metastatici derivati dal tumore primario. Quindi mediante la prioritizzazione degli antigeni basali per l'IVAC si può considerevolmente incrementare la possibilità che l'IVAC sia in grado di eliminare l'intero tumore e non solo una parte del tumore.

Se esistono tumori secondari e questi sono stati anch'essi campionati, un albero evolutivo di tutti i tumori può essere stimato. Questo può migliorare la robustezza dell'albero e permettere il rilevamento di mutazioni basali in tutti i tumori.

3. Identificare gli antigeni che si estendono massimamente per il/i tumore/i

Un altro approccio per ottenere antigeni che coprono massimamente tutti i siti tumorali è prendere diverse biopsie dal tumore. Una strategia sarebbe selezionare antigeni identificati mediante l'analisi di NGS per essere presenti in tutte le biopsie. Per migliorare le probabilità di identificare mutazioni basali, un'analisi filogenetica sulla base di mutazioni di singole cellule da tutte le biopsie può essere eseguita.

Nel caso di metastasi, possono essere ottenute biopsie da tutti i tumori e possono essere scelte mutazioni identificate attraverso l'NGS che sono comuni a tutti i tumori.

4. Utilizzare le CTC per prioritizzare gli antigeni che inibiscono le metastasi

Si ritiene che i tumori metastatici siano derivati da cellule singole. Quindi mediante genotipizzazione delle cellule singole estratte da tumori differenti di un dato paziente in combinazione con la genotipizzazione delle cellule tumorali circolanti (CTC) del paziente, si può ricostruire la storia evolutiva del cancro. L'attesa è di osservare il tumore metastatico che si evolve dal tumore originale attraverso un sottotipo di CTC derivato dal tumore primario.

Sotto (metodo imparziale per identificare, per contare e per sondare geneticamente le CTC) descriviamo un'estensione del suddetto metodo di genotipizzazione a singola cellula ad ampio genoma a volumi elevati per un isolamento ed un'analisi genomica imparziali delle CTC. Utilizzando l'analisi descritta sopra, si può quindi ricostruire un albero filogenetico del tumore di innesco, delle CTC e dei tumori secondari derivanti dalle metastasi (se esistono). Sulla base di questo albero si possono identificare mutazioni (secondarie o scatenanti) che sono avvenute nel momento in cui o strettamente dopo che le CTC si sono prima staccate dal tumore primario. L'attesa è che i genomi delle CTC derivanti dal tumore primario siano evolutivamente più simili ai genomi del tumore primario rispetto ai genomi del tumore secondario. Inoltre, è atteso che i genomi delle CTC derivanti dal tumore primario conterranno mutazioni uniche che sono fissate nei tumori secondari, o che saranno probabilmente fissate se si formeranno tumori secondari in futuro. Queste mutazioni uniche possono essere prioritizzate per l'IVAC per avere come bersaglio (o impedire) le metastasi.

Il vantaggio di prioritizzare le mutazioni delle CTC contro le mutazioni basali primarie è che gli antigeni derivati dalle CTC possono mobilitare i linfociti T specificamente per avere come bersaglio le metastasi, e quindi saranno un braccio indipendente dai linfociti T diretti verso il tumore primario (utilizzando antigeni differenti). Inoltre, se vi sono pochi tumori secondari (o nessun tumore secondario), allora è atteso che la possibilità di un'elusione immunitaria dagli antigeni derivati dalle CTC sia inferiore poiché la probabilità per l'elusione tumorale dovrebbe aumentare con il numero di cellule cancerose recanti il dato antigene.

5. Identificare gli antigeni copresenti sulla stessa cellula (il “cocktail” IVAC)

Si ritiene che il tumore evolva per sopprimere mutazioni a causa della pressione selettiva del sistema immunitario e della terapia. I vaccini contro il cancro diretti verso antigeni multipli che sono copresenti sulla stessa cellula e che sono anche frequenti nel tumore hanno una maggiore possibilità di scavalcare i meccanismi di evasione del tumore e quindi riducono la possibilità di una ricaduta. Tali “vaccini cocktail” sarebbero analoghi alla terapia antiretrovirale di combinazione per i pazienti HIV+. Le mutazioni copresenti possono essere identificate mediante l’analisi filogenetica o ispezionando l’allineamento delle SNV di tutte le cellule.

Inoltre, secondo l’invenzione, un metodo imparziale per identificare, per contare e per sondare geneticamente le CTC può essere utilizzato. Tale approccio può comprendere i seguenti passaggi:

1. ottenere una biopsia di tumore/i e determinare l’atlante delle mutazioni somatiche.
- 2.

Opzione 1: selezionare $N \geq 96$ mutazioni per un’ulteriore indagine sulla base di schemi di prioritizzazione precedentemente stabiliti.

Opzione 2: svolgere un saggio su singole cellule (si veda il metodo descritto sopra di genotipizzazione a singola cellula ad ampio genoma a volumi elevati) seguito da un’analisi filogenetica per selezionare $N \geq 96$ mutazioni basali primarie e possibilmente mutazioni più recenti per massimizzare la diversità. Le mutazioni precedenti sono utili per identificare le CTC (si veda sotto), e queste ultime per generare un’analisi filogenetica (si veda la sezione “Identificare gli antigeni copresenti sulla stessa cellula (il “cocktail” IVAC”)).

3. Ottenere sangue intero dal paziente con il cancro.
4. Lisare i globuli rossi.
5. Rimuovere i globuli bianchi eliminando le cellule CD45+ (per esempio attraverso la separazione, sfere magnetiche coniugate ad un anticorpo anti-CD45, ecc.) per arricchire le CTC.
6. Rimuovere il DNA libero mediante digestione con DNasi. L’origine del DNA libero può essere DNA presente nel sangue o DNA dalle cellule morte.
7. Separare le cellule rimanenti in provette per PCR, svolgere l’STA (sulla base delle mutazioni selezionate) e setacciare su Fluidigm (metodo descritto sopra di genotipizzazione a singola cellula ad ampio genoma a volumi elevati). Le CTC dovrebbero in generale essere positive per SNV multiple.
8. Le cellule identificate come cancerose (=CTC) possono quindi essere ulteriormente analizzate filogeneticamente sulla base del pannello di SNV

setacciate (si veda la sezione “Identificare gli antigeni copresenti sulla stessa cellula (il “cocktail” IVAC”)).

È anche possibile combinare questo metodo con precedenti metodi stabiliti per le CTC isolate. Per esempio, si possono separare le cellule EpCAM+, o le cellule positive per citocheratine (Rao CG. et al. (2005) International journal of oncology 27: 49; Allard WJ. et al. (2004) Clinical Cancer Research 10: da 6897 a 6904). Queste CTC presunte possono quindi essere verificate/profilate sul Fluidigm/NGS per derivare le loro mutazioni.

Questo metodo può essere utilizzato per contare le CTC. Poiché il metodo non si basa su un marcatore particolare, il quale può o può non essere espresso dalle cellule cancerose, ma piuttosto sul profilo di mutazione di mutazioni somatiche cancerose uniche del paziente, questo è un metodo imparziale per rilevare e per enumerare le CTC.

Secondo l’invenzione, può essere utilizzato un approccio riguardante la ricostruzione filogenetica del tumore sulla base della genotipizzazione a singola cellula per arricchire le mutazioni scatenanti (“filtraggio filogenetico”).

In una forma di realizzazione di questo approccio, viene eseguita un’analisi filogenetica pantumorale per recuperare le mutazioni scatenanti.

Per esempio, possono essere rilevate le mutazioni scatenanti da $n=1$ tumori.

Nella sezione sopra “Identificare antigeni basali primari sulla base dell’analisi ad albero radicato” descriviamo un metodo per recuperare sequenze ancestrali e/o per identificare cellule che hanno sequenze vicine alla radice dell’albero. È atteso che il numero di mutazioni in queste sequenze sia significativamente minore rispetto al numero di mutazioni nel campione totale del cancro poiché per definizione queste sono sequenze vicine alla radice dell’albero. Quindi, selezionando sequenze vicine alla radice dell’albero è atteso che molte mutazioni di passaggio siano “filogeneticamente eliminate”.

Questa procedura ha il potenziale di arricchire considerevolmente le mutazioni scatenanti. Le mutazioni scatenanti possono quindi essere utilizzate per identificare/selezionare un trattamento per un paziente o possono essere utilizzate come guida per nuove terapie.

In un altro esempio, le mutazioni scatenanti da $n>1$ tumori di un dato tipo possono essere rilevate.

Ricostruendo le mutazioni basali primarie da molti tumori di un tipo particolare si può considerevolmente incrementare la possibilità di rilevare mutazioni scatenanti. Poiché le sequenze basali vicino alla radice dell’albero eliminano molte mutazioni di passaggio, è atteso che il rapporto segnale/rumore nel rilevare mutazioni scatenanti incrementi considerevolmente. Questo metodo quindi ha il potenziale per rilevare (1) una mutazione scatenante meno frequente (2) mutazioni scatenanti frequenti da meno campioni.

In un’altra forma di realizzazione dell’approccio riguardante la ricostruzione filogenetica del tumore sulla base della genotipizzazione a singola cellula per arricchire le mutazioni scatenanti (“filtraggio filogenetico”), viene eseguita un’analisi filogenetica per recuperare le metastasi che

causano mutazioni scatenanti.

Nella sezione sopra “utilizzare le CTC per prioritizzare gli antigeni che inibiscono le metastasi” descriviamo un metodo di rilevamento di mutazioni associate alle CTC. Questo metodo può anche essere utilizzato per arricchire le mutazioni scatenanti che conducono alle metastasi. Per esempio, mediante una mappatura della filogenia combinata del tumore di innesco, dei tumori secondari e delle CTC, le CTC derivate dal tumore primario dovrebbero connettersi tra i cladi dei tumori primari secondari. Una tale analisi filogenetica può aiutare a determinare con precisione le mutazioni uniche a questa transizione tra tumori di innesco e secondari. Una frazione di queste mutazioni può essere costituita da mutazioni scatenanti. Inoltre, confrontando le mutazioni uniche delle CTC da differenti esempi dello stesso cancro (cioè, $n > 1$ tumore), si può ulteriormente arricchire le uniche mutazioni scatenanti che causano le metastasi.

Secondo l'invenzione, può essere utilizzata l'analisi filogenetica per identificare i tumori primari contro secondari.

Nel caso di metastasi, se tutti i tumori sono campionati, un albero radicato può essere utilizzato per predire l'ordine temporale in cui sono comparsi i tumori: quale tumore è il tumore primario (i nodi più vicini alla radice dell'albero) e quali tumori sono i più recenti. Questo può essere utile in casi in cui è difficile determinare quale tumore è il primario.

Nel contesto della presente invenzione, il termine “RNA” riguarda una molecola che comprende almeno un residuo ribonucleotidico e che è in modo preferibile completamente o sostanzialmente costituita da residui ribonucleotidici. “Ribonucleotide” riguarda un nucleotide con un gruppo idrossile in posizione 2' di un gruppo β -D-ribofuranosile. Il termine “RNA” comprende RNA a doppio filamento, RNA a singolo filamento, RNA isolato quale RNA parzialmente o completamente purificato, RNA essenzialmente puro, RNA sintetico, e RNA generato per via ricombinante quale RNA modificato che si differenzia da RNA presente in natura per l'aggiunta, la delezione, la sostituzione e/o l'alterazione di uno o più nucleotidi. Tali alterazioni possono includere l'aggiunta di materiale non nucleotidico, quale alla/alle estremità di un RNA o internamente, per esempio ad uno o più nucleotidi dell'RNA. I nucleotidi in molecole di RNA possono comprendere anche nucleotidi non standard, quali nucleotidi non comunemente presenti in natura o nucleotidi o deossinucleotidi sintetizzati chimicamente. Questi RNA alterati possono essere indicati come analoghi o analoghi di RNA presente in natura.

Secondo la presente invenzione, il termine “RNA” comprende e riguarda preferibilmente l'“mRNA”. Il termine “mRNA” significa “RNA messaggero” e riguarda un “trascritto” che viene generato utilizzando uno stampo di DNA e codifica per un peptide o per un polipeptide. Tipicamente, un mRNA comprende un 5'-UTR, una regione codificante per una proteina, ed un 3'-UTR. L'mRNA possiede solo un'emivita limitata nelle cellule ed in vitro. Nel contesto della presente invenzione, l'mRNA può essere generato mediante trascrizione in vitro da uno stampo

di DNA. La metodologia di trascrizione in vitro è nota alla persona esperta. Per esempio, vi è una varietà di kit per la trascrizione in vitro disponibile in commercio.

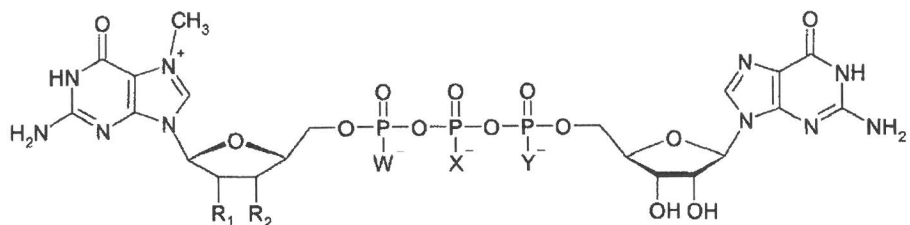
Secondo l'invenzione, la stabilità e l'efficienza della traduzione dell'RNA possono essere modificate come richiesto. Per esempio, l'RNA può essere stabilizzato e la sua traduzione aumentata mediante una o più modifiche aventi effetti stabilizzanti e/o incrementanti l'efficienza della traduzione dell'RNA. Tali modifiche vengono descritte, per esempio, in PCT/EP2006/009448. Allo scopo di incrementare l'espressione dell'RNA utilizzato secondo la presente invenzione, esso può essere modificato all'interno della regione codificante, cioè la sequenza codificante per il peptide o per la proteina espressi, preferibilmente senza alterare la sequenza del peptide o della proteina espressi, in modo tale da incrementare il contenuto in GC per incrementare la stabilità dell'mRNA e per eseguire un'ottimizzazione dei codoni e, così, aumentare la traduzione nelle cellule. Il termine "modifica" nel contesto dell'RNA utilizzato nella presente invenzione comprende qualsiasi modifica di un RNA che non è naturalmente presente nel suddetto RNA.

In una forma di realizzazione dell'invenzione, l'RNA utilizzato secondo l'invenzione non ha trifosfati al 5' senza cappuccio. La rimozione di tali trifosfati al 5' senza cappuccio può essere ottenuta trattando l'RNA con una fosfatasi.

L'RNA secondo l'invenzione può avere ribonucleotidi modificati allo scopo di incrementare la sua stabilità e/o di diminuire la citotossicità. Per esempio, in una forma di realizzazione, nell'RNA utilizzato secondo l'invenzione la 5-metilcitidina è sostituita parzialmente o completamente, preferibilmente completamente, al posto della citidina. In alternativa o in aggiunta, in una forma di realizzazione, nell'RNA utilizzato secondo l'invenzione la pseudouridina è sostituita parzialmente o completamente, preferibilmente completamente, al posto dell'uridina.

In una forma di realizzazione, il termine "modifica" riguarda il fornire all'RNA un cappuccio al 5' o un analogo del cappuccio al 5'. Il termine "cappuccio al 5'" si riferisce ad una struttura di copertura trovata all'estremità al 5' di una molecola di mRNA ed in generale consiste di un nucleotide guanosina connesso all'mRNA attraverso un insolito legame trifosfato da 5' a 5'. In una forma di realizzazione, questa guanosina è metilata in posizione 7. Il termine "cappuccio al 5' convenzionale" si riferisce ad un cappuccio al 5' di RNA presente in natura, preferibilmente al cappuccio di 7-metilguanosina (m⁷G). Nel contesto della presente invenzione, il termine "cappuccio al 5'" comprende un analogo del cappuccio al 5' che assomiglia alla struttura del cappuccio dell'RNA e viene modificato per possedere la capacità di stabilizzare l'RNA e/o di aumentare la traduzione dell'RNA se attaccato a questo, preferibilmente in vivo e/o in una cellula.

Preferibilmente, l'estremità al 5' dell'RNA comprende una struttura di copertura avente la seguente formula generale:



in cui R_1 ed R_2 sono indipendentemente idrossi o metossi e W^- , X^- e Y^- sono indipendentemente ossigeno, zolfo, selenio, o BH_3 . In una forma di realizzazione preferita, R_1 ed R_2 sono idrossi e W^- , X^- e Y^- sono ossigeno. In un'ulteriore forma di realizzazione preferita, uno di R_1 ed R_2 , preferibilmente R_1 è idrossi e l'altro è metossi e W^- , X^- e Y^- sono ossigeno. In un'ulteriore forma di realizzazione preferita, R_1 ed R_2 sono idrossi ed uno di W^- , X^- e Y^- , preferibilmente X^- è zolfo, selenio, o BH_3 , preferibilmente zolfo, mentre gli altri sono ossigeno. In un'ulteriore forma di realizzazione preferita, uno di R_1 ed R_2 , preferibilmente R_2 è idrossi e l'altro è metossi ed uno di W^- , X^- e Y^- , preferibilmente X^- è zolfo, selenio, o BH_3 , preferibilmente zolfo mentre gli altri sono ossigeno.

Nella formula sopra, il nucleotide sul lato destro è connesso alla catena dell'RNA attraverso il suo gruppo 3'.

Quelle strutture a cappuccio in cui almeno uno di W^- , X^- e Y^- è zolfo, cioè che hanno un gruppo funzionale fosforotioato, esistono in differenti forme diastereoisomeriche tutte le quali sono qui comprese. Inoltre, la presente invenzione comprende tutti i tautomeri e gli stereoisomeri della formula sopra.

Per esempio, la struttura a cappuccio avente la struttura sopra in cui R_1 è metossi, R_2 è idrossi, X^- è zolfo e W^- e Y^- sono ossigeno esiste in due forme diastereoisomeriche (R_p e S_p). Queste possono essere risolte mediante HPLC a fase inversa e vengono chiamate D1 e D2 secondo il loro ordine di eluizione dalla colonna dell'HPLC a fase inversa. Secondo l'invenzione, l'isomero D1 di $m_2^{7,2'-O}GppspG$ è particolarmente preferito.

Il fornire un RNA con un cappuccio al 5' o con un analogo del cappuccio al 5' può essere ottenuto mediante trascrizione in vitro di uno stampo di DNA in presenza del suddetto cappuccio al 5' o analogo del cappuccio al 5', in cui il suddetto cappuccio al 5' è cotrascrizionalmente incorporato nel filamento di RNA generato, o l'RNA può essere generato, per esempio, mediante trascrizione in vitro, ed il cappuccio al 5' può essere attaccato all'RNA post-trascrizionalmente utilizzando enzimi di incapsulazione, per esempio, enzimi di incapsulazione di virus del vaiolo vaccino.

L'RNA può comprendere ulteriori modifiche. Per esempio, un'ulteriore modifica dell'RNA utilizzato nella presente invenzione può essere un'estensione o un troncamento della coda di poli(A) presente in natura o un'alterazione delle regioni non tradotte al 5' o al 3' (UTR) quale un'introduzione di un'UTR che non è correlata alla regione codificante del suddetto RNA, per esempio, lo scambio dell'UTR al 3' esistente con o l'inserimento di una o più, preferibilmente due copie di una 3'-UTR derivata da un gene della globina, quale alfa2-globina, alfa1-globina, beta-

globina, preferibilmente beta-globina, più preferibilmente beta-globina umana.

L'RNA avente una sequenza di poliA non mascherata è tradotto più efficientemente rispetto all'RNA avente una sequenza di poliA mascherata. Il termine "coda di poli(A)" o "sequenza di poliA" riguarda una sequenza di residui adenil (A) che tipicamente è posizionata all'estremità al 3' di una molecola di RNA e "sequenza di poliA non mascherata" significa che la sequenza di poliA all'estremità al 3' di una molecola di RNA termina con un'A della sequenza di poliA e non è seguita da nucleotidi diversi dall'A localizzata all'estremità al 3', cioè a valle, della sequenza di poliA. Inoltre, una lunga sequenza di poliA di circa 120 coppie di basi risulta in una stabilità e in un'efficienza della traduzione ottimali dei trascritti di RNA.

Quindi, allo scopo di incrementare la stabilità e/o l'espressione dell'RNA utilizzato secondo la presente invenzione, esso può essere modificato in modo tale da essere presente in combinazione con una sequenza di poliA, avente preferibilmente una lunghezza da 10 a 500, più preferibilmente da 30 a 300, ancora più preferibilmente da 65 a 200 e specialmente da 100 a 150 residui di adenosina. In una forma di realizzazione specialmente preferita la sequenza di poliA ha una lunghezza di approssimativamente 120 residui di adenosina. Per incrementare ulteriormente la stabilità e/o l'espressione dell'RNA utilizzato secondo l'invenzione, la sequenza di poliA può essere non mascherata.

Inoltre, l'incorporazione di una regione non tradotta al 3' (UTR) nella regione non tradotta al 3' di una molecola di RNA può portare ad un potenziamento dell'efficienza della traduzione. Un effetto sinergico può essere ottenuto incorporando due o più di tali regioni non tradotte al 3'. Le regioni non tradotte al 3' possono essere autologhe o eterologhe all'RNA in cui esse vengono introdotte. In una forma di realizzazione particolare la regione non tradotta al 3' è derivata dal gene della β -globina umana.

Una combinazione delle modifiche sopra, cioè l'incorporazione di una sequenza di poliA, lo smascheramento di una sequenza di poliA e l'incorporazione di una o più regioni non tradotte al 3', ha un'influenza sinergica sulla stabilità dell'RNA e sull'incremento dell'efficienza della traduzione.

Il termine "stabilità" dell'RNA riguarda l'"emivita" dell'RNA. L'"emivita" riguarda il periodo di tempo che è richiesto per eliminare metà dell'attività, della quantità, o del numero di molecole. Nel contesto della presente invenzione, l'emivita di un RNA è indicativa della stabilità del suddetto RNA. L'emivita di RNA può influenzare la "durata dell'espressione" dell'RNA. Ci si può attendere che l'RNA avente un'emivita lunga sarà espresso per un periodo di tempo prolungato.

Ovviamente, se secondo la presente invenzione si desidera diminuire la stabilità e/o l'efficienza della traduzione dell'RNA, è possibile modificare l'RNA in modo tale da interferire con la funzione di elementi come descritto sopra aumentando la stabilità e/o l'efficienza della traduzione

dell'RNA.

Il termine "espressione" viene utilizzato secondo l'invenzione nel suo significato più generale e comprende la produzione di RNA e/o di peptidi o di polipeptidi, per esempio mediante trascrizione e/o traduzione. In riferimento all'RNA, il termine "espressione" o "traduzione" riguarda in particolare la produzione di peptidi o di polipeptidi. Esso comprende anche l'espressione parziale di acidi nucleici. Inoltre, l'espressione può essere transitoria o stabile.

Secondo l'invenzione, il termine espressione comprende inoltre un'"espressione aberrante" o un'"espressione anormale". "Espressione aberrante" o "espressione anormale" significa secondo l'invenzione che l'espressione è alterata, preferibilmente aumentata, in confronto a un riferimento, per esempio uno stato in un soggetto non avente una malattia associata all'espressione aberrante o anormale di una certa proteina, per esempio, un antigene tumorale. Un aumento dell'espressione si riferisce ad un aumento di almeno il 10%, in particolare almeno il 20%, almeno il 50% o almeno il 100%, o superiore. In una forma di realizzazione, l'espressione si trova solo in un tessuto malato, mentre l'espressione in un tessuto sano è repressa.

Il termine "specificamente espressa" significa che una proteina è espressa essenzialmente solo in un tessuto o un organo specifico. Per esempio, un antigene tumorale espresso specificamente nella mucosa gastrica significa che la suddetta proteina è principalmente espressa nella mucosa gastrica e non viene espressa in altri tessuti o non viene espressa ad un grado significativo in altri tipi di tessuti o di organi. Così, una proteina che è esclusivamente espressa nelle cellule della mucosa gastrica e ad un grado significativamente minore in qualsiasi altro tessuto, quale i testicoli, è specificamente espressa nelle cellule della mucosa gastrica. In alcune forme di realizzazione, un antigene tumorale può anche essere specificamente espresso in condizioni normali in più di un tipo di tessuto o di organo, quale in 2 o 3 tipi di tessuto o di organi, ma preferibilmente in non più di 3 differenti tipi di tessuti o di organi. In questo caso, l'antigene tumorale è quindi specificamente espresso in questi organi. Per esempio, se un antigene tumorale viene espresso in condizioni normali preferibilmente ad un grado approssimativamente uguale nel polmone e nello stomaco, il suddetto antigene tumorale è specificamente espresso nel polmone e nello stomaco.

Nel contesto della presente invenzione, il termine "trascrizione" riguarda un procedimento, in cui il codice genetico in una sequenza di DNA è trascritto nell'RNA. Successivamente, l'RNA può essere tradotto in proteina. Secondo la presente invenzione, il termine "trascrizione" comprende la "trascrizione in vitro", in cui il termine "trascrizione in vitro" riguarda un procedimento in cui l'RNA, in particolare mRNA, è sintetizzato in vitro in un sistema privo di cellule, utilizzando preferibilmente appropriati estratti cellulari. Preferibilmente, vettori di clonaggio vengono applicati per la generazione di trascritti. Questi vettori di clonaggio sono in generale indicati come vettori di trascrizione e sono secondo la presente

invenzione compresi nel termine "vettore". Secondo la presente invenzione, l'RNA utilizzato nella presente invenzione è preferibilmente RNA trascritto in vitro (IVT-RNA) e può essere ottenuto mediante trascrizione in vitro di uno stampo appropriato di DNA. Il promotore per controllare la trascrizione può essere qualsiasi promotore per qualsiasi RNA polimerasi. Esempi particolari di RNA polimerasi sono le RNA polimerasi di T7, di T3 e di SP6. Preferibilmente, la trascrizione in vitro secondo l'invenzione viene controllata da un promotore di T7 o di SP6. Uno stampo di DNA per la trascrizione in vitro può essere ottenuto clonando un acido nucleico, in particolare cDNA, ed introducendolo in un vettore appropriato per la trascrizione in vitro. Il cDNA può essere ottenuto mediante la trascrizione inversa di RNA.

Il termine "traduzione" secondo l'invenzione riguarda un procedimento nei ribosomi di una cellula mediante il quale un filamento di RNA messaggero dirige l'assemblaggio di una sequenza amminoacidica per fare un peptide o un polipeptide.

Sequenze di controllo dell'espressione o sequenze di regolazione, le quali secondo l'invenzione possono essere legate funzionalmente con un acido nucleico, possono essere omologhe o eterologhe rispetto all'acido nucleico. Una sequenza codificante ed una sequenza di regolazione sono legate insieme "funzionalmente" se sono legate insieme covalentemente, così che la trascrizione o la traduzione della sequenza codificante sia sotto il controllo o sotto l'influenza della sequenza di regolazione. Se la sequenza codificante deve essere tradotta in una proteina funzionale, con un legame funzionale di una sequenza di regolazione con la sequenza codificante, l'induzione della sequenza di regolazione conduce ad una trascrizione della sequenza codificante, senza causare uno spostamento della fase di lettura nella sequenza codificante o un'incapacità della sequenza codificante di essere tradotta nella proteina o nel peptide desiderati.

Il termine "sequenza di controllo dell'espressione" o "sequenza di regolazione" comprende, secondo l'invenzione, promotori, sequenze di legame al ribosoma ed altri elementi di controllo, i quali controllano la trascrizione di un acido nucleico o la traduzione dell'RNA derivato. In alcune forme di realizzazione dell'invenzione, le sequenze di regolazione possono essere controllate. La struttura precisa delle sequenze di regolazione può variare in relazione alla specie o in relazione al tipo cellulare, ma comprende in generale sequenze non trascritte al 5' e non tradotte al 5' e al 3', le quali sono coinvolte nell'inizio della trascrizione o della traduzione, quali TATA box, sequenza di incapsulazione, sequenza CAAT e simili. In particolare, sequenze di regolazione non trascritte al 5' comprendono una regione del promotore che comprende una sequenza del promotore per il controllo trascrizionale del gene funzionalmente legato. Le sequenze di regolazione possono comprendere anche sequenze potenziatrici o sequenze attivatrici a monte.

Preferibilmente, secondo l'invenzione, l'RNA da esprimere in una cellula viene introdotto nella suddetta cellula. In una forma di realizzazione dei metodi secondo l'invenzione, l'RNA che deve essere introdotto in una cellula viene ottenuto mediante trascrizione in vitro di uno stampo

appropriato di DNA.

Secondo l'invenzione, termini quali "RNA in grado di esprimere" e "RNA codificante" vengono qui utilizzati in modo intercambiabile ed in relazione ad un particolare peptide o polipeptide significando che l'RNA, se presente nell'ambiente appropriato, preferibilmente all'interno di una cellula, può essere espresso per produrre il suddetto peptide o polipeptide. Preferibilmente, l'RNA secondo l'invenzione è in grado di interagire con il meccanismo cellulare di traduzione a dare il peptide o il polipeptide che è in grado di esprimere.

Termini quali "trasferire", "introdurre" o "trasfettare" vengono qui utilizzati in modo intercambiabile e riguardano l'introduzione di acidi nucleici, in particolare acidi nucleici esogeni o eterologhi, in particolare RNA in una cellula. Secondo la presente invenzione, la cellula può far parte di un organo, di un tessuto e/o di un organismo. Secondo la presente invenzione, la somministrazione di un acido nucleico è ottenuta come acido nucleico nudo o in combinazione con un reagente di somministrazione. Preferibilmente, la somministrazione di acidi nucleici è in forma di acidi nucleici nudi. Preferibilmente, l'RNA viene somministrato in combinazione con sostanze stabilizzanti quali inibitori di RNasi. La presente invenzione prevede anche l'introduzione ripetuta di acidi nucleici nelle cellule per permettere l'espressione prolungata per periodi prolungati di tempo.

Le cellule possono essere trasfettate con qualsiasi veicolante con cui l'RNA può essere associato, per esempio formando complessi con l'RNA o formando vescicole in cui l'RNA è incluso o incapsulato, risultando in una stabilità aumentata dell'RNA in confronto all'RNA nudo. Veicolanti utili secondo l'invenzione includono, per esempio, veicolanti contenenti lipidi quali lipidi cationici, liposomi, in particolare liposomi cationici, e micelle, e nanoparticelle. I lipidi cationici possono formare complessi con acidi nucleici caricati negativamente. Qualsiasi lipide cationico può essere utilizzato secondo l'invenzione.

Preferibilmente, l'introduzione di RNA che codifica per un peptide o per un polipeptide in una cellula, in particolare in una cellula presente in vivo, porta all'espressione del suddetto peptide o polipeptide nella cellula. In forme di realizzazione particolari, si preferisce l'indirizzamento a bersaglio degli acidi nucleici a cellule particolari. In tali forme di realizzazione, un veicolante che viene applicato per la somministrazione dell'acido nucleico ad una cellula (per esempio un retrovirus o un liposoma), mostra una molecola di indirizzamento al bersaglio. Per esempio, una molecola quale un anticorpo che è specifico per una proteina superficiale di membrana sulla cellula bersaglio o un ligando per un recettore sulla cellula bersaglio può essere incorporato nel veicolante dell'acido nucleico o può essere legato a questo. Nel caso in cui l'acido nucleico venga somministrato mediante liposomi, proteine che si legano ad una proteina superficiale di membrana che è associata all'endocitosi possono essere incorporate nella formulazione liposomiale allo scopo di permettere l'indirizzamento a bersaglio e/o l'assorbimento. Tali proteine comprendono proteine del capsido o loro frammenti che sono specifici per un tipo cellulare particolare, anticorpi contro proteine che sono internalizzate, proteine che sono dirette

verso una posizione intracellulare ecc.

Secondo la presente invenzione, il termine "peptide" si riferisce a sostanze comprendenti due o più, preferibilmente 3 o più, preferibilmente 4 o più, preferibilmente 6 o più, preferibilmente 8 o più, preferibilmente 10 o più, preferibilmente 13 o più, preferibilmente 16 o più, preferibilmente 21 o più e fino a preferibilmente 8, 10, 20, 30, 40 o 50, in particolare 100 amminoacidi uniti covalentemente mediante legami peptidici. Il termine "polipeptide" o "proteina" si riferisce a peptidi grandi, preferibilmente a peptidi con più di 100 residui amminoacidici, ma in generale i termini "peptide", "polipeptide" e "proteina" sono sinonimi e vengono qui utilizzati in modo intercambiabile.

Secondo l'invenzione, il termine "cambiamento di sequenza" in relazione a peptidi oppure a proteine riguarda varianti per inserzione amminoacidica, varianti per aggiunta amminoacidica, varianti per delezione amminoacidica e varianti per sostituzione amminoacidica, preferibilmente varianti per sostituzione amminoacidica. Tutte queste variazioni di sequenza secondo l'invenzione possono potenzialmente creare nuovi epitopi.

Le varianti per inserzione amminoacidica comprendono inserzioni di singoli o due o più amminoacidi in una particolare sequenza amminoacidica.

Le varianti per aggiunta amminoacidica comprendono fusioni aminoterminali e/o carbossiterninali di uno o più amminoacidi, come 1, 2, 3, 4 o 5, o più amminoacidi.

Le varianti per delezione amminoacidica sono caratterizzate dalla rimozione di uno o più amminoacidi dalla sequenza, come mediante rimozione di 1, 2, 3, 4 o 5, o più amminoacidi.

Le varianti per sostituzione amminoacidica sono caratterizzate da almeno un residuo nella sequenza che viene rimosso e da un altro residuo che viene inserito al suo posto.

Il termine "derivato" significa secondo l'invenzione che un'entità particolare, in particolare una sequenza particolare, è presente nell'oggetto da cui è derivato, in particolare un organismo o una molecola. Nel caso di sequenze amminoacidiche, specialmente regioni di sequenza specifiche, "derivato" in particolare significa che la sequenza amminoacidica rilevante è derivata da una sequenza amminoacidica in cui è presente.

Il termine "cellula" o "cellula ospite" è preferibilmente una cellula intatta, cioè una cellula con una membrana intatta che non ha rilasciato i suoi normali componenti intracellulari quali enzimi, organelli, o materiale genetico. Una cellula intatta è preferibilmente una cellula vitale, cioè una cellula vivente in grado di eseguire le sue normali funzioni metaboliche. Preferibilmente il suddetto termine riguarda secondo l'invenzione qualsiasi cellula che può essere trasformata o trasfettata con un acido nucleico esogeno. Il termine "cellula" comprende secondo l'invenzione cellule procariotiche (per esempio E. coli) o cellule eucariotiche (per esempio cellule dendritiche, linfociti B, cellule di CHO, cellule COS, cellule K562,

cellule HEK293, cellule HELA, cellule di lievito, e cellule di insetto). L'acido nucleico esogeno può trovarsi all'interno della cellula (i) liberamente disperso come tale, (ii) incorporato in un vettore ricombinante, o (iii) integrato nel genoma della cellula ospite o nel DNA mitocondriale. Le cellule di mammifero sono particolarmente preferite, quali cellule dall'uomo, dai topi, dai criceti, dai maiali, dalle capre, e dai primati. Le cellule possono essere derivate da un numero elevato di tipi di tessuto ed includono cellule e linee cellulari primarie. Esempi specifici includono cheratinociti, leucociti del sangue periferico, cellule staminali di midollo osseo, e cellule staminali embrionali. In ulteriori forme di realizzazione, la cellula è una cellula presentante l'antigene, in particolare una cellula dendritica, un monocita, o un macrofago.

Una cellula che comprende una molecola di acido nucleico esprime preferibilmente il peptide o il polipeptide codificato dall'acido nucleico.

Il termine "espansione clonale" si riferisce ad un procedimento in cui un'entità specifica viene moltiplicata. Nel contesto della presente invenzione, il termine viene utilizzato preferibilmente nel contesto di una risposta immunologica in cui i linfociti sono stimolati mediante un antigene, proliferano, e lo specifico linfocita riconoscendo il suddetto antigene viene amplificato. Preferibilmente, l'espansione clonale conduce alla differenziazione dei linfociti.

Termini quali "ridurre" o "inibire" riguardano la capacità di causare una diminuzione complessiva, preferibilmente del 5% o più, del 10% o più, del 20% o più, più preferibilmente del 50% o più, ed il più preferibilmente del 75% o più, nel livello. Il termine "inibire" o frasi simili comprende un'inibizione completa o essenzialmente completa, cioè una riduzione a zero o essenzialmente a zero.

