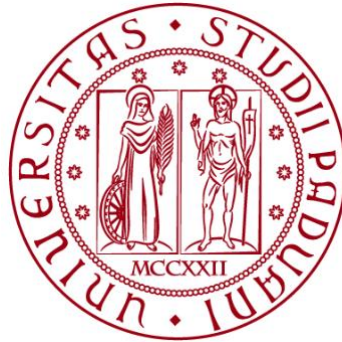


UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI BIOLOGIA

Corso di Laurea in Biologia Molecolare



ELABORATO DI LAUREA

**Rilevamento precoce delle metastasi – utilizzo delle
cellule MIC come marcatori dinamici**

**Tutor: Prof. Luigi Leanza
Dipartimento di Biologia**

**Laureanda: Elisa Picardi
Matricola: 2008062**

ANNO ACCADEMICO 2022/2023

INDICE

ABSTRACT	3
1. INTRODUZIONE	5
1.1 Il problema delle metastasi e il loro rilevamento	5
1.2 Approcci per la rilevazione delle metastasi	5
1.3 Segnali precoci di metastasi: le cellule MIC e i nano-sensori	7
1.4 Differenze tra cellule TIC (inizianti tumore) e MIC.....	7
2. MATERIALI e METODI	9
2.1.1 La spettroscopia Raman.....	9
2.1.2 L'analisi SERS.....	9
2.2.1 Fabbricazione dei nanosensori auto-funzionalizzati	10
2.2.1.1 L'ossido di grafene	10
2.2.1.2 I nanosensori	10
2.2.2 Caratterizzazione dei nanosensori auto-funzionalizzanti	11
2.3 Citofluorimetria a flusso (FACTS) e sorting	11
2.3.1 Colture cellulari.....	12
2.3.2 Cell sorting con flusso citometrico	12
2.3.3 Studio dell'uptake delle cellule	12
2.4 Analisi SERS delle cellule	12
2.5 Machine learning e modello PLS-R.....	13
2.6 Analisi statistiche	13
3. RISULTATI	15
3.1 Identificazione delle cellule inizianti metastasi (MIC).....	15
3.1.1 Investigazione delle MIC con SERS	15
3.2 Uptake dei nanosensori autofunzionalizzati in TIC e MIC	16
3.3 Caratterizzazione delle MIC attraverso la SERS	17
2.6.1 Determinazione dello stato intracellulare delle cellule MIC.....	19
2.6.2 Determinazione dello stato intracellulare delle MIC utilizzando la proliferazione cellulare	19
2.6.3 Differenziazione fenotipica cellulare per la rilevazione delle MIC	20
2.6.4 Analisi del microambiente tumorale per la differenziazione delle MIC	20
2.6.5 Applicazione dei nanosensori per differenziare le MIC in una popolazione eterogenea di sferoidi.....	20
2.7 Previsione delle metastasi attraverso il fenotipo cellulare e il machine learning 21	

2.8 Predizione della metastasi cancerosa da differenti tessuti di origine.....	22
4. DISCUSSIONE.....	23
BIBLIOGRAFIA.....	25
APPENDICE.....	26

ABSTRACT

La capacità del cancro di creare metastasi è un problema clinico urgente: i trattamenti correnti sono inefficaci per il cancro metastatico dato che le convenzionali tecniche di imaging non sono in grado di individuarlo nei primi stadi. La possibilità di prevedere in anticipo la capacità del cancro di metastatizzare sarebbe un valido mezzo per migliorare la prognosi del paziente: è necessario utilizzare l'eterogenicità fenotipica delle cellule tumorali come un marcatore diagnostico per le metastasi tumorali. In questa tesi è stata identificata una rara subpopolazione di cellule nel tumore che può segnalare l'inizio del processo di formazione delle metastasi, note come cellule inizianti metastasi (MIC): queste risultano essere un eccellente marcatore per il cancro metastatico indipendentemente dal tipo di tumore, dal tessuto di origine e dallo stadio della neoplasia. Al fine di fornire la sensibilità necessaria per la predizione del cancro metastatico, è stato usato un nano sensore auto-funzionalizzato, altamente attivo alla SERS, che permette il rilevamento delle cellule MIC utilizzando le proprietà dei processi biologici intracellulari e le caratteristiche del microambiente tumorale. Le cellule MIC possono quindi essere dei marcatori universali per la diagnosi, la previsione e la prognosi dei tumori metastatici.