Termini quali "incrementare", "potenziare", "promuovere" o "prolungare" riguardano preferibilmente un aumento, un potenziamento, una promozione o un prolungamento di circa almeno il 10%, preferibilmente almeno il 20%, preferibilmente almeno il 30%, preferibilmente almeno il 40%, preferibilmente almeno il 50%, preferibilmente almeno l'80%, preferibilmente almeno il 100%, preferibilmente almeno il 200% ed in particolare almeno il 300%. Questi termini possono riguardare anche un aumento, un potenziamento, una promozione o un prolungamento da zero o da un livello non misurabile o non rilevabile ad un livello maggiore di zero o ad un livello che è misurabile o rilevabile.

Gli agenti, le composizioni ed i metodi qui descritti possono essere utilizzati per trattare un soggetto con una malattia, per esempio, una malattia caratterizzata dalla presenza di cellule malate esprimenti un antigene e presentanti un peptide antigenico. In particolare, malattie preferite sono le malattie cancerose. Gli agenti, le composizioni ed i metodi qui descritti possono anche essere utilizzati per l'immunizzazione o per la vaccinazione per prevenire una malattia qui descritta.

Secondo l'invenzione, il termine "malattia" si riferisce a qualsiasi stato patologico, comprese le malattie cancerose, in particolare quelle forme di malattie cancerose qui descritte.

Il termine "normale" si riferisce allo stato sano o alle condizioni in un soggetto o tessuto sano, cioè, condizioni non patologiche, in cui "sano" significa preferibilmente non canceroso.

"Malattia riguardante le cellule esprimenti un antigene" significa secondo l'invenzione che l'espressione dell'antigene nelle cellule di un tessuto o di un organo malato è rilevata. L'espressione nelle cellule di un tessuto o di un organo malato può essere aumentata rispetto allo stato in un tessuto o in un organo sano. Un aumento si riferisce ad un aumento di almeno il 10%, in particolare di almeno il 20%, almeno il 50%, almeno il 100%, almeno il 200%, almeno il 500%, almeno il 1000%, almeno il 10000% o anche superiore. In una forma di realizzazione, l'espressione si trova solo in un tessuto malato, mentre l'espressione in un tessuto sano è repressa. Secondo l'invenzione, le malattie riguardanti o che sono associate a cellule esprimenti un antigene includono le malattie cancerose.

Il cancro (termine medico: neoplasma maligno) è una classe di malattie in cui un gruppo di cellule mostra crescita incontrollata (divisione oltre i limiti normali), invasione (intrusione e distruzione di tessuti adiacenti), e qualche volta metastasi (diffusione ad altre posizioni nell'organismo attraverso la linfa o il sangue). Queste tre proprietà maligne dei tumori li differenziano dai tumori benigni, i quali sono autolimitati, e non invadono o metastatizzano. La maggior parte dei tumori forma un tumore ma alcuni, come la leucemia, non lo fanno.

Tumore maligno è essenzialmente sinonimo di cancro. Entità maligna, neoplasma maligno, e tumore maligno sono essenzialmente sinonimi di cancro.

Secondo l'invenzione, il termine "tumore" o "malattia di tipo tumorale" si riferisce ad una crescita anormale delle cellule (chiamate cellule neoplastiche, cellule tumorigeniche o cellule tumorali) preferibilmente formante un rigonfiamento o una lesione. Con "cellula tumorale" si intende una cellula anormale che cresce per mezzo di una rapida, incontrollata proliferazione cellulare e continua a crescere dopo che gli stimoli che hanno dato inizio alla nuova crescita cessano. I tumori mostrano una mancanza parziale o completa di organizzazione strutturale e di coordinazione funzionale con il tessuto normale, e solitamente formano una distinta massa di tessuto, la quale può essere benigna, premaligna o maligna.

Un tumore benigno è un tumore che è privo di tutte e tre le proprietà maligne di un cancro. Così, per definizione, un tumore benigno non cresce in modo illimitato, aggressivo, non invade i tessuti circostanti, e non si diffonde ai tessuti non adiacenti (metastatizzare).

Il neoplasma è una massa anormale di tessuto come risultato di neoplasia. La neoplasia (nuova crescita in greco) è la proliferazione anormale di cellule. La crescita delle cellule supera, e non è coordinata con quella dei tessuti normali intorno ad essa. La crescita persiste nello stesso modo eccessivo anche dopo la cessazione degli stimoli. Essa solitamente causa un grumo o un tumore. I neoplasmi possono essere benigni, premaligni o maligni.

“Crescita di un tumore” o “crescita tumorale” secondo l’invenzione riguarda la tendenza di un tumore ad incrementare la sua dimensione e/o la tendenza di cellule tumorali a proliferare.

Per gli scopi della presente invenzione, i termini “cancro” e “malattia oncologica” vengono utilizzati in modo intercambiabile ai termini “tumore” e “malattia di tipo tumorale”.

I tumori sono classificati per il tipo di cellula che assomiglia al tumore e, quindi, il tessuto supposto essere l’origine del tumore. Questi sono l’istologia e la posizione, rispettivamente.

Il termine “cancro” secondo l’invenzione comprende leucemie, seminomi, melanomi, teratomi, linfomi, neuroblastomi, gliomi, cancro rettale, cancro dell’endometrio, cancro del rene, cancro del surrene, cancro della tiroide, cancro del sangue, cancro della pelle, cancro del cervello, cancro della cervice, cancro dell’intestino, cancro del fegato, cancro del colon, cancro dello stomaco, cancro dell’intestino, cancro della testa e del collo, cancro dell’apparato gastrointestinale, cancro dei linfonodi, cancro dell’esofago, cancro del colon-retto, cancro del pancreas, cancro dell’orecchio, del naso e della gola (ENT), cancro della mammella, cancro della prostata, cancro dell’utero, cancro dell’ovaio e cancro dei polmoni e loro metastasi. Loro esempi sono carcinomi dei polmoni, carcinomi della mammella, carcinomi della prostata, carcinomi del colon, carcinomi a cellule renali, carcinomi della cervice, o metastasi dei tipi di cancro o di tumore descritti sopra. Il termine cancro secondo l’invenzione comprende anche metastasi tumorali e ricaduta del cancro.

I principali tipi di cancro del polmone sono carcinoma polmonare a piccole cellule (CPPC) e carcinoma polmonare non a piccole cellule (CPNPC).

Vi sono tre principali sottotipi dei carcinomi del polmone non a piccole cellule: carcinoma polmonare a cellule squamose, adenocarcinoma, e carcinoma polmonare a grandi cellule. Gli adenocarcinomi spiegano approssimativamente il 10% dei tumori polmonari. Questo cancro solitamente è visto perifericamente nei polmoni, rispetto al cancro del polmone a piccole cellule e al cancro del polmone a cellule squamose, i quali entrambi tendono ad essere localizzati più centralmente.

Il cancro della pelle è una crescita maligna sulla pelle. I tumori della pelle più comuni sono il cancro a cellule basali, il cancro a cellule squamose, ed il melanoma. Il melanoma maligno è un serio tipo di cancro della pelle. È dovuto alla crescita incontrollata delle cellule del pigmento, chiamate melanociti.

Secondo l’invenzione, un “carcinoma” è un tumore maligno derivato dalle cellule epiteliali. Questo gruppo rappresenta i tumori più comuni, comprese le forme comuni di tumore della mammella, della prostata, del polmone e del colon.

Il “carcinoma bronchiale” è un carcinoma del polmone, che si pensa derivi dall’epitelio dei bronchioli terminali, in cui il tessuto neoplastico si

estende lungo le pareti alveolari e cresce in piccole masse all'interno degli alveoli. La mucina può essere dimostrata in alcune delle cellule e nel materiale negli alveoli, il quale comprende anche cellule denudate.

L'“adenocarcinoma” è un cancro che ha origine nel tessuto ghiandolare. Questo tessuto è anche parte di una categoria più ampia di tessuto noto come tessuto epiteliale. Il tessuto epiteliale comprende pelle, ghiandole ed una varietà di altro tessuto che riveste la cavità e gli organi del corpo. L'epitelio è derivato embriologicamente dall'ectoderma, dall'endoderma e dal mesoderma. Per essere classificate come adenocarcinoma, le cellule non devono essere necessariamente parte di una ghiandola, a condizione che esse abbiano proprietà secretorie. Questa forma di carcinoma può presentarsi in alcuni mammiferi superiori, compreso l'uomo. Adenocarcinomi ben differenziati tendono ad assomigliare al tessuto ghiandolare da cui sono derivati, mentre quelli scarsamente differenziati possono non farlo. Mediante la marcatura delle cellule da una biopsia, un patologo determinerà se il tumore è un adenocarcinoma o qualche altro tipo di cancro. Gli adenocarcinomi possono presentarsi in molti tessuti del corpo a causa della natura ubiquitaria delle ghiandole all'interno dell'organismo. Anche se ciascuna ghiandola può non secernere la stessa sostanza, finché vi è una funzione esocrina della cellula, è considerata ghiandolare e la sua forma maligna è quindi chiamata adenocarcinoma. Gli adenocarcinomi maligni invadono altri tessuti e spesso metastatizzano dato abbastanza tempo per farlo. L'adenocarcinoma dell'ovaio è il tipo più comune di carcinoma dell'ovaio. Esso comprende gli adenocarcinomi sierosi e mucinosi, l'adenocarcinoma a cellule chiare e l'adenocarcinoma dell'endometrio.

Il carcinoma a cellule renali noto anche come cancro a cellule renali o adenocarcinoma a cellule renali è un cancro del rene che ha origine nel rivestimento interno del tubulo convoluto prossimale, i condotti molto piccoli nel rene che filtrano il sangue e rimuovono i prodotti di scarto. Il carcinoma a cellule renali è di gran lunga il tipo più comune di cancro del rene negli adulti ed il più letale di tutti i tumori dell'apparato genitourinario. Sottotipi distinti di carcinoma a cellule renali sono il carcinoma a cellule renali a cellule chiare ed il carcinoma papillare a cellule renali. Il carcinoma a cellule renali a cellule chiare è la forma più comune di carcinoma a cellule renali. Quando viste al microscopio, le cellule che costituiscono il carcinoma a cellule renali a cellule chiare appaiono molto chiare o trasparenti. Il carcinoma papillare a cellule renali è il secondo sottotipo più comune. Questi tumori formano piccole proiezioni tipo dito (chiamate papille) in alcuni, se non nella maggior parte, dei tumori.

Il linfoma e la leucemia sono tumori maligni derivati da cellule ematopoietiche (che formano il sangue).

Il tumore blastico o blastoma è un tumore (solitamente maligno) che somiglia ad un tessuto immaturo o embrionale. molti di questi tumori sono maggiormente comuni nei bambini.

Con “metastasi” si intende la diffusione di cellule cancerose dal loro sito originale ad un'altra parte del corpo. La formazione di metastasi è un

procedimento molto complesso e dipende dal distacco di cellule maligne dal tumore primario, dall'invasione della matrice extracellulare, dalla penetrazione delle membrane basali endoteliali per entrare nella cavità e nei vasi corporei, e quindi, dopo essere stato trasportato mediante il sangue, dall'infiltrazione degli organi bersaglio. Infine, la crescita di un nuovo tumore, cioè un tumore secondario o un tumore metastatico, al sito bersaglio dipende dall'angiogenesi. Le metastasi tumorali spesso avvengono anche dopo la rimozione del tumore primario poiché cellule o componenti tumorali possono rimanere e sviluppare un potenziale metastatico. In una forma di realizzazione, il termine "metastasi" secondo l'invenzione riguarda "metastasi a distanza" che riguarda una metastasi che è distante dal tumore primario e dal sistema linfonodale regionale. Le cellule di un tumore secondario o metastatico sono come quelle nel tumore originale. Questo significa, per esempio, che, se il cancro dell'ovaio metastatizza al fegato, il tumore secondario è costituito da cellule ovariche anormali, non da cellule epatiche anormali. Il tumore nel fegato è quindi chiamato cancro metastatico dell'ovaio, non cancro del fegato.

Nel cancro dell'ovaio, le metastasi possono avvenire nei seguenti modi: mediante contatto o estensione diretti, esso può invadere il tessuto nelle vicinanze o gli organi localizzati vicino o intorno all'ovaio, quali le tube di Falloppio, l'utero, la vescica, il retto, ecc.; attecchendo o distaccandosi nella cavità addominale, il quale è il modo più comune attraverso il quale il cancro dell'ovaio si diffonde. Le cellule cancerose rompono la superficie della massa ovarica e "cadono" su altre strutture nell'addome quali il fegato, lo stomaco, il colon o il diaframma; liberandosi dalla massa ovarica, invadendo i vasi linfatici e quindi viaggiando ad altre aree del corpo oppure ad organi distanti quali il polmone o il fegato; liberandosi dalla massa ovarica, invadendo il sistema sanguigno e viaggiando ad altre aree del corpo oppure ad organi distanti.

Secondo l'invenzione, il cancro metastatico dell'ovaio comprende il cancro nelle tube di Falloppio, il cancro in organi dell'addome quale il cancro nell'intestino, il cancro nell'utero, il cancro nella vescica, il cancro nel retto, il cancro nel fegato, il cancro nello stomaco, il cancro nel colon, il cancro nel diaframma, il cancro nei polmoni, il cancro nel rivestimento interno dell'addome o delle pelvi (peritoneo), ed il cancro nel cervello. In modo simile, cancro metastatico del polmone si riferisce ad un cancro che si è diffuso dai polmoni a siti distanti e/o diversi nell'organismo e comprende il cancro nel fegato, il cancro nelle ghiandole surrenali, il cancro nelle ossa, ed il cancro nel cervello.

Il termine "cellule tumorali circolanti" o "CTC" riguarda cellule che si sono staccate da un tumore primario o da metastasi tumorali e circolano nel circolo sanguigno. Le CTC possono costituire nuclei per la successiva crescita di ulteriori tumori (metastasi) in tessuti differenti. Le cellule tumorali circolanti si trovano in frequenze nell'ordine da 1 a 10 CTC per ml di sangue intero in pazienti con malattia metastatica. Sono stati sviluppati metodi di ricerca per isolare le CTC. Diversi metodi di ricerca sono stati descritti nell'arte per isolare le CTC, per esempio tecniche che utilizzano il fatto che le cellule epiteliali esprimono comunemente la proteina di adesione cellulare EpCAM, la quale è assente nelle cellule normali del sangue.

La cattura basata su sfere immunomagnetiche implica il trattamento di campioni di sangue con l'anticorpo per EpCAM che è stato coniugato con particelle magnetiche, cui segue la separazione delle cellule marcate in un campo magnetico. Le cellule isolate vengono quindi marcate con l'anticorpo per un altro marcatore epiteliale, la citocheratina, così come un comune marcatore leucocitario CD45, in modo tale da distinguere le rare CTC dai globuli bianchi contaminanti. Questo approccio robusto e semiautomatizzato identifica le CTC con una resa media di approssimativamente 1 CTC/ml ed una purezza dello 0,1% (Allard et al., 2004: Clin. Cancer Res. 10, da 6897 a 6904). Un secondo metodo per isolare le CTC utilizza un dispositivo di cattura delle CTC a base microfluidica che implica il far fluire il sangue intero attraverso una camera rivestita con 80.000 microposizioni che sono state rese funzionali da un rivestimento con l'anticorpo per EpCAM. Le CTC sono quindi marcate con anticorpi secondari contro la citocheratina o con marcatori specifici per un tessuto, quali PSA nel cancro della prostata o HER2 nel tumore della mammella e sono visualizzate mediante scansione automatizzata di microposizioni in piani multipli lungo coordinate tridimensionali. Le piastrine di CTC sono in grado di identificare cellule tumorali circolanti positive per la citocheratina in pazienti con una resa media di 50 cellule/ml e purezza variabile dall'1% all'80% (Nagrath et al., 2007: Nature 450, da 1235 a 1239). Un'altra possibilità per isolare le CTC è utilizzare la prova per le cellule tumorali circolanti (CTC) CellSearch™ da Veridex, LLC (Raritan, New Jersey) che cattura, identifica, e conta le CTC in una provetta di sangue. Il sistema CellSearch™ è una metodologia approvata dalla Food and Drug Administration (FDA) degli Stati Uniti per contare le CTC nel sangue intero che si basa su una combinazione di marcatura immunomagnetica e microscopia digitale automatizzata. Vi sono altri metodi per isolare le CTC descritti nella letteratura tutti i quali possono essere utilizzati in combinazione con la presente invenzione.

Una ricaduta o una recidiva avviene quando una persona è nuovamente affetta da una condizione che l'aveva interessata in passato. Per esempio, se un paziente ha sofferto di una malattia tumorale, ha ricevuto un trattamento riuscito della suddetta malattia e sviluppa ancora la suddetta malattia la suddetta malattia recentemente sviluppata può essere considerata come ricaduta o recidiva. Tuttavia, secondo l'invenzione, una ricaduta o una recidiva di una malattia di tipo tumorale può avvenire ma non necessariamente avviene al sito della malattia tumorale originaria. Così, per esempio, se un paziente ha sofferto di tumore dell'ovaio ed ha ricevuto un trattamento riuscito una ricaduta o una recidiva può essere la comparsa di un tumore dell'ovaio o la comparsa di un tumore ad un sito differente dall'ovaio. Una ricaduta o una recidiva di un tumore comprende anche situazioni in cui un tumore si presenta ad un sito differente dal sito tumorale originale così come al sito tumorale originale. Preferibilmente, il tumore originale per cui il paziente ha ricevuto un trattamento è un tumore primario ed il tumore ad un sito differente dal sito tumorale originale è un tumore secondario o metastatico.

Con "trattare" si intende somministrare un composto o una composizione come qui descritti ad un soggetto allo scopo di impedire o di eliminare

una malattia, compresa la riduzione della dimensione di un tumore o del numero di tumori in un soggetto; l'arresto o il rallentamento di una malattia in un soggetto; l'inibizione o il rallentamento dello sviluppo di una nuova malattia in un soggetto; la diminuzione della frequenza o della gravità dei sintomi e/o delle recidive in un soggetto che attualmente ha o che precedentemente ha avuto una malattia; e/o il prolungamento, cioè l'incremento della durata della vita del soggetto. In particolare, il termine "trattamento di una malattia" comprende curare, accorciare la durata, migliorare, prevenire, rallentare o inibire la progressione o il peggioramento, o prevenire o ritardare la comparsa di una malattia o dei suoi sintomi.

Con "essere a rischio" si intende un soggetto, cioè un paziente, che viene identificato come avente una probabilità superiore al normale di sviluppare una malattia, in particolare il cancro, in confronto alla popolazione complessiva. Inoltre, un soggetto che ha avuto, o che attualmente ha, una malattia, in particolare il cancro, è un soggetto che ha un rischio aumentato di sviluppare una malattia, poiché tale soggetto può continuare a sviluppare una malattia. I soggetti che attualmente hanno, o che hanno avuto, un cancro hanno anche un rischio aumentato di metastasi tumorali.

Il termine "immunoterapia" riguarda un trattamento riguardante l'attivazione di una reazione immunitaria specifica. Nel contesto della presente invenzione, termini quali "proteggere", "impedire", "profilattico", "preventivo", o "protettivo" riguardano la prevenzione o il trattamento di entrambe la recidiva e/o la propagazione di una malattia in un soggetto e, in particolare, il minimizzare la possibilità che un soggetto svilupperà una malattia o il ritardare lo sviluppo di una malattia. Per esempio, una persona a rischio per un tumore, come descritto sopra, sarebbe un candidato per la terapia per prevenire un tumore.

Una somministrazione profilattica di un'immunoterapia, per esempio, una somministrazione profilattica della composizione qui descritta, preferibilmente protegge il ricevente dallo sviluppo di una malattia. Una somministrazione terapeutica di un'immunoterapia, per esempio, una somministrazione terapeutica della composizione qui descritta, può condurre all'inibizione del/la progresso/crescita della malattia. Questo comprende il rallentamento del/la progresso/crescita della malattia, in particolare una distruzione della progressione della malattia, la quale conduce preferibilmente all'eliminazione della malattia.

L'immunoterapia può essere eseguita utilizzando qualsiasi di una varietà di tecniche, in cui agenti qui forniti funzionano per rimuovere le cellule malate da un paziente. Tale rimozione può avvenire come risultato dell'aumento o dell'induzione di una risposta immunitaria in un paziente specifico per un antigene o per una cellula esprimente un antigene.

All'interno di alcune forme di realizzazione, l'immunoterapia può essere immunoterapia attiva, in cui il trattamento si basa sulla stimolazione in vivo del sistema immunitario endogeno dell'ospite per reagire nei confronti delle cellule malate con la somministrazione di agenti che modificano la risposta immunitaria (quali polipeptidi ed acidi nucleici come qui forniti).

Gli agenti e le composizioni qui forniti possono essere utilizzati da soli o in combinazione con regimi terapeutici convenzionali quali l'intervento chirurgico, l'irradiazione, la chemioterapia e/o il trapianto di midollo osseo (autologo, singenico, allogenico o non correlato).

Il termine "immunizzazione" o "vaccinazione" descrive il procedimento di trattare un soggetto allo scopo di indurre una risposta immunitaria per ragioni terapeutiche o profilattiche.

Il termine "in vivo" riguarda la situazione in un soggetto.

I termini "soggetto", "individuo", "organismo" o "paziente" vengono utilizzati in modo intercambiabile e riguardano i vertebrati, preferibilmente i mammiferi. Per esempio, i mammiferi nel contesto della presente invenzione sono l'uomo, primati diversi dall'uomo, animali addomesticati quali cani, gatti, pecore, bovini, capre, maiali, cavalli ecc., animali da laboratorio quali topi, ratti, conigli, cavie, ecc. così come animali in cattività quali animali da zoo. Il termine "animale" come qui utilizzato comprende anche l'uomo. Il termine "soggetto" può includere anche un paziente, cioè un animale, preferibilmente un essere umano avente una malattia, preferibilmente una malattia come qui descritta.

Il termine "autologo" viene utilizzato per descrivere qualsiasi cosa che è derivata dallo stesso soggetto. Per esempio, "trapianto autologo" si riferisce ad un trapianto di tessuto o di organi derivati dallo stesso soggetto. Tali procedure sono vantaggiose poiché superano la barriera immunologica che altrimenti risulta nel rigetto.

Il termine "eterologo" viene utilizzato per descrivere qualcosa composto da multipli elementi differenti. Come esempio, il trasferimento del midollo osseo di un individuo in un individuo differente costituisce un trapianto eterologo. Un gene eterologo è un gene derivato da una fonte diversa dal soggetto.

Come parte della composizione per un'immunizzazione o per una vaccinazione, preferibilmente uno o più agenti come qui descritti vengono somministrati insieme ad uno o più adiuvanti per indurre una risposta immunitaria o per aumentare una risposta immunitaria. Il termine "adiuvante" riguarda composti che prolunga od accresce o accelera una risposta immunitaria. La composizione qui descritta esercita preferibilmente il suo effetto senza l'aggiunta di adiuvanti. Ancora, la composizione della presente domanda può contenere qualsiasi adiuvante noto. Gli adiuvanti comprendono un gruppo eterogeneo di composti quali emulsioni oleose (per esempio adiuvanti di Freund), composti minerali (quali allume), prodotti batterici (quali la tossina di Bordetella pertussis), liposomi, e complessi immunostimolanti. Esempi di adiuvanti sono monofosforil-lipide-A (MPL SmithKline Beecham). Saponine quali QS21 (SmithKline Beecham), DQS21 (SmithKline Beecham; WO 96/33739), QS7, QS17, QS18, e QS-L1 (So et al., 1997, Mol. Cells 7: da 178 a 186), adiuvanti incompleti di Freund, adiuvanti completi di Freund, vitamina E, montanide, allume, oligonucleotidi CpG (Krieg et al., 1995, Nature 374: da 546 a 549), e varie emulsioni acqua in olio che vengono preparate da oli biologicamente

degradabili quali squalene e/o tocoferolo.

Altre sostanze che stimolano una risposta immunitaria del paziente possono anch'esse essere somministrate. È possibile, per esempio, utilizzare citochine in una vaccinazione, a causa delle loro proprietà regolatrici sui linfociti. Tali citochine comprendono, per esempio, interleuchina-12 (IL-12) che è stata mostrata incrementare le azioni protettive dei vaccini (si confronti Science 268: da 1432 a 1434, 1995), GM-CSF ed IL-18.

Vi è un certo numero di composti che aumentano una risposta immunitaria e che possono quindi essere utilizzati in una vaccinazione. I suddetti composti comprendono molecole costimolatorie fornite in forma di proteine o di acidi nucleici quali B7-1 e B7-2 (CD80 e CD86, rispettivamente).

Secondo l'invenzione, un "campione tumorale" è un campione quale un campione corporeo contenente cellule tumorali o cancerose quali le cellule tumorali circolanti (CTC), in particolare un campione di tessuto, compresi fluidi corporei, e/o un campione cellulare. Secondo l'invenzione, un "campione non tumorigenico" è un campione quale un campione corporeo non contenente cellule tumorali o cancerose quali le cellule tumorali circolanti (CTC), in particolare un campione di tessuto, compresi fluidi corporei, e/o un campione cellulare. Tali campioni corporei possono essere ottenuti nel modo convenzionale come mediante biopsia tissutale, compresa la biopsia mediante bisturi circolare, e prendendo sangue, aspirato bronchiale, espettorato, urina, feci od altri fluidi corporei. Secondo l'invenzione, il termine "campione" comprende anche campioni trattati quali frazioni o isolati di campioni biologici, per esempio isolati di acidi nucleici di cellule.

I principi terapeuticamente attivi, i vaccini e le composizioni qui descritti possono essere somministrati attraverso qualsiasi via convenzionale, comprese per iniezione o per infusione. La somministrazione può essere realizzata, per esempio, per via orale, per via endovenosa, per via intraperitoneale, per via intramuscolare, per via sottocutanea o per via transdermica. In una forma di realizzazione, la somministrazione viene realizzata per via intranodale come mediante iniezione in un linfonodo. Altre forme di somministrazione prevedono la trasfezione in vitro di cellule presentanti l'antigene quali le cellule dendritiche con gli acidi nucleici qui descritti cui segue la somministrazione delle cellule presentanti l'antigene.

Gli agenti qui descritti vengono somministrati in quantità efficaci. Una "quantità efficace" si riferisce alla quantità che raggiunge una reazione desiderata o un effetto desiderato da sola o insieme ad ulteriori dosi. Nel caso di un trattamento di una malattia particolare o di una condizione particolare, la reazione desiderata riguarda preferibilmente l'inibizione del decorso della malattia. Questo comprende il rallentamento del progresso della malattia e, in particolare, l'interruzione o l'inversione del progresso della malattia. La reazione desiderata in un trattamento di una malattia o di una condizione può anche essere il ritardo della comparsa o una prevenzione della comparsa della suddetta malattia o della suddetta condizione.

Una quantità efficace di un agente qui descritto dipenderà dalla condizione da trattare, dalla gravità della malattia, dai singoli parametri del

paziente, compresi l'età, la condizione fisiologica, la dimensione e il peso, dalla durata del trattamento, dal tipo di terapia di accompagnamento (se presente), dalla specifica via di somministrazione e da fattori simili. Conseguentemente, le dosi somministrate degli agenti qui descritti possono dipendere da vari tali parametri. Nel caso in cui una reazione in un paziente sia insufficiente con una dose iniziale, dosi più elevate (o dosi efficacemente più elevate ottenute mediante una differente via di somministrazione più localizzata) possono essere utilizzate.

Secondo la presente invenzione, le composizioni farmaceutiche sono preferibilmente sterili e contengono una quantità efficace del principio terapeuticamente attivo per generare la reazione desiderata o l'effetto desiderato.

Secondo la presente invenzione, le composizioni farmaceutiche sono in generale somministrate in quantità farmaceuticamente compatibili ed in una preparazione farmaceuticamente compatibile. Il termine "farmaceuticamente compatibile" si riferisce ad un materiale non tossico che non interagisce con l'azione del principio attivo della composizione farmaceutica. Preparazioni di questo tipo possono contenere solitamente sali, sostanze tampone, conservanti, veicolanti, sostanze supplementari che potenziano l'immunità quali adiuvanti, per esempio oligonucleotidi CpG, citochine, chemochine, saponina, GM-CSF e/o RNA e, dove appropriato, altri composti terapeuticamente attivi. Quando utilizzati in medicina, i sali dovrebbero essere farmaceuticamente compatibili. Tuttavia, i sali che non sono farmaceuticamente compatibili possono essere utilizzati per preparare sali farmaceuticamente compatibili e sono inclusi secondo l'invenzione. Sali farmacologicamente e farmaceuticamente compatibili di questo tipo comprendono in un modo non limitante quelli preparati dai seguenti acidi: acidi cloridrico, bromidrico, solforico, nitrico, fosforico, maleico, acetico, salicilico, citrico, formico, malonico, succinico, e simili. Sali farmaceuticamente compatibili possono anche essere preparati come sali di metalli alcalini o sali di metalli alcalino terrosi, quali sali di sodio, sali di potassio o sali di calcio.

Secondo la presente invenzione, una composizione farmaceutica può comprendere un veicolante farmaceuticamente compatibile. Il termine "veicolante" si riferisce ad un componente organico o inorganico, di una natura naturale o sintetica, in cui il principio attivo viene combinato allo scopo di facilitare l'applicazione. Secondo l'invenzione, il termine "veicolante farmaceuticamente compatibile" comprende uno o più riempitivi solidi o liquidi, diluenti o sostanze incapsulanti compatibili, i quali sono adatti per la somministrazione ad un paziente. Secondo la presente invenzione, i componenti della composizione farmaceutica sono solitamente in modo tale che non avvenga alcuna interazione che danneggi sostanzialmente l'efficacia farmaceutica desiderata.

Secondo la presente invenzione, le composizioni farmaceutiche possono contenere sostanze tampone adatte quali acido acetico in un sale, acido citrico in un sale, acido boricco in un sale ed acido fosforico in un sale.

Le composizioni farmaceutiche possono, dove appropriato, contenere anche conservanti adatti quali cloruro di benzalconio, clorobutanolo,

parabene e timerosal.

Le composizioni farmaceutiche sono solitamente fornite in una forma di dosaggio uniforme e possono essere preparate in un modo noto come tale. Secondo la presente invenzione, le composizioni farmaceutiche possono essere in forma di capsule, compresse, losanghe, soluzioni, sospensioni, sciroppi, elisir o in forma di emulsione, per esempio.

Composizioni adatte per la somministrazione parenterale comprendono solitamente una preparazione sterile acquosa o non acquosa del principio attivo, la quale è preferibilmente isotonica al sangue del ricevente. Esempi di veicolanti e di solventi compatibili sono la soluzione di Ringer e la soluzione isotonica di cloruro di sodio. Inoltre, solitamente oli non volatili, sterili vengono utilizzati come mezzo di soluzione o di sospensione.

La presente invenzione viene descritta in dettaglio mediante le figure e gli esempi sotto, i quali vengono utilizzati solo per scopi illustrativi e non sono intesi essere limitanti. A causa della descrizione e degli esempi, ulteriori forme di realizzazione che sono incluse in modo simile nell'invenzione sono accessibili al tecnico esperto.

FIGURE

Figura 1:

parte superiore: procedimento per scoprire e per prioritizzare probabili mutazioni somatiche immunogeniche in campioni di masse tumorali. Parte inferiore: procedimento come applicato al sistema di B16 e di Black6.

Figura 2: mutazione convalidata di esempio in Kif18b

Una mutazione identificata nel gene Kif18b mediante sequenziamento dell'esoma con NGS che è stata confermata mediante sequenziamento secondo Sanger. Nelle cellule selvatiche, la sequenza è T/T. Nelle cellule tumorali, la sequenza è una miscela di T/G.

Figura 3: reattività immunologica nei confronti delle sequenze mutate

I topi (n=5) sono stati immunizzati due volte (giorno 0, giorno 7) con sequenze peptidiche mutate (100 µg+50 µg di poli(I:C); s.c.). Il giorno 12 i topi sono stati sacrificati e le cellule della milza raccolte. Un ELISPOT per IFN-γ è stato eseguito utilizzando 5x10⁵ cellule della milza/pozzetto come effettori e 5x10⁴ cellule dendritiche di midollo osseo caricate con peptidi (2µg/ml per 2 ore a 37°C e CO₂ al 5%) come cellule bersaglio. Le cellule effettrici di milza sono state testate nei confronti del peptide mutato, del peptide selvatico e di un peptide di controllo (nucleoproteina del virus della stomatite vescicolare, VSV-NP, amminoacidi da 52 a 59). Viene mostrato il numero medio misurato di macchie da cui le macchie di sfondo nei confronti della VSV-NP sono state sottratte per ogni topo (cerchi vuoti: topi immunizzati con il peptide selvatico; riquadri pieni: topi immunizzati con i peptidi mutati). I dati vengono mostrati per ciascun topo e viene descritta la media±ESM.

Figura 4: beneficio di sopravvivenza per i topi vaccinati con la sequenza peptidica mutata recentemente identificata

Le cellule B16-F10 ($7,5 \times 10^4$) sono state inoculate per via sottocutanea il giorno 0. I topi sono stati vaccinati con il peptide 30 (Jerini Peptide Technologies (Berlino); 100 μg di peptide+50 μg di poli(I:C) s.c. (Invivogen)) il giorno -4, il giorno +2, il giorno +9. Il gruppo di controllo ha ricevuto solo poli(I:C) (50 μg s.c.). La crescita tumorale è stata monitorata fino al giorno +16 *, $p < 0,05$ nel test dei ranghi logaritmici (di Mantel-Cox).

Figura 5:

(A) esempi di un'espressione aumentata della proteina (sinistra eGFP, destra luciferasi) con RNA ottimizzato per la stabilità e per l'efficienza traduzionale.

(B) Esempio di espansione poliepitopica di linfociti T CD8+ e CD4+ specifici per l'antigene con RNA ottimizzato per l'indirizzamento efficace dell'antigene (si veda il riferimento Kreiter, Konrad, Sester et al., Cancer Immunol. Immunother. 56: da 1577 a 1587, 2007). (C) Esempio di una prova preclinica di efficacia antitumorale in un modello di melanoma B16 utilizzando un vaccino a RNA che codifica per un singolo epitopo (OVA-SIINFEKL).

I dati di sopravvivenza sono stati ottenuti per i topi trattati con il vaccino da solo o con il vaccino in combinazione con l'adiuvante. (D) Progettazione di un vaccino individualizzato, polineoepitopico. Il veicolo del vaccino integra elementi funzionali per un'espressione aumentata e per un'immunogenicità ottimizzata. Fino a 30 epitopi mutati che sono distanziati mediante gruppi di collegamento possono essere integrati per molecola nel loro contesto naturale di sequenza.

Figura 6: progettazione del costrutto

(A) Diagramma schematico di un costrutto di RNA poliepitopico. Cap: analogo del cappuccio; 5'UTR: regione 5' non tradotta; L: gruppo di collegamento; Seq. 1: sequenza di RNA codificante per un peptide contenente amminoacidi mutati; 3'UTR: sequenza non tradotta al 3'; poliA: coda di poliA. (B) Sequenza dei costrutti di RNA codificanti per 2 sequenze amminoacidiche comprendenti un amminoacido mutato da B16-F10. Il codone di inizio e di arresto così come il peptide segnale e la sequenza MITD non sono parte del disegno schematico che è simboleggiato da "...".

Figura 7: funzionalità dell'RNA poliepitopico

(Da A a C) Dati per l'ELISPOT per IFN- γ utilizzando 5×10^5 cellule della milza per pozzetto come effettori e 5×10^4 BMDC come cellule bersaglio. Le BMDC sono state caricate con il peptide (2 $\mu\text{g}/\text{ml}$ per 2 ore a 37°C e CO₂ al 5%) o trasfettate con l'RNA (20 μg) mediante elettroporazione. L'RNA di controllo era eGFP (riquadro a sinistra) o un costrutto di RNA codificante per 2 peptidi non correlati contenenti un amminoacido mutato separati da un gruppo di collegamento. I dati vengono mostrati come media \pm ESM. (A) Vengono mostrati i dati per il peptide con mutazione 30, per il peptide selvatico 30 e per l'RNA codificante per le mutazioni 30 e 31. (B) Vengono mostrati i dati per il peptide con la mutazione 12, per il peptide selvatico 12 e per l'RNA codificante per le mutazioni 12 e 39. (C) Viene descritta una scansione rappresentativa di un ELISPOT da un singolo topo della lettura mostrata in

(B).