1. INTRODUZIONE

1.1 Il problema delle metastasi e il loro rilevamento

Le metastasi derivate dal cancro sono causa al 90% delle morti dovute a tumore: il miglior modo per ridurre la mortalità dovuta alle metastasi è attraverso la diagnosi precoce di queste; tuttavia, la comprensione della loro formazione rimane ridotta ed è quindi fondamentale sviluppare la capacità di individuare e prevedere la formazione delle metastasi a livello precoce per migliorare la sopravvivenza del paziente, la gestione della malattia e ridurre la mortalità. In particolare, la plasticità Epiteliale-Mesenchimale (EMP) gioca un ruolo fondamentale nella formazione delle metastasi: l'EMP è definita come la capacità delle cellule di mantenere la propria plasticità tra i due stati di transizione EMT (epiteliale-Mesenchimale) e MET (Mesenchimale-Epiteliale). Queste transizioni cellulari permettono ad alcune cellule tumorali di migrare dal tumore primario per invadere un sito secondario e l'EMP è associata con la migrazione, l'invasione, la colonizzazione e la resistenza ai farmaci. La transizione EMT, in particolare, presenta diverse caratteristiche: diminuzione delle adesioni cellula-cellula, alterazioni del citoscheletro, delle giunzioni strette e degli emidesmosomi, perdita della polarità baso-apicale, promozione della migrazione delle cellule attraverso il rimodellamento della matrice cellulare e infine colonizzazione di siti secondari. I fattori di trascrizione che attivano l'EMT, influenzati da condizioni come l'ipossia, contribuiscono alla progressione del tumore inducendo la resistenza all'apoptosi e mantenendo le proprietà delle cellule staminali nei tumori. Queste prendono il nome di "Cancer stem cells" (CSCs) e sono un gruppo di cellule capaci di auto rinnovarsi, iniziare i tumori e andare incontro a differenziazioni eterogenee (Vi Ley Chin and Chooi Ling Lim). La possibilità di prevedere in modo precoce la formazione delle metastasi è impedita in modo critico dalla mancanza di tecniche che presentano una sensibilità a livello della singola cellula, dalla mancanza di marcatori distinguibili e dall'elevata eterogeneità dei tumori metastatici. Quindi da un lato è necessario individuare un marcatore imparziale che possa essere utilizzato per il rilevamento precoce delle metastasi, in secondo luogo bisogna sviluppare un metodo ultrasensibile con elevata specificità per eseguire una diagnosi precoce.

1.2 Approcci per la rilevazione delle metastasi

Il rilevamento delle metastasi è principalmente eseguito con due approcci: tecniche di imaging non invasive, come la tomografia computerizzata a raggi X (TC) e la tomografia a emissione di positroni; e marcatori molecolari, come i geni, i miRNA, gli lncRNA e la metilazione del DNA. In particolare, è stato osservato come gli lncRNA possono interagire con i miRNA e come queste interazioni giochino un ruolo cruciale nella formazione delle metastasi attraverso la regolazione di eventi cruciali, in particolare l'EMT.

Recentemente è stato riportato che l'EMT è regolato da fattori post trascrizionali, come miRNA e lncRNA che esercitano la loro influenza regolando gli effettori, i fattori di trascrizione e le vie di segnalazione. I membri della famiglia miR200 sono i miRNA correlati all'EMT più ampiamente studiati: la diminuzione delle loro espressioni sopprime l'E-caderina, tipica dell'identità epiteliale di una cellula, e inizia l'EMT attraverso la modificazione dell'attività di specifici fattori di trascrizione; al contrario le espressioni di questi miRNA nelle cellule mesenchimali inducono il MET aumentando l'E-caderina e diminuendo l'N-caderina, secondo un doppio loop di feedback negativo che controlla il processo EMT/MET attraverso l'espressione bilanciata della famiglia miR-200 e dei fattori di trascrizione di cui sopra nell'ambiente tumorale. I ruoli funzionali degli lncRNA nelle metastasi del cancro sono meno noti rispetto a quelli dei miRNA: gli lncRNA funzionali influenzano l'EMT nella metastasi del cancro attraverso la regolazione dell'espressione genica a vari livelli, compresa la modifica della cromatina, la trascrizione e l'elaborazione post trascrizione. Gli lncRNA svolgono le loro funzioni regolando la trascrizione attraverso un'ampia gamma di meccanismi e diversi lncRNA possono fungere da trascrizioni antisense formando dei duplex con i loro corrispondenti mRNA per indurre o inibire la loro traduzione e quindi influenzare l'EMT. In alcuni casi la stabilità del lncRNA verrà ridotta a causa dell'interazione con specifici miRNA; in altri casi gli lncRNA, noti anche come RNA endogeni concorrenti (ceRNA), possono trattenere i miRNA lontano dagli mRNA bersaglio agendo da antagonisti. Inoltre, alcuni lncRNA possono produrre dei miRNA causando la repressione degli mRNA bersaglio. Tuttavia, ulteriori indagini sono necessarie per chiarire la relazione tra lncRNA e miRNA che possono interagire in modi diversi a seconda dei tumori considerati. (Ming-xin Cao et al.).