Figura 8: due forme di realizzazione di vaccini polineoepitopici a RNA che mostrano epitopi di giunzione

Il vaccino a RNA può essere costruito con (parte superiore) o senza (parte inferiore) gruppi di collegamento tra i peptidi codificanti con mutazione. I buoni epitopi includono quelli che includono la mutazione somatica (“*”) e si legano a molecole dell’MHC. Gli cattivi epitopi includono epitopi che si legano a molecole dell’MHC ma contengono parti di due peptidi (parte inferiore) o parti di peptide e sequenze di collegamento (parte superiore).

Figura 9: scoperta e caratterizzazione del “mutanoma colpibile farmacologicamente dei linfociti T”

(A) Il diagramma di flusso dà una visione d’insieme della procedura sperimentale a partire da campioni di B16-F10 e di C57BL/6 al risultato dell’ELISPOT. (B) Viene mostrato il numero di risultati per ciascun passaggio di valutazione ed il procedimento per la selezione di mutazioni per la prova di convalida del DNA e di immunogenicità. Le mutazioni scelte per la prova di convalida e di immunogenicità sono state quelle previste essere immunogeniche ed in geni espressi a RPKM>10. (C) Il mutanoma colpibile farmacologicamente dei linfociti T è stato mappato al genoma di B16-F10. Gli anelli dall’esterno verso l’interno rappresentano i seguenti sottoinsiemi: (1) presente in tutte le triplici copie, (2) hanno un FDR<0,05, (3) sono localizzate in regioni codificanti per la proteina, (4) causano variazioni non sinonime, (5) sono localizzate in geni espressi, e (6) sono nella serie convalidata. Cromosomi di topo (cerchio esterno), densità genica (verde), espressione genica (verde(bassa)/giallo/rosso(elevata)), e mutazioni somatiche (arancione).

Figura 10: risposte immunitarie provocate in vivo dalla vaccinazione dei topi con una mutazione rappresentante peptidi sintetici lunghi

(A,B) Analisi mediante ELISPOT per IFN- γ di linfociti T effettori da topi vaccinati con peptidi codificanti per una mutazione. Le colonne rappresentano le medie (\pm ESM) di 5 topi per gruppo. Gli asterischi indicano differenze statisticamente significative di reattività nei confronti del peptide con la mutazione e selvatico (test di t di Student; valore di $p<0,05$). (A) Gli splenociti di topi vaccinati sono stati stimolati di nuovo con BMDC trasfettate con il peptide codificante per la mutazione utilizzato per la vaccinazione, con il peptide selvatico corrispondente e con un peptide di controllo non rilevante (VSV-NP). (B) Per l’analisi della reattività dei linfociti T nei confronti delle mutazioni processate a livello endogeno splenociti di topi vaccinati sono stati stimolati di nuovo con BMDC trasfettate con l’RNA di controllo (eGFP) o con un RNA codificante per la mutazione indicata. (C) Mutazione 30 (gene Kif18B, proteina Q6PFD6, mutazione p.K739N). Traccia del sequenziamento secondo Sanger e sequenza della mutazione (parte superiore). Domini proteici e posizione della mutazione (parte inferiore).

Figura 11: effetti antitumorali di vaccini di peptidi mutati in topi con tumori B16-F10 che crescono in modo aggressivo

(A) I topi C57BL/6 (n=7) hanno ricevuto un inoculo di $7,5 \times 10^4$ cellule B16-F10 s.c. nel fianco dei topi. I giorni 3 e 10 dopo l’inoculo del tumore i topi

sono stati vaccinati con 100 µg di peptide MUT30 o MUT44+50 µg di poli(I:C) o con adiuvante da solo. **(B)** I topi C57BL/6 (n=5) hanno ricevuto un'immunizzazione di 100 µg di peptide MUT30+50 µg di poli(I:C) il giorno -4. Il giorno 0 $7,5 \times 10^4$ cellule B16-F10 sono state inoculate s.c. nel fianco dei topi. Le immunizzazioni di mantenimento con peptide MUT30 (+poli(I:C)) sono state fatte i giorni 2 e 9. Grafico della sopravvivenza di Kaplan-Meier (sinistra). Cinetiche di crescita del tumore (destra).

Figura 12: la vaccinazione con RNA codificanti per una mutazione conduce a risposte di linfociti T CD4+ e CD8+

Dati di analisi della marcatura delle citochine intracellulari per l'IFN-γ in linfociti T effettori CD4+ e CD8+ da topi vaccinati con RNA codificanti per una mutazione. Gli RNA sono stati codificanti per 1 (monoepitopo, riga superiore), 2 (biepitopo, riga intermedia), o 16 (poliepitopo, riga inferiore) mutazioni differenti. I punti rappresentano le medie di 3 topi per gruppo. Gli asterischi indicano differenze statisticamente significative di reattività nei confronti del peptide con la mutazione e di controllo (VSV-NP) (test di t di Student; valore di $p < 0,05$). I grafici della FACS mostrano effettori dall'animale massimamente secernente IFN-γ per ciascuna mutazione ed indicano il fenotipo della risposta dei linfociti T.

Figura 13: la vaccinazione con RNA poliepitopico codificante per una mutazione conduce risposte dei linfociti T nei confronti di diverse mutazioni

Analisi mediante ELISPOT per IFN-γ di linfociti T effettori da topi vaccinati con un poliepitopo codificante per una mutazione comprendente 16 differenti mutazioni. Le colonne rappresentano le medie (\pm ESM) di 3 topi per gruppo. La fotografia mostra pozzetti in triplice copia di cellule da un animale di esempio stimolato di nuovo con i peptidi indicati.

Figura 14: la vaccinazione con 5 differenti epitopi modello codificati da un RNA conduce a risposte immunitarie nei confronti di tutti gli epitopi codificati

A) Analisi mediante ELISPOT per IFN-γ di linfociti T effettori da topi vaccinati con un poliepitopo modello codificante per una mutazione comprendente 5 differenti epitopi modello (SIINFEKL, Trp2, VSV-NP, Inf-NP, OVA di classe II). Gli splenociti sono stati stimolati di nuovo con i peptidi indicati. Le macchie rappresentano le medie di pozzetti in triplice copia da 5 topi per gruppo. B) Marcatura pentamerica di linfociti del sangue di un topo di controllo e di un topo immunizzato con il poliepitopo modello. Le cellule CD8+ marcate in modo pentamerico per Inf-NP sono specifiche per il peptide Inf-NP.

Figura 15: una mutazione che induce i linfociti T CD4+ può indurre un potente effetto antitumorale contro il melanoma B16-F10 in sinergia con un debole epitopo dei linfociti T CD8+

I topi C57BL/6 (n=8) hanno ricevuto un inoculo di 1×10^5 cellule B16-F10 s.c. nel fianco dei topi. I giorni 3, 10 e 17 dopo l'inoculo del tumore i topi sono stati vaccinati con 100 µg di MUT30, di Trp2 o di entrambi i peptidi+50 µg di poli(I:C). A) Sono mostrate le cinetiche medie di crescita tumorale di

ciascun gruppo. Il giorno 28 i valori medi tra i gruppi di trattamento singolo e gli animali non trattati e il gruppo di combinazione sono statisticamente differenti (test di Mann-Whitney, valore di $p < 0,05$). B) Grafico di sopravvivenza di Kaplan-Meyer dei gruppi differenti. Le curve di sopravvivenza dei topi vaccinati con MUT30 e con MUT30+Trp2 sono statisticamente differenti (test dei ranghi logaritmici, valore di $p = 0,0029$).

Figura 16: panoramica di procedimento per trovare mutazioni somatiche in B16

I numeri per i singoli passaggi vengono forniti come esempio per un campione di B16, in confronto ad un campione di Black6. "Esoni" si riferisce alle coordinate esoniche definite da tutti i trascritti RefSeq codificanti per proteine.

Figura 17: diagramma di Venn che mostra i numeri di variazioni somatiche in esoni codificanti per proteine, trovati mediante il singolo strumento informatico, due o tutti e tre gli strumenti informatici, rispettivamente

I numeri sono stati calcolati dopo filtraggio e rappresentano il consenso di tutti e tre i campioni.

Figura 18: A esempi di singole variazioni nucleotidiche trovate: una mutazione somatica trovata in tutti e tre i campioni di B16 (sinistra), una mutazione non somatica trovata in tutti i campioni di B16 e di Black6 (in mezzo) ed una mutazione trovata solo in un campione di Black6 (destra).

B La distribuzione calcolata degli FDR per la serie di dati da cui le mutazioni convalidate sono state selezionate; la distribuzione è visualizzata come curva ROC media stimata con le barre grigie che danno l'intervallo di confidenza al 95% per la media in entrambe le dimensioni in posizioni uniformemente campionate. La media è stata ottenuta dalla distribuzione delle curve ROC stimate degli FDR per tutte le possibili 18 combinazioni (si veda il testo).

Figura 19: A Curve ROC stimate per il confronto dei tre differenti strumenti informatici (duplici copie, copertura 38x). **B** Curve ROC stimate per il confronto di differenti profondità medie di sequenziamento (samtools, nessuna replicazione). 38x indica la copertura ottenuta mediante l'esperimento, mentre altre coperture sono state sottocampionate a partire da questi dati. **C** Curve ROC stimate che visualizzano l'effetto della replicazione dell'esperimento (copertura 38x, samtools). **D** Curve ROC stimate per differenti protocolli di sequenziamento (samtools, nessuna replicazione). Le curve sono state calcolate utilizzando i risultati della libreria di 2x100 nucleotidi.

Figura 20: A dieci mutazioni convalidate con i minimi FDR, scelte utilizzando la serie ottimale di parametri su una serie finale di 2396 variazioni. Nessuna di queste mutazioni è presente in dbSNP (versione 128; gruppo genomico mm9). **B** Quantità relativa di variazioni trovate nella stessa serie di dati come A per un dato limite di FDR, rappresentata separatamente per tutte le varianti nella serie di dati e le mutazioni convalidate. Per chiarezza visiva vengono mostrati solo i valori di FDR dallo 0% al 10%.

Figura 21: attività antitumorale di un vaccino ad RNA polipeptidico codificante per la mutazione

I topi C57BL/6 (n=10) hanno ricevuto un inoculo di 1×10^5 cellule B16-F10 s.c. nel fianco dei topi. I giorni 3, 6, 10, 17 e 21 dopo l'inoculo del tumore i topi sono stati vaccinati con un RNA polipeptidico formulato con un reagente di trasfezione dell'RNA liposomiale. Il gruppo di controllo ha ricevuto i liposomi senza RNA. La figura mostra il grafico di sopravvivenza di Kaplan-Meyer dei gruppi differenti. Curve di sopravvivenza statisticamente differenti (test dei ranghi logaritmici, valore di $p=0,0008$).

ESEMPI

Le tecniche ed i metodi qui utilizzati vengono qui descritti o realizzati in un modo noto come tale e come descritto, per esempio, in Sambrook et al., Molecular Cloning: A Laboratory Manual, seconda edizione (1989) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York. Tutti i metodi comprendenti l'uso dei kit e dei reagenti sono realizzati secondo le informazioni dei produttori se non indicato specificamente.

Esempio 1: rilevamento e prioritizzazione delle mutazioni

Noi prima dimostriamo l'ottenimento del profilo di sequenza di campioni tumorali e normali per identificare le mutazioni somatiche in un modo imparziale. Noi dimostriamo questo non solo per campioni di masse tumorali ma anche, per la prima volta, dimostriamo la capacità di identificare mutazioni da singole cellule tumorali circolanti. In seguito, prioritizziamo le mutazioni per l'inclusione in un vaccino polineoepitopico sulla base dell'immunogenicità prevista della mutazione e dimostriamo che le mutazioni identificate sono infatti immunogeniche.

Rilevamento delle mutazioni

Razionale per utilizzare le CTC: il rilevamento di cellule tumorali circolanti (CTC) dal sangue periferico di pazienti con il cancro è un riconosciuto marcatore prognostico indipendente per il decorso clinico dei tumori (Pantel et al., Trends Mol Med 2010; 16(9): da 398 a 406). Per molti anni, la significatività clinica delle CTC è stata oggetto di intensa ricerca scientifica e clinica in oncologia. È stato mostrato che il rilevamento delle CTC nel sangue di pazienti con cancro metastatico della mammella, della prostata e del colon-retto ha rilevanza prognostica, fornendo ulteriori informazioni alle tecniche di visualizzazione convenzionali e ad altri biomarcatori prognostici tumorali. Campioni di sangue sequenziali prelevati da un paziente prima di, durante uno stadio iniziale, e dopo il trattamento con un agente terapeutico (sistemico o mirato) fornisce informazioni sulla risposta al trattamento o sull'insuccesso del trattamento. L'analisi molecolare delle CTC resistenti ai farmaci può fornire un'ulteriore comprensione dei meccanismi di resistenza (per esempio mutazioni in vie specifiche di trasduzione del segnale o perdita di espressione dei bersagli) in singoli pazienti. Un'ulteriore possibilità dall'ottenimento del profilo e dalla caratterizzazione genetica delle CTC è l'identificazione di nuovi bersagli tumorali per lo sviluppo di nuove terapie mirate. Questa nuova strategia diagnostica viene indicata come "biopsia tumorale liquida." Poiché questo ottenimento del profilo potrebbe essere velocemente e ripetutamente fatto, richiedendo solo sangue del paziente e nessun intervento chirurgico,

questo fornirebbe una vista in “tempo reale” dello stato tumorale.

Mutazioni dalle cellule tumorali: noi dimostriamo la nostra capacità di identificare mutazioni utilizzando le cellule B16 di melanoma, la cattura dell'esoma per estrarre regioni codificanti per le proteine, il sequenziamento di nuova generazione utilizzando il nostro HiSeq 2000, seguito dall'analisi bioinformatica utilizzando il nostro programma di accesso “iCAM” (figura 1). Noi identifichiamo 2448 mutazioni non sinonime e ne scegliamo 50 per conferma. Noi siamo stati in grado confermare tutte le 50 mutazioni somatiche.

Ciò che segue è un esempio dell'impatto sulla proteina di una scoperta delle mutazioni somatiche nelle cellule B16 di melanoma:

Kif18b, NM_197959, esone 3

Mutazione (+15 amminoacidi)

SPSKPSFQEFVDWENVSPELNSTDQPFLPS

Selvatico (+15 amminoacidi)

SPSKPSFQEFVDWEKVSPELNSTDQPFLPS

Mutazioni da singole cellule tumorali circolanti (CTC): poi, siamo stati in grado di identificare le mutazioni somatiche specifiche per il tumore dall'ottenimento del profilo mediante NGS di RNA dalle CTC singole. Cellule B16 marcate di melanoma sono state iniettate per via endovenosa in code di topo, i topi sono stati sacrificati, è stato prelevato il sangue dai cuori, le cellule separate per recuperare cellule B16 marcate circolanti (CTC), l'RNA estratto, una sintesi di cDNA basata su SMART e un'amplificazione non specifica eseguite, seguite dal saggio di RNA-Seq mediante NGS e successiva analisi dei dati (sotto).

Noi abbiamo profilato otto singole CTC e mutazioni somatiche identificate. Inoltre, in otto cellule su otto, mutazioni somatiche precedentemente identificate sono state identificate. In multipli casi, i dati hanno mostrato un'eterogeneità a livello di singola cellula. Per esempio, in posizione 144078227 sul cromosoma 2 (gruppo mm9), nel gene Snx15, due cellule hanno mostrato il nucleotide di riferimento (C) mentre due cellule hanno mostrato il nucleotide mutato (T).

Questo dimostra che noi siamo in grado di ottenere il profilo di singole CTC per identificare le mutazioni somatiche, un fondamentale cammino ad un iVAC (vaccino individualizzato) in “tempo reale”, in cui i pazienti sono profilati ripetutamente ed i risultati riflettono lo stato attuale del paziente piuttosto che lo stato ad un punto temporale precoce. Inoltre, questo dimostra che noi siamo in grado di identificare le mutazioni somatiche eterogenee che sono presenti in un sottoinsieme di cellule tumorali, permettendo la valutazione della frequenza di mutazione, come per l'identificazione di mutazioni principali e di mutazioni rare.

Metodi

Campioni: per l'esperimento di ottenimento del profilo, i campioni hanno incluso campioni di coda da 5 a 10 mm da topi C57BL/6 ("Black6") e cellule di topo B16-F10 altamente aggressive di melanoma ("B16"), le quali sono originariamente derivate da topi Black6.

Le cellule tumorali circolanti (CTC) sono state generate utilizzando cellule B16 di melanoma marcate in modo fluorescente. Le cellule B16 sono state risospese in PBS ed un volume uguale di soluzione di CFSE preparata di recente (5 μ M in PBS) è stato aggiunto alle cellule. Il campione è stato miscelato delicatamente mediante vortice cui segue un'incubazione per 10 minuti a temperatura ambiente. Per fermare la reazione di marcatura, la quantità uguale di PBS contenente FSC al 20% è stata aggiunta al campione e miscelata delicatamente mediante vortice. Dopo 20 minuti di incubazione a temperatura ambiente, le cellule sono state lavate due volte utilizzando PBS. Infine, le cellule sono state risospese in PBS ed iniettate per via endovenosa (e.v.) nei topi. Dopo 3 minuti, i topi sono stati sacrificati e il sangue raccolto.

Gli eritrociti dai campioni di sangue sono stati lisati aggiungendo 1,5 ml di soluzione PharmLyse preparata di fresco (Beckton Dickinson) per 100 μ l di sangue. Dopo un passaggio di lavaggio, 7-AAD è stata aggiunta al campione ed incubata per 5 minuti a temperatura ambiente. L'incubazione è stata seguita da due passaggi di lavaggio ed il campione è stato risospeso in 500 μ l di PBS.

Le cellule B16 circolanti marcate con CFSE sono state separate con un separatore di cellule Aria I (BD). Le cellule singole sono state separate su piastre da 96 pozzetti con fondo a V preparate con 50 μ l/pozzetto di tampone RLT (Qiagen). Dopo aver finito la separazione le piastre sono state conservate a -80°C fino a che l'estrazione dell'acido nucleico e la preparazione del campione non sono cominciate.

Estrazione dell'acido nucleico e preparazione del campione: gli acidi nucleici dalle cellule B16 (DNA ed RNA) e dal tessuto caudale di Black6 (DNA) sono stati estratti utilizzando un kit Qiagen DNeasy per sangue e per tessuto (DNA) ed un kit Qiagen RNeasy Micro (RNA).

Per le CTC singole separate, l'RNA è stato estratto ed una sintesi di cDNA basata su SMART ed un'amplificazione non specifica eseguite. L'RNA dalle cellule CTC separate è stato estratto con il kit RNeasy Micro (Qiagen, Hilden, Germania) secondo le istruzioni del fornitore. Un protocollo modificato di BD SMART è stato utilizzato per la sintesi del cDNA: trascrittasi inversa MINT (Evrogen, Mosca, Russia) è stata combinata con un innesco oligo(dT)-T-lungo per l'attivazione della reazione di sintesi del primo filamento e con un TS-breve (Eurogentec S.A., Seraing, Belgio) introducendo una sequenza oligo(riboG) per permettere la creazione di uno stampo esteso mediante l'attività trasferasica terminale della trascrittasi inversa e per un cambio dello stampo [Chenchik, A., Y. et al. 1998. Generation and use of high quality cDNA from small amounts of total RNA by SMART PCR. In Gene Cloning and Analysis by RT-PCR. P. L. J. Siebert, edizioni BioTechniques Books, Massachusset, Natick. da 305 a 319]. Il primo filamento di cDNA sintetizzato secondo le istruzioni del produttore è stato sottoposto a 35 cicli di amplificazione con 5 U di DNA polimerasi

PfuUltra Hotstart ad alta fedeltà (Stratagene, La Jolla, California) e con un innesco TS 0,48 μ M per la PCR in presenza di dNTP 200 μ M (condizioni di ciclizzazione: 2 minuti a 95°C, 30 secondi a 94°C, 30 secondi a 65°C, 1 minuto a 72°C, estensione finale di 6 minuti a 72°C). L'amplificazione riuscita dei geni delle CTC è stata controllata con inneschi specifici per monitorare l'actina e la GAPDH.

Sequenziamento di nuova generazione, sequenziamento del DNA: la cattura esomica per il risequenziamento del DNA è stata eseguita utilizzando il saggio di cattura a base di soluzione Agilent Sure-Select [Gnirke A et al: Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat. Biotechnol. 2009, 27: da 182 a 189], in questo caso progettato per catturare tutte le regioni codificanti per proteine di topo.

Brevemente, 3 μ g di DNA genomico purificato sono stati frammentati da 150 a 200 coppie di basi utilizzando un dispositivo ad ultrasuoni Covaris S2. I frammenti di gDNA sono stati riparati alle estremità utilizzando la DNA polimerasi T4, la DNA polimerasi di Klenow e fosforilati al 5' utilizzando la chinasi polinucleotidica T4. I frammenti di gDNA con estremità non appuntite sono stati adenilati al 3' utilizzando il frammento di Klenow (da 3' a 5' eso meno). Adattatori ad estremità appaiate con singola sporgenza a T al 3' di Illumina sono stati ligati ai frammenti di gDNA utilizzando un rapporto molare di 10:1 di adattatore e inserto di DNA genomico utilizzando la DNA ligasi T4. I frammenti di gDNA ligati all'adattatore sono stati arricchiti precattura e sequenze specifiche di cellule del flusso sono state aggiunte utilizzando inneschi di PCR Illumina PE 1.0 e 2.0 e polimerasi Herculase II (Agilent) utilizzando 4 cicli di PCR.

500 ng di frammenti di gDNA ligati all'adattatore arricchiti mediante PCR sono stati ibridati con esche biotilate SureSelect di una libreria di RNA dell'intero esoma di topo di Agilent per 24 ore a 65°C. I complessi ibridati gDNA/esca di RNA sono stati rimossi utilizzando sfere magnetiche rivestite di streptavidina. I complessi gDNA/esca di RNA sono stati lavati e le esche di RNA scisse durante l'eluizione in tampone di eluizione SureSelect lasciando i frammenti catturati di gDNA arricchiti mediante PCR ligati all'adattatore. I frammenti di gDNA sono stati amplificati mediante PCR dopo la cattura utilizzando la DNA polimerasi Herculase II (Agilent) e gli inneschi di PCR SureSelect GA per 10 cicli.

Tutti i passaggi di rimozione sono stati fatti utilizzando un volume 1,8x di sfere magnetiche AMPure XP (Agencourt). Tutti i controlli di qualità sono stati fatti utilizzando un saggio Qubit HS di Invitrogen e la dimensione dei frammenti è stata determinata utilizzando un saggio del DNA 2100 Bioanalyzer HS di Agilent.

Librerie di gDNA arricchite in esoma sono state raggruppate sul cBot utilizzando il kit Truseq SR cluster v2.5 utilizzando 7pM e 50 coppie di basi sono state sequenziate sull'Illumina HiSeq2000 utilizzando un kit Truseq SBS-HS da 50 coppie di basi.

Sequenziamento di nuova generazione, sequenziamento dell'RNA (RNA-Seq): librerie indicizzate di cDNA da mRNA-Seq sono state preparate da 5

µg di RNA totale utilizzando una versione modificata del protocollo di mRNA-seq di Illumina. L'mRNA è stato isolato utilizzando sfere magnetiche Seramag Oligo(dT) (Thermo Scientific). L'mRNA isolato è stato frammentato utilizzando cationi bivalenti e calore risultando in frammenti variabili da 160 a 220 coppie di basi. L'mRNA frammentato è stato convertito in cDNA utilizzando inneschi casuali e SuperScriptII (Invitrogen) cui segue la sintesi del secondo filamento utilizzando la DNA polimerasi I e l'RNasi H. Il cDNA è stato riparato alle estremità utilizzando la DNA polimerasi T4, la DNA polimerasi di Klenow e fosforilato al 5' utilizzando la chinasi polinucleotidica T4. I frammenti di cDNA con estremità non appuntite sono stati adenilati al 3' utilizzando il frammento di Klenow (da 3' a 5' eso meno). Adattatori multipli specifici con singola sporgenza a T al 3' di Illumina sono stati ligati utilizzando un rapporto molare di 10:1 di adattatore e inserto di cDNA utilizzando la DNA ligasi T4.

Le librerie di cDNA sono state purificate e selezionate per dimensione da 200 a 220 coppie di basi utilizzando il gel E-Gel SizeSelect al 2% (Invitrogen). L'arricchimento, l'aggiunta di un indice a sei basi di Illumina e di sequenze specifiche di cellule del flusso sono stati fatti mediante PCR utilizzando la DNA polimerasi Phusion (Finnzymes). Tutti i passaggi di rimozione sono stati fatti utilizzando un volume 1,8x di sfere magnetiche AgencourtAMPure XP. Tutti i controlli di qualità sono stati fatti utilizzando un saggio Qubit HS di Invitrogen e la dimensione dei frammenti è stata determinata utilizzando un saggio del DNA 2100 Bioanalyzer HS di Agilent.

Le librerie di RNA-Seq indicizzate sono state raggruppate sul cBot utilizzando il kit Truseq SR cluster v2.5 utilizzando 7pM e 50 coppie di basi sono state sequenziate sull'Illumina HiSeq2000 utilizzando un kit Truseq SBS-HS da 50 coppie di basi.

CTC: per l'ottenimento del profilo di RNA-Seq delle CTC, è stata utilizzata una versione modificata di questo protocollo in cui da 500 ng a 700 ng di cDNA amplificato mediante SMART sono stati utilizzati, adattatori con estremità appaiate sono stati ligati e l'arricchimento mediante PCR è stato fatto utilizzando inneschi di PCR Illumina PE 1.0 e 2.0.

Analisi dei dati di NGS, espressione genica: per determinare i valori di espressione, le letture di sequenza in uscita da campioni di RNA dall'Illumina HiSeq 2000 sono state preprocessate secondo il protocollo standard di Illumina. Questo comprende il filtraggio di letture a bassa qualità e la demultiplazione. Per l'analisi trascrittomico mediante RNA-Seq, le letture di sequenza sono state allineate alla sequenza genomica di riferimento [Mouse Genome Sequencing Consortium. Initial sequencing and comparative analysis of the mouse genome. Nature, 420, da 520 a 562 (2002)] utilizzando bowtie (versione 0.12.5) [Langmead B. et al. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10: R25] utilizzando parametri "-v2-best" per gli allineamenti del genoma e parametri preimpostati per gli allineamenti dei trascritti. Le coordinate di allineamento sono state confrontate alle coordinate esoniche dei trascritti di RefSeq [Pruitt KD. et al. NCBI Reference

Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Res. 1 gennaio 2005; 33 (uscita della banca dati): da D501 a D504] e per ciascun trascritto le conte degli allineamenti sovrapposti sono state registrate. Le letture delle sequenze non allineabili alla sequenza genomica sono state allineate ad una banca dati di tutte le possibili sequenze di giunzione esone-esone dei trascritti di RefSeq. Le conte delle letture che si allineano alle giunzioni di splicing sono state aggregate alle rispettive conte dei trascritti ottenute nel passaggio precedente e normalizzate rispetto a RPKM (numero di letture che mappano per kilobase del modello esonico per milione di letture mappate [Mortazavi, A. et al. (2008). Mapping and quantifying mammalian transcriptomes by rna-seq. Nat. Methods, 5(7): da 621 a 628]) per ciascun trascritto. Sia i valori di espressione genica che di espressione esonica sono stati calcolati sulla base del numero normalizzato di letture sovrapposte a ciascun gene o esone, rispettivamente.

Scoperta delle mutazioni, massa di tumore: letture di 50 nucleotidi a singola estremità dall'Illumina HiSeq 2000 sono state allineate utilizzando bwa (versione 0.5.8c) [Li H. e Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler Transform. Bioinformatics, 25: da 1754 a 1760] utilizzando opzioni preimpostate al gruppo genomico mm9 di riferimento di topo. Le letture ambigue - quelle letture mappate in posizioni multiple del genoma - sono state rimosse, gli allineamenti rimanenti sono stati separati, indicizzati e convertiti in un formato binario e compresso (BAM) ed i punteggi della qualità di lettura convertiti dai punteggi di qualità dell'Illumina standard phred+64 a quelli del Sanger standard utilizzando script a conchiglia.

Per ciascuna corsia di sequenziamento, le mutazioni sono state identificate utilizzando tre programmi informatici: compresi samtools (versione 0.1.8) [Li H. Improving SNP discovery by base alignment quality. Bioinformatics. 15 Aprile 2011; 27(8): da 1157 a 1158. Pubblicazione elettronica 13 febbraio 2011], GATK (versione 1.0.4418) [McKenna A. et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. Settembre 2010; 20(9): da 1297 a 1303. Pubblicazione elettronica 19 luglio 2010], e SomaticSniper (<http://genome.wustl.edu/software/somaticsniper>). Per samtools, sono stati utilizzati le opzioni e i criteri dei filtri raccomandati dall'autore, compreso un primo ciclo di filtraggio, copertura massima 200. Per un secondo ciclo di filtraggio di samtools, il minimo punteggio di qualità degli indel è stato 50, la qualità minima delle mutazioni puntiformi è stata 30. Per chiamare le mutazioni con GATK, noi abbiamo seguito le linee guida della pratica migliore progettate dall'autore sul manuale dell'utente di GATK (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit). Il passaggio di ricalibrazione del punteggio delle varianti è stato omesso e sostituito con l'opzione di filtraggio stringente. Per chiamare le mutazioni con SomaticSniper, sono state utilizzate le opzioni preimpostate e solo le mutazioni predette con un "punteggio somatico" di 30 o superiore sono state considerate ulteriormente.

Scoperta delle mutazioni, CTC: come per il procedimento iCAM della massa di tumore, letture di 50 nucleotidi a singola estremità dall'Illumina HiSeq 2000 sono state allineate utilizzando bwa (versione 0.5.8c) [5]) utilizzando opzioni preimpostate al gruppo genomico mm9 di riferimento di topo. Poiché le letture di NGS delle CTC sono state derivate dal saggio di RNA-Seq, le letture sono state anche allineate alle sequenze trascrittomiche, comprese le giunzioni esone-esone, utilizzando bowtie (sopra). Utilizzando tutti gli allineamenti, le sequenze nucleotidiche dalle letture sono state confrontate sia con il genoma di riferimento che con le mutazioni di B16 derivate dalla massa tumorale. Le mutazioni identificate sono state valutate sia utilizzando gli script perl che manualmente utilizzando il programma informatico samtools e l'IGV (Visualizzatore Integrato del Genoma) per visualizzare i risultati. Il risultato della "scoperta delle mutazioni" è l'identificazione di mutazioni somatiche nelle cellule tumorali, dal campione ai dati di NGS ad un elenco di mutazioni. Nei campioni di B16, abbiamo identificato 2448 mutazioni somatiche utilizzando il risequenziamento dell'esoma.

Prioritizzazione delle mutazioni

Poi, noi dimostriamo una possibilità di una linea diretta per la prioritizzazione delle mutazioni per l'inclusione del vaccino. Questo metodo, chiamato "pipeline per il rilevamento di singole mutazioni tumorali" (iCAM) identifica e priorizza le mutazioni somatiche attraverso una serie di passaggi incorporanti multipli algoritmi e metodi bioinformatici innovativi. Il risultato di questo procedimento è un elenco di mutazioni somatiche, priorizzate sulla base della probabile immunogenicità.

Identificazione delle mutazioni somatiche: le mutazioni vengono identificate utilizzando tre algoritmi differenti, per entrambi i campioni di B16 e di Black6 (scoperta delle mutazioni, sopra). Il primo passaggio di iCAM è combinare gli elenchi dei risultati da ciascun algoritmo per generare un elenco ad elevata confidenza di mutazioni somatiche. GATK e samtools riportano varianti in un campione rispetto ad un genoma di riferimento. Per selezionare mutazioni ad elevata confidenza con pochi falsi positivi per un dato campione (cioè, tumore o normale), vengono selezionate mutazioni che vengono identificate in tutte le repliche. Quindi, vengono scelte varianti che sono presenti nel campione tumorale ma non presenti nel campione normale. SomaticSniper riporta automaticamente potenziali variazioni somatiche da coppie di dati tumorali e normali. Noi abbiamo ulteriormente filtrato i risultati attraverso l'intersezione dei risultati ottenuti dalle repliche. Per rimuovere più chiamate false positive possibili, noi abbiamo incrociato l'elenco di mutazioni derivate dall'uso di tutti e tre gli algoritmi e da tutte le repliche. Il passaggio finale per ciascuna mutazione somatica è assegnare un valore di confidenza (valore di p) per ciascuna mutazione sulla base della profondità di copertura, della qualità degli SNP, della qualità del consenso e della qualità di mappatura.

Impatto delle mutazioni: l'impatto delle mutazioni somatiche consenso filtrate viene determinato mediante uno script all'interno della pipeline

iCaM per le mutazioni. Innanzitutto, le mutazioni che avvengono in regioni genomiche che non sono uniche all'interno del genoma, come avvengono per alcuni paraloghi a proteine e pseudogeni, sono escluse dall'analisi poiché le letture di sequenza che si allineano a posizioni multiple vengono rimosse. Secondo, viene determinato se la mutazione avvenga in un trascritto. Terzo, viene determinato se la mutazione avvenga in una regione codificante per proteine. Quarto, la sequenza del trascritto è tradotta con e senza la mutazione per determinare se vi sia una modifica nella sequenza amminoacidica.

Espressione delle mutazioni: la pipeline di iCaM seleziona mutazioni somatiche che si trovano in geni e in esoni che vengono espressi nelle cellule tumorali. I livelli di espressione vengono determinati attraverso RNA-Seq di NGS di cellule tumorali (sopra). Il numero di letture che si sovrappongono ad un gene e ad un esone indica i livelli di espressione. Queste conte sono normalizzate rispetto a RPKM (letture per kilobase del modello esonico per milione di letture mappate, [Mortazavi A. et al. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods. Luglio 2008; 5(7): da 621 a 628. Pubblicazione elettronica 30 maggio 2008]) e quelle espresse oltre 10 RPKM vengono scelte.

Legame all'MHC: per determinare la probabilità che un epitopo contenente il peptide mutato si leghi ad una molecola dell'MHC, la pipeline di iCaM processa una versione modificata del programma di predizione dell'MHC dalla banca dati degli epitopi immunitari (<http://www.iedb.org/>). L'installazione locale comprende modifiche per ottimizzare il flusso di dati attraverso l'algoritmo. Per i dati di B16 e di Black6, la predizione è stata eseguita utilizzando tutti gli alleli disponibili dell'MHC di classe I di Black6 e tutti gli epitopi per le rispettive lunghezze peptidiche. Vengono scelte mutazioni che cadono in un epitopo classificato nel 95mo percentile della distribuzione del punteggio di predizione dei dati di addestramento di IEDB (<http://mhcbindingpredictions.immuneepitope.org/dataset.html>), considerando tutti gli alleli dell'MHC e tutti gli epitopi potenziali che si sovrappongono alla mutazione.

Criteri di selezione delle mutazioni: le mutazioni somatiche vengono scelte mediante i seguenti criteri: a) hanno un contenuto di sequenza unico, b) identificate mediante tutti e tre i programmi, c) elevata confidenza di mutazione, d) cambiamento non sinonimo della proteina, e) espressione elevata del trascritto, f) e predizione di legame favorevole dell'MHC di classe I.

Il risultato di questo procedimento è un elenco di mutazioni somatiche, priorizzate sulla base della probabile immunogenicità. Nelle cellule B16 di melanoma, vi sono 2448 mutazioni somatiche. 1247 di queste mutazioni si trovano in trascritti genici. Di queste, 734 causano variazioni non sinonime delle proteine. Di queste, 149 sono in geni espressi nelle cellule tumorali. Di queste, si prevede che 102 di queste mutazioni espresse non sinonime siano presentate sulle molecole dell'MHC. Queste 102 mutazioni probabilmente immunogeniche sono quindi passate alla conferma delle mutazioni (sotto).

Conferma delle mutazioni

Le mutazioni somatiche dal risequenziamento esomico del DNA sono state confermate mediante ciascuno dei due metodi, il risequenziamento della regione mutata e l'analisi dell'RNA-Seq.