Per quanto riguarda le tecniche di imaging non invasive, anche queste non sono adatte per il rilevamento delle metastasi precoci: data l'elevata eterogeneità tra le cellule cancerogene e la dimensione ridotta delle metastasi, è difficile identificare queste popolazioni cellulari attraverso le tecniche convenzionali. Ricerche più estese hanno identificato numerosi marcatori epigenetici associati con le metastasi derivanti dai tumori, si tratta però di marcatori altamente dipendenti dal tipo di tumore che non forniscono un modello universale per la diagnosi precoce delle metastasi. Convenzionalmente, la capacità di un tumore primario di formare delle metastasi è predetta sulla base dell'assunzione che il tumore metastatico sia come il tumore primario dal punto di vista molecolare, tuttavia, questa assunzione non tiene conto dell'eterogeneità dei tumori e ostacola la ripetibilità dei modelli di predizione. Questi fattori rendono i marcatori epigenetici di cui sopra dei marcatori instabili per la previsione delle metastasi. Sarebbe quindi bisogno di un marcatore universale per le metastasi cancerogene che può essere utilizzato per la diagnosi precoce.

1.3 Segnali precoci di metastasi: le cellule MIC e i nano-sensori

È stata identificata una rara popolazione di cellule all'interno dei tumori che presentano segnali precoci di metastasi: si tratta delle cellule inizianti metastasi (MIC) che costituiscono una sub-popolazione di cellule unica, distinta e non specifica in base al cancro e potrebbero essere utilizzate per rilevare le metastasi quando si trovano nello stadio di singola cellula.

La natura dinamica delle metastasi giustifica l'esistenza di un marcatore come le MIC che rappresenta gli eventi pro-metastatici comuni a diversi tumori solidi: queste cellule hanno il potenziale di essere marcatori universali e imparziali sia per il rilevamento sia per la previsione della formazione di metastasi. La natura non rilevabile delle MIC necessita di un metodo ultrasensibile e non distruttivo che possa garantire una rappresentazione completa di queste cellule: i metodi molecolari non invasivi possono fornire una nuova prospettiva per la previsione e la diagnosi della formazione di metastasi, come nel caso della tecnica SERS. Dunque, per studiare questa specifica popolazione di cellule tra le cellule eterogenee che costituiscono il microambiente tumorale, c'è bisogno di un sensore che possa entrare in modo efficace in questo ambiente e che possa essere internalizzato nella cellula di interesse e rimanere integro nell'ambiente stesso. L'ossido di grafene (dimensioni <5nm) presenta un'elevata biocompatibilità e rimane non reattivo nel microambiente tumorale; inoltre presenta un'alta efficienza di internalizzazione il che lo rende un ideale candidato per essere applicato per il rilevamento precoce e la predizione del cancro metastatico, utilizzando come sonda le MIC. In questo studio sono stati affrontati il rilevamento precoce e la predizione della formazione di metastasi su due fronti: da un lato è stata impiegata per la prima volta una rara popolazione di cellule all'interno di un tumore, conosciute come cellule inizianti metastasi (MIC), come marcatore imparziale per la diagnosi precoce; in secondo piano, questo studio è il primo a creare un nano-sensore auto-funzionalizzato altamente sensibile per rilevare le MIC. Le diagnosi accurate e le predizioni delle metastasi sono state ottenute dopo aver efficientemente internalizzato il nano-sensore nelle MIC per amplificare i segnali che portano alla diagnosi. Il nano-sensore è stato sintetizzato attraverso la pulsazione di un laser "ultrashort" che interagisce con un substrato di carbonio in un ambiente ricco di ossigeno. La dimensione del nano-sensore auto-funzionalizzato (circa 2nm), insieme alla presenza di numerosi gruppi funzionali dell'ossigeno sulla superficie della sonda, hanno permesso un'efficace internalizzazione dello stesso sensore da parte delle cellule.

1.4 Differenze tra cellule TIC (inizianti tumore) e MIC

I modelli preclinici di cellule del cancro al pancreas e del cancro ai polmoni sono stati utilizzati per accertare la natura non-cancro specifica delle cellule MIC e la possibilità di utilizzarle come marcatori universali per la diagnosi delle metastasi e per la previsione della loro formazione.

L'analisi SERS è stata eseguita in una popolazione di cellule selezionata per investigare le differenze significative che caratterizzano le TIC e le MIC. Le caratteristiche predominanti che possono aiutare nella diagnosi sono state trovate nello stato proliferativo, nella plasticità fenotipica delle cellule, nel DNA intracellulare delle MIC e delle TIC. Il microambiente tumorale è stato analizzato tramite il secretoma tumorale che comprende i metaboliti, i nutrienti e le molecole responsabili della comunicazione tra cellule. I risultati hanno dimostrato che il nano-sensore auto-funzionalizzato permetteva di distinguere le cellule TIC dalle cellule MIC in una popolazione eterogenea senza ambiguità. I dati ottenuti dalla popolazione selezionata sono stati utilizzati per il training di un modello per la previsione della formazione delle metastasi utilizzando l'algoritmo "Support Vector Machine": il modello presenta un'accuratezza nella revisione del 99.62% per una popolazione eterogenea di TIC e MIC di cellule derivanti da tessuti diversi.