Per la conferma delle mutazioni mediante il risequenziamento, una regione genomica contenente la mutazione è stata amplificata mediante PCR standard da 50 ng di entrambi il DNA tumorale ed il DNA normale di controllo. La dimensione dei prodotti amplificati è stata nell'intervallo da 150 a 400 nucleotidi. La specificità della reazione è stata controllata caricando il prodotto di PCR sul dispositivo Qiagel (Qiagen). I prodotti di PCR sono stati purificati utilizzando il kit di purificazione per PCR minElute (Qiagen). Specifici prodotti di PCR sono stati sequenziati utilizzando il metodo standard di sequenziamento di Sanger (Eurofins), seguito dall'analisi dell'elettroferogramma.

La conferma delle mutazioni è stata anche realizzata attraverso l'esame dell'RNA tumorale. I valori di espressione dei geni e degli esoni tumorali sono stati generati dall'RNA-Seq (NGS di RNA), il quale genera sequenze nucleotidiche che sono state mappate a trascritti e contate. Abbiamo esaminato i dati stessi delle sequenze per identificare mutazioni nel campione tumorale [Berger MF. et al. Integrative analysis of the melanoma transcriptome. Genome Res. Aprile 2010; 20(4): da 413 a 427. Pubblicazione elettronica 23 febbraio 2010], fornendo una conferma indipendente delle mutazioni somatiche identificate derivate dal DNA.

Tabella 1: elenco di geni contenenti le 50 mutazioni convalidate

Geni contenenti le 50 mutazioni somatiche identificate e confermate, con l'annotazione riguardante il simbolo del gene, il nome del gene, e la localizzazione e la funzione previste.

ID	simbolo	nome del gene di Entrez	posizione
NM_021895	ACTN4	actinina, alfa 4	citoplasma
NM_028840	ARMC1	1 contenente una ripetizione di armadillo	non nota
NM_029291	ASCC2	subunità 2 del complesso 1 cointegratore del segnale di attivazione	non nota
NM_024184	ASF1B	omologo B della funzione 1 di antisilenziamento di ASF1 (S. cerevisiae)	nucleo
NM_138679	ASH1L	simile ad ash1 (assente, piccolo, o omeotico) (Drosophila)	nucleo

NM_015804	ATP11A	ATPasi, classe VI, tipo 11A	membrana plasmatica
NM_009730	ATRN	attrattina	spazio extracellulare
NM_028020	CPSF3L	simile al fattore 3 specifico per la scissione e per la poliadenilazione	nucleo
NM_010017	DAG1	distroglicano 1 (glicoproteina 1 associata alla distrofina)	membrana plasmatica
NM_015735	DDB1	proteina 1 legante il DNA specifica per il danno, 127 kDa	nucleo
NM_001080981	DDX23	polipeptide 23 con DEAD (Asp-Glu-Ala-Asp) box	nucleo
NM_054046	DEF8	omologo 8 differenzialmente espresso in FDCP (topo)	non nota
NM_019965	DNAJB12	omologo a DnaJ (Hsp40), sottofamiglia B, membro 12	citoplasma
NM_011262	DPF2	D4, famiglia 2 a dita di zinco e doppio PHD	nucleo
NM_007907	EEF2	fattore 2 eucariotico di allungamento della traduzione	citoplasma
NM_001081286	FAT1	omologo 1 del soppressore tumorale FAT (Drosophila)	membrana plasmatica
NM_173182	FNDC3B	dominio fibronectinico di tipo III contenente 3B	non nota
NM_008057	FZD7	omologo 7 di frizzled (Drosophila)	membrana plasmatica
NM_201617	GNAS	locus del complesso GNAS	membrana plasmatica
NM_030035	GOLGB1	golgina B1	citoplasma
NM_011365	ITSN2	intersettina 2	citoplasma
NM_029841	KIAA2013	KIAA2013	non nota
NM_197959	KIF18B	membro 18B della famiglia chinesinica	non nota
NM_145479	KLHL22	simile a kelch 22 (Drosophila)	non nota
NM_018810	MKRN1	proteina 1 con dito ad anello di makorina	non nota
NM_001170785	MTHFD1L	simile alla metilentetraidrofolato deidrogenasi (dipendente dal NADP+) 1	citoplasma
NM_133947	NUMA1	proteina 1 dell'apparato mitotico nucleare	nucleo

NM_178884	OBSL1	simile all'oscurina 1	non nota
NM_008765	ORC2	complesso di riconoscimento dell'origine, subunità 2	nucleo
NM_023209	PBK	chinasi legante PDZ	citoplasma
NM_033594	PCDHGA11	sottofamiglia A della protocaderina gamma, 11	membrana plasmatica
NM_025951	PI4K2B	fosfatidilinositolo 4-chinasi tipo 2 beta	citoplasma
NM_011961	PLOD2	procollagene-lisina, 2-ossoglutarato 5-diossigenasi 2	citoplasma
NM_023200	PPP1R7	proteinfosfatasi 1, subunità regolatoria (inibitrice) 7	nucleo
NM_008986	PTRF	polimerasi I e fattore di rilascio dei trascritti	nucleo
NM_011240	RANBP2	proteina 2 legante RAN	nucleo
NM_009438	RPL13A	proteina ribosomiale L13a	citoplasma
NM_009113	S100A13	proteina S100 A13 che lega il calcio	citoplasma
NM_001081203	SBNO1	omologo 1 di strawberry notch (Drosophila	non nota
NM_009153	SEMA3B	dominio sema, dominio immunoglobulinico (Ig), breve dominio basico, secreto (semaforina) 3B	spazio extracellulare
NM_026912	SNX15	nessina 15 di separazione	citoplasma
NM_024225	SNX5	nessina 5 di separazione	citoplasma
NM_008188	THUMPD3	3 contenente il dominio THUMP	non nota
NM_133352	TM9SF3	membro 3 della superfamiglia transmembrana 9	citoplasma
NM_177296	TNPO3	trasportina 3	citoplasma
NM_011640	TP53	proteina tumorale p53	nucleo
NM_023279	TUBB3	tubulina, beta 3	citoplasma
NM_029896	WDR82	dominio di ripetizione WD 82	non nota
NM_025830	WWP2	proteina ligasi 2 dell'ubiquitina E3 contenente il dominio WW	citoplasma

NM_001081056	XPOT	esportina, tRNA (recettore di esportazione nucleare per nucleo tRNA)
--------------	------	--

Esempio 2: l'algoritmo di selezione IVAC permette il rilevamento di mutazioni immunogeniche

Per indagare se risposte specifiche dei linfociti T potessero essere indotte nei confronti delle mutazioni confermate da cellule di melanoma B16-F10, topi vergini C57BL/6 (n=5/peptide) sono stati immunizzati due volte (giorno 0, giorno 7) per via sottocutanea con 100 µg di peptide (+50 µg di poli(I:C) come adiuvante) comprendente la sequenza amminoacidica mutata o selvatica (si veda la tabella 2). Tutti i peptidi hanno avuto una lunghezza di 27 amminoacidi con l'amminoacido mutato/selvatico in posizione centrale. Il giorno 12 i topi sono stati sacrificati e le cellule della milza sono state raccolte. Come metodo di lettura è stato eseguito un ELISPOT per IFN-γ utilizzando 5x10⁵ cellule della milza/pozzetto come effettori e 5x10⁴ cellule dendritiche di midollo osseo caricate con peptidi (2 µg/ml) come cellule bersaglio. Le cellule effettrici di milza sono state testate nei confronti del peptide mutato, del peptide selvatico e di un peptide di controllo (nucleoproteina del virus della stomatite vescicolare, VSV-NP). Con 44 sequenze testate abbiamo osservato che 6 di esse hanno indotto un'immunità dei linfociti T diretta solo contro la sequenza mutata ma non contro il peptide selvatico (figura 3). I dati dimostrano che le mutazioni identificate e priorizzate possono essere utilizzate per indurre l'immunità specifica dei linfociti T contro il tumore dopo essere state utilizzate come vaccino peptidico nei topi vergini all'antigene.

Tabella 2: elenco di sequenze mutate che hanno indotto una reattività dei linfociti T specifica per il peptide mutato contro il peptide selvatico. Lo scambio amminoacidico è marcato con la sottolineatura.

numero	ID RefSeq	tipo di sequenza	sequenza peptidica	reattività dei linfociti T (topi)
12	NM_00107750,	mutata	TPPPEEAMPFEFNG <u>PA</u> QGDHSQPPLQV	5/5
	NM_010309, NM_201618, NM_201617			
16	NM_008188	selvatica	TPPPEEAMPFEFNE <u>PA</u> QGDHSQPPLQV	0/5
		mutata	RVTCNRAGEKHCF <u>S</u> SNEAARDFGGAIQ	3/5
		selvatica	RVTCNRAGEKHCF <u>S</u> SNEAARDFGGAIQ	0/5

20	NM_023279	mutata	FRRKAFLHWYTGE <u>A</u> MDEMEFTEAESNM	5/5
		selvatica	FRRKAFLHWYTGE <u>G</u> MDEMEFTEAESNM	1/5
30	NM_197959	mutata	PSKPSFQEFVDWEN <u>V</u> SPELNSTDQPFL	5/5
		selvatica	PSKPSFQEFVDWE <u>K</u> VSPELNSTDQPFL	1/5
34	NM_145479	mutata	HLTQQLDTYILKN <u>V</u> AFSRTDKYRQLP	3/5
		selvatica	HLTQQLDTYILKN <u>F</u> AFSRTDKYRQLP	0/5
36	NM_133352	mutata	CGTAFFINFAIY <u>H</u> HASRAIPFGTMVA	5/5
		selvatica	CGTAFFINFAIY <u>I</u> HASRAIPFGTMVA	0/5

Esempio 3: le mutazioni identificate possono fornire un'immunità antitumorale terapeutica

Allo scopo di convalidare se le mutazioni identificate abbiano il potenziale di conferire l'immunità antitumorale dopo la vaccinazione a topi non trattati abbiamo investigato questa domanda con il peptide per la mutazione numero 30 che è stata mostrata indurre una reattività dei linfociti T selettiva per la mutazione. Le cellule B16-F10 ($7,5 \times 10^4$) sono state inoculate per via sottocutanea il giorno 0. I topi sono stati vaccinati con il peptide 30 (si veda la tabella 1; 100 μ g di peptide+50 μ g di poli(I:C) s.c.) il giorno -4, il giorno +2, e il giorno +9. Il gruppo di controllo ha ricevuto solo poli(I:C) (50 μ g s.c.). La crescita tumorale è stata monitorata a giorni alterni. Il giorno +16 abbiamo osservato che solo 1 topo su 5 nel gruppo del vaccino peptidico aveva sviluppato un tumore mentre nel gruppo di controllo 4 topi su 5 hanno mostrato una crescita tumorale.

I dati dimostrano che una sequenza peptidica che incorpora una mutazione specifica di B16-F10 può conferire un'immunità antitumorale che è efficientemente in grado di distruggere le cellule tumorali (si veda la figura 4). Poiché B16-F10 è una linea cellulare tumorale altamente aggressiva la scoperta che la metodologia applicata per identificare e per prioritizzare le mutazioni ha condotto infine alla selezione di una mutazione che è già potente da sola come vaccino è un'importante verifica teorica per l'intero procedimento.

Esempio 4: dati che supportano la presentazione dell'antigene polipeptidico

Mutazioni convalidate da regioni codificanti per le proteine di un paziente costituiscono il gruppo da cui i candidati possono essere scelti per l'assemblaggio del modello del vaccino polineoepitopico da utilizzare come precursore per la produzione in GMP del vaccino a RNA. Cassette vettoriali adatte come struttura del vaccino erano già state descritte (Holtkamp, S. et al., Blood, 108: da 4009 a 4017, 2006; Kreiter, S. et al., Cancer Immunol. Immunother., 56: da 1577 a 1587, 2007; Kreiter, S. et al., J. Immunol., 180: da 309 a 318, 2008). Le cassette vettoriali preferite sono modificate nelle loro regioni codificanti e non tradotte (UTR) e garantiscono una traduzione massimizzata della proteina codificata per periodi

prolungati (Holtkamp, S. et al., Blood, 108: da 4009 a 4017, 2006; Kuhn, A. N. et al., Gene Ther., 17: da 961 a 971, 2010). Inoltre, la struttura vettoriale contiene moduli di direzionamento dell'antigene per l'espansione simultanea di linfociti T citotossici così come helper (Kreiter, S. et al., Cancer Immunol. Immunother., 56: da 1577 a 1587, 2007; Kreiter, S. et al., J. Immunol., 180: da 309 a 318, 2008; Kreiter, S. et al., Cancer Research, 70 (22), da 9031 a 9040, 2010 (figura 5). In modo importante, abbiamo dimostrato che tale vaccino a RNA può essere utilizzato per presentare multipli epitopi dell'MHC di classe I e di classe II simultaneamente.

Le sequenze di vaccino polineoepitopico ad RNA di IVAC sono costruite da filamenti di fino a 30 amminoacidi che includono la mutazione nel centro. Queste sequenze sono connesse testa-coda attraverso brevi gruppi di collegamento per formare un vaccino polineoepitopico codificante per fino a 30 o più mutazioni scelte e per le loro regioni fiancheggianti. Questi inserti individualmente personalizzati specifici per un paziente sono a codoni ottimizzati e clonati nella struttura di RNA descritta sopra. Il controllo di qualità di tali costrutti comprende la trascrizione e l'espressione in vitro nelle cellule per la convalida della trascrizione e della traduzione funzionali. L'analisi della traduzione sarà eseguita con anticorpi nei confronti del dominio carbossiterminale di indirizzamento a bersaglio.

Esempio 5: verifica teorica scientifica per il costrutto polineoepitopico di RNA

Il concetto di polineoepitopo di RNA si basa su un lungo mRNA trascritto in vitro che consiste di sequenze disposte consecutivamente codificanti per i peptidi mutati connesse mediante sequenze di collegamento (si veda la figura 6). Le sequenze codificanti vengono scelte tra le mutazioni non sinonime e sono sempre costruite del codone per l'amminoacido mutato fiancheggiato da regioni da 30 a 75 coppie di basi dal contesto della sequenza originale. La sequenza di collegamento codifica per gli amminoacidi che sono preferibilmente non trattati dal macchinario cellulare di processamento degli antigeni. I costrutti per la trascrizione in vitro si basano sul vettore pST1-A120 contenente un promotore T7, una sequenza tandem di beta-globina al 3' UTR ed una coda di poli(A) di 120 coppie di basi, i quali sono stati mostrati incrementare la stabilità e l'efficienza traduzionale dell'RNA aumentando quindi la capacità stimolatoria dei linfociti T dell'antigene codificato (Holtkamp S. et al., Blood 2006; PMID: 16940422). Inoltre, un frammento di peptide segnale dell'MHC di classe I ed i domini transmembrana e citosolico comprendenti il codone di arresto (segnale di movimentazione dell'MHC di classe I o MITD) fiancheggianti una sequenza di policollegamento per clonare gli epitopi sono stati inseriti (Kreiter S. et al., J. Immunol., 180: da 309 a 318, 2008). È stato mostrato che gli ultimi incrementano la presentazione dell'antigene, aumentando quindi l'espansione di linfociti T CD8+ e CD4+ specifici per l'antigene e migliorando le funzioni effettrici.

Per una prima verifica teorica, sono stati utilizzati vettori biepitopici, cioè codificanti per un polipeptide contenente due epitopi mutati. Le sequenze a codoni ottimizzati codificanti per (i) un epitopo mutato da 20 a 50 amminoacidi, (ii) un gruppo di collegamento ricco in glicina/serina, (iii) un

secondo epitopo mutato da 20 a 50 amminoacidi, e (iv) un ulteriore gruppo di collegamento ricco in glicina/serina - fiancheggiati da siti di riconoscimento adatti per le endonucleasi di restrizione per essere clonati nel costrutto basato su pST1 come descritto sopra - sono state progettate e sintetizzate da un fornitore commerciale (Geneart, Regensburg, Germania). Dopo la verifica della sequenza, queste sono state clonate nella struttura del vettore basato su pST1 per ottenere costrutti come descritto nella figura 6.

I plasmidi basati su pST1-A120 come descritto sopra sono stati linearizzati con un'endonucleasi di restrizione di classe IIs. I DNA plasmidici linearizzati sono stati purificati mediante estrazione con fenolo cloroformio e precipitazione con etanolo. I DNA vettoriali linearizzati sono stati quantificati spettrofotometricamente e sottoposti a trascrizione in vitro essenzialmente come descritto da Pokrovskaya e Gurevich (1994, Anal. Biochem. 220: da 420 a 423). Un analogo del cappuccio è stato aggiunto alla reazione di trascrizione per ottenere RNA con le strutture al cappuccio al 5' modificate in modo corrispondente. Nelle reazioni, il GTP era presente a 1,5mM, mentre l'analogo del cappuccio era presente a 6,0mM. Tutti gli altri NTP erano presenti a 7,5mM. Alla fine della reazione di trascrizione, il DNA del vettore linearizzato è stato digerito con 0,1 U/ μ l di TURBO DNasi (Ambion, Austin/Texas, Stati Uniti) per 15 minuti a 37°C. Gli RNA sono stati purificati da queste reazioni utilizzando il kit MEGAclean (Ambion, Austin/Texas, Stati Uniti) come da protocollo del produttore. La concentrazione e la qualità dell'RNA sono state valutate mediante spettrofotometria ed analisi su un bioanalizzatore 2100 (Agilent, Santa Clara, California, Stati Uniti).

Allo scopo di provare che una sequenza che incorpora un amminoacido mutato e che è fiancheggiata al 5' così come al 3' dalla sequenza di collegamento può essere processata, presentata e riconosciuta dai linfociti T specifici per un antigene abbiamo utilizzato i linfociti T da topi vaccinati con il peptide come cellule effettrici. In un ELISPOT per IFN- γ abbiamo testato se i linfociti T indotti dalla vaccinazione peptidica come descritto sopra sono in grado di riconoscere le cellule bersaglio (cellule dendritiche di midollo osseo, BMDC) stimulate con il peptide (2 μ g/ml per 2 ore a 37°C e CO₂ al 5%) o trasfettate con l'RNA (20 μ g prodotti come descritto sopra) mediante elettroporazione. Come esemplificato nella figura 7 per le mutazioni 12 e 30 (si veda la tabella 2) abbiamo potuto osservare che il costrutto di RNA è in grado di dare origine all'epitopo riconosciuto dai linfociti T specifici per le mutazioni.

Con i dati forniti abbiamo potuto dimostrare che un polineoepitopo codificato dall'RNA comprendente un gruppo di collegamento ricco in glicina/serina può essere tradotto e processato nelle cellule presentanti l'antigene conducendo alla presentazione del corretto epitopo che è riconosciuto dai linfociti T specifici per un antigene.

Esempio 6: progettazione di vaccini polineoepitopici - la rilevanza del gruppo di collegamento

Il costrutto polineoepitopico di RNA contiene un costrutto di sostegno in cui vengono posti multipli peptidi codificanti per mutazioni somatiche

connessi con una sequenza peptidica di collegamento. Oltre all'ottimizzazione dei codoni e ad una stabilità e un'efficienza traduzionale aumentate dell'RNA a causa della struttura, una forma di realizzazione del vaccino polineoepitopico a RNA contiene gruppi di collegamento progettati per incrementare la presentazione di peptidi antigenici all'MHC di classe I e II e per diminuire la presentazione di epitopi dannosi.

Gruppo di collegamento: la sequenza di collegamento è stata progettata per collegare multipli peptidi contenenti mutazioni. Il gruppo di collegamento dovrebbe permettere la creazione e la presentazione dell'epitopo con la mutazione mentre impedire la creazione di epitopi dannosi, quali quelli generati alla sutura di giunzione tra peptidi adiacenti o tra la sequenza di collegamento e i peptidi endogeni. Questi epitopi "di giunzione" possono non solo competere con gli epitopi intesi per essere presentati sulla superficie cellulare, diminuendo l'efficacia del vaccino, ma potrebbero generare una reazione autoimmune indesiderata. Così, abbiamo progettato la sequenza di collegamento per a) evitare di creare peptidi "di giunzione" che si legano a molecole dell'MHC, b) evitare il processamento proteasomico per creare peptidi "di giunzione", c) essere efficientemente tradotti e processati dal proteasoma.

Per evitare la creazione di peptidi "di giunzione" che si legano a molecole dell'MHC, abbiamo confrontato sequenze di collegamento differenti. La glicina, per esempio, inibisce il forte legame nelle posizioni dei solchi di legame all'MHC [Abastado JP. et al., J. Immunol. 1 ottobre 1993; 151(7): da 3569 a 3575]. Abbiamo esaminato sequenze multiple del gruppo di collegamento e lunghezze multiple del gruppo di collegamento e calcolato il numero di peptidi "di giunzione" che si legano a molecole dell'MHC. Abbiamo utilizzato strumenti informatici dalla banca dati degli epitopi immunitari (IEDB, <http://www.immuneepitope.org/>) per calcolare la probabilità che una data sequenza peptidica contenga un ligando che si legherà alle molecole dell'MHC di classe I.

Nel modello B16, abbiamo identificato 102 mutazioni somatiche non sinonime espresse previste per essere presentate sulle molecole dell'MHC di classe I. Utilizzando le 50 mutazioni confermate, abbiamo progettato in modo computazionale costrutti vaccinici differenti, compresi l'uso di nessun gruppo di collegamento o l'uso di sequenze di collegamento differenti, e calcolato il numero di peptidi "di giunzione" dannosi utilizzando l'algoritmo IEDB (figura 8).

La tabella 5 mostra i risultati di diversi gruppi di collegamento differenti, lunghezze del gruppo di collegamento differenti, e l'uso di nessun gruppo di collegamento e di cinque gruppi di collegamento. Il numero di peptidi di giunzione che legano l'MHC varia da 2 a 91 per le predizioni degli epitopi di 9 amminoacidi e di 10 amminoacidi (parte superiore e intermedia). La dimensione del gruppo di collegamento influenza il numero di peptidi di giunzione (parte inferiore). Per questa sequenza, la minore quantità di epitopi di 9 amminoacidi è prevista per la sequenza di collegamento di 7 amminoacidi GGSGGGG.

Il gruppo di collegamento 1 e il gruppo di collegamento 2 utilizzati nei costrutti vaccinici polineopeptidici a RNA testati sperimentalmente (si veda sotto) avevano anche avuto un numero favorevolmente basso di neoepitopi giunzionali previsti. Questo è vero per le predizioni di enameri e di decameri.

Questo dimostra che la sequenza del gruppo di collegamento è criticamente importante per la creazione di cattivi epitopi di legame all'MHC. Inoltre, la lunghezza della sequenza di collegamento influenza il numero di cattivi epitopi di legame all'MHC. Troviamo che sequenze che sono ricche in G impediscono la creazione di ligandi di legame all'MHC.

Tabella 3: impatto del gruppo di collegamento (epitopi di 10 amminoacidi). Numero previsto di cattivi epitopi definiti come epitopi di legame all'MHC di classe I che contengono sequenze di giunzione, per ciascun gruppo di collegamento peptidico. Qui, sono considerati gli epitopi di 10 amminoacidi. I gruppi di collegamento ricchi in glicina hanno la minore quantità di epitopi di giunzione.

gruppo di collegamento	# di cattivi epitopi (10 amminoacidi)
nessuno	14
TSLNALLNAH	54
SIINFEKL	65
SSSSSSSSS	85
GGGGGGGGG	6
GGSGGGSGG (gruppo di collegamento 1)	8
GGSGGGSGG (gruppo di collegamento 2)	9

Tabella 4: impatto della parte del gruppo di collegamento (epitopi di 9 amminoacidi). Numero previsto di cattivi epitopi, definiti come epitopi di legame all'MHC di classe I che contengono sequenze di giunzione, per ciascun gruppo di collegamento peptidico. Qui, sono considerati gli epitopi di 9 amminoacidi. I gruppi di collegamento ricchi in glicina hanno la minore quantità di epitopi di giunzione.

gruppo di collegamento	# di cattivi epitopi (9 amminoacidi)
nessuno	17
TSLNALLNAH	83
SIINFEKL	64

SSSSSSSSS	33
GGGGGGGGG	2
GGSGGGSGG (gruppo di collegamento 1)	4
GGSGGGSGG (gruppo di collegamento 2)	3

Tabella 5: impatto della parte del gruppo di collegamento. Numero previsto di cattivi epitopi, definiti come epitopi di legame all'MHC di classe I che contengono sequenze di giunzione, per ciascun gruppo di collegamento peptidico. Qui, sono considerati gli epitopi di 9 amminoacidi. Parte superiore: numero di epitopi di giunzione di 9 amminoacidi per nessun gruppo di collegamento e per 5 diversi gruppi di collegamento. Parte intermedia: numero di epitopi di giunzione di 10 amminoacidi per nessun gruppo di collegamento e per 5 diversi gruppi di collegamento. Parte inferiore: numero di epitopi di giunzione di 99 amminoacidi per gruppi di collegamento simili di lunghezze differenti. I gruppi di collegamento ricchi in glicina hanno la minore quantità di epitopi di giunzione.

sequenza di collegamento	# di epitopi di giunzione (9 amminoacidi)
nessuna	17
TSLNALLNA	91
SIINFEKL	64
SSSSSSSSS	33
GGGGGGGGG	2
GGSGGGSGG	4

sequenza di collegamento	# di epitopi di giunzione (10 amminoacidi)
nessuna	14
TSLNALLNA	63
SIINFEKL	65
SSSSSSSSS	85
GGGGGGGGG	6
GGSGGGSGG	9

sequenza di collegamento	# di epitopi di giunzione (9 amminoacidi)
GGSGG	5
GGSGGG	4
GGSGGGG	2
GGSGGGGS	7
GGSGGGGSG	4
GGSGGGGSGG	4

Per evitare il processamento proteasomico che può creare peptidi “di giunzione”, abbiamo esplorato l’uso di amminoacidi differenti nel gruppo di collegamento. Le sequenze ricche in glicina danneggiano il processamento proteasomico [Hoyt MA et al. (2006). EMBO J 25 (8): da 1720 a 1729; Zhang M. e Coffino P. (2004) J. Biol. Chem. 279 (10): da 8635 a 8641]. Così sequenze di collegamento ricche in glicina agiscono per minimizzare il numero di peptidi contenenti un gruppo di collegamento che possono essere processati dal proteasoma.

Il gruppo di collegamento dovrebbe permettere ai peptidi contenenti mutazioni di essere efficientemente tradotti e processati dal proteasoma. Gli amminoacidi glicina e serina sono flessibili [Schlessinger A e Rost B., Proteins. 1 ottobre 2005; 61(1): da 115 a 126]; la loro inclusione in un gruppo di collegamento risulta in una proteina più flessibile. Abbiamo incorporato glicina e serina nel gruppo di collegamento per incrementare la flessibilità della proteina che dovrebbe permettere una traduzione ed un processamento mediante il proteasoma più efficienti, a sua volta permettendo un migliore accesso ai peptidi antigenici codificati.

Così, il gruppo di collegamento dovrebbe essere ricco in glicina per impedire la creazione di cattivi epitopi di legame all’MHC; dovrebbe impedire la capacità del proteasoma di processare i peptidi di collegamento, i quali possono essere realizzati attraverso l’inclusione di glicina; e dovrebbe essere flessibile per incrementare l’accesso a peptidi contenenti mutazioni, i quali possono essere realizzati attraverso la combinazione degli amminoacidi glicina e serina. Quindi, in una forma di realizzazione del costrutto vaccinicco, le sequenze GGSGGGGSGG e GGSGGGGSGS sono preferibilmente incluse come sequenze di collegamento.

Esempio 7: vaccino polineoepitopico a RNA

I costrutti vaccinici polineoepitopici a RNA sono basati sul vettore pST1-A120 contenente un promotore T7, una sequenza tandem di beta-globina al 3’ UTR ed una coda di poli(A) di 120 coppie di basi, i quali sono stati mostrati incrementare la stabilità e l’efficienza traduzionale dell’RNA aumentando quindi la capacità stimolatoria dei linfociti T dell’antigene codificato (Holtkamp S. et al., Blood 2006; PMID: 16940422). Inoltre, sono

stati inseriti un frammento di peptide segnale dell'MHC di classe I ed i domini transmembrana e citosolico comprendenti il codone di arresto (segnale di movimentazione dell'MHC di classe I o MITD) fiancheggiante una sequenza di policollegamento per clonare gli epitopi (Kreiter S. et al., J. Immunol., 180: da 309 a 318, 2008). È stato mostrato che gli ultimi incrementano la presentazione dell'antigene, aumentando quindi l'espansione di linfociti T CD8+ e CD4+ specifici per l'antigene e migliorando le funzioni effettrici.

Per fornire costrutti polineoepitopici a RNA per le 50 mutazioni identificate e convalidate di B16-F10 sono stati generati 3 costrutti di RNA. Il costrutto consiste di sequenze a codoni ottimizzati codificanti per (i) un epitopo mutato di 25 amminoacidi, (ii) un gruppo di collegamento ricco in glicina/serina, (iii) ripetizioni di una sequenza epitopica mutata cui segue un gruppo di collegamento ricco in glicina/serina. La catena di un epitopo mutato contenente sequenze e gruppi di collegamento è fiancheggiata da siti di riconoscimento adatti per le endonucleasi di restrizione per essere clonati nel costrutto basato su pST1 come descritto sopra. I costrutti vaccinici sono stati progettati e sintetizzati da GENEART. Dopo la verifica della sequenza, questi sono stati clonati nella struttura del vettore basato su pST1 per ottenere i costrutti vaccinici polineoepitopici a RNA.

Descrizione dell'approccio clinico

L'applicazione clinica coprirà i passaggi seguenti:

i pazienti reclutabili devono acconsentire all'analisi del DNA mediante sequenziamento di nuova generazione.

Il campione tumorale ottenuto da procedure diagnostiche di routine (tessuto fissato in formalina incluso in paraffina) e le cellule del sangue periferico saranno ottenuti ed utilizzati per l'analisi delle mutazioni come descritto.

Le mutazioni scoperte saranno confermate.

Il vaccino sarà progettato sulla base della prioritizzazione. Per i vaccini a RNA un modello di plasmide principale sarà generato mediante sintesi e clonazione genica.

I plasmidi verranno utilizzati per la produzione di RNA di grado clinico, per il controllo di qualità e per il rilascio del vaccino a RNA.

Il prodotto farmaceutico vaccinale sarà mandato al rispettivo centro di studio per l'applicazione clinica.

Il vaccino a RNA può essere utilizzato come vaccino nudo in tampone di formulazione o incapsulato in nanoparticelle o in liposomi per l'iniezione diretta per esempio in linfonodi, s.c., e.v., i.m.. In alternativa, il vaccino a RNA può essere utilizzato per la trasfezione in vitro per esempio di cellule dendritiche per il trasferimento adottivo.

L'intero procedimento clinico richiede meno di 6 settimane. La "fase di attesa" tra il consenso informato del paziente e la disponibilità del farmaco sarà affrontata con attenzione dal protocollo di studio clinico, compreso il permettere al regime di trattamento standard di essere continuato fino a

che il prodotto farmaceutico da investigare è disponibile.

Esempio 8: identificazione di mutazioni tumorali e loro utilizzo per la vaccinazione contro i tumori

Abbiamo applicato il risequenziamento esomico mediante NGS per la scoperta delle mutazioni nella linea cellulare B16-F10 di melanoma di topo e identificato 962 mutazioni somatiche puntiformi non sinonime, 563 in geni espressi. Le potenziali mutazioni scatenanti avvengono in geni oncosoppressori classici (Pten, Trp53, Tp63, Pml) e in geni coinvolti in vie di trasduzione del segnale protooncogeniche che controllano la proliferazione cellulare (per esempio Mdm1, Pdgfra), l'adesione cellulare e la migrazione (per esempio Fdz7, Fat1) o l'apoptosi (Casp9). Inoltre, B16-F10 porta mutazioni in Aim1 e Trrap che sono stati descritti precedentemente essere frequentemente alterati nel melanoma umano.

L'immunogenicità e la specificità di 50 mutazioni convalidate sono state saggiate utilizzando topi C57BL/6 immunizzati con peptidi lunghi codificanti per gli epitopi mutati. Un terzo (16/50) di essi è stato mostrato essere immunogenico. Di questi, il 60% ha provocato risposte immunitarie preferibilmente dirette contro la sequenza mutata in confronto alla sequenza selvatica.

Abbiamo testato l'ipotesi in modelli tumorali di trapianto. L'immunizzazione con i peptidi ha conferito il controllo del tumore in vivo in configurazioni protettive e terapeutiche, qualificando gli epitopi mutati contenenti singole sostituzioni amminoacidiche come vaccini efficaci.

Animali

I topi C57BL/6 (Jackson Laboratories) sono stati mantenuti secondo le linee di condotta federali e statali sulla ricerca animale all'Università di Magonza.

Cellule

La linea cellulare B16-F10 di melanoma è stata acquistata nel 2010 dall'American Type Culture Collection (prodotto: ATCC CRL-6475, numero di lotto: 58078645). Passaggi precoci (terzo, quarto) di cellule sono stati utilizzati per gli esperimenti tumorali. Le cellule sono state solitamente testate per Mycoplasma. Una nuova autenticazione delle cellule non è stata eseguita dal momento della ricezione.

Sequenziamento di nuova generazione

Estrazione dell'acido nucleico e preparazione del campione: il DNA e l'RNA da cellule B16-F10 sfuse e il DNA da tessuto caudale di C57BL/6 sono stati estratti in triplice copia utilizzando un kit Qiagen DNeasy per sangue e per tessuto (per DNA) ed un kit Qiagen RNeasy Micro (per RNA).

Sequenziamento esomico del DNA: la cattura esomica per il risequenziamento del DNA è stata eseguita in triplice copia utilizzando il saggio di cattura a base di soluzione di topo Sure-Select di Agilent (Gnirke A et al., Nat. Biotechnol. 2009; 27: da 182 a 189), progettato per catturare tutte le

regioni codificanti per le proteine di topo. 3 µg di DNA genomico purificato (gDNA) sono stati frammentati da 150 a 200 coppie di basi utilizzando un dispositivo ad ultrasuoni Covaris S2. I frammenti sono stati riparati alle estremità e fosforilati al 5' e adenilati al 3' secondo le istruzioni del produttore. Gli adattatori con estremità appaiate di Illumina sono stati ligati ai frammenti di gDNA utilizzando un rapporto molare di 10:1 di adattatore e gDNA. Sequenze specifiche arricchite di cellule precattura e fluenti sono state aggiunte utilizzando inneschi di PCR Illumina PE 1.0 e 2.0 per 4 cicli di PCR. 500 ng di frammenti di gDNA ligati all'adattatore arricchiti mediante PCR sono stati ibridati con esche biotinilate SureSelect di una libreria di RNA dell'intero esoma di topo di Agilent per 24 ore a 65°C. I complessi ibridati gDNA/esca di RNA sono stati rimossi utilizzando sfere magnetiche rivestite di streptavidina, lavati e le esche di RNA scisse durante l'eluizione in tampone di eluizione SureSelect. Questi frammenti di gDNA eluiti sono stati amplificati mediante 10 cicli di PCR dopo la cattura. Le librerie di gDNA arricchite in esoma sono state raggruppate sul cBot utilizzando il kit Truseq SR cluster v2.5 utilizzando 7pM e 50 coppie di basi sono state sequenziate sull'Illumina HiSeq2000 utilizzando un kit Truseq SBS-HS da 50 coppie di basi.

Ottenimento del profilo "trascrittomico" dell'espressione genica dell'RNA (RNA-Seq): librerie indicizzate di cDNA da mRNA-seq sono state preparate in triplice copia, da 5 µg di RNA totale (protocollo modificato di mRNA-seq di Illumina). L'mRNA è stato isolato utilizzando sfere magnetiche Seramag Oligo(dT) (Thermo Scientific) e frammentato utilizzando cationi bivalenti e calore. I frammenti risultanti (da 160 a 220 coppie di basi) sono stati convertiti in cDNA utilizzando inneschi casuali e SuperScriptII (Invitrogen) cui segue la sintesi del secondo filamento utilizzando la DNA polimerasi I e l'RNasi H. Il cDNA è stato riparato alle estremità, fosforilato al 5' e adenilato al 3' secondo le istruzioni del produttore. Adattatori multipli specifici con singola sporgenza a T al 3' di Illumina sono stati ligati con DNA ligasi T4 utilizzando un rapporto molare di 10:1 di adattatore e inserto di cDNA. Le librerie di cDNA sono state purificate e selezionate per dimensione da 200 a 220 coppie di basi (gel E-gel SizeSelect al 2%, Invitrogen). L'arricchimento, l'aggiunta di un indice a sei basi di Illumina e di sequenze specifiche di cellule del flusso sono stati fatti mediante PCR utilizzando la DNA polimerasi Phusion (Finnzymes). Tutti i passaggi di rimozione fino a questo passaggio sono stati fatti con un volume 1,8x di sfere magnetiche AgencourtAMPure XP. Tutti i controlli di qualità sono stati fatti utilizzando un saggio Qubit HS di Invitrogen e la dimensione dei frammenti è stata determinata utilizzando un saggio del DNA 2100 Bioanalyzer HS di Agilent. Le librerie indicizzate di RNA-Seq sono state raggruppate e sequenziate come descritto sopra.

Analisi dei dati di NGS, espressione genica: le letture di sequenza in uscita da campioni di RNA sono state preprocessate secondo il protocollo standard di Illumina, compreso il filtraggio per letture di bassa qualità. Le letture di sequenza sono state allineate alla sequenza genomica di riferimento mm9 (Waterston RH et al., Nature 2002; 420: da 520 a 562) con bowtie (versione 0.12.5) (Langmead B et al., Genome Biol 2009; 10:

R25). Per gli allineamenti del genoma, sono stati permessi due appaiamenti errati e solo l'allineamento migliore ("-v2-best") è stato registrato; per gli allineamenti del trascrittoma sono stati utilizzati i parametri preimpostati. Le letture non allineabili alla sequenza genomica sono state allineate ad una banca dati di tutte le possibili sequenze di giunzione esone-esone di trascritti RefSeq (Pruitt KD et al., Nucleic Acids Res 2007; 35: da D61 a D65). I valori di espressione sono stati determinati incrociando le coordinate di lettura con quelle dei trascritti RefSeq, contando le letture sovrapposte esone e giunzione, e normalizzando rispetto alle unità di espressione di RPKM (letture che mappano per kilobase del modello esonico per milione di letture mappate) (Mortazavi A et al., Nat. Methods 2008; 5: da 621 a 628).

Analisi dei dati di NGS, scoperta delle mutazioni somatiche: le mutazioni somatiche sono state identificate come descritto nell'esempio 9. 50 letture nucleotidiche (nt), a singola estremità sono state allineate al genoma di topo di riferimento mm9 utilizzando bwa (opzioni preimpostate, versione 0.5.8c) (Li H e Durbin R, Bioinformatics 2009; 25: da 1754 a 1760). Le letture ambigue mappate in posizioni multiple del genoma sono state rimosse. Le mutazioni sono state identificate utilizzando tre programmi informatici: samtools (versione 0.1.8) (Li H, Bioinformatics 2011; 27: da 1157 a 1158), GATK (versione 1.0.4418) (McKenna A et al., Genome Res 2010; 20: da 1297 a 1303), e SomaticSniper (<http://genome.wustl.edu/software/somaticsniper>) (Ding L et al., Hum Mol Genet 2010; 19: da R188 a R196). Alle variazioni potenziali identificate in tutte le triplici copie di B16-F10 è stato assegnato un valore di confidenza del "tasso di scoperta falsa" (FDR) (si confronti l'esempio 9).

Selezione, convalida, e funzione delle mutazioni

Selezione: le mutazioni avevano dovuto soddisfare i seguenti criteri per essere selezionate: (i) presenti in tutte le triplici copie di B16-F10 ed assenti in tutte le triplici copie di C57BL/6, (ii) $FDR \leq 0,05$, (iii) omogenee in C57BL/6, (iv) avvengono in un trascritto RefSeq, e (v) inducono le variazioni non sinonime ad essere conteggiate come mutazione autentica. La selezione per la prova di convalida e di immunogenicità ha richiesto che le mutazioni fossero geni espressi (RPKM medio sulle repliche > 10).

Convalida: le mutazioni derivate dal DNA sono state classificate come convalidate se confermate mediante sequenziamento secondo Sanger o le letture di RNA-Seq di B16-F10. Tutte le varianti selezionate sono state amplificate da 50 ng di DNA da cellule B16-F10 e da tessuto caudale di C57BL/6 utilizzando inneschi fiancheggiati, prodotti visualizzati (sistema QIAxcel, Qiagen) e purificati (kit di purificazione per PCR QIAquick, Qiagen). L'amplicone della dimensione attesa è stato asportato dal gel, purificato (kit di estrazione su gel QIAquick, Qiagen) e sottoposto a sequenziamento secondo Sanger (Eurofins MWG Operon, Ebersberg, Germania) con l'innesco in avanti utilizzato per l'amplificazione mediante PCR.

Impatto funzionale: i programmi SIFT (Kumar P et al., Nat Protoc 2009; 4: da 1073 a 1081) e POLYPHEN-2 (Adzhubei IA et al., Nat. Methods

2010; 7: da 248 a 249), i quali predicono la significatività funzionale di un amminoacido sulla funzione proteica sulla base della localizzazione dei domini proteici e della conservazione delle sequenze tra le specie, sono stati impiegati per verificare l'impatto delle mutazioni selezionate. Gli strumenti IPA di Ingenuity sono stati utilizzati per dedurre la funzione genica.

Peptidi sintetici ed adiuvanti

Tutti i peptidi comprese ovalbumina di classe I (OVA₂₅₈₋₂₆₅), di classe II (OVA di classe II₃₃₀₋₃₃₈), nucleoproteina dell'influenza (Inf-NP₃₆₆₋₃₇₄), nucleoproteina del virus della stomatite vescicolare (VSV-NP₅₂₋₅₉) e proteina 2 correlata alla tirosinasi (Trp₂₁₈₀₋₁₈₈) sono stati acquistati da Jerini Peptide Technologies (Berlino, Germania). I peptidi sintetici erano lunghi 27 amminoacidi con l'amminoacido mutato (MUT) o selvatico (WT) in posizione 14. Acido poliinosinico:policitidilico (poli(I:C), Invitrogen) è stato utilizzato come adiuvante iniettato per via sottocutanea. Un pentamero di MHC specifico per il peptide Inf-NP₃₆₆₋₃₇₄ è stato acquistato da ProImmune Ltd..

Immunizzazione dei topi

A topi femmina C57BL/6 appaiati per età sono stati iniettati per via sottocutanea 100 µg di peptide e 50 µg di poli(I:C) formulati in PBS (volume totale di 200 µl) nel fianco laterale (5 topi per gruppo). Ogni gruppo è stato immunizzato il giorno 0 e il giorno 7 con due differenti peptidi codificanti per una mutazione, un peptide per fianco. Dodici giorni dopo l'iniezione iniziale i topi sono stati sacrificati e gli splenociti sono stati isolati per le prove immunologiche.

In alternativa, a topi femmina C57BL/6 appaiati per età sono stati iniettati per via endovenosa 20 µg di RNA trascritto in vitro formulati con 20 µl di Lipofectamina™ RNAiMAX (Invitrogen) in PBS in un volume totale di iniezione di 200 µl (3 topi per gruppo). Ogni gruppo è stato immunizzato i giorni 0, 3, 7, 14 e 18. Ventitré giorni dopo l'iniezione iniziale i topi sono stati sacrificati e gli splenociti sono stati isolati per le prove immunologiche. Le sequenze di DNA rappresentanti una (monoepitopo), due (biepitopo), o 16 mutazioni (poliepitopo) sono state costruite utilizzando 50 amminoacidi (aa) con la mutazione in posizione 25 (biepitopo) o 27 amminoacidi con la mutazione in posizione 14 (monoepitopo e poliepitopo), sono state separate mediante un gruppo di collegamento di glicina/serina di 9 amminoacidi e clonate nella struttura di pST1-2BgUTR-A120 (Holtkamp et al., Blood 2006; 108: da 4009 a 4017). La trascrizione in vitro da questo stampo e la purificazione sono state descritte precedentemente (Kreiter et al., Cancer Immunol Immunother 2007; 56: da 1577 a 1587).

Saggio di immunospot enzimatico

Un saggio di immunospot enzimatico (ELISPOT) (Kreiter S et al., Cancer Res. 2010; 70: da 9031 a 9040) e la generazione di cellule dendritiche singeniche derivate dal midollo osseo (BMDC) come stimolatori sono stati descritti precedentemente (Lutz MB et al., J. Immunol. Methods 1999;

223: da 77 a 92). Le BMDC sono state stimulate con il peptide (2 µg/ml), o trasfettate con RNA trascritto in vitro (IVT) codificante per la mutazione indicata o per l'RNA di controllo (eGFP-RNA). Le sequenze rappresentanti due mutazioni, ciascuna comprendente 50 amminoacidi con la mutazione in posizione 25 e separate da un gruppo di collegamento di glicina/serina di 9 amminoacidi sono state clonate nella struttura di pST1-2BgUTR-A120 (Holtkamp S et al., Blood 2006; 108: da 4009 a 4017). La trascrizione in vitro da questo stampo e la purificazione sono state descritte precedentemente (Kreiter S et al., Cancer Immunol Immunother 2007; 56: da 1577 a 1587). Per il saggio, 5×10^4 BMDC ingegnerizzate con il peptide o con l'RNA sono state coincubate con 5×10^5 splenociti isolati di fresco in una piastra da microtitolazione rivestita con un anticorpo anti-IFN- γ (10 µg/ml, clone AN18; Mabtech). Dopo 18 ore a 37°C, la secrezione di citochine è stata rilevata con un anticorpo anti-IFN- γ (clone R4-6A2; Mabtech). I numeri delle macchie sono stati contati ed analizzati con l'analizzatore ImmunoSpot® S5 Versa ELISPOT, il programma di acquisizione delle immagini ImmunoCapture™ e il programma di analisi ImmunoSpot® versione 5. L'analisi statistica è stata eseguita mediante test di t di Student e test di Mann-Whitney (prova non parametrica). Le risposte sono state considerate significative, quando la prova ha dato un valore di $p < 0,05$ ed i numeri medi di macchie sono stati > 30 macchie/ 5×10^5 cellule effettrici. Le reattività sono state valutate mediante i numeri medi di macchie (-: <30; +: >30; ++: >50; +++: >200 macchie/pozzetto).

Saggio delle citochine intracellulari

Le aliquote degli splenociti preparate per il saggio ELISPOT sono state sottoposte all'analisi della produzione di citochine mediante citometria intracellulare a flusso. A questo scopo 2×10^6 splenociti per campione sono stati piastrati nel terreno di coltura (RPMI+FCS al 10%) arricchito con l'inibitore del Golgi brefeldina A (10 µg/ml) in una piastra da 96 pozzetti. Le cellule da ciascun animale sono state stimulate di nuovo per 5 ore a 37°C con 2×10^5 BMDC stimulate con il peptide. Dopo l'incubazione le cellule sono state lavate con PBS, risospese in 50 µl di PBS e marcate extracellularmente con i seguenti anticorpi antitopo per 20 minuti a 4°C: anti-CD4 FITC, anti-CD8 APC-Cy7 (BD Pharmingen). Dopo l'incubazione le cellule sono state lavate con PBS e successivamente risospese in 100 µl di soluzione Cytofix/Cytoperm (BD Bioscience) per 20 minuti a 4°C per la permeabilizzazione della membrana esterna. Dopo la permeabilizzazione le cellule sono state lavate con tampone Perm/Wash (BD Bioscience), risospese in 50 µl/campione in tampone Perm/Wash e marcate intracellularmente con i seguenti anticorpi antitopo per 30 minuti a 4°C: anti-IFN- γ PE, anti-TNF- α PE-Cy7, anti-IL2 APC (BD Pharmingen). Dopo il lavaggio con tampone Perm/Wash le cellule sono state risospese in PBS contenente paraformaldeide all'1% per l'analisi di citometria a flusso. I campioni sono stati analizzati utilizzando un citometro BD FACSCanto™ II e FlowJo (versione 7.6.3).

Modello tumorale di melanoma B16

Per gli esperimenti tumorali di vaccinazione $7,5 \times 10^4$ cellule di melanoma B16-F10 sono state inoculate s.c. nei fianchi di topi C57BL/6. Nella configurazione profilattica, l'immunizzazione con un peptide specifico per una mutazione è stata eseguita 4 giorni prima dell'inoculo del tumore e i giorni 2 e 9 dopo l'inoculo del tumore. Per l'esperimento terapeutico il vaccino peptidico è stato somministrato i giorni 3 e 10 dopo l'iniezione tumorale. Le dimensioni del tumore sono state misurate ogni tre giorni e i topi sono stati sacrificati quando il diametro tumorale ha raggiunto 15 mm.

In alternativa, per gli esperimenti tumorali di vaccinazione 1×10^5 cellule di melanoma B16-F10 sono state inoculate s.c. nei fianchi di topi femmina C57BL/6 appaiati per età. La vaccinazione peptidica è stata eseguita i giorni 3, 10 e 17 dopo l'inoculo del tumore con 100 µg di peptide e 50 µg di poli(I:C) formulati in PBS (volume totale di 200 µl) iniettati per via sottocutanea nel fianco laterale. Le immunizzazioni a RNA sono state eseguite utilizzando 20 µg di RNA codificante per mutazioni trascritto in vitro formulato con 20 µl di Lipofectamina™ RNAiMAX (Invitrogen) in PBS in un volume totale di iniezione di 200 µl. Come controllo ad un gruppo di animali è stato iniettato RNAiMAX (Invitrogen) in PBS. Gli animali sono stati immunizzati i giorni 3, 6, 10, 17 e 21 dopo l'inoculo del tumore. Le dimensioni del tumore sono state misurate ogni tre giorni utilizzando un calibro e i topi sono stati sacrificati quando il diametro tumorale ha raggiunto 15 mm.

Identificazione di mutazioni non sinonime nel melanoma di topo B16-F10

Il nostro obiettivo è stato identificare le mutazioni somatiche puntiformi potenzialmente immunogeniche nel melanoma di topo B16-F10 mediante NGS e testarle per l'immunogenicità in vivo mediante vaccinazione peptidica dei topi misurando le risposte provocate dei linfociti T mediante saggio ELISPOT (figura 9A). Abbiamo sequenziato gli esomi del genoma selvatico di sfondo di C57BL/6 e delle cellule B16-F10, ciascuno con estrazioni e catture in triplice copia. Per ciascun campione, sono state generate più di 100 milioni di letture di 50 nucleotidi a singola estremità. Di queste l'80% si allinea unicamente al genoma di topo mm9 e il 49% si allinea sul bersaglio, dimostrando un arricchimento riuscito del bersaglio e risultando in una copertura di oltre 20 volte per il 70% dei nucleotidi del bersaglio in ciascuno dei campioni in triplice copia. L'RNA-Seq delle cellule B16-F10, anch'esso profilato in triplice copia, ha generato una media di 30 milioni di letture di 50 nucleotidi a singola estremità, di cui l'80% si allinea al trascrittoma murino.

Le letture di DNA (cattura degli esomi) da B16-F10 e da C57BL/6 sono state analizzate per identificare le mutazioni somatiche. L'analisi della variazione del numero di copie (Sathirapongsasuti JF et al., Bioinformatics 2011; 27: da 2648 a 2654) ha dimostrato amplificazioni e delezioni di DNA in B16-F10, compresa la delezione omozigote del soppressore tumorale Cdkn2a (inibitore 2A delle chinasi dipendenti da ciclina, p16Ink4A).

Focalizzandoci sulle mutazioni puntiformi per identificare possibili mutazioni immunogeniche, abbiamo identificato 3570 mutazioni somatiche puntiformi a $FDR \leq 0,05$ (figura 9B). La classe più frequente di mutazioni è stata rappresentata dalle transizioni C>T/G>A, tipicamente risultanti dalla luce ultravioletta (Pfeifer GP et al., *Mutat Res* 2005; 571: da 19 a 31). Di queste mutazioni somatiche, 1392 avvengono nei trascritti, con 126 mutazioni in regioni non tradotte. Delle 1266 mutazioni in regioni codificanti, 962 causano variazioni non sinonime delle proteine e 563 di queste avvengono in geni espressi (figura 9B).

Assegnazione di mutazioni identificate a geni veicolanti e convalida

Notevolmente, molti dei geni mutati (962 geni contenenti mutazioni somatiche puntiformi non sinonime) sono stati precedentemente associati ai fenotipi tumorali. Le mutazioni sono state trovate in geni soppressori tumorali stabiliti, compresi Pten, Trp53 (chiamato anche p53), e Tp63. In Trp53, il miglior soppressore tumorale stabilito (Zilfou JT et al., *Cold Spring Harb Perspect Biol* 2009; 1: a001883), la mutazione da asparagina ad acido aspartico in posizione 127 della proteina (p.N127D) è posizionata nel dominio di legame del DNA ed è prevista alterare la funzione da SIFT. Pten conteneva due mutazioni (p.A39V, p.T131P), le quali sono entrambe previste avere un impatto dannoso sulla funzione proteica. La mutazione p.T131P è adiacente ad una mutazione (p.R130M) mostrata diminuire l'attività fosfataseica (Dey N et al., *Cancer Res.* 2008; 68: da 1862 a 1871). Inoltre, le mutazioni sono state trovate in geni associati a vie metaboliche di riparazione del DNA, quali Brca2 (tumore della mammella 2, comparsa precoce), Atm (ataxia teleangectasia mutata), Ddb1 (proteina 1 legante il DNA specifica per il danno) e Rad9b (omologo B di RAD9). Inoltre, le mutazioni avvengono in altri geni associati al tumore, compresi Aim1 (soppressore tumorale "assente nel melanoma 1"), Flt1 (oncogene Vegf1, tirosin chinasi 1 correlata a fms), Pml (soppressore tumorale "leucemia promielocitica"), Fat1 ("omologo 1 al soppressore tumorale FAT"), Mdm1 (proteina nucleare legante TP53), Mta3 (famiglia 1 associata alle metastasi, membro 3), e Alk (tirosin chinasi recettoriale del linfoma anaplastico). Abbiamo trovato una mutazione in p.S144F in Pdgfra (recettore del fattore di crescita derivato dalle piastrine, polipeptide alfa), una tirosin chinasi recettoriale legata alla membrana cellulare della via metabolica di MAPK/ERK, precedentemente identificata nei tumori (Verhaak RG et al., *Cancer Cell* 2010; 17: da 98 a 110). Una mutazione avviene in p.L222V in Casp9 (caspasi 9, peptidasi cisteinica correlata all'apoptosi). CASP9 scinde proteoliticamente la poli(ADP-ribosio) polimerasi (PARP), regola l'apoptosi, ed è stata legata a diversi tumori (Hajra KM et al., *Apoptosis* 2004; 9: da 691 a 704). La mutazione che abbiamo trovato può potenzialmente avere un impatto su PARP e sulla trasduzione del segnale apoptotico. Nel modo più interessante, nessuna mutazione è stata trovata in Braf, c-Kit, Kras o Nras. Tuttavia, sono state identificate mutazioni in Rassf7 (proteina associata a RAS) (p.S90R), Ksr1 (soppressore chinasi di ras 1) (p.L301V), e Atm (via metabolica della PI3K) (p.K91T), le quali sono tutte previste avere un impatto significativo sulla funzione proteica. Trrap (proteina associata al dominio di trasformazione/trascrizione) è stata

identificata per prima quest'anno in campioni umani di melanoma come nuovo potenziale bersaglio del melanoma (Wei X et al., Nat Genet 2011; 43: da 442 a 446). In B16-F10, una mutazione di Trrap avviene in p.K2783R ed è previsto che disturbi il dominio chinasi sovrapposto FAT correlato alla fosfatidilinositolo chinasi (PIK).

Dalle 962 mutazioni non sinonime identificate utilizzando l'NGS, abbiamo selezionato 50 mutazioni, comprese 41 con FDR<0,05, per la convalida basata sulla PCR e per la prova di immunogenicità. I criteri di selezione sono stati la posizione in un gene espresso (RPKM>10) e l'immunogenicità prevista. Notevolmente, siamo stati in grado di convalidare tutte le 50 mutazioni (tabella 6, figura 9B).

Tabella 6: mutazioni scelte per la convalida. Da sinistra: ID assegnato, simbolo del gene, sostituzione e posizione amminoacidica, nome del gene, localizzazione subcellulare e tipo previsti (Ingenuity).

ID	simbolo	cambiamento	nome del gene di Entrez	localizzazione subcellulare	tipo
MUT1	Fzd7	p.G304A	recettore 7 della famiglia frizzled	membrana plasmatica	recettore accoppiato a proteine G
MUT2	Xpot	p-I830S	esportina, tRNA (recettore di nucleo esportazione nucleare per tRNA)	nucleo	altro
MUT3	Ranbp2	p.Q2871H	proteina 2 legante RAN	nucleo	enzima
MUT4	Dnajb12	p.P54T	omologo a DnaJ (Hsp40), sottofamiglia B, membro 12	citoplasma	altro
MUT5	Eef2	p.G795A	fattore 2 eucariotico di allungamento della traduzione	citoplasma	regolatore della traduzione
MUT6	Ptrf	p.D382G	polimerasi I e fattore di rilascio dei trascritti	nucleo	regolatore della trascrizione
MUT7	Trp53	p.N128D	proteina tumorale p53	nucleo	regolatore della trascrizione

MUT8	Ddx23	p.V602A	polipeptide 23 con DEAD (Asp-Glu-Ala-Asp) box	nucleo	enzima
MUT9	Golgb1	p.E2855D	golgina B1	citoplasma	altro
MUT10	Pcdhgal1	p.G82R	sottofamiglia A della protocaderina gamma, 11	membrana plasmatica	altro
MUT11	Snx15	p.E211G	nessina 15 di separazione	citoplasma	trasportatore
MUT12	Gnas	p.S112G	locus complesso GNAS (proteina legante il nucleotide guaninico, stimolante alfa)	membrana plasmatica	enzima
MUT13	Fndc3b	p.C561W	dominio fibronectinico di tipo III contenente 3B	citoplasma	altro
MUT14	Sbno1	p.P309T	omologo 1 di strawberry notch (Drosophila)	non nota	enzima
MUT15	Pi4k2b	p.R344Q	fosfatidilinositolo 4-chinasi tipo 2 beta	citoplasma	chinasi
MUT16	Thumpd3	p.T243S	3 contenente il dominio THUMP	non nota	altro
MUT17	Tnp3	p.G504A	trasportina 3	citoplasma	altro
MUT18	Numa1	p.Q447K	proteina 1 dell'apparato mitotico nucleare	nucleo	altro
MUT19	Wwp2	p.E742K	proteina ligasi 2 dell'ubiquitina E3 contenente il dominio WW	citoplasma	enzima
MUT20	Tubb3	p.G402A	tubulina, beta 3	citoplasma	altro
MUT21	Atp11a	p.R522S	ATPasi, classe VI, tipo 11A	membrana plasmatica	trasportatore
MUT22	Asf1b	p.A141P	omologo B della funzione 1 di antisilenziamento di ASF1 (S.	nucleo	altro

			cerevisiae)		
MUT23	Wdr82	p.I221L	dominio di ripetizione WD 82	nucleo	altro
MUT24	Dag1	p.P425A	distroglicano 1 (glicoproteina 1 associata alla distrofina)	membrana plasmatica	recettore transmembrana
MUT25	Plod2	p.F530V	procollagene-lisina, 2-ossoglutarato 5- diossigenasi 2	citoplasma	enzima
MUT26	Orc2	p.F278V	complesso di riconoscimento dell'origine, subunità 2	nucleo	altro
MUT27	Obs1	p.T1764M	simile all'oscurina 1	non nota	altro
MUT28	Ppp1r7	p.L170P	proteinfosfatasi 1, subunità regolatoria (inibitrice) 7	nucleo	fosfatasi
MUT29	Mthfd11	p.F294V	simile alla metilentetraidrofolato deidrogenasi 1 (dipendente dal NADP+)	citoplasma	enzima
MUT30	Kif18b	p.K739N	membro 18B della famiglia chinesinica	non nota	altro
MUT31	Ascc2	p.A59G	subunità 2 del complesso 1 cointegratore del segnale di attivazione	non nota	altro
MUT32	Itsn2	p.S1551R	intersettina 2	citoplasma	altro
MUT33	Pbk	p.V145D	chinasi legante PDZ	citoplasma	chinasi
MUT34	Klhl22	p.F179V	simile a kelch 22 (Drosophila)	non nota	altro
MUT35	Ddb1	p.L438I	proteina 1 legante il DNA specifica per il danno, 127 kDa	nucleo	altro
MUT36	Tm9sf3	p.Y382H	membro 3 della superfamiglia transmembrana 9	citoplasma	trasportatore

MUT37	Dpf2	p.F275V	D4, famiglia 2 a dita di zinco e doppio PHD	nucleo	altro
MUT38	Atrn	p.5745N	attrattina	spazio extracellulare	altro
MUT39	Snx5	p.R373Q	nessina 5 di separazione	citoplasma	trasportatore
MUT40	Armc1	p.585I	1 contenente una ripetizione di armadillo	citoplasma	altro
MUT41	Ash11	p.L632I	simile ad ash1 (assente, piccolo, o omeotico) (Drosophila)	nucleo	regolatore della trascrizione
MUT42	S100a13	p.S18C	proteina S100 A13 che lega il calcio	citoplasma	altro
	2510039O18				
MUT43	Rik	p.E391K	KIAA2013	non nota	altro
MUT44	Cpsf31	p.D314N	simile al fattore 3 specifico per la scissione e per la poliadenilazione	nucleo	altro
MUT45	Mkml	p.N346Y	proteina 1 con dito ad anello di makorina	non nota	altro
MUT46	Actn4	p.F835V	actinina, alfa 4	citoplasma	altro
MUT47	Rpl13a	p.A24G	proteina ribosomiale L13a	citoplasma	altro
MUT48	Def8	p.R255G	omologo 8 differenzialmente espresso in FDCP (topo)	non nota	altro
MUT49	Fat1	p.I1940M	omologo 1 al soppressore tumorale FAT (Drosophila)	membrana plasmatica	altro
MUT50	Sema3b	p.L663V	dominio sema, dominio immunoglobulinico (Ig), breve dominio basico, secreto, (semaforina)	spazio extracellulare	altro

La figura 9C mostra le posizioni dei cromosomi, la densità genica, l'espressione genica, le mutazioni, e le mutazioni filtrate di B16-F10 (anelli interni).

Prova in vivo di immunogenicità con peptidi lunghi rappresentanti le mutazioni

Per fornire antigeni per la prova di immunogenicità di queste mutazioni, abbiamo impiegato peptidi lunghi che hanno molti vantaggi rispetto ad altri peptidi per l'immunizzazione (Melief CJ e van der Burg SH, Nat Rev Cancer 2008; 8: da 351 a 360). I peptidi lunghi sono in grado di indurre linfociti T CD8+ così come CD4+ specifici per l'antigene (Zwaveling S et al., Cancer Res. 2002; 62: da 6187 a 6193; Bijker MS et al., J. Immunol. 2007; 179: da 5033 a 5040). Inoltre, i peptidi lunghi richiedono un processamento per essere presentati sulle molecole dell'MHC. Tale assorbimento è fatto il più efficientemente dalle cellule dendritiche, le quali sono ottimali per l'attivazione di una potente risposta dei linfociti T. I peptidi della dimensione corretta, al contrario, non richiedono di essere tagliati e sono caricati in modo esogeno su tutte le cellule esprimenti molecole dell'MHC, compresi linfociti B e T non attivati, conducendo all'induzione di tolleranza e al fratricidio (Toes RE et al., J. Immunol. 1996; 156: da 3911 a 3918; Su MW et al., J. Immunol. 1993; 151: da 658 a 667). Per ciascuna delle 50 mutazioni convalidate, abbiamo progettato peptidi lunghi 27 amminoacidi con l'amminoacido mutato o selvatico posizionato centralmente. Così, qualsiasi potenziale epitopo dell'MHC di classe I e di classe II lungo da 8 a 14 amminoacidi recante la mutazione potrebbe essere processato da questo precursore peptidico. Come adiuvante per la vaccinazione peptidica abbiamo utilizzato poli(I:C) che è noto promuovere la presentazione incrociata ed incrementa l'efficacia del vaccino (Datta SK et al., J. Immunol. 2003; 170: da 4102 a 4110; Schulz O et al., Nature 2005; 433: da 887 a 892). Le 50 mutazioni sono state testate in vivo su topi per l'induzione di linfociti T. Grandiosamente, è stato trovato che 16 peptidi codificanti per mutazioni su 50 provocano risposte immunitarie in topi immunizzati. I linfociti T indotti hanno mostrato modelli differenti di reattività (tabella 7).

Tabella 7: riassunto delle reattività dei linfociti T determinate in seguito alla vaccinazione con il peptide codificante per la mutazione.

L'analisi statistica è stata eseguita mediante test di t di Student e test di Mann-Whitney (prova non parametrica). Le risposte sono state considerate significative quando la prova ha dato un valore di $p < 0,05$ ed i numeri medi di macchie sono stati >30 macchie/ 5×10^5 cellule effettrici. Le reattività sono state valutate mediante i numeri medi di macchie -;<30; +;>30; ++;>50; +++>200 macchie/pozzetto.

mutazione	simbolo del gene	reattività nei confronti della mutazione	reattività nei confronti del selvatico	mutazione	simbolo del gene	reattività nei confronti della mutazione	reattività nei confronti

							del selvatico
MUT01	Fzd7	-	-	MUT26	Orc2	-	-
MUT02	Xpot	-	-	MUT27	Obsl1	-	-
MUT03	Ranbp2	-	-	MUT28	Ppp1r7	+	+
MUT04	Dnajb12	-	-	MUT29	Mthfd11	+	-
MUT05	Eef2	+++	+++	MUT30	Kif18b	+++	-
MUT06	Ptrf	-	-	MUT31	Ascc2	-	-
MUT07	Trp53	-	-	MUT32	Itsn2	-	-
MUT08	Ddx23	-	-	MUT33	Pbk	-	-
MUT09	Golgb1	-	-	MUT34	Klhl22	-	-
MUT10	Pcdhga11	-	-	MUT35	Ddb1	-	-
MUT11	Snx15	-	-	MUT36	Tm9sf3	+	-
MUT12	Gnas	+	-	MUT37	Dpf2	-	-
MUT13	Fndc3b	-	-	MUT38	Atrn	-	-
MUT14	Sbno1	-	-	MUT39	Snx5	-	-
MUT15	Pi4k2b	-	-	MUT40	Armc1	-	-
MUT16	Thumpd3	-	-	MUT41	Ash11	-	-
MUT17	Tnpo3	+++	++	MUT42	S100a13	-	-
MUT18	Numa1	-	-	MUT43	Rik	-	-
MUT19	Wwp2	-	-	MUT44	Cpsf31	+++	++
MUT20	Tubb3	+++	-	MUT45	Mkru1	++	++
MUT21	Atp11a	-	-	MUT46	Actn4	++	+
MUT22	Asf1b	++	++	MUT47	Rpl13a	-	-

MUT23	Wdr82	-	-	MUT48	Def8	++	++
MUT24	Dag1	++	+	MUT49	Fat1	-	-
MUT25	Plod2	+++	++	MUT50	Sema3b	+++	++

Undici peptidi hanno indotto una risposta immunitaria preferibilmente riconoscendo l'epitopo mutato. Questo è esemplificato per topi immunizzati con le mutazioni 30 (MUT30, Kif18b) e 36 (MUT36, Plod2) (figura 10A). La prova mediante ELISPOT ha rivelato forti risposte immunitarie specifiche per una mutazione senza una reattività incrociata nei confronti del peptide selvatico o di un peptide di controllo non correlato (VSV-NP). Con cinque peptidi, comprese le mutazioni 05 (MUT05, Eef2) e 25 (MUT25, Plod2) (figura 10A), sono state ottenute risposte immunitarie con riconoscimento paragonabile sia del peptide mutato che selvatico. La maggioranza dei peptidi mutati non è stata in grado di indurre risposte significative dei linfociti T come esemplificato dalle mutazioni 01 (MUT01, Fzd7), 02 (MUT02, Xpot), e 07 (MUT07, Trp53). Le risposte immunitarie indotte da diverse delle mutazioni scoperte sono state ben all'interno dell'intervallo di immunogenicità (500 macchie/5x10⁵ cellule) generato immunizzando topi come controllo positivo con un epitopo descritto dell'MHC di classe I dall'antigene tumorale del melanoma di topo proteina 2 correlata alla tirosina (Trp2180-188, figura 10A) (Bloom MB et al., Exp Med 1997; 185: da 453 a 459; Schreurs MW et al. Cancer Res. 2000; 60: da 6995 a 7001). Per i peptidi scelti che inducono una forte risposta dei linfociti T specifici per una mutazione, abbiamo confermato il riconoscimento immunitario mediante un approccio indipendente. Invece di peptidi lunghi, l'RNA trascritto in vitro (IVT RNA) codificante per i frammenti peptidici mutati MUT17, MUT30 e MUT44 è stato utilizzato per la lettura immunologica. Le BMDC trasfettate con un RNA codificante per una mutazione o con un RNA irrilevante sono servite come cellule presentanti l'antigene (APC) in un saggio ELISPOT, mentre le cellule della milza di topi immunizzati sono servite come popolazione cellulare effettrice. Le BMDC trasfettate con mRNA codificante per MUT17, MUT30 e MUT44 sono state specificamente e fortemente riconosciute dagli splenociti di topi immunizzati con i rispettivi peptidi lunghi (figura 10B). È stata registrata una reattività significativamente inferiore nei confronti delle BMDC trasfettate con l'RNA di controllo, la quale è probabilmente dovuta all'attivazione non specifica delle BMDC da parte dell'RNA a singolo filamento (test di t di Student; MUT17: p=0,0024, MUT30: p=0,0122, MUT44: p=0,0075). Questi dati confermano che i linfociti T indotti specifici per la mutazione in effetti riconoscono gli epitopi processati a livello endogeno. Due mutazioni che inducono un riconoscimento preferito di epitopi mutati sono nei geni Actn4 e Kif18b. La mutazione somatica in ACTN4 (actinina, alfa 4) è in p.F835V nel dominio proteico "EF-hand" di legame al calcio. Anche se entrambi SIFT e POLYPHEN predicano un impatto significativo di questa mutazione sulla funzione proteica, il gene non è un oncogene stabilito. Tuttavia, linfociti T specifici per la mutazione nei confronti di ACTN4 sono stati recentemente associati ad un risultato positivo per il paziente (Echchakir H et al., Cancer Res. 2001; 61: da 4078

a 4083). KIF18B (membro 18B della famiglia chinesinica) è una chinesina con attività motoria microtubulare e legante ATP e nucleotidi che è coinvolta nella regolazione della divisione cellulare (Lee YM et al., Gene 2010; 466: da 16 a 25) (figura 10C). La sequenza di DNA in posizione codificante p.K739 è omogenea nel riferimento C57BL/6, mentre le letture di DNA di B16-F10 rivelano una mutazione somatica eterozigote. Entrambi i nucleotidi sono stati rilevati nelle letture di RNA-Seq di B16-F10 e convalidati mediante sequenziamento secondo Sanger. KIF18B non è stato precedentemente associato ad un fenotipo tumorale. La mutazione p.K739N non è posizionata in un dominio proteico funzionale o conservato noto (figura 10C, parte inferiore) e così molto probabilmente è una mutazione secondaria piuttosto che una principale. Questi esempi suggeriscono una mancanza di correlazione tra la capacità di indurre una risposta immunitaria che riconosce una mutazione ed una rilevanza funzionale o immunologica.

Valutazione in vivo dell'attività antitumorale dei candidati vaccinali

Per valutare se le risposte immunitarie provocate in vivo si traducano in effetti antitumorali in topi affetti da tumore, abbiamo scelto MUT30 (mutazione in Kif18b) e MUT44 come esempi. Era stato mostrato che queste mutazioni inducono una forte reazione immunitaria preferibilmente nei confronti del peptide mutato e sono processate a livello endogeno (figura 10A, 10B). Il potenziale terapeutico della vaccinazione con i peptidi mutati è stato esplorato immunizzando i topi con MUT30 o MUT44 ed adiuvante 3 e 10 giorni dopo l'innesto con $7,5 \times 10^5$ B16-F10. La crescita dei tumori è stata inibita da entrambe le vaccinazioni peptidiche rispetto al gruppo di controllo (figura 11A). Poiché B16-F10 è un tumore che cresce in modo molto aggressivo, abbiamo inoltre testato le risposte immunitarie protettive. I topi sono stati immunizzati con il peptide MUT30, hanno ricevuto un'inoculazione s.c. di $7,5 \times 10^5$ cellule B16-F10 4 giorni dopo e sono stati stimolati con MUT30 2 e 9 giorni dopo l'induzione del tumore. Sono state osservate una protezione tumorale completa e una sopravvivenza del 40% dei topi trattati con MUT30, mentre tutti i topi nel gruppo di controllo trattato sono deceduti entro 44 giorni (figura 11B sinistra). In quei topi, sviluppati i tumori nonostante l'immunizzazione con MUT30, la crescita dei tumori è stata più lenta risultando in un allungamento della sopravvivenza media a 6 giorni rispetto al gruppo di controllo (figura 11B destra). Questi dati implicano che già la vaccinazione contro una mutazione singola è in grado di conferire effetti antitumorali.