2. MATERIALI e METODI

2.1.1 La spettroscopia Raman

La spettroscopia Raman è utilizzata per studiare le interazioni della radiazione con la materia impiegando la diffusione inelastica della luce da parte della materia. Quando un fotone di luce interagisce con una molecola, può indurre una transizione di breve durata a uno stato di energia virtuale o superiore. Se la molecola rilascia una quantità di energia diversa da quella del fotone incidente, si parla di diffusione di Raman: a causa del moto nucleare, l'energia può essere trasferita al fotone incidente o dal fotone incidente, attraverso la diffusione Stokes o anti-Stokes. Nella diffusione Stokes, l'energia viene trasferita alla molecola attraverso il moto nucleare all'interno della molecola stessa: si ha un fotone diffuso con energia e frequenza ridotte; al contrario, nella diffusione anti Stokes, l'energia viene trasferita al fotone perché le molecole si trovano inizialmente in uno stato di energia superiore. Nella ricerca, la diffusione Stokes è la più osservata e il segnale è intrinsecamente debole. Nella diffusione Raman si usa un laser per fornire luce coerente e monocromatica per eccitare le molecole; successivamente uno spettrometro difatti o viene usato per disperdere la luce su rilevatori altamente sensibili per generare uno spettro. Uno spettro Raman è un grafico dell'intensità della luce diffusa in numeri d'onda, in rapporto all'eccitazione laser incidente e ogni molecola, in base alla vibrazione dei legami molecolari specifici, ha una caratteristica impronta spettroscopica. L'intensità dello spettro Raman è correlata alla concentrazione delle molecole e l'interpretazione degli spettri può essere applicata a una serie di campioni per conoscerne la composizione chimica. Nelle analisi biochimiche, la spettroscopia di Raman è particolarmente efficace quando viene combinata con la microscopia ottica per produrre un'immagine composta che combina informazioni molecolari e morfologiche. Queste tecniche di microspettroscopia possono fornire informazioni sulle proprietà di campioni eterogenei e potrebbero essere usate come alternative non invasive e non distruttive all'istologia, oltre a permettere di discriminare tra tipi cellulari.

2.1.2 L'analisi SERS

La SERS (surface-enhanced Raman scattering) è una tecnica di rilevamento che consente di amplificare i segnali Raman di diversi ordini di grandezza. Questa amplificazione avviene principalmente attraverso l'interazione elettromagnetica della luce con i metalli, che produce notevoli amplificazioni del campo laser attraverso delle eccitazioni generalmente note come risonanze plasmoniche. Per beneficiare di queste amplificazioni, le molecole devono essere adsorbite sulla superficie metallica, o comunque devono essere molto

vicine ad essa. L'amplificazione del segnale è fornita dalle risonanze nel substrato metallico: si tratta di una serie di effetti associati all'interazione della radiazione elettromagnetica con i metalli che compaiono nell'effetto SERS sotto forma di nanostrutture metalliche.

2.2.1 Fabbricazione dei nanosensori auto-funzionalizzati

2.2.1.1 L'ossido di grafene

L'ossido di grafene potrebbe costituire la base di una strategia per impedire potenzialmente le metastasi, agendo come adsorbente per enzimi pro-metastatici, attraverso i suoi residui cationici e idrofilici. Sono stati condotti degli esperimenti per valutare gli effetti tossici dell'ossido di grafene (GO) sulle linee cellulari di cancro al polmone (A549 e SKMES-1) a diverse concentrazioni di GO: è stata osservata una leggera, ma significativa riduzione della vitalità cellulare in entrambe le linee cellulari a concentrazioni di 250 e 500 $\mu\text{g/mL}$ di GO dopo 24 ore di esposizione. Il GO ha dimostrato quindi di ridurre la vitalità cellulare e di indurre principalmente l'apoptosi a concentrazioni elevate, suggerendo il suo potenziale come agente tumorale (Tabish et al., 2019).

2.2.1.2 I nanosensori

I nanosensori auto-funzionalizzati sono stati sintetizzati utilizzando un laser UPL attraverso il processo di ionizzazione in presenza di ossigeno gassoso: questo processo produce nanosensori della morfologia e della taglia desiderate, inoltre i nanosensori presentano le proprietà chimiche desiderate grazie all'impatto delle pulsazioni laser sulla superficie di grafite. Il trasferimento di energia, dagli elettroni nella banda di valenza agli elettroni nella banda di conduzione, causa una decomposizione della grafite in specie del carbonio ioniche e neutre, sotto forma di nube in espansione. L'introduzione di ossigeno gassoso in questa nube porta ad avere un'espansione più lenta della nube stessa.

Inoltre, la temperatura della nube aumenta a causa dell'effetto schermante dell'ossigeno: in questo modo si aumenta il potenziale di ionizzazione della nube stessa, con la conseguente formazione di una particella di dimensioni minore.

La morfologia e la distribuzione della dimensione dei nano sensori è stata determinata utilizzando la microscopia a trasmissione elettronica: si osserva come la distribuzione delle dimensioni di campione si trova nella curva normale della distribuzione, con una dimensione media di 2.09 nm e una deviazione standard di 0.7 nm.

Si utilizza la spettroscopia fotoelettronica a raggi X (XPS), riportata in **figura 1** per studiare la composizione chimica superficiale dei nanosensori: si osserva come questi siano composti prevalentemente da carbonio (79.83%) e ossigeno (18.45%).

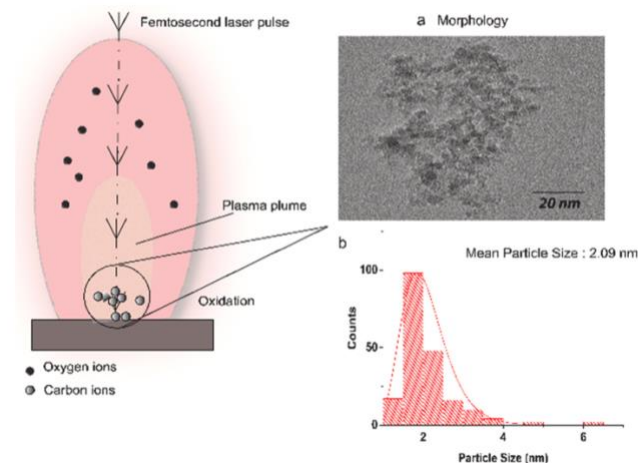


Figura 1. In figura a viene mostrata la morfologia dei nanosensori attraverso la TEM; l'istogramma riporta la distribuzione della dimensione dei nanosensori autofunzionalizzati.

2.2.2 Caratterizzazione dei nanosensori auto-funzionalizzanti

La caratterizzazione ottica dello spettro Raman è stata ottenuta utilizzando lo spettrometro confocale Renishaw Invia equipaggiato con un microscopio ad epifluorescenza DMI6000. La lunghezza d'onda di 785 nm è stata utilizzata per ottenere lo spettro Raman al fine di caratterizzare i nanosensori auto-funzionalizzati.

2.3 Citofluorimetria a flusso (FACTS) e sorting

Si tratta di una tecnica utilizzata per distinguere le cellule TIC dalle cellule MIC negli sferoidi tumorali: è una tecnica multi-parametrica condotta su cellule o particelle microscopiche in sospensione in un fluido di trasporto. Tra i diversi parametri si trovano il volume cellulare, la complessità morfologica, il contenuto dei pigmenti, degli acidi nucleici, delle proteine e degli antigeni di superficie e intracellulari.

Si considerano anche le attività enzimatiche, il pH, flussi di calcio, potenziale di membrana e produzione di ROS. Si tratta di una tecnica che dura pochi secondi e fornisce molti dati su migliaia di cellule al secondo. Attraverso una focalizzazione idrodinamica, sulla base della differenza di pressione tra il campione e il fluido di trasporto, le cellule vengono messe in fila: la luce laser incidente viene riemessa, in parte viene assorbita e in parte viene trasmessa: nello specifico, la radiazione emessa può essere una radiazione diffusa o portare a light scattering. Acquisendo il light scattering frontale a 0° (forward scatter) e laterale a 90° (side scatter) rispetto alla direzione del fascio incidente, si ottengono informazioni sulle dimensioni, la forma e l'omogeneità della popolazione. Grazie alla presenza dei Sorter è possibile recuperare analizzare

i campioni per poi separarli sulla base di una delle caratteristiche (morfologiche o fluorescenza).

2.3.1 Colture cellulari

Le cellule del cancro pancreatico (ASPC-1) e le cellule del cancro ai polmoni (H69-AR), ottenute dall'American Type Culture Collection (ATCC), sono state coltivate separatamente utilizzando il terreno RPMI, contenente 10% di siero fetale bovino (FBS) e una combinazione di penicillina e streptomina a 37° in un'atmosfera con il 5% di CO₂. Una volta raggiunta la confluenza, le cellule sono state trasferite in un nuovo contenitore con un rapporto di 1:3; successivamente sono state raccolte utilizzando una soluzione di Tripsina-EDTA e centrifugate a 1000 rpm per 5 minuti, in modo da raccogliere in un pellet. Il mezzo di coltura MammoCult medium è stato utilizzato per coltivare le cellule staminali tumorali in modo che si aggregassero a formare degli sferoidi (CSC), utili per lo studio delle cellule stesse e del loro comportamento.