Immunizzazione con un RNA codificante per una mutazione

Le 50 mutazioni convalidate dalla linea cellulare B16-F10 di melanoma sono state utilizzate per costruire differenti vaccini a RNA. Le sequenze di DNA rappresentanti una (monoepitopo), due (biepitopo), o 16 differenti mutazioni (poliepitopo), sono state costruite utilizzando 50 amminoacidi (aa) con la mutazione in posizione 25 (biepitopo) o 27 amminoacidi con la mutazione in posizione 14 (monoepitopo e poliepitopo) e sono state separate mediante un gruppo di collegamento di glicina/serina di 9 amminoacidi. Questi costrutti sono stati clonati nella struttura di pST1-2BgUTR-

A120 per la trascrizione in vitro dell'mRNA (Holtkamp et al., Blood 2006; 108: da 4009 a 4017).

Per testare la capacità in vivo di indurre risposte di linfociti T nei confronti dei differenti vaccini a RNA gruppi di tre topi C57BL/6 sono stati immunizzati mediante formulazione dell'RNA con RNAiMAX Lipofectamine e successiva iniezione endovenosa. Dopo 5 immunizzazioni i topi sono stati sacrificati e gli splenociti sono stati analizzati per le risposte dei linfociti T specifici per la mutazione utilizzando la marcatura delle citochine intracellulari e l'analisi mediante ELISPOT per IFN- γ dopo ristimolazione con il corrispondente peptide codificante per la mutazione o con il peptide di controllo (VSV-NP).

La figura 12 mostra un esempio per ciascuna progettazione di vaccini. Nella riga superiore i topi sono stati vaccinati con l'RNA monoepitopo codificante per MUT30 (mutazione in Kif18b), il quale induce linfociti T CD4+ specifici per MUT30 (si veda il grafico esemplificativo di FACS). Nella riga in mezzo il grafico ed il grafico di FACS mostrano l'induzione di linfociti T CD4+ specifici per MUT08 (mutazione in Ddx23) dopo l'immunizzazione con il biepitopo codificante per MUT33 e MUT08. Nella riga inferiore i topi sono stati immunizzati con un poliepitopo codificante per 16 differenti mutazioni comprese MUT08, MUT33 e MUT27 (si veda la tabella 8). Il grafico ed il grafico di FACS illustrano che i linfociti T reattivi per MUT27 sono di un fenotipo CD8.

Tabella 8. Panoramica delle mutazioni e dei nomi dei geni codificati da vaccini a RNA monoepitopici, biepitopici e poliepitopici.

costrutto	mutazione codificata	annotazione del gene
monoepitopo	MUT30	<i>Kif18b</i>
biepitopo	MUT33	<i>Pbk</i>
	MUT08	<i>Ddx23</i>
	MUT01	<i>Fzd7</i>
	MUT02	<i>Xpot</i>
	MUT03	<i>Ranbp2</i>
	MUT04	<i>Dnajb12</i>
	MUT05	<i>Eef2</i>
	MUT06	<i>Ptrf</i>
	MUT07	<i>Trp53</i>

poliepitopo	MUT08	<i>Ddx23</i>
	MUT26	<i>Orc2</i>
	MUT27	<i>Obsl1</i>
	MUT28	<i>Ppp1r7</i>
	MUT29	<i>Mthfd11</i>
	MUT30	<i>Kif18b</i>
	MUT31	<i>Ascc2</i>
	MUT32	<i>Itn2</i>
	MUT33	<i>Pbk</i>

Lo stesso poliepitopo è stato utilizzato per generare i dati mostrati nella figura 13. Il grafico mostra i dati di ELISPOT dopo la ristimolazione degli splenociti con i peptidi di controllo (VSV-NP), MUT08, MUT27 e MUT33, dimostrando che il vaccino poliepitopico può indurre risposte specifiche dei linfociti T nei confronti delle diverse mutazioni differenti.

Presi insieme i dati mostrano la possibilità di indurre linfociti T specifici per la mutazione utilizzando monoepitopi, biepitopi e poliepitopi codificati dall'RNA. Inoltre, i dati mostrano l'induzione di linfociti T CD4+ e CD8+ e l'induzione di diverse specificità differenti da un costrutto.

Immunizzazione con epitopi modello

Per caratterizzare ulteriormente la progettazione dei vaccini poliepitopici a RNA è stata costruita una sequenza di DNA, la quale ha incluso cinque differenti epitopi modello noti comprendenti un epitopo dell'MHC di classe II (ovalbumina di classe I (SIINFEKL), di classe II (OVA di classe II), nucleoproteina dell'influenza (Inf-NP), nucleoproteina del virus della stomatite vescicolare (VSV-NP) e proteina 2 correlata alla tirosinasi (Trp2)). Gli epitopi sono stati separati con lo stesso gruppo di collegamento di glicina/serina di 9 amminoacidi utilizzato per il poliepitopo con mutazione. Questo costrutto è stato clonato nella struttura di pST1-2BgUTR-A120 per la trascrizione in vitro dell'mRNA.

L'RNA trascritto in vitro è stato utilizzato per vaccinare cinque topi C57BL/6 mediante immunizzazione intranodale (quattro immunizzazioni con 20 µg di RNA nei linfonodi inguinali). Cinque giorni dopo l'ultima immunizzazione i campioni di sangue e di splenociti sono stati presi dai topi per l'analisi. La figura 14A mostra l'analisi mediante ELISPOT per IFN-γ degli splenociti ristimolati con i peptidi indicati. Può essere chiaramente visto che tutti e tre gli epitopi dell'MHC di classe I (SIINFEKL, Trp2 e VSV-NP) inducono un numero molto elevato di linfociti T CD8+ specifici

per l'antigene. Inoltre, l'epitopo dell'MHC di classe II OVA di classe II induce una forte risposta dei linfociti T CD4+. Il quarto epitopo dell'MHC di classe I è stato analizzato mediante la marcatura di linfociti T CD8+ specifici per Inf-NP con un complesso pentamerico MHC-peptide (pentamero) marcato in modo fluorescente (figura 14B).

Questi dati dimostrano che la progettazione del poliepitopo utilizzando il gruppo di collegamento di glicina/serina per separare epitopi immunogenici differenti dell'MHC di classe I e di classe II è in grado di indurre linfociti T specifici nei confronti di ogni epitopo codificato, indipendentemente dalla sua immunodominanza.

Risposta antitumorale dopo la terapia con un vaccino poliepitopico a RNA codificante per una mutazione

Lo stesso poliepitopo che è stato analizzato nella figura 13 per l'immunogenicità è stato utilizzato per indagare l'attività antitumorale degli RNA codificanti per mutazioni nei confronti delle cellule tumorali B16-F10. In dettaglio, ai gruppi di topi C57BL/6 (n=10) sono state inoculate per via sottocutanea 1×10^5 cellule di melanoma B16-F10 nel fianco. I giorni 3, 6, 10, 17 e 21 i topi sono stati immunizzati con l'RNA poliepitopico utilizzando un reagente di trasfezione liposomiale. Al gruppo di controllo sono stati iniettati i liposomi da soli.

La figura 21 mostra le curve di sopravvivenza dei gruppi, rivelando una sopravvivenza media fortemente migliorata di 27 giorni con 1 su 10 topi che sopravvive senza tumore in confronto ad una sopravvivenza media di 18,5 giorni nel gruppo di controllo.

Risposta antitumorale dopo la terapia con una combinazione di peptide mutato e normale

L'attività antitumorale delle mutazioni convalidate è stata valutata mediante un esperimento tumorale terapeutico in vivo utilizzando il MUT30 come vaccino peptidico. In dettaglio, ai gruppi di topi C57BL/6 (n=8) sono state inoculate per via sottocutanea 1×10^5 cellule di melanoma B16-F10 nel fianco. I giorni 3, 10 e 17 i topi sono stati immunizzati utilizzando poli(I:C) come adiuvante con MUT30, con la proteina 2 correlata alla tirosinasi (Trp₂₁₈₀₋₁₈₈) oppure con una combinazione di entrambi i peptidi. Trp2 è un epitopo noto di CD8+ espresso dalle cellule di melanoma B16-F10.

La figura 15A mostra la crescita tumorale media dei gruppi. Può essere chiaramente visto che fino al giorno 28 la crescita tumorale è quasi completamente inibita nel gruppo che è stato immunizzato con la combinazione dell'epitopo noto dei linfociti T CD8+ e di MUT30 che induce i linfociti T CD4+. L'epitopo noto Trp2 da solo non è sufficiente a fornire un buon effetto antitumorale in questa configurazione, ma entrambi i gruppi con singola terapia (MUT30 e Trp2) forniscono ancora un'inibizione della crescita tumorale rispetto al gruppo non trattato dall'inizio dell'esperimento fino al giorno 25. Questi dati vengono rafforzati dalle curve di sopravvivenza mostrate nella figura 15B. Chiaramente la sopravvivenza media è aumentata dai topi a cui sono stati iniettati i peptidi singoli, con 1/8 dei topi che sopravvive nel gruppo con la vaccinazione

con Trp2. Inoltre, il gruppo trattato con entrambi i peptidi mostra una sopravvivenza media anche migliore con 2/8 dei topi che sopravvive.

Presi insieme entrambi gli epitopi agiscono in un modo sinergico a dare un forte effetto antitumorale.

Esempio 9: struttura per il rilevamento di mutazioni somatiche sulla base della confidenza e per l'applicazione a cellule B16-F10 di melanoma

L'NGS è imparziale in quanto esso permette una scoperta a volumi elevati di variazioni all'interno di un intero genoma o di regioni mirate, quali esoni codificanti per proteine.

Tuttavia, anche se in modo rivoluzionario, la piattaforma di NGS è ancora incline a errori conducendo a risultati erronei di variazione. Inoltre, la qualità dei risultati dipende dai parametri del progetto sperimentale e dalle metodologie di analisi. Anche se i risultati di variazione includono tipicamente punteggi progettati per differenziare le variazioni vere dagli errori, né l'utilità di questi punteggi né la loro interpretazione per quanto riguarda l'ottimizzazione degli esperimenti sono completamente comprese. Questo è particolarmente vero confrontando gli stati dei tessuti, come confrontando il tumore e il tessuto normale per le mutazioni somatiche. Come conseguenza, i ricercatori sono forzati a basarsi sull'esperienza personale per determinare i parametri sperimentali e le soglie arbitrarie di filtraggio per selezionare le mutazioni. Il nostro studio mira a) a stabilire una struttura per confrontare i parametri e i metodi per identificare le mutazioni somatiche e b) ad assegnare un valore di confidenza alle mutazioni identificate. Sequenziamo i campioni in triplice copia dai topi C57BL/6 e dalla linea cellulare B16-F10 di melanoma. Utilizzando questi dati, abbiamo formulato il tasso di scoperta falsa delle mutazioni somatiche rilevate, una misura che quindi usiamo per valutare il programma di scoperta delle mutazioni esistenti e i protocolli di laboratorio.

Vari fattori sperimentali ed algoritmici contribuiscono al tasso di falsi positivi per le variazioni trovate mediante NGS [Nothnagel, M. et al., Hum. Genet. 23 febbraio 2011 [pubblicazione online precedente la stampa]]. Le fonti di errore includono artefatti di PCR, influenze nell'innesco [Hansen, K.D., et al., Nucleic. Acids. Res. 38, e131 (2010); Taub, M.A. et al., Genome Med. 2, 87 (2010)] e arricchimento mirato [Bainbridge, M.N. et al., Genome Biol. 11, R62 (2010)], effetti delle sequenze [Nakamura, K. et al., Acids Res. (2011) pubblicato prima online il 16 maggio 2011 doi:10.1093/nar/gkr344], chiamata delle basi che causa errori di sequenza [Kircher, M. et al., Genome Biol. 10, R83 (2009). Pubblicazione elettronica 14 agosto 2009] e allineamento di lettura [Lassmann, T. et al., Bioinformatics 27, da 130 a 131 (2011)], causando una variazione nella copertura ed errori di sequenziamento che influenzano l'ulteriore analisi a valle, per esempio variante trovata intorno a indel [Li, H, Bioinformatics 27, da 1157 a 1158 (2011)].

Nessun modello statistico generale è stato descritto per descrivere l'impatto di differenti fonti di errore sulle chiamate delle mutazioni somatiche; solo singoli aspetti sono coperti senza rimuovere tutte le influenze. Metodi computazionali recenti per misurare la quantità attesa di chiamate di

mutazioni false positive includono l'utilizzo del rapporto di transizione/trasversione di una serie di variazioni [Zhang, Z., Gerstein, M., *Nucleic Acids Res* 31, da 5338 a 5348 (2003); DePristo, M.A. et al., *Nature Genetics* 43, da 491 a 498 (2011)], l'apprendimento della macchina [DePristo, M.A. et al., *Nature Genetics* 43, da 491 a 498 (2011)] e gli errori ereditati lavorando con genomi di famiglia [Ewen, K.R. et al., *Am. J. Hum. Genet.* 67, da 727 a 736 (2000)] o campioni raggruppati [Druley, T.E. et al., *Nature Methods* 6, da 263 a 265 (2009); Bansal, V., *Bioinformatics* 26, da 318 a 324 (2010)]. Per scopi di ottimizzazione, Druley et al. [Druley, T.E. et al., *Nature Methods* 6, da 263 a 265 (2009)] si sono affidati a brevi frammenti di una sequenza plasmidica, i quali tuttavia potrebbero non essere rappresentativi per il campione. Per una serie di singole variazioni nucleotidiche (SNV) e di esperimenti scelti, è realizzabile un confronto con le SNV identificate mediante altre tecniche [Van Tassel, C.P. et al., *Nature Methods* 5, da 247 a 252 (2008)] ma è difficile valutare in termini di nuove mutazioni somatiche.

Utilizzando un progetto di sequenziamento dell'esoma come esempio, proponiamo il calcolo di un tasso di scoperta falsa (FDR) sulla base di dati di NGS da soli. Il metodo non è solo applicabile alla selezione e alla prioritizzazione di bersagli diagnostici e terapeutici, ma supporta anche lo sviluppo di algoritmi e di metodi permettendoci di definire raccomandazioni guidate dalla confidenza per esperimenti simili.

Per scoprire le mutazioni, il DNA dal tessuto caudale di tre topi C57BL/6 (Black6) (compagni di progenie) e il DNA da cellule di melanoma B16-F10 (B16), in triplice copia, sono stati singolarmente arricchiti per esoni codificanti per proteine (Agilent Sure Select per esoma intero di topo), risultando in 6 campioni. L'RNA è stato estratto dalle cellule B16 in triplice copia. Le letture di 50 nucleotidi a singola estremità (1x50 nucleotidi) e di 100 nucleotidi ad estremità appaiate (2x100 nucleotidi) sono state generate su un Illumina HiSeq 2000. Ciascun campione è stato caricato in una corsia individuale, risultando in una media di 104 milioni di letture per corsia. Le letture di DNA sono state allineate al genoma di topo di riferimento utilizzando bwa [Li, H. Durbin, R., *Bioinformatics* 25, da 1754 a 1760 (2009)] e le letture di RNA sono state allineate con bowtie [Langmead, B. et al., *Genome Biol.* 10, R25 (2009)]. Una copertura media di 38 volte per il 97% delle regioni mirate è stata ottenuta per le librerie 1x50 nucleotidi, mentre l'esperimento di 2x100 nucleotidi ha dato una copertura media di 165 volte per il 98% delle regioni mirate.

Le variazioni somatiche sono state indipendentemente identificate utilizzando i pacchetti di programmi informatici SAMtools [Li, H. et al., *Bioinformatics* 25, da 2078 a 2079 (2009)], GATK [DePristo, M.A. et al., *Nature Genetics* 43, da 491 a 498 (2011)] e SomaticSNiPer [Ding, L. et al., *Hum. Mol. Genet.* (2010) pubblicato prima online il 15 settembre 2010] (figura 16) confrontando le singole variazioni nucleotidiche trovate in campioni di B16 rispetto ai loci corrispondenti nei campioni di Black6 (le cellule B16 sono state originariamente derivate da un topo Black6). Le potenziali mutazioni sono state filtrate secondo le raccomandazioni dai rispettivi autori dei programmi informatici (SAMtools e GATK) o selezionando un'appropriata soglia inferiore per il punteggio somatico di SomaticSNiPer, rispettivamente.

Per creare un tasso di scoperta falsa (FDR) per la scoperta delle mutazioni, abbiamo prima incrociato i siti di mutazione e ottenuto 1.355 mutazioni somatiche di qualità elevata come consenso tra tutti e tre i programmi (figura 17). Tuttavia, le differenze osservate nei risultati degli strumenti informatici applicati sono sostanziali. Per evitare conclusioni erranee, abbiamo sviluppato un metodo per assegnare un FDR a ciascuna mutazione utilizzando le repliche. Le ripetizioni tecniche di un campione dovrebbero generare risultati identici e qualsiasi mutazione rilevata in questo “confronto stesso contro stesso” è un falso positivo. Così, per determinare il tasso di scoperta falsa per rilevamento di mutazione somatica in un campione tumorale rispetto ad un campione normale (“confronto tumorale”), noi possiamo utilizzare una ripetizione tecnica del campione normale come riferimento per stimare il numero di falsi positivi.

La figura 18A mostra esempi di variazioni trovate nei dati di Black6/B16, compresa una mutazione somatica (sinistra), una variazione non somatica rispetto al riferimento (in mezzo), e un possibile falso positivo (destra). Ciascuna mutazione somatica può essere associata ad un punteggio di qualità Q. Il numero di falsi positivi nel confronto tumorale indica un certo numero di falsi positivi nel confronto stesso contro stesso. Così, per una data mutazione con punteggio di qualità Q rilevata nel confronto tumorale, stimiamo il tasso di scoperta falsa calcolando il rapporto delle mutazioni stesso contro stesso con un punteggio di Q o migliore del numero complessivo di mutazioni trovate nel confronto tumorale con un punteggio di Q o migliore.

Uno stimolo emerge nel definire Q poiché la maggior parte delle strutture per il rilevamento delle mutazioni computa multipli punteggi di qualità. Qui, noi applichiamo un classificatore a foresta casuale [Breiman, L., Statist. Sci. 16, da 199 a 231 (2001)] per combinare punteggi multipli in un singolo punteggio di qualità Q. Ci riferiamo alla sezione dei metodi per i dettagli riguardanti i dettagli del punteggio di qualità e del calcolo di FDR.

Una potenziale influenza nel confrontare i metodi è la copertura differenziale; abbiamo così normalizzato il tasso di scoperta falsa per la copertura:

$$FDR(Q) = \frac{\#SNV \text{ stesso contro stesso con punteggio} \geq Q}{\#SNV \text{ tumorale con punteggio} \geq Q} \times \frac{\# \text{confronto tumorale di copertura comune}}{\# \text{confronto stesso contro stesso di copertura comune}}$$

Abbiamo calcolato la copertura comune contando tutte le basi del genoma di riferimento che sono coperte da entrambi il campione tumorale e normale o da entrambi i campioni “stesso contro stesso”, rispettivamente.

Valutando il numero di falsi positivi e di positivi a ciascun FDR (si vedano i metodi), generiamo caratteristiche curve operative del ricevitore (ROC) e calcoliamo l’AUC (area sotto la curva) per ciascun metodo di scoperta delle mutazioni, permettendo così un confronto di strategie per la scoperta delle mutazioni (figura 18B).

Inoltre, la selezione dei dati di riferimento potrebbe influenzare il calcolo degli FDR. Utilizzando i dati disponibili di Black6/B16 è possibile creare 18 tripletti (combinazioni di Black6 contro Black6 e di Black6 contro B16). Confrontando le risultanti distribuzioni di FDR per le serie di mutazioni

somatiche, i risultati sono corrispondenti (figura 18B).

Utilizzando questa definizione di un tasso di scoperta falsa, abbiamo stabilito una struttura generica per valutare l'influenza di numerosi parametri sperimentali ed algoritmici sulla serie risultante di mutazioni somatiche. In seguito, applichiamo questa struttura per studiare l'influenza di strumenti informatici, la copertura, il sequenziamento ad estremità appaiate ed il numero di repliche tecniche sull'identificazione delle mutazioni somatiche.

Prima, la scelta dello strumento informatico ha un chiaro impatto sulle mutazioni somatiche identificate (figura 19A). Sui dati testati, SAMtools produce l'arricchimento massimo di veri positivi in una serie di mutazioni somatiche classificate mediante l'FDR. Tuttavia, abbiamo notato che tutti gli strumenti offrono molti parametri e punteggi di qualità per le singole mutazioni. Qui, abbiamo utilizzato le impostazioni predefinite come specificato dagli sviluppatori dell'algoritmo; ci aspettiamo che i parametri possano essere ottimizzati ed evidenziamo che la struttura di FDR qui definita è progettata per funzionare e per valutare una tale ottimizzazione.

Per l'esperimento descritto di sequenziamento di B16, abbiamo sequenziato ciascun campione in una corsia di flusso cellulare individuale e ottenuto una copertura di base media della regione bersaglio di 38 volte per i campioni singoli. Tuttavia, questa copertura potrebbe non essere richiesta per ottenere una serie ugualmente buona di mutazioni somatiche, probabilmente riducendo i costi. Inoltre, l'impatto della profondità di copertura sul rilevamento di SNV dell'intero genoma è stato discusso recentemente [Ajay, S.S. et al., Genome Res. 21, da 1498 a 1505 (2011)]. Allo scopo di studiare l'effetto della copertura sui dati esonici di cattura, abbiamo sottocampionato il numero di letture di sequenza allineate per ogni libreria di 1x50 nucleotidi per generare una copertura approssimativa di 5, 10 e 20 volte, rispettivamente, e quindi abbiamo riapplicato gli algoritmi di chiamata delle mutazioni. Come atteso, una copertura superiore risulta in una migliore serie di mutazioni somatiche (cioè meno falsi positivi), sebbene il miglioramento da 20 volte la copertura al massimo sia marginale (figura 19B).

È lineare simulare e classificare differenti impostazioni sperimentali utilizzando i dati e la struttura disponibili. Confrontando le duplici copie alle triplici copie, le triplici copie non offrono un beneficio rispetto alle duplici copie (figura 19C), mentre le duplici copie offrono un chiaro miglioramento in confronto a uno studio senza alcuna replica. In termini del rapporto di mutazioni somatiche nelle serie di dati, vediamo un arricchimento ad un FDR del 5% dal 24,2% per un esperimento senza repliche al 71,2% per le duplici copie e all'85,8% per le triplici copie. Nonostante l'arricchimento, utilizzare l'intersezione delle triplici copie rimuove più mutazioni con un basso FDR rispetto a quelle con un elevato FDR, come indicato dall'AUC inferiore delle ROC e dallo spostamento della curva a sinistra (figura 19C): la specificità è leggermente aumentata al costo di una sensibilità inferiore.

L'ulteriore libreria sequenziata di 2x100 nucleotidi è stata utilizzata per simulare una libreria di 1x100 nucleotidi, due librerie di 2x50 nucleotidi e due librerie di 1x50 nucleotidi, rispettivamente, mediante rimozione in silico della seconda lettura e/o delle estremità al 3' e al 5' delle letture, risultando in un totale di 5 librerie simulate. Queste librerie sono state confrontate utilizzando gli FDR calcolati di mutazioni predette (figura 19D). Nonostante la copertura media molto più alta (più di 77 contro 38), le mutazioni somatiche trovate utilizzando le librerie di 2x50 nucleotidi al 5' e di 1x100 nucleotidi hanno un'AUC delle ROC inferiore e così una peggiore distribuzione dell'FDR rispetto alla libreria di 1x50 nucleotidi. Questo fenomeno risulta dall'accumulo di elevate mutazioni di FDR in regioni a bassa copertura poiché le serie di basse mutazioni di FDR trovate sono estremamente simili. La conseguenza è che la lunghezza di sequenziamento ottimale è così piccola che le basi sequenziate sono concentrate intorno alle sequenze sonda di cattura (perdendo potenzialmente informazioni sullo stato somatico delle mutazioni in regioni non coperte, però) o dovrebbe essere vicina alla lunghezza del frammento (2x100 nucleotidi=lunghezza totale di 200 nucleotidi per frammenti di circa 250 nucleotidi nel nostro caso), riempiendo efficacemente le interruzioni di copertura. Questo è anche supportato dall'AUC della ROC della libreria di 2x50 nucleotidi al 3' (simulata utilizzando solo le estremità al 3' della libreria di 2x100 nucleotidi) che è superiore a quella della libreria di 2x50 nucleotidi al 5' (simulata utilizzando solo le estremità al 5' della libreria di 2x100 nucleotidi) nonostante la qualità di base inferiore delle estremità di lettura al 3'.

Queste osservazioni ci permettono di definire le procedure di pratica migliore per la scoperta di mutazioni somatiche. Attraverso tutti i parametri valutati, una copertura di 20 volte in entrambi i campioni ed utilizzando un duplicato tecnico arriva vicina ai risultati ottimali in questi campioni relativamente omogenei, anche considerando nel mentre i costi. Una libreria di 1x50 nucleotidi che risulta in approssimativamente 100 milioni di letture sembra essere la scelta più pragmatica per raggiungere questa copertura. Questo rimane vero attraverso tutti i possibili abbinamenti delle serie di dati. Abbiamo applicato retrospettivamente quelle impostazioni dei parametri, non utilizzato alcun ulteriore filtraggio delle chiamate delle varianti grezze, e calcolato gli FDR per 50 mutazioni scelte dall'intersezione di tutti e tre i metodi come mostrato nella figura 17. Tutte le mutazioni sono state confermate mediante una combinazione di risequenziamento secondo Sanger e delle letture di sequenza di B16 mediante RNA-Seq. 44 di quelle mutazioni sarebbero state trovate utilizzando un limite di FDR del 5% (figura 20). Come controllo negativo, abbiamo risequenziato i loci di 44 mutazioni predette con elevati FDR (>50%) ed esaminato le rispettive sequenze nei dati di RNA-Seq. Abbiamo trovato che 37 di queste mutazioni non sono state convalidate mentre i rimanenti sette loci di mutazioni potenziali sono stati entrambi non coperti dalle letture di RNA-Seq ed hanno reso nella reazione di non sequenziamento.

Anche se noi mostriamo l'applicazione della struttura a quattro questioni specifiche, non è per niente limitata a questi parametri, ma può essere applicata per studiare l'influenza di tutti i parametri sperimentali o algoritmici, per esempio l'influenza del programma di allineamento, la scelta di

una metrica di mutazione, o la scelta del fornitore per la selezione esomica.

Abbiamo eseguito tutti gli esperimenti su un set di esperimenti su cellule di melanoma B16; tuttavia, il metodo non è ristretto a questi dati. Il solo requisito è la disponibilità di una serie di dati di riferimento 'stesso contro stesso', significando che almeno una singola ripetizione tecnica di un campione non tumorale dovrebbe essere eseguita per ciascun nuovo protocollo. Anche se i nostri esperimenti indicano che il metodo è robusto in relazione alla scelta della ripetizione tecnica all'interno di certi limiti, così che una ripetizione non sia necessariamente richiesta in ogni singolo esperimento. Tuttavia, il metodo richiede che le varie misure di qualità siano paragonabili tra la serie di dati di riferimento e le serie di dati rimanenti.

All'interno di questo contributo, abbiamo aperto la strada ad una struttura statistica per un rilevamento guidato dal tasso di scoperta falsa di mutazioni somatiche. Questa struttura non è solo applicabile per la selezione di un bersaglio diagnostico o terapeutico, ma permette anche un confronto generico dei passaggi di un protocollo sperimentale e computazionale sui dati di verità generati quasi grezzi. Qui, abbiamo applicato questa idea per prendere decisioni sul protocollo per quanto riguarda gli strumenti informatici, la copertura, le repliche così come il sequenziamento ad estremità appaiate.

Metodi

Cattura e sequenziamento della libreria

Sequenziamento di nuova generazione, sequenziamento del DNA: la cattura esomica per il risequenziamento del DNA è stata eseguita utilizzando il saggio di cattura a base di soluzione Agilent Sure-Select [Gnirke, A., et al., Nat. Biotechnol. 27, da 182 a 189 (2009)], in questo caso progettato per catturare tutti gli esoni di topo noti. 3 µg di DNA genomico purificato sono stati frammentati da 150 a 200 nucleotidi utilizzando un dispositivo ad ultrasuoni Covaris S2. I frammenti di gDNA sono stati riparati alle estremità utilizzando la DNA polimerasi T4, la DNA polimerasi di Klenow e fosforilati al 5' utilizzando la chinasi polinucleotidica T4. I frammenti di gDNA con estremità non appuntite sono stati adenilati al 3' utilizzando il frammento di Klenow (da 3' a 5' eso meno). Gli adattatori ad estremità appaiate con singola sporgenza a T al 3' di Illumina sono stati ligati ai frammenti di gDNA utilizzando un rapporto molare di 10:1 di adattatore e inserto di DNA genomico utilizzando la DNA ligasi T4. I frammenti di gDNA ligati all'adattatore sono stati arricchiti precattura e sequenze specifiche di cellule del flusso sono state aggiunte utilizzando inneschi di PCR Illumina PE 1.0 e 2.0 e la polimerasi Herculase II (Agilent) utilizzando 4 cicli di PCR.

500 ng di frammenti di gDNA ligati all'adattatore arricchiti mediante PCR sono stati ibridati con esche biotinate SureSelect di una libreria di RNA dell'intero esoma di topo di Agilent per 24 ore a 65°C. I complessi ibridati gDNA/esca di RNA sono stati rimossi utilizzando sfere magnetiche

rivestite di streptavidina. I complessi gDNA/esca di RNA sono stati lavati e le esche di RNA scisse durante l'eluizione in tampone di eluizione SureSelect lasciando i frammenti catturati di gDNA arricchiti mediante PCR ligati all'adattatore. I frammenti di gDNA sono stati amplificati mediante PCR dopo la cattura utilizzando la DNA polimerasi Herculase II (Agilent) e gli inneschi di PCR SureSelect GA per 10 cicli.

I lavaggi sono stati eseguiti utilizzando un volume 1,8x di sfere magnetiche AMPure XP (Agencourt). Per i controlli di qualità abbiamo utilizzato il saggio Qubit HS di Invitrogen e la dimensione del frammento è stata determinata utilizzando un saggio del DNA 2100 Bioanalyzer HS di Agilent.

Le librerie di gDNA arricchite in esoma sono state raggruppate sul cBot utilizzando il kit Truseq SR cluster v2.5 utilizzando 7pM e sequenziate sull'Illumina HiSeq2000 utilizzando il kit Truseq SBS.

Analisi dei dati esomici

Le letture di sequenza sono state allineate utilizzando bwa (versione 0.5.8c) [Li, H. Durbin, R., Bioinformatics 25, da 1754 a 1760 (2009)] utilizzando opzioni preimpostate al gruppo genomico mm9 di riferimento di topo [Genome Sequencing Consortium, Nature 420, da 520 a 562 (2002)]. Le letture ambigue - quelle letture mappate in posizioni multiple del genoma come fornite dal risultato di bwa - sono state rimosse. Gli allineamenti rimanenti sono stati separati, indicizzati e convertiti in un formato binario e compresso (BAM) ed i punteggi della qualità di lettura convertiti dai punteggi di qualità dell'Illumina standard phred+64 a quelli del Sanger standard utilizzando script a conchiglia.

Per ciascuna corsia di sequenziamento, le mutazioni sono state identificate utilizzando tre programmi informatici: SAMtools pileup (versione 0.1.8) [Li, H. et al., Bioinformatics 25, da 2078 a 2079 (2009)], GATK (versione 1.0.4418) [DePristo, M.A. et al., Nature Genetics 43, da 491 a 498 (2011)], e SomaticSniper [Ding, L. et al., Hum. Mol. Genet (2010) pubblicato prima online il 15 settembre 2010]. Per SAMtools, sono stati utilizzati le opzioni e i criteri dei filtri raccomandati dall'autore (http://sourceforge.net/apps/mediawiki/SAMtools/index.php?title=SAM_FAQ; visionato nel settembre 2011), compreso un primo ciclo di filtraggio, copertura massima 200. Per un secondo ciclo di filtraggio con SAMtools, il minimo punteggio di qualità degli indel è stato 50, la qualità minima delle mutazioni puntiformi è stata 30. Per chiamare le mutazioni con GATK, noi abbiamo seguito le linee guida della pratica migliore progettate dall'autore sul manuale dell'utente di GATK (http://www.broadinstitute.org/gsa/wiki/index.php/The_Genome_Analysis_Toolkit; visionato nell'ottobre 2010). Per ciascun campione è stato eseguito un riallineamento locale intorno ai siti indel seguito da una ricalibrazione qualitativa di base. Il modulo UnifiedGenotyper è stato applicato ai risultanti file dei dati di allineamento. Quando richiesto, i polimorfismi noti del dbSNP [Sherry, S.T. et al., Nucleic Acids Res. 29, da 308 a 311 (2009)] (versione 128 per mm9) sono stati forniti ai singoli passaggi. Il passaggio di ricalibrazione del punteggio delle varianti è stato omesso e sostituito con l'opzione di filtraggio stringente. Per chiamare le mutazioni con SomaticSniper, sono state utilizzate le opzioni preimpostate e solo le

mutazioni predette con un “punteggio somatico” di 30 o superiore sono state considerate ulteriormente. Inoltre, per ciascun locus potenzialmente mutato abbiamo richiesto una copertura non zero nel tessuto normale e rimosso tutte le mutazioni localizzate in sequenze ripetitive come definito dalla traccia RepeatMasker dell’UCSC Genome Browser per il gruppo mm9 del genoma di topo [Fujita, P.A. et al., *Nucleic Acids Res.* 39, da 876 a 882 (2011)].

RNA-Seq

Librerie indicizzate di cDNA da mRNA-seq sono state preparate da 5 µg di RNA totale utilizzando una versione modificata del protocollo di mRNA-seq di Illumina. L’ mRNA è stato isolato utilizzando sfere magnetiche SeramagOligo(dT) (Thermo Scientific). L’ mRNA isolato è stato frammentato utilizzando cationi bivalenti e calore risultando in frammenti variabili da 160 a 200 coppie di basi. L’ mRNA frammentato è stato convertito in cDNA utilizzando inneschi casuali e SuperScriptII (Invitrogen) cui segue la sintesi del secondo filamento utilizzando la DNA polimerasi I e l’RNasi H. Il cDNA è stato riparato alle estremità utilizzando la DNA polimerasi T4, la DNA polimerasi di Klenow e fosforilato al 5’ utilizzando la chinasi polinucleotidica T4. I frammenti di cDNA con estremità non appuntite sono stati adenilati al 3’ utilizzando il frammento di Klenow (da 3’ a 5’ eso meno). Gli adattatori multipli specifici con singola sporgenza a T al 3’ di Illumina sono stati ligati sui frammenti di cDNA utilizzando la DNA ligasi T4. Le librerie di cDNA sono state purificate e selezionate per dimensione a 300 coppie di basi utilizzando il gel E-Gel SizeSelect al 2% (Invitrogen). L’arricchimento, l’aggiunta di un indice a sei basi di Illumina e di sequenze specifiche di cellule del flusso sono stati fatti mediante PCR utilizzando la DNA polimerasi Phusion (Finnzymes). Tutti i passaggi di rimozione sono stati eseguiti utilizzando un volume 1,8x di sfere magnetiche AMPure XP Agencourt.

Le librerie indicizzate di RNA-seq sono state raggruppate sul cBot utilizzando il kit Truseq SR cluster v2.5 utilizzando 7pM e sequenziate sull’Illumina HiSeq2000 utilizzando il kit Truseq SBS.