2.3.2 Cell sorting con flusso citometrico

La marcatura dei marcatori di superficie cellulare (CD44, CD133 e citocherarina, ottenuti da Miltenyi Biotec) è stata eseguita nel terreno DMEM con il 5% di FBS e le cellule marcate sono state mantenute su ghiaccio per 15 minuti. Successivamente sono state lavate per rimuovere i marcatori di specie cellulare e sono state sospese in un tampone per la separazione cellulare citofluorimetrica. Le cellule morte sono state escluse con la colorazione DAPI, consentendo l'utilizzo di singole cellule vive ed efficienti per ulteriori analisi.

2.3.3 Studio dell'uptake delle cellule

Gli sferoidi CSC sono stati incubati con i nanosensori in una piastra Petri con 3 mL di terreno di coltura per 24 ore; le cellule selezionate con la citometria a basso flusso sono state incubate separatamente con nanosensori autofunzionalizzati, sottoposti a loro volta a citometria per caratterizzare la percentuale di assorbimento dei nanosensori stessi, basata sulla fluorescenza dei campioni.

2.4 Analisi SERS delle cellule

Le cellule sono state incubate con i nanosensori auto-funzionalizzati per 24 ore e il terreno utilizzato è stato aspirato. Successivamente le cellule sono state lavate tre volte con PBS all'1% per rimuovere ogni traccia di terreno rimasto, e incubate in 4 mL di soluzione di paraformaldeide al 4% per 20 minuti a temperatura ambiente. Dopo l'incubazione, la soluzione di paraformaldeide è stata rimossa e le cellule sono state lavate con PBS e poi fatte seccare all'aria prima dell'analisi SERS. Gli spettri Raman delle cellule sono stati raccolti utilizzando un microscopio confocale Renishaw con una lunghezza d'onda mantenuta a 785 nm. Il tempo di acquisizione è stato mantenuto costante a 10 secondi e per ogni spettro sono state fatte tre acquisizioni in modo da

assicurare la riproducibilità e l'uniformità dei segnali. La mappatura Raman a livello di singola cellula è stata eseguita su cellule appositamente fissate utilizzando il microscopio con ingrandimento 20X: la mappa ottenuta per una singola cellula aveva un passo di 0.1×0.1 .

2.5 Machine learning e modello PLS-R

Il modello PLS-R è una tecnica di regressione multivariata usata nell'analisi dei dati e nella modellazione statistica: è particolarmente utile quando si sono molte variabili indipendenti tra loro che però presentano una co-linearità. Il PLS-R funziona cercando di stabilire una relazione tra le variabili indipendenti e la variabile dipendente, attraverso una combinazione lineare delle variabili indipendenti che prende il nome di componente latente. Questa viene determinata in modo da essere altamente correlata alla variabile dipendente, riducendo al minimo la perdita di informazioni. Il PLS-R è quindi uno strumento fondamentale nell'analisi dei dati multivariati e nella costruzione di modelli di regressione che consentono di gestire efficacemente dati complessi e correlati.

2.6 Analisi statistiche

Tutti gli esperimenti sono stati condotti in triplicato e i dati degli spettri SERS, elaborati da uno specifico software, sono stati impiegati per determinare le differenze tra le cellule TIC e le cellule MIC; da questi dati è stato ottenuto il dataset di addestramento dell'algoritmo, che tiene conto degli spettri SERS delle cellule ordinate tramite FACS. In particolare, sono stati utilizzati gli spettri da 20 cellule per condurre l'analisi di apprendimento automatico con il modello Support Vector Machine per testare l'accuratezza della rilevazione. Gli spettri della popolazione eterogenea non ordinata sono stati utilizzati come dati di test e la dimensione del campione è stata mantenuta costante.

3. RISULTATI

3.1 Identificazione delle cellule inizianti metastasi (MIC)

Un esempio di plasticità cellulare nella progressione del tumore è il processo di plasticità Epiteliale-Mesenchimale (EMP), che comprende sia la transizione Epiteliale-Mesenchimale (EMT), sia il processo inverso di transizione Mesenchimale-Epiteliale (MET), primo passo nella cascata di metastasi-invasione. Il meccanismo principale di EMT-MET avviene attraverso la perdita funzionale e il cambiamento di proteine legate alle funzioni cellulari, come l'E-caderina, la Vimentina e la citocheratina mediante diversi meccanismi genetici, come mutazioni, ipermetilazione e repressione dei fattori di trascrizione. La principale differenza tra cellule TIC e MIC è l'espressione di marcatori epiteliali, come l'E-caderina, la citocheratina e l'EpCAM: per lo studio della sottopopolazione MIC viene utilizzata la co-espressione di diversi marcatori di superficie. In particolare, si osserva come la CD44 sia essenziale per la crescita invasiva, mentre CD133 è coinvolta nel mantenimento delle proprietà di auto-rinnovamento delle cellule che iniziano il tumore; inoltre, tra i marcatori epiteliali associati all'inizio del tumore, l'espressione della citocheratina è associata a una prognosi sfavorevole in diversi tumori solidi, dal momento che la sua sovra-regolazione è correlata a cambiamenti morfologici e fenotipici nelle lesioni maligne. Si tratta dell'unico marcatore epiteliale che viene acquisito dalle cellule che hanno subito una differenziazione epiteliale.

3.1.1 Investigazione delle MIC con SERS

Per investigare le cellule MIC con SERS, prima di tutto è necessario isolarle: il potenziale metodo per identificare le cellule TIC con capacità tumorigeniche, e quindi distinguerle dalle MIC, consiste nell'utilizzare cellule CD44+, CD133+ e citocheratina+ (CK+), corrispondenti rispettivamente a un fenotipo simile alle cellule staminali (CD44 e CD133), e alle cellule che hanno subito la transizione dal fenotipo Mesenchimale a quello epiteliale (indicate dalla citocheratina). Le cellule sferoidi TIC, separate mediante citofluorimetria, comprendono una popolazione eterogenea di sferoidi pancreatici TIC e sferoidi TIC polmonari: questi ultimi comprendono il 31,12% di cellule CD44+. Queste cellule sono state ulteriormente utilizzate per ottenere le cellule doppio-positivo (CD44+/ CD133+), costituendo il 15,54% della popolazione e denominate come cellule che iniziano il tumore polmonare (L-TIC). Nello specifico, la popolazione di cellule staminali del cancro con capacità tumorigeniche (L-MIC) costituisce solo l'1,53% della popolazione TIC totale e comprende cellule CD44+, CD133+ e CK+. Si può notare che gli sferoidi TIC pancreatici comprendono il 26,79% di cellule CD44+ e anche queste sono state utilizzate per ottenere cellule doppio positive, costituenti il 29,76% della popolazione, e denominate cellule che iniziano il tumore pancreatico (P-TIC). La popolazione di cellule staminali del cancro pancreatico con capacità tumorigeniche (L-MIC), comprendente cellule CD44+, CD133+ e CK+,

costituisce solo lo 0,82% della popolazione TIC totale. Di conseguenza, il basso rapporto di TIP con capacità tumorigeniche nella popolazione cellulare rende estremamente difficile rilevare la presenza di queste cellule. Le cellule sono state separate utilizzando FACS e utilizzate per ulteriori analisi per caratterizzare le differenze tra le MIC e le TIC, come viene illustrato in **figura 2**.

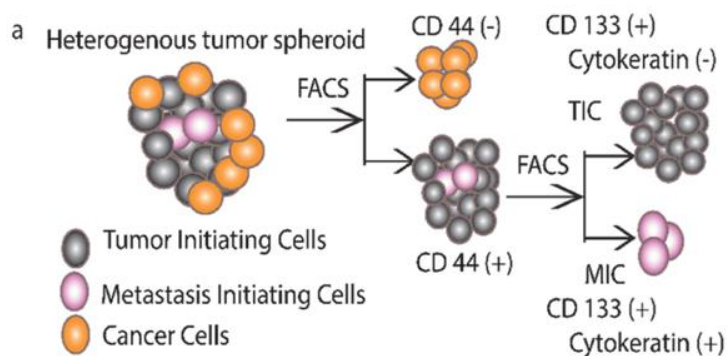


Figura 2. In figura viene riportato il meccanismo di identificazione delle cellule MIC negli sferoidi: una prima suddivisione permette di distinguere tra le cellule cancerose vere e proprie, prive di CD 44, e le cellule iniziatrici, sia del tumore sia delle metastasi, dotate di CD 44. Queste ultime, proprio attraverso FACS, vengono ulteriormente suddivise in TIC e MIC: entrambe presentano CD 133, mentre la citocheratina è presente esclusivamente a livello delle MIC.