I dati grezzi in uscita dall’HiSeq sono stati sottoposti a processo secondo il protocollo standard di Illumina, comprese la rimozione delle letture di bassa qualità e la demultiplazione. Le letture di sequenza sono state quindi allineate alla sequenza genomica di riferimento [Mouse Genome Sequencing Consortium, *Nature* 420, da 520 a 562 (2002)] utilizzando bowtie [Langmead, B. et al., *Genome Biol.* 10, R25 (2009)]. Le coordinate di allineamento sono state confrontate alle coordinate esoniche dei trascritti di RefSeq [Pruitt, K.D. et al., *Nucleic Acids Res.* 33, da 501 a 504 (2005)] e per ciascun trascritto le conte degli allineamenti sovrapposti sono state registrate. Le letture di sequenza non allineate alla sequenza genomica sono state allineate ad una banca dati di tutte le possibili sequenze di giunzione esone-esone dei trascritti di RefSeq [Pruitt, K.D. et al., *Nucleic Acids Res.* 33, da 501 a 504 (2005)]. Le coordinate di allineamento sono state confrontate con le coordinate di RefSeq di esoni e giunzioni,

le letture contate, e normalizzate rispetto a RPKM (numero di letture che mappano per nucleotide kilobase di trascritto per milione di letture mappate [Mortazavi, A. et al., Nat. Methods 5, da 621 a 628 (2008)]) per ciascun trascritto.

Convalida delle SNV

Abbiamo scelto le SNV per la convalida mediante il risequenziamento secondo Sanger e l'RNA. Sono state identificate le SNV che sono state previste da tutti e tre i programmi, non sinonime, e trovate in trascritti aventi un RPKM minimo di 10. Di queste, abbiamo scelto le 50 con i massimi punteggi di qualità degli SNP come forniti dai programmi. Come controllo negativo, sono state scelte 44 SNV che hanno un FDR del 50% o più, sono presenti in solo un campione di linea cellulare e sono previste solo da un programma di chiamata delle mutazioni. Utilizzando il DNA, le varianti selezionate sono state convalidate mediante amplificazione con PCR delle regioni utilizzando 50 ng di DNA, cui segue sequenziamento secondo Sanger (Eurofins MWG Operon, Ebersberg, Germania). Le reazioni sono riuscite per 50 e 32 loci di controlli positivi e negativi, rispettivamente. La convalida è anch'essa stata fatta mediante esame delle letture dell'RNA-Seq tumorale.

Calcolo degli FDR e apprendimento della macchina

Calcolo del punteggio di qualità a foresta casuale: punteggi multipli in uscita di algoritmi di chiamata delle mutazioni comunemente utilizzati (DePristo, M.A. et al., Nature Genetics 43, da 491 a 498 (2011), Li, H. et al., Bioinformatics 25, da 2078 a 2079 (2009), Ding, L. et al., Hum. Mol. Genet (2010) pubblicato prima online il 15 settembre 2010), i quali influenzano tutti potenzialmente la qualità della chiamata della mutazione. Questi includono - ma senza limitazione - la qualità della base d'interesse come assegnata dallo strumento, l'allineamento di qualità per questa posizione, il numero di letture che copre questa posizione o un punteggio per la differenza tra i due genomi rispetto a questa posizione. Per il calcolo del tasso di scoperta falsa noi richiediamo un ordinamento delle mutazioni, tuttavia questa non è direttamente realizzabile per tutte le mutazioni poiché noi potremmo avere informazioni contrastanti dai vari punteggi di qualità.

Noi utilizziamo la strategia seguente per raggiungere un ordinamento completo. In un primo passaggio, applichiamo una definizione molto rigorosa di superiorità assumendo che una mutazione ha una qualità migliore rispetto ad un'altra se e solo se è superiore in tutte le categorie. Così una serie di proprietà di qualità $S=(s_1, \dots, s_n)$ è preferibile a $T=(t_1, \dots, t_n)$, indicato mediante $S>T$, se $s_i>t_i$ per tutti $i=1, \dots, n$. Definiamo un FDR intermedio (IFDR) come segue

$$IFDR(T) = \frac{\#SNV \text{ stesso contro stesso con punteggio } S>T}{\#SNV \text{ tumorali con punteggio } S>T} \times \frac{\#confronto \text{ tumorale di copertura comune}}{\#confronto \text{ stesso contro stesso di copertura comune}}$$

Tuttavia, consideriamo l'IFDR solo come passaggio intermedio poiché in molti casi strettamente correlati, non è realizzabile alcun confronto e così noi non beneficiamo della vasta quantità di dati disponibili. Così, traiamo vantaggio dalla buona proprietà di generalizzazione della regressione a

foresta casuale [Breiman, L., Statist. Sci. 16, da 199 a 231 (2001)] e istruiamo una foresta casuale come implementata in R (squadra principale di sviluppo di R. R: A language and environment for statistical computing. Fondazione R per il calcolo statistico, Vienna, Austria, 2010, Liaw, A., Wiener, M., R News 2, da 18 a 22 (2002)).

Per le mutazioni in entrata m con n proprietà di qualità ciascuna, l'intervallo di valori per ciascuna proprietà è stato determinato e fino a valori di p sono stati campionati con suddivisione uniforme di questo intervallo; quando la serie di valori per una proprietà di qualità è stata più piccola di p , è stata utilizzata questa serie invece della serie campionata. Quindi è generata ciascuna possibile combinazione di valori di qualità campionati o scelti, la quale risulta in un massimo di p^n punti di dati nello spazio di qualità n -dimensionale. Un campione casuale dell'1% di questi punti ed i corrispondenti valori di IFDR sono stati utilizzati come predittore e risposta, rispettivamente, per l'addestramento a foresta casuale.

Il risultante punteggio di regressione è il nostro punteggio di qualità Q generalizzato; esso può essere considerato come una combinazione localmente pesata dei singoli punteggi di qualità. Esso permette un confronto diretto del valore singolo di due qualsiasi mutazioni ed il calcolo del reale tasso di scoperta falsa:

$$FDR(Q) = \frac{\#SNV \text{ stesso contro stesso con punteggio} \geq Q}{\#SNV \text{ tumorale con punteggio} \geq Q} \times \frac{\#confronto \text{ tumorale di copertura comune}}{\#confronto \text{ stesso contro stesso di copertura comune}}$$

Per l'addestramento del modello a foresta casuale utilizzato per creare i risultati per questo studio, abbiamo calcolato l'IFDR del campione sulle mutazioni somatiche di tutti i campioni prima di selezionare il sottoinsieme casuale dell'1%. Questo assicura la mappatura dell'intero spazio di qualità disponibile ai valori di FDR. Abbiamo utilizzato le proprietà di qualità "qualità degli SNP", "profondità di copertura", "qualità del consenso" e "qualità di mappatura di RMS" (SAMtools, $p=20$); "qualità degli SNP", "profondità di copertura", "confidenza della variante/profondità non filtrata" e "qualità di mappatura di RMS" (GATK, $p=20$); o "qualità degli SNP", "profondità di copertura", "qualità del consenso", "qualità di mappatura di RMS" e "punteggio somatico" (SomaticSNiPer, $p=12$), rispettivamente. I valori differenti di p garantiscono una dimensione della serie di grandezza paragonabile.

Calcolo di copertura comune: il numero di possibili chiamate delle mutazioni può introdurre un'importante influenza nella definizione di un tasso di scoperta falsa. Solo se abbiamo lo stesso numero di posizioni possibili dove le mutazioni possono avvenire per il nostro confronto tumorale e per il nostro confronto stesso contro stesso, il numero di mutazioni chiamate è paragonabile e può servire come base per un calcolo del tasso di scoperta falsa. Per correggere questa influenza potenziale, noi usiamo il rapporto di copertura comune. Come copertura comune noi definiamo il numero di basi con copertura di almeno uno in entrambi i campioni che vengono utilizzati per la chiamata della mutazione. Noi calcoliamo la copertura comune singolarmente per il confronto tumorale così come per il confronto stesso contro stesso.

Stima delle ROC

Le curve operanti stimate dal ricevente (ROC) caratteristiche e l'area corrispondente sotto la curva (AUC) sono utili per organizzare i classificatori e per visualizzare la loro prestazione [Fawcett, T., Pattern Recogn. Lett. 27, da 861 a 874 (2006)]. Noi estendiamo questo concetto per valutare la prestazione di procedure sperimentali e computazionali. Tuttavia, la rappresentazione dei grafici di ROC richiede una conoscenza di tutti gli esempi veri e falsi positivi (TP e FP) in una serie di dati, informazione che è solitamente non data e dura da stabilire per dati a volumi elevati (quali i dati di NGS). Così, noi usiamo gli FDR calcolati per stimare i rispettivi tassi di TP e di FP e rappresentiamo un grafico di ROC e calcoliamo un'AUC. L'idea centrale è che l'FDR di una mutazione singola nella serie di dati dà la proporzione di quanto questa mutazione contribuisca alla somma delle mutazioni TP/FP, rispettivamente. Inoltre, per un elenco di assegnazioni casuali a TP e ad FP, l'AUC risultante delle ROC sarà uguale a 0,5 con il nostro metodo, indicando una predizione completamente casuale. Iniziamo con due condizioni:

$$FDR = \frac{FPR}{FPR + TPR}$$

e

$$FPR + TPR = 1$$

con FPR e TPR che sono i rapporti richiesti falsi positivi veri positivi, rispettivamente, per la mutazione data, definendo il punto corrispondente nello spazio di ROC. [1] e [2] possono essere riarrangiate a

$$TPR = 1 - FPR$$

e

$$FPR = FDR$$

Per ottenere una curva ROC stimata, le mutazioni nella serie di dati sono separate dall'FDR e per ciascuna mutazione viene disegnato un punto ai valori cumulativi di TPR e di FPR fino a questa mutazione, divisa per la somma di tutti i valori di TPR e di TPR, rispettivamente. L'AUC è calcolata raggruppando le aree di tutti i trapezoidi consecutivi tra la curva e l'asse delle x.

ELENCO DELLE SEQUENZE

<110> biotech AG et al.
 <120> INDIVIDUALIZED VACCINES FOR CANCER
 <130> 674-75PCT
 <150> PCT/EP2011/002576
 <151> 2011-05-24
 <150> PCT/EP2012/000006
 <151> 2012-01-02
 <160> 19
 <170> Patent version 3.5
 <210> 1
 <211> 15
 <212> PRT
 <213> Artificial sequence
 <220>
 <221> Linker sequence
 <220>
 <221> REPEAT
 <222> (1)..(3)
 <223> Portion of sequence repeated a times, wherein a is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20
 <220>
 <221> REPEAT
 <222> (4)..(6)
 <223> Portion of sequence repeated b times, wherein b is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20
 <220>
 <221> REPEAT
 <222> (7)..(9)
 <223> Portion of sequence repeated c times, wherein c is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20
 <220>
 <221> REPEAT
 <222> (10)..(12)
 <223> Portion of sequence repeated d times, wherein d is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20
 <220>
 <221> REPEAT
 <222> (13)..(15)
 <223> Portion of sequence repeated e times, wherein e is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, or 20
 <220>
 <221> MISC_FEATURE
 <222> (1)..(15)

<223> a + b + c + d + e are different from 0 and preferably are 2 or more, 3 or more, 4 or more or 5 or more

<400> 1

Gly Gly Ser Gly Ser Ser Gly Gly Gly Ser Ser Gly Gly Ser Gly
1 5 10 15

<210> 2
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker sequence

<400> 2

Gly Gly Ser Gly Gly Gly Gly Ser Gly
1 5

<210> 3
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<223> Epitope sequence

<400> 3

Ser Ile Ile Asn Phe Glu Lys Leu
1 5

<210> 4
<211> 30
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated peptide sequence

<400> 4

Ser Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Asn Val
1 5 10 15

Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu Pro Ser
20 25 30

<210> 5
<211> 30
<212> PRT
<213> Artificial Sequence

<220>
<223> Wildtype peptide sequence

<400> 5

Ser Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Lys Val
1 5 10 15

Marco Giovanni Mari
USBM - CPI-090

Ser Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu Pro Ser
20 25 30

<210> 6
<211> 27
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated peptide sequence

<400> 6
Thr Pro Pro Pro Glu Glu Ala Met Pro Phe Glu Phe Asn Gly Pro Ala
1 5 10 15

Gln Gly Asp His Ser Gln Pro Pro Leu Gln Val
20 25

<210> 7
<211> 27
<212> PRT
<213> Artificial Sequence

<220>
<223> Wildtype peptide sequence

<400> 7
Thr Pro Pro Pro Glu Glu Ala Met Pro Phe Glu Phe Asn Gly Pro Ala
1 5 10 15

Gln Gly Asp His Ser Gln Pro Pro Leu Gln Val
20 25

<210> 8
<211> 27
<212> PRT
<213> Artificial Sequence

<220>
<223> Mutated peptide sequence

<400> 8
Arg Val Thr Cys Asn Arg Ala Gly Glu Lys His Cys Phe Ser Ser Asn
1 5 10 15

Glu Ala Ala Arg Asp Phe Gly Gly Ala Ile Gln
20 25

<210> 9
<211> 27
<212> PRT
<213> Wildtype peptide sequence

<400> 9
Arg Val Thr Cys Asn Arg Ala Gly Glu Lys His Cys Phe Thr Ser Asn
1 5 10 15

Glu Ala Ala Arg Asp Phe Gly Gly Ala Ile Glu
20 25

<210> 10
<211> 27
<212> PKT
<213> Artificial Sequence

<220>
<223> mutated peptide sequence

<400> 10

Phe Arg Arg Lys Ala Phe Leu His Trp Tyr Thr Gly Glu Ala Met Asp
1 5 10 15

Glu Met Glu Phe Thr Glu Ala Glu Ser Asn Met
20 25

<210> 11
<211> 27
<212> PKT
<213> Artificial Sequence

<220>
<223> wildtype peptide sequence

<400> 11

Phe Arg Arg Lys Ala Phe Leu His Trp Tyr Thr Gly Glu Gly Met Asp
1 5 10 15

Glu Met Glu Phe Thr Glu Ala Glu Ser Asn Met
20 25

<210> 12
<211> 27
<212> PKT
<213> Artificial Sequence

<220>
<223> Mutated peptide sequence

<400> 12

Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Asn Val Ser
1 5 10 15

Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu
20 25

<210> 13
<211> 27
<212> PKT
<213> Artificial Sequence

<220>
<223> wildtype peptide sequence

<400> 13
 Pro Ser Lys Pro Ser Phe Gln Glu Phe Val Asp Trp Glu Lys Val Ser
 1 5 10 15
 Pro Glu Leu Asn Ser Thr Asp Gln Pro Phe Leu
 20 25
 <210> 14
 <211> 27
 <212> PRT
 <213> Artificial sequence
 <220>
 <221> Mutated peptide sequence
 <400> 14
 His Leu Thr Gln Gln Leu Asp Thr Tyr Ile Leu Lys Asn Val Val Ala
 1 5 10 15
 Phe Ser Arg Thr Asp Lys Tyr Arg Gln Leu Pro
 20 25
 <210> 15
 <211> 27
 <212> PRT
 <213> Artificial sequence
 <220>
 <221> Wildtype peptide sequence
 <400> 15
 His Leu Thr Gln Gln Leu Asp Thr Tyr Ile Leu Lys Asn Phe Val Ala
 1 5 10 15
 Phe Ser Arg Thr Asp Lys Tyr Arg Gln Leu Pro
 20 25
 <210> 16
 <211> 27
 <212> PRT
 <213> Artificial sequence
 <220>
 <221> Mutated peptide sequence
 <400> 16
 Cys Gly Thr Ala Phe Phe Ile Asn Phe Ile Ala Ile Tyr His His Ala
 1 5 10 15
 Ser Arg Ala Ile Pro Phe Gly Thr Met Val Ala
 20 25
 <210> 17
 <211> 27
 <212> PRT
 <213> Artificial sequence

Marco Giovanni Mari
 USBM-CPI-090

```

<210>
<223> wildtype peptide sequence
<400> 17
Cys Gly Thr Ala Phe Phe Ile Asn Phe Ile Ala Ile Tyr Tyr His Ala
1          5          10          15

Ser Arg Ala Ile Pro Phe Gly Thr Met Val Ala
          20          25

<210> 18
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker sequence
<400> 18
Gly Gly Ser Gly Gly Gly Gly
1          5

<210> 19
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker sequence
<400> 19
Thr Ser Leu Asn Ala Leu Leu Asn Ala His
1          5          10

<210> 20
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker sequence
<400> 20
Ser Ser Ser Ser Ser Ser Ser Ser Ser Ser
1          5          10

<210> 21
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<223> Linker sequence
<400> 21
Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
1          5          10

```


 Marco Giovanni Mari
 USBA - CPI-090

<210> 22
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 22
Gly Gly Ser Gly Gly Gly Ser Gly Gly
1 5 10

<210> 23
<211> 10
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 23
Gly Gly Ser Gly Gly Gly Ser Gly Gly Gly
1 5 10

<210> 24
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 24
Thr Ser Leu Asp Ala Leu Leu Asp Ala
1 5

<210> 25
<211> 8
<212> PRT
<213> Artificial Sequence

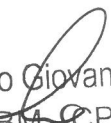
<220>
<221> Linker sequence

<400> 25
Ser Ser Ser Ser Ser Ser Ser Ser
1 5

<210> 26
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 26

Marco Giovanni Mari
USBM  CPI-090

Gly Gly Gly Gly Gly Gly Gly Gly Gly Gly
1 5

<210> 27
<211> 9
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 27

Gly Gly Ser Gly Gly Gly Ser Gly Gly
1 5

<210> 28
<211> 5
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 28

Gly Gly Ser Gly Gly
1 5

<210> 29
<211> 6
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 29

Gly Gly Ser Gly Gly Gly
1 5

<210> 30
<211> 7
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

<400> 30

Gly Gly Ser Gly Gly Gly Gly
1 5

<210> 31
<211> 8
<212> PRT
<213> Artificial Sequence

<220>
<221> Linker sequence

Marco Giovanni Mari
USBM-RCPI-090

```

<400> 31
Gly Gly Ser Gly Gly Gly Ser
1 5

<210> 32
<211> 10
<212> PRT
<213> Artificial Sequence
<220>
<221> Linker sequence
<400> 32
Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
1 5 10

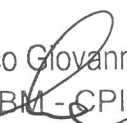
<210> 33
<211> 14
<212> DNA
<213> Artificial Sequence
<220>
<221> Oligonucleotide
<400> 33
ggaaacttt tccc 14

<210> 34
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<221> oligonucleotide
<400> 34
ccggggaaac tttttccag t 21

<210> 35
<211> 21
<212> DNA
<213> Artificial Sequence
<220>
<221> Oligonucleotide
<400> 35
ccggggaaac gttttccag t 21

<210> 36
<211> 60
<212> PRT
<213> Artificial Sequence
<220>
<221> Artificial construct
<400> 36
Asa Thr Thr Phe Asn Val Ser Glu Glu Ser Pro Ser Lys Pro Ser Phe
1 5 10 15

```


 Marco Giovanni Mari
 USBM - CPI-090

Gln Glu Phe Val Asp Trp Glu Asn Val Ser Pro Glu Leu Asn Ser Thr
 20 25 30
 Asp Gln Pro Phe Leu Pro Ser Ala Pro Val Phe Ile Phe Thr Lys Gly
 35 40 45
 Arg Lys Gly Gly Ser Gly Gly Gly Ser Gly Gly
 50 55 60
 <210> 37
 <211> 60
 <212> PRT
 <213> Artificial sequence
 <220>
 <223> Artificial construct
 <400> 37
 Tyr Phe Val Leu Tyr Lys Pro Pro Pro Lys Asp Asn Ile Pro Ala Leu
 1 5 10 15
 Val Glu Glu Tyr Leu Glu Arg Gly Asn Phe Val Ala Asn Asp Leu Asp
 20 25 30
 Trp Leu Leu Ala Leu Pro His Asp Lys Phe Trp Cys Gln Val Ile Phe
 35 40 45
 Asp Glu Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
 50 55 60
 <210> 38
 <211> 60
 <212> PRT
 <213> Artificial sequence
 <220>
 <223> Artificial construct
 <400> 38
 Phe Glu Gln Leu Ser Glu Ser Ala Lys Glu Glu Leu Ile Asn Phe Lys
 1 5 10 15
 Arg Lys Arg Val Ala Ala Phe Gln Lys Asn Leu Ile Glu Met Ser Glu
 20 25 30
 Leu Glu Ile Lys His Ala Arg Asn Asn Val Ser Leu Leu Gln Ser Cys
 35 40 45
 Ile Asp Gly Gly Ser Gly Gly Gly Gly Ser Gly Gly
 50 55 60
 <210> 39
 <211> 60
 <212> PRT
 <213> Artificial sequence
 <220>
 <223> Artificial construct
 <400> 39
 Tyr Gly Ala Ser Pro Gly Leu Gly Ala Tyr Thr Pro Pro Pro Glu Glu
 1 5 10 15
 Ala Met Pro Phe Glu Phe Asn Gly Pro Ala Gln Gly Asp His Ser Gln
 20 25 30
 Pro Pro Leu Gln Val Pro Asp Leu Ala Pro Gly Gly Pro Glu Ala Leu
 35 40 45
 Val Pro Gly Gly Ser Gly Gly Gly Ser Gly Gly Ser
 50 55 60

["INDIVIDUALIZED VACCINES FOR CANCER" = VACCINI INDIVIDUALIZZATI PER IL CANCRO

"PatentIn version 3.5" = PatentIn versione 3.5

"Artificial Sequence" = sequenza artificiale

“Linker sequence” = sequenza di collegamento

“REPEAT” = RIPETERE

“Portion of sequence repeated a times, wherein a is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20” = porzione di sequenza ripetuta a volte, dove a è indipendentemente un numero scelto tra 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

o 20

“Portion of sequence repeated b times, wherein b is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20” = porzione di sequenza ripetuta b volte, dove b è indipendentemente un numero scelto tra 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

o 20

“Portion of sequence repeated c times, wherein c is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20” = porzione di sequenza ripetuta c volte, dove c è indipendentemente un numero scelto tra 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

o 20

“Portion of sequence repeated d times, wherein d is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20” = porzione di sequenza ripetuta d volte, dove d è indipendentemente un numero scelto tra 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

o 20

“Portion of sequence repeated e times, wherein e is independently a number selected from 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19 or 20” = porzione di sequenza ripetuta e volte, dove e è indipendentemente un numero scelto tra 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19

o 20

“MISC_FEATURE” = caratteristica MISC

“a+b+c+d+e are different from 0 and preferably are 2 or more, 3 or more, 4 or more or 5 or more” = a+b+c+d+e sono diversi da 0 e sono preferibilmente 2

o più, 3 o più, 4 o più o 5 o più

“Epitope sequence” = sequenza epitopica

“Mutated peptide sequence” = sequenza del peptide mutato

“wildtype peptide sequence” = sequenza del peptide selvatico

“oligonucleotide” = oligonucleotide

“Artificial construct” = costruito artificiale]

RIVENDICAZIONI

1. Vaccino individualizzato contro il cancro per uso in un metodo di trattamento di un paziente con il cancro, il suddetto metodo comprendendo i passaggi:
 - (a) fornire il suddetto vaccino individualizzato contro il cancro mediante un metodo comprendente i passaggi:
 - (aa) identificare le mutazioni somatiche specifiche per il cancro in un campione tumorale del paziente con il cancro per fornire una firma delle mutazioni tumorali del paziente; e
 - (ab) fornire un vaccino a RNA che comprende la firma delle mutazioni tumorali ottenute nel passaggio (aa), in cui il vaccino a RNA che comprende la firma delle mutazioni del paziente comprende RNA codificante per un polipeptide poliepitopico ricombinante comprendente neoepitopi basati su mutazioni; e
 - (b) somministrare il suddetto vaccino individualizzato contro il cancro al paziente.
2. Vaccino individualizzato contro il cancro per l'uso secondo la rivendicazione 1, in cui il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro comprende l'identificazione della firma delle mutazioni tumorali dell'esoma di una o più cellule cancerose.
3. Vaccino individualizzato contro il cancro per l'uso secondo la rivendicazione 1 o 2, in cui il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro comprende il sequenziamento a singola cellula di una o più cellule cancerose.
4. Vaccino individualizzato contro il cancro per l'uso secondo la rivendicazione 3, in cui le cellule cancerose sono cellule tumorali circolanti.
5. Vaccino individualizzato contro il cancro per l'uso secondo una qualsiasi delle rivendicazioni da 1 a 4, in cui il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro implica l'utilizzo del sequenziamento di nuova generazione (NGS).
6. Vaccino individualizzato contro il cancro per l'uso secondo una qualsiasi delle rivendicazioni da 1 a 5, in cui il passaggio di identificazione delle mutazioni somatiche specifiche per il cancro comprende il sequenziamento del DNA e/o dell'RNA genomici del campione tumorale.
7. Vaccino individualizzato contro il cancro per l'uso secondo una qualsiasi delle rivendicazioni da 1 a 6, in cui il polipeptide poliepitopico ricombinante comprende fino a 30 neoepitopi basati su mutazioni.
8. Vaccino individualizzato contro il cancro per l'uso secondo la rivendicazione 7, in cui il polipeptide poliepitopico ricombinante comprende inoltre epitopi non contenenti mutazioni somatiche specifiche per il cancro che vengono espresse dalle cellule cancerose.
9. Vaccino individualizzato contro il cancro per l'uso secondo la rivendicazione 7 o 8, in cui gli epitopi sono nel loro contesto naturale di sequenza in modo tale da formare una sequenza vaccinale, in cui la sequenza vaccinale comprende preferibilmente 20 o più amminoacidi.

10. Vaccino individualizzato contro il cancro per l'uso secondo una qualsiasi delle rivendicazioni da 1 a 9, in cui i neoepitopi, gli epitopi e/o le sequenze vaccinali sono allineati testa-coda e/o sono distanziati mediante gruppi di collegamento.

* * * * *

Si attesta la perfetta conformità della traduzione che precede.


Marco Giovanni Mari
USBM - CPI-090

LEGENDA DELLE TAVOLE DEI DISEGNI

TAVOLA 1/25

FIGURA 1

“Fig.” = figura

“Tumor cells” = cellule tumorali

“RNA profiling with NGS” = profilo dell’RNA con NGS

“Normal cells” = cellule normali

“DNA profiling with NGS” = profilo del DNA con NGS

“Somatic mutation identification” = identificazione della mutazione somatica

“Mutation in transcript” = mutazione nel trascritto

“Mutation non-synonymous” = mutazione non sinonima

“Mutation expressed” = mutazione espressa

“Mutation immunogenic” = mutazione immunogenica

“Mutation validation” = convalida della mutazione

“B16 melanoma” = melanoma B16

“Black6 cells” = cellule Black6

“DNA exon profiling with NGS” = profilo esonico del DNA con NGS

“50 of 50 tested” = 50 di 50 testati

“Mutations for vaccine desing” = mutazioni per la progettazione di un vaccino

TAVOLA 2/25

FIGURA 2

“Fig.” = figura

“Validated mutation in Kif18b” = mutazione convalidata in Kif18b

“Wild type” = tipo selvatico

“Tumor cells” = cellule tumorali

Marco Giovanni Mari
USBM - CPT-090

TAVOLA 3/25

FIGURA 3

“Fig.” = figura

“IFN γ Spots/mouse” = macchie di IFN γ /topo

TAVOLA 4/25

FIGURA 4

“Fig.” = figura

“Tumor free mice” = topi privi di tumore

“PolyI:C” = poliI:C

“Peptide Mut-30+PolyI:C” = peptide Mut-30+poliI:C

TAVOLA 5/25

FIGURA 5

“Fig.” = figura

“Standard RNA” = RNA standard

“Optimized RNA” = RNA ottimizzato

“IFN γ spots” = macchie di IFN γ

“pp65 overlapping peptides 1-138” = peptidi 1-138 sovrapposti di pp65

“Survival Rate in %” = tasso di sopravvivenza in %

“Untreated control” = controllo non trattato

“Adjuvant alone” = adiuvante da solo

“RNA vaccine” = vaccino a RNA

“RNA+adjuvant” = RNA+adiuvante

“Days after start of immunisation” = giorni dopo l’inizio dell’immunizzazione

“Seq.” = sequenza

“Targeting signal” = segnale di indirizzamento

“poly-A” = poli-A

“RNA backbone” = scheletro dell’RNA

“Vaccine sequence” = sequenza del vaccino

TAVOLA 6/25

FIGURA 6

“Fig.” = figura

“Seq.” = sequenza

“poly-A” = poli-A

“RNA backbone” = scheletro dell’RNA

“Vaccine sequence” = sequenza del vaccino

“Linker” = gruppo di collegamento

TAVOLA 7/25

FIGURA 7

“Fig.” = figura

“IFN γ Spots/5x10⁵ spleen cells” = macchie di IFN γ /5x10⁵ cellule della milza

“Peptide” = peptide

“RNA control” = controllo dell’RNA

“Peptide control” = controllo del peptide

TAVOLA 8/25

FIGURA 8

“Fig.” = figura

“Construct with linker” = costruito con gruppo di collegamento

“Peptide” = peptide

“Linker” = gruppo di collegamento

“aa” = amminoacidi

“Good epitope (in peptide)” = buon epitopo (nel peptide)

“Bad epitope (overlaps junction)” = cattivo epitopo (si sovrappone ad una giunzione)

“Construct without linker” = costruito senza gruppo di collegamento

TAVOLA 9/25

FIGURA 9

“Fig.” = figura

“Mutation discovery” = scoperta della mutazione

“DNA and RNA sequencing” = sequenziamento del DNA e dell’RNA

“Sequence analysis and mutation identification” = analisi di sequenza ed identificazione della mutazione

“Normal” = normale

“Tumor” = tumore

“Mutation prioritization” = prioritizzazione della mutazione

“Mutation validation” = convalida della mutazione

“Immunogenicity testing” = prova di immunogenicità

“Validated mutation in Kif18b” = mutazione convalidata in Kif18b

“Wild type” = tipo selvatico

“Tumor cells” = cellule tumorali

“Validated mutations” = mutazioni convalidate

“Immunize with peptide containing mutation” = immunizzare con il peptide contenente la mutazione

“ELISPOT mutated and wild type” = ELISPOT mutato e selvatico

“RNA profiling with NGS” = profilo dell’RNA con NGS

“Normal cells” = cellule normali

“DNA exon profiling with NGS” = profilo esonico del DNA con NGS

“Somatic mutations” = mutazioni somatiche

“Mutation in transcript” = mutazione nel trascritto

“Mutation non-synonymous” = mutazione non sinonima

“Gene with mutation expressed” = gene con la mutazione espressa

“Mutations validated in DNA” = mutazioni convalidate nel DNA

“Mutations immunogenic” = mutazioni immunogeniche

“B16F10 melanoma” = melanoma B16F10

“C57BL/6 mice” = topi C57BL/6

“50 of 50” = 50 di 50

“22 of 50” = 22 di 50

“14,462 genes expressed” = 14.462 geni espressi

TAVOLA 10/25

FIGURA 9

“Fig.” = figura

“B16F10 genome” = genoma di B16F10

TAVOLA 11/25

FIGURA 10

“Fig.” = figura

“positive control” = controllo positivo

“no reactivity” = nessuna reattività

“reactivity against MUT” = reattività nei confronti di MUT

“reactivity against MUT+WT” = reattività nei confronti di MUT+WT

“Trp2 peptide” = peptide di Trp2

“no peptide” = nessun peptide

“mouse” = topo

“control peptide” = peptide di controllo

“MUT peptide” = peptide MUT

“WT peptide” = peptide WT

“IFN γ Spots/5x10⁵ splenocytes” = macchie di IFN γ /5x10⁵ splenociti

“control RNA” = RNA di controllo

“MUT RNA” = RNA MUT

TAVOLA 12/25

FIGURA 10

“Fig.” = figura

“B16F10 tumor cells” = cellule tumorali B16F10

“Domain (Kinesin-motor)” = dominio (motore chinesiaco)

“Nucleotide binding” = legame al nucleotide

“Coiled coil (Potential)” = spirale avvolta (potenziale)

“Motif (Nuclear localization signal)” = motivo strutturale (segnale di localizzazione nucleare)

TAVOLA 13/25

FIGURA 11

“Fig.” = figura

“Tumor size” = dimensione del tumore

“poly(I:C) alone” = poli(I:C) da solo

“vacc.” = vaccino

“Days after tumor inoculation” = giorni dopo l’inoculo del tumore

“Percent survival” = sopravvivenza percentuale

TAVOLA 14/25

FIGURA 12

“Fig.” = figura

“Vaccinated with Monoepitope RNA” = vaccinato con RNA monoepitopico

“% of CD4+/IFN- γ + cells” = % di cellule CD4+/IFN- γ +

“Vaccinated with Biepitope RNA” = vaccinato con RNA biepitopico

“Vaccinated with Polyepitope RNA” = vaccinato con RNA poliepitopico

“% of CD8+/IFN- γ + cells” = % di cellule CD8+/IFN- γ +

TAVOLA 15/25

FIGURA 13

“Fig.” = figura

“IFN γ spots/5x10⁵ splenocytes” = macchie di IFN γ /5x10⁵ splenociti

“Restimulation with:” = ristimolazione con:

“Vaccinated with Polyepitope RNA” = vaccinato con RNA poliepitopico

TAVOLA 16/25

FIGURA 14

“Fig.” = figura

“spots/5x10⁵ splenocytes” = macchie/5x10⁵ splenociti

“no peptide” = nessun peptide

“OVA class II” = OVA di classe II

“control” = controllo

“Inf-NP Pentamer” = pentamero Inf-NP

“model Polyepitope” = poliepitopo modello

TAVOLA 17/25

FIGURA 15A

“Fig.” = figura

“Tumor size” = dimensione del tumore

“Days after tumor inoculation” = giorni dopo l’inoculo del tumore

“untreated” = non trattato

TAVOLA 18/25

FIGURA 15B

“Fig.” = figura

“Percent survival” = sopravvivenza percentuale

“Days after tumor inoculation” = giorni dopo l’inoculo del tumore

“untreated” = non trattato

TAVOLA 19/25

FIGURA 16

“Fig.” = figura

“104,671,967 sequence reads” = 104.671.967 letture di sequenza

“alignment to genome with BWA, sorting, indexing” = allineamento al genoma con BWA, selezione, indicizzazione

“83,367,584 non-ambiguous alignments (38.89% on target)” = 83.367.584 allineamenti non ambigui (38,89% sul bersaglio)

“local realignment” = riallineamento locale

“base quality score recalibration” = ricalibrazione del punteggio di qualità di base

“SNV calling” = chiamata delle SNV

“indel calling” = chiamata degli indel

“variation call (SNVs and indels)” = chiamata di variazione (SNV e indel)

“somatic variation call” = chiamata di variazione somatica

“variation calls” = chiamate di variazione

“tumor versus normal comparison” = confronto tumore contro normale

“somatic variation calls” = chiamate di variazione somatica

“in exons” = negli esoni

“filtering” = filtraggio

“final selection” = selezione finale

TAVOLA 20/25

FIGURA 17

“Fig.” = figura

TAVOLA 21/25

FIGURA 18

“Fig.” = figura

“black 6” = black 6

“somatic” = somatico

“non somatic” = non somatico

“same vs. same” = stesso contro stesso

TAVOLA 22/25

FIGURA 18

“Fig.” = figura

“ratio of true variations” = rapporto delle variazioni vere

“ratio of false positives” = rapporto dei falsi positivi

TAVOLA 23/25

FIGURA 19

“Fig.” = figura

“ratio of true variations” = rapporto delle variazioni vere

“ratio of false positives” = rapporto dei falsi positivi

“coverage” = copertura

“no replicates” = nessun replicato

“duplicate” = duplicato

“triplicate” = triplicato

TAVOLA 24/25

FIGURA 20

“Fig.” = figura

“Chromosome” = cromosoma

“Position” = posizione

“Reference allele” = allele di riferimento

“Sample allele(s)” = allele/i del campione

“relative rank [%]” = posizione relativa [%]

“Influence of FDR” = influenza dell’FDR

“% found” = % trovata

“% of validated mutations found” = % di mutazioni convalidate trovate

“% of all mutations found” = % di tutte le mutazioni trovate

TAVOLA 25/25

FIGURA 21

“Fig.” = figura

“Percent survival” = sopravvivenza percentuale

“Days after tumor inoculation” = giorni dopo l’inoculo del tumore

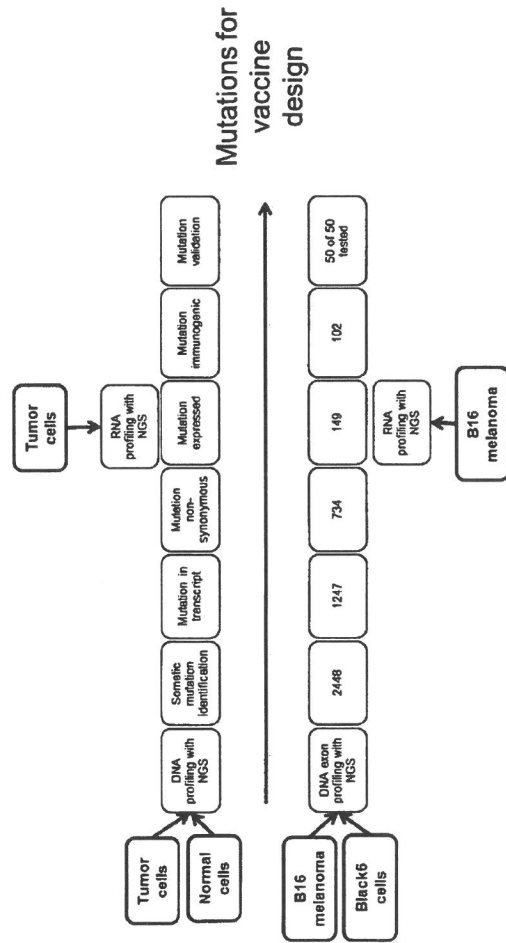
“Polyepitope RNA” = RNA poliepitopico

“Liposome alone” = liposoma da solo

Si attesta la perfetta conformità della traduzione che precede.

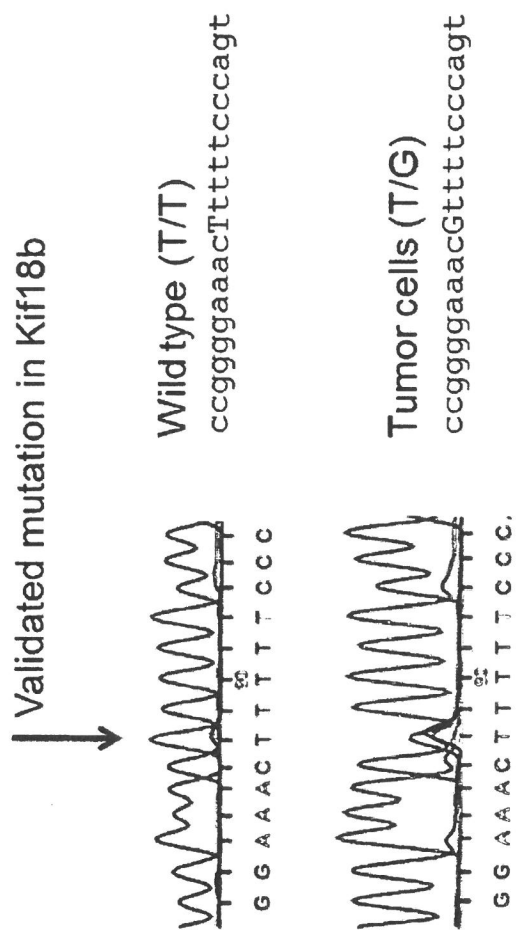

Marco Giovanni Mari
USBM - CPI-090

Fig. 1



Marco Giovanni Marin
USBM-CPI-090

Fig. 2



Marco Giovanni Mari
USBM CPL090

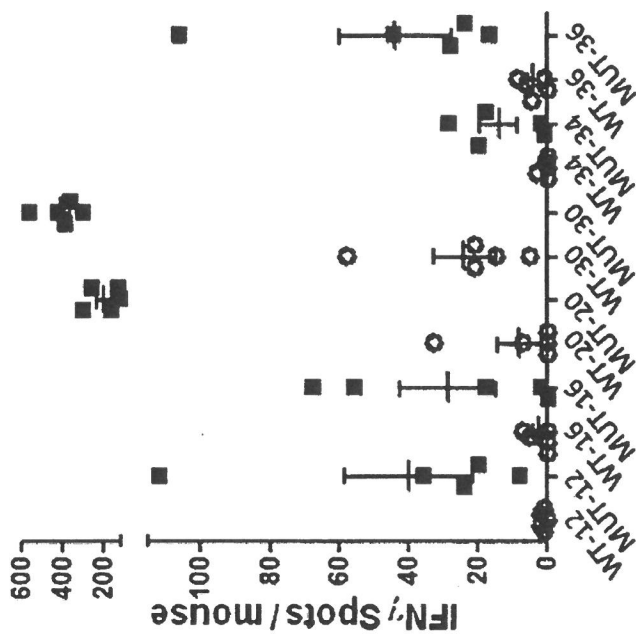


Fig. 3

Marco Giovanni Mari
USBM - CPI-090

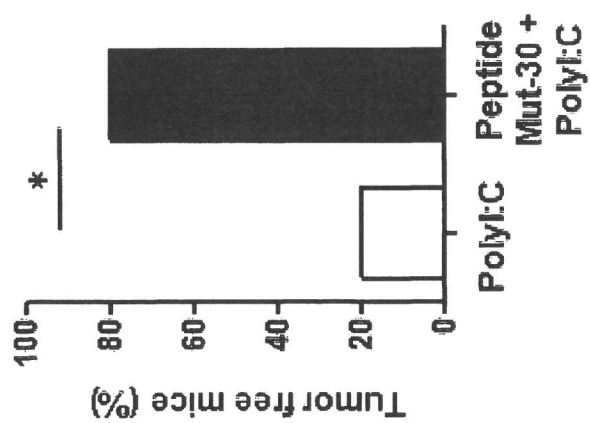


Fig. 4

Marco Giovanni Mari
USBM - SPI-090

Fig. 5

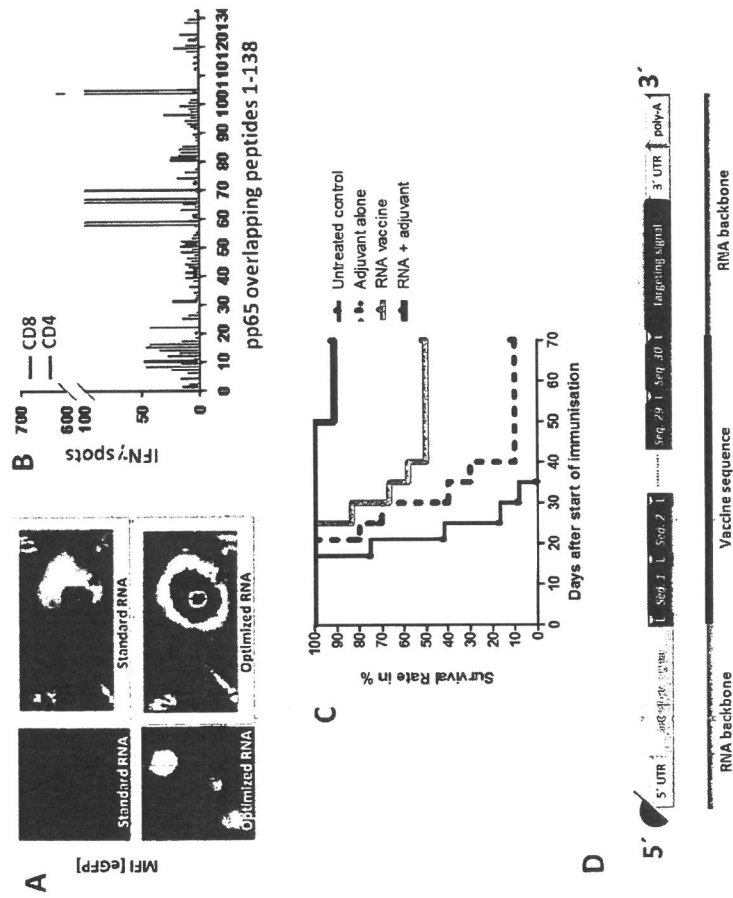


Fig. 6

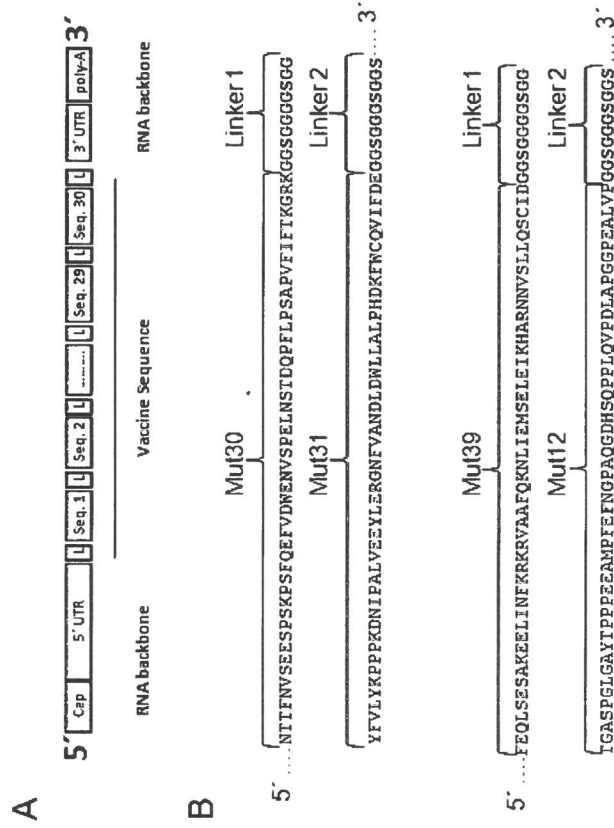
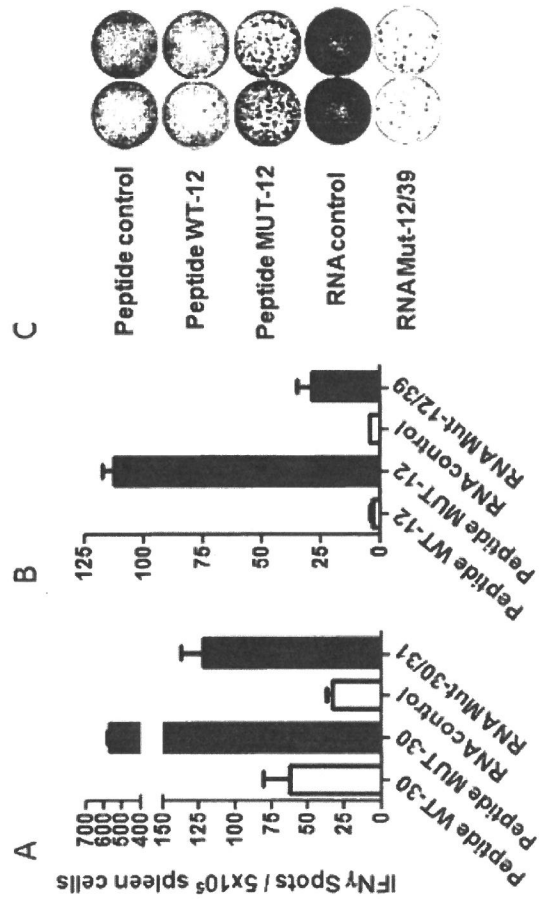


Fig. 7



Marco Giovanni Mari
USBM-CPI090

Fig. 8

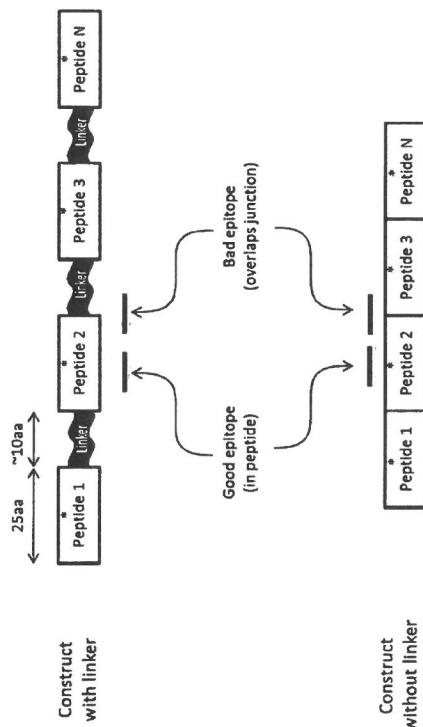
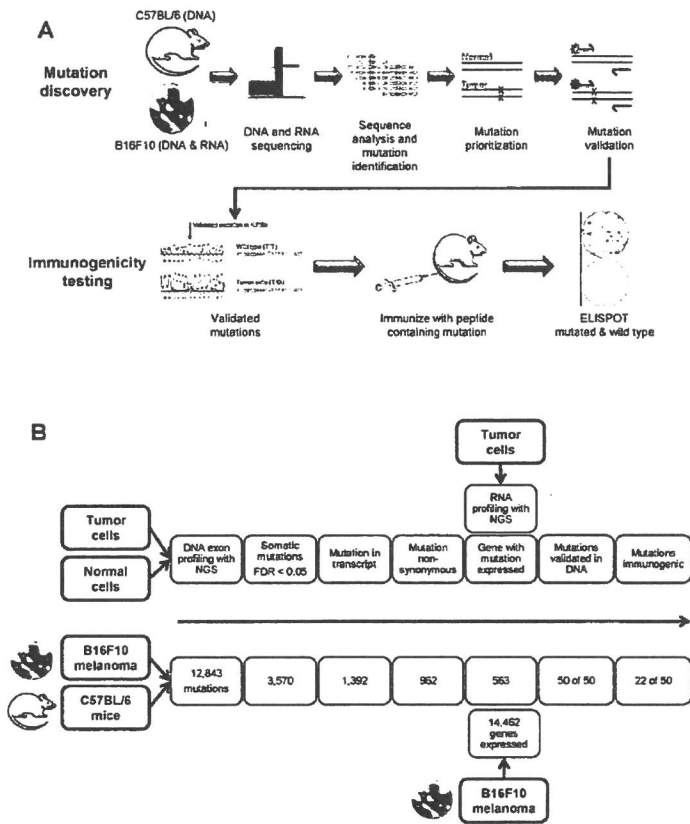


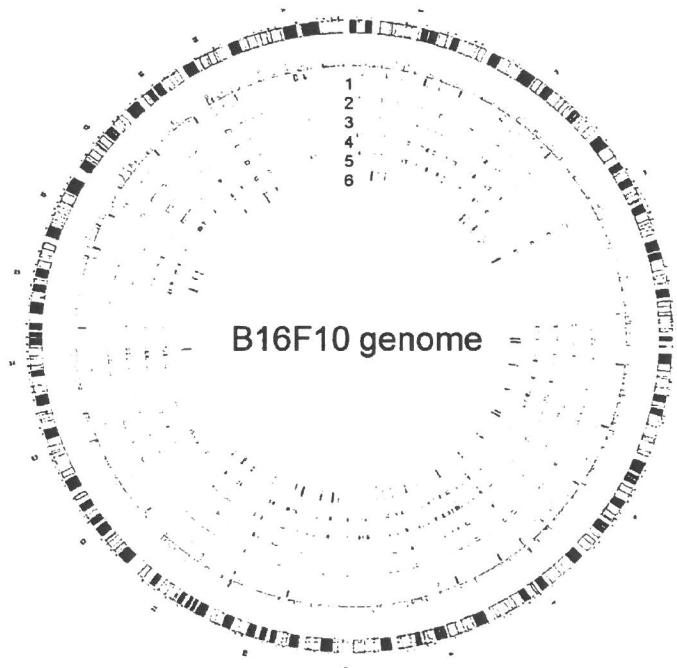
Fig. 9



Marco Giovanni Mari
 USBM - CPI-090

Fig. 9

C



Marco Giovanni Mari
USBM - CPI-090

Fig. 10

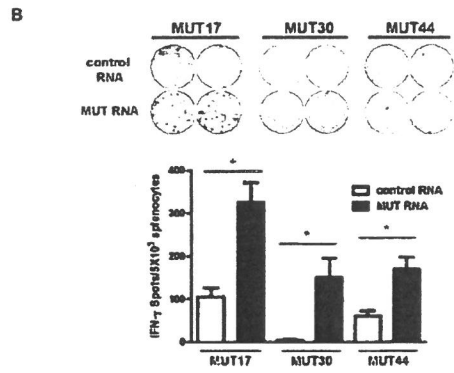
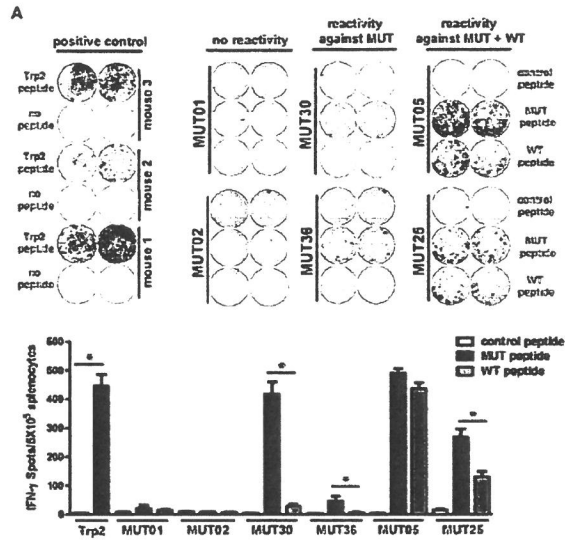
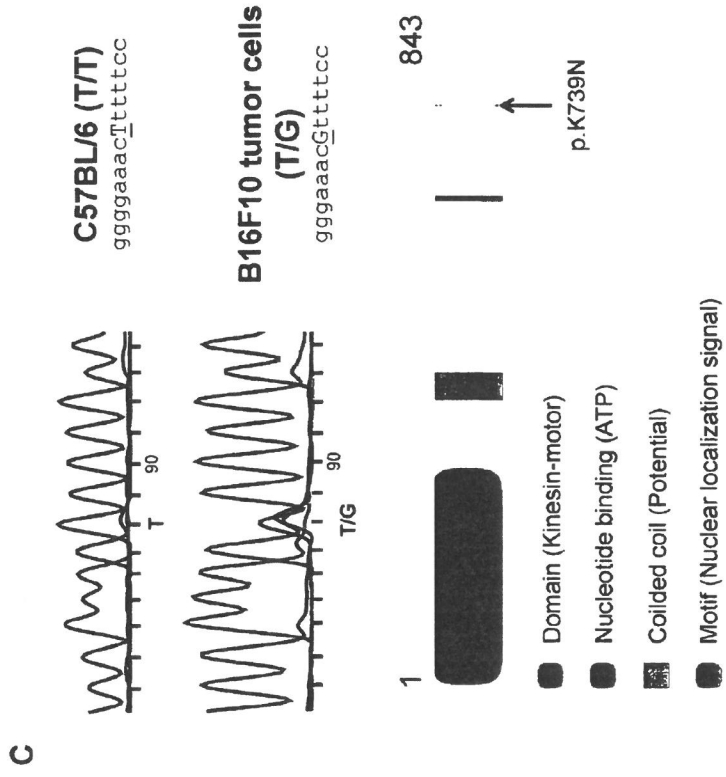
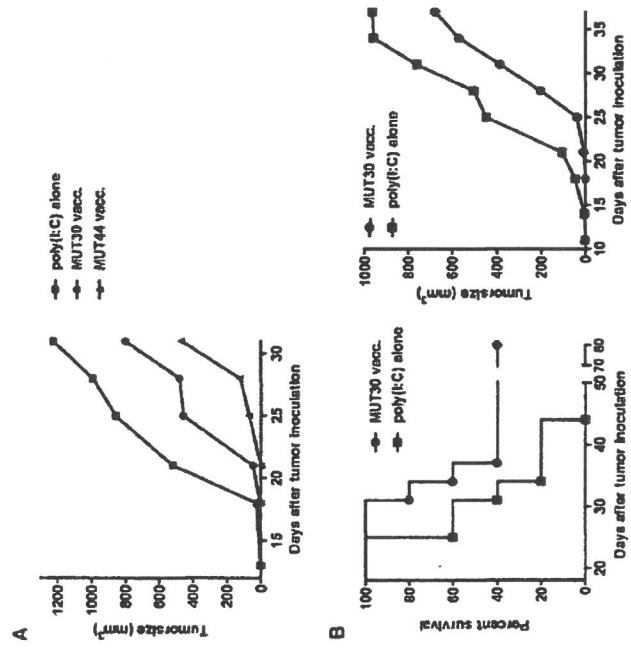


Fig. 10



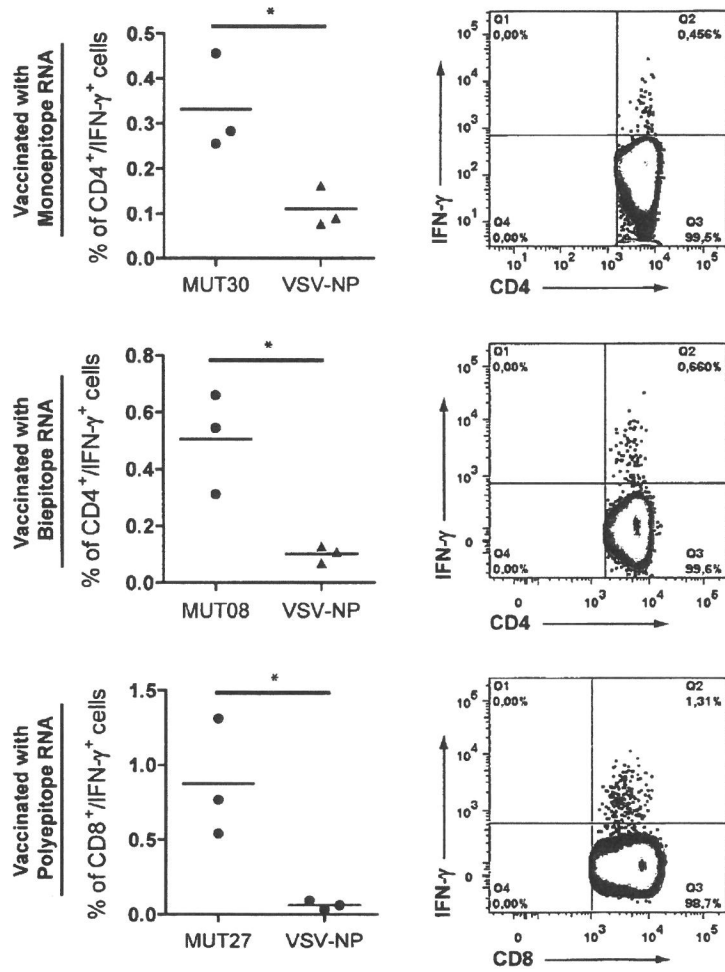
Marco Giovanni Mari
USBM - CPL 090

Fig. 11



Marco Giovanni Mari
USBM SCPI 090

Fig. 12



Marco Giovanni Mari
 USBM CPI-090

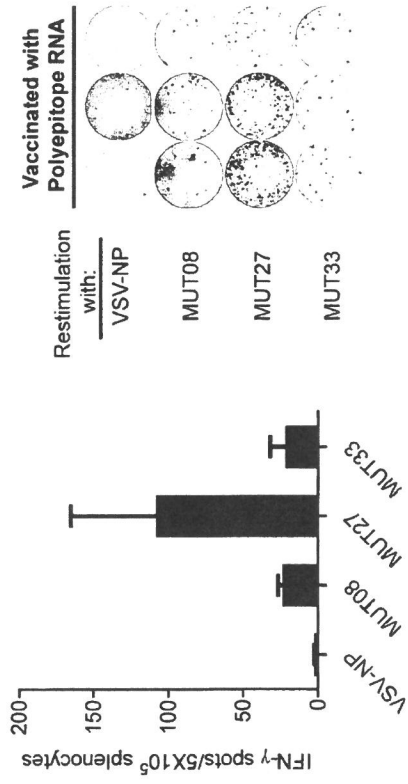
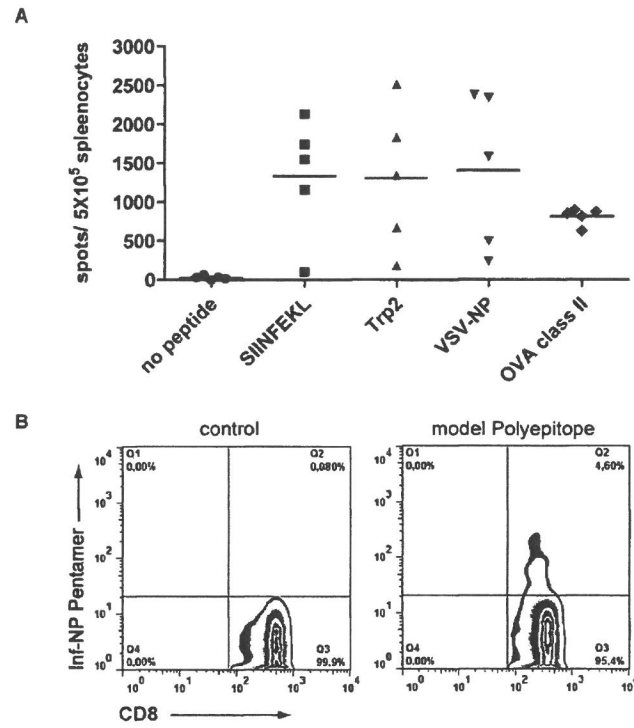


Fig. 13

Marco Giovanni Mari
USBM - CPI-090

Fig. 14



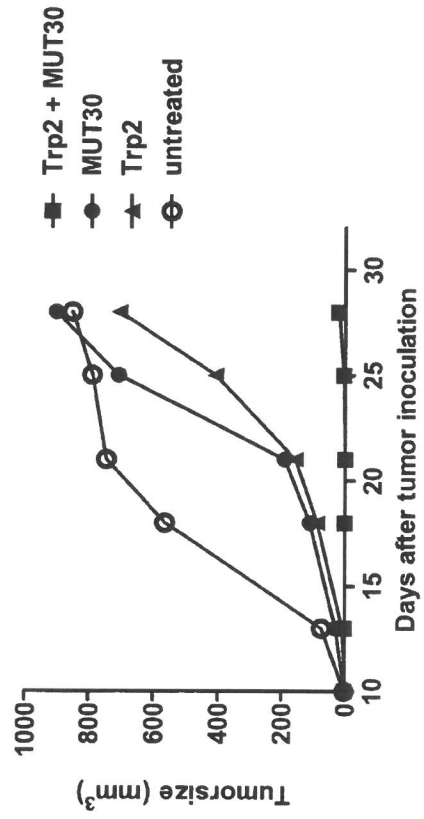


Fig. 15 A

Marco Giovanni Mari
USBM - GPI-090

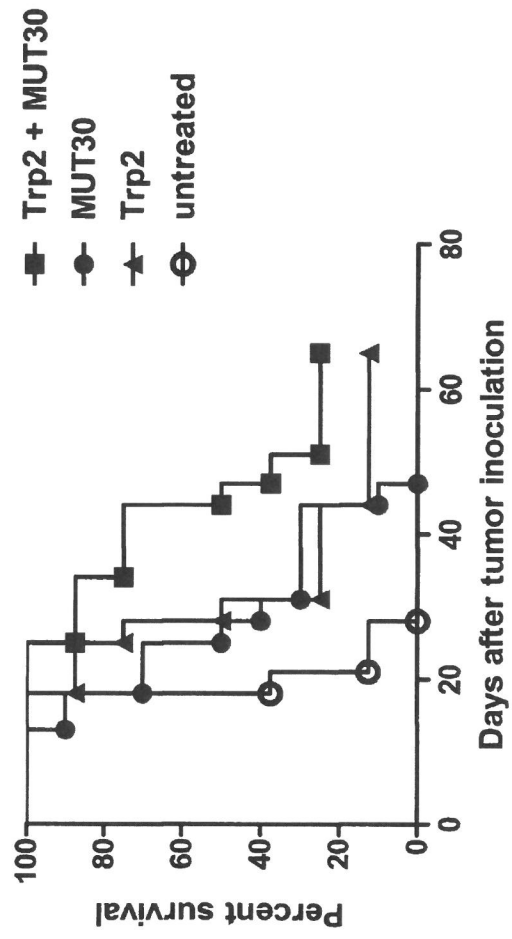
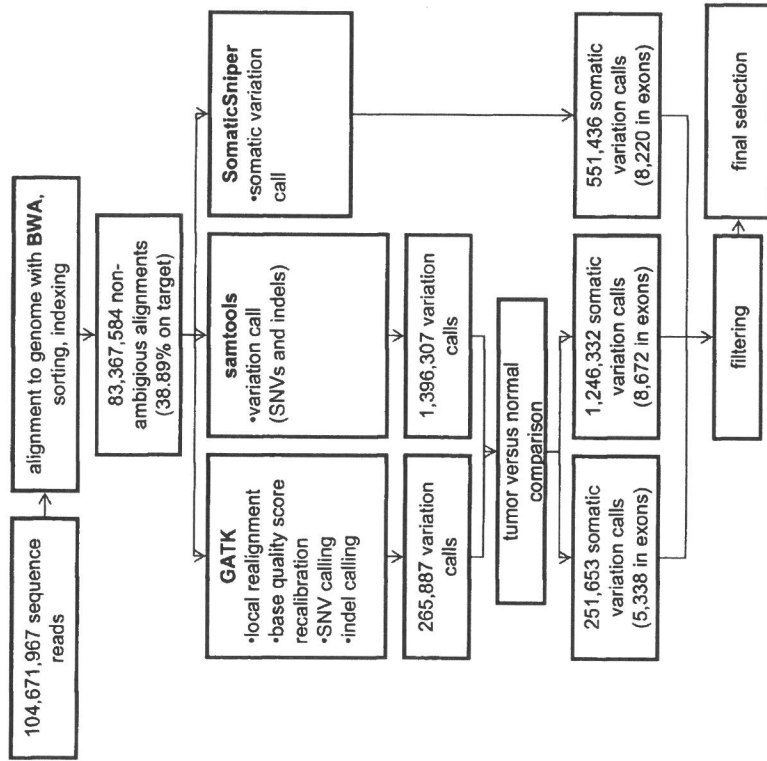


Fig. 15 B

Fig. 16



Marco Giovanni Mari
USBM - 02PL090

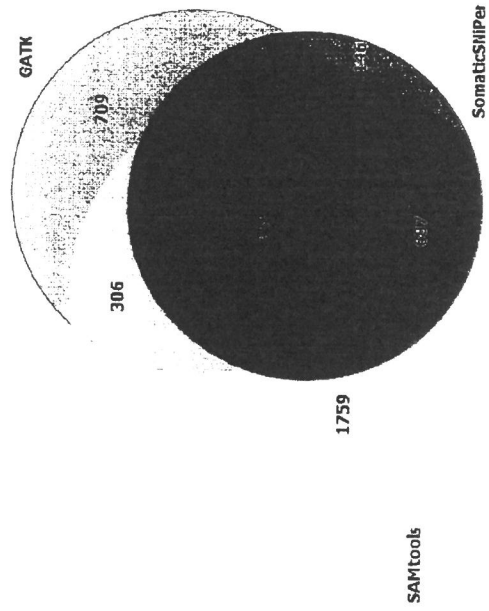


Fig. 17

Marco Giovanni Mari
USFBI - CPI - 090

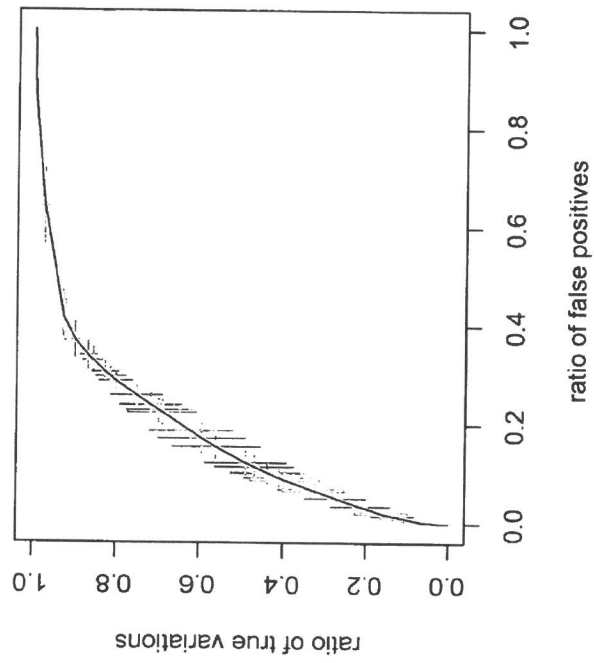
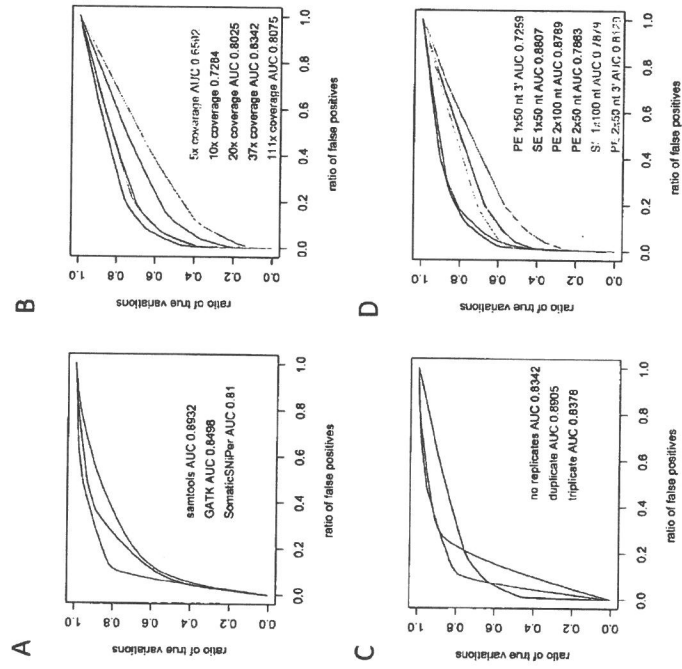


Fig. 18

B

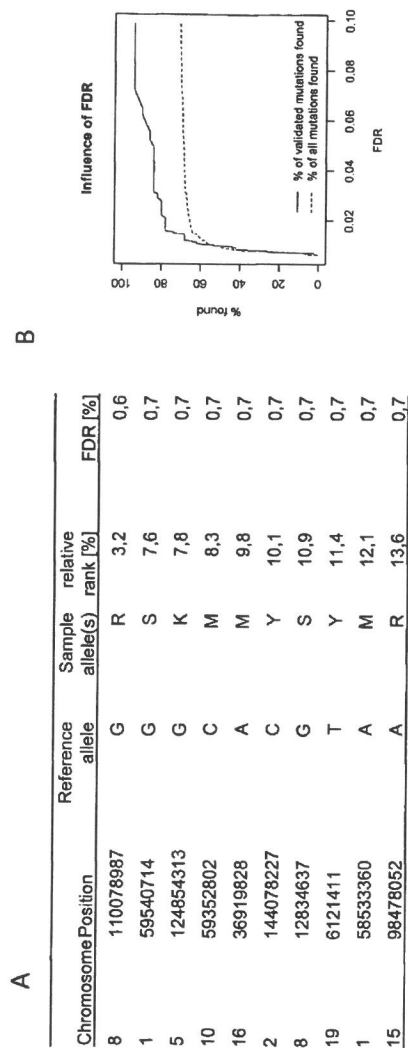
Marco Giovanni Mari
USBN - CPI-090

Fig. 19



Marco Giovanni Mari
USBM CRL-090

Fig. 20



Marco Giovanni Mari
USBM CPI-090

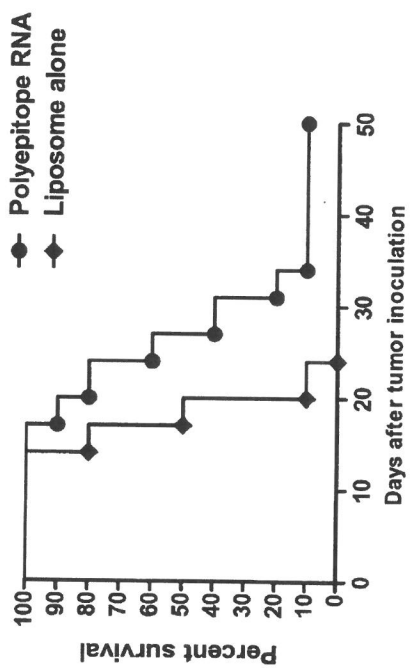


Fig. 21

Marco Giovanni Mari
USBM - GPI-090