3.2 Uptake dei nanosensori autofunzionalizzati in TIC e MIC

L'analisi SERS delle cellule richiede l'uptake dei nanosensori autofunzionalizzati sia dalle cellule TIC sia dalle MIC: successivamente viene condotta un'analisi di citofluorimetria a flusso per investigare l'internalizzazione dei nanosensori stessi. La FACS, utilizzata per verificare l'uptake della sonda, è una tecnica altamente sensibile per quantificare l'uptake delle particelle di nanosensori con una proprietà intrinseca data dalla fluorescenza: il carico dei nanosensori per cellula è caratteristicamente espresso come il numero di eventi in cui la fluorescenza associata con questi nanosensori corrisponde all'emissione caratteristica a 525 nm (BV-510). La fluorescenza distinta dei sensori autofunzionalizzati permette il rilevamento e la quantificazione dell'uptake della sonda con una sensibilità di singola cellula: in particolare l'internalizzazione del sensore è rilevabile con il side scattering, che è proporzionale alla granularità e alla complessità della cellula. Le cellule TIC con un nanosensore internalizzato presentano un'elevata intensità di BV-510, sia che derivino dalle cellule tumorali pancreatiche sia che derivino da quelle del polmone. Inoltre, si può osservare che le cellule MIC mostrano un'augmentata intensità di fluorescenza rispetto alle cellule TIC e alle cellule di controllo, come

riportato in **figura 3**: questo potrebbe essere attribuito a un aumentato uptake dei nanosensori da parte delle cellule MIC.

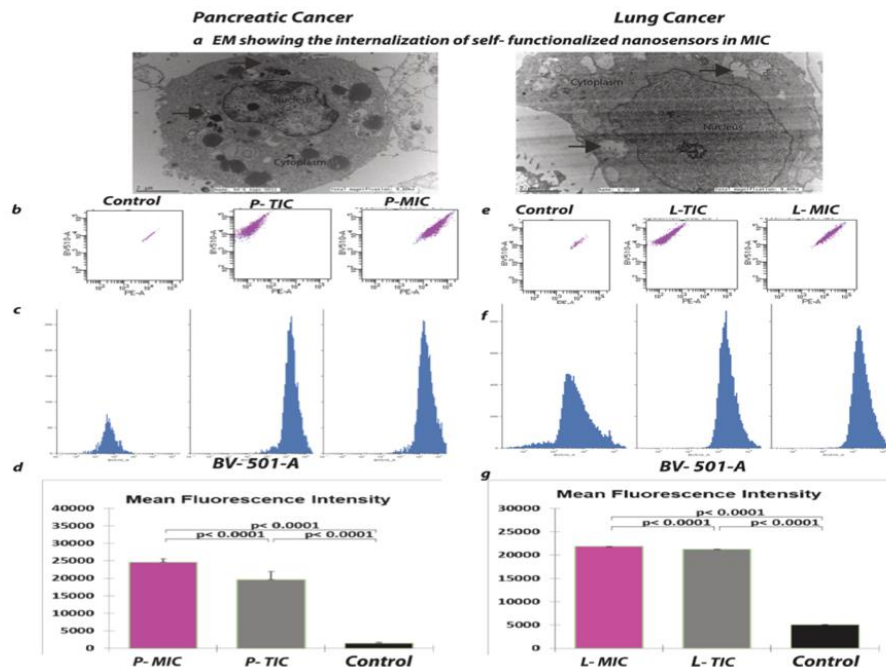


Figura 3. In figura a) l'immagine con microscopia elettronica permette di mostrare la presenza di nanosensori nella regione citoplasmatica delle MIC. In figura b) e in figura e) si osservano dei dot plot dell'uptake dei nanosensori auto-funzionalizzati, rispettivamente nelle cellule del cancro al pancreas e delle cellule del cancro ai polmoni. I grafici c) e d) confrontano l'uptake di nanosensori nelle MIC, nelle TIC e nei controlli delle cellule del pancreas; i grafici f) e g) confrontano l'uptake dei nanosensori nelle cellule dei polmoni.

3.3 Caratterizzazione delle MIC attraverso la SERS

Per stimare l'attività SERS intracellulare dei nanosensori, sono state analizzate le differenze dello spettro SERS tra le cellule MIC con nanosensori, le cellule TIC con nanosensori e le cellule prive di nanosensori, utilizzate come controllo. Si osserva come le MIC mostrano un notevole aumento del segnale rispetto alle cellule TIC; inoltre, le cellule di controllo non mostrano dei segnali particolarmente evidenti che vadano a indicare la presenza di molecole intracellulari. L'elevata attività SERS a livello intracellulare potrebbe essere attribuita a molteplici fattori, compresa la dimensione ridotta dei nanosensori, la presenza di gruppi funzionali sulla superficie dei sensori e la presenza di difetti che amplificano ulteriormente l'attività SERS. Le dimensioni inferiori a 5 nm dell'ossido di grafene permettono di fornire un alto livello di amplificazione del segnale delle molecole di DNA e la presenza di gruppi funzionali sulla superficie dei nanosensori porta all'interazione di questi con

proteine, lipidi e molecole di RNA intracellulari. Dopo l'assorbimento dei nanosensori auto-funzionalizzanti da parte di TIC e MIC, sono stati ottenuti gli spettri SERS delle cellule, riportati in **figura 4**, e queste sono state successivamente ordinate attraverso FACS e trattate con i nanosensori per 24 ore e poi fissate con la formaldeide.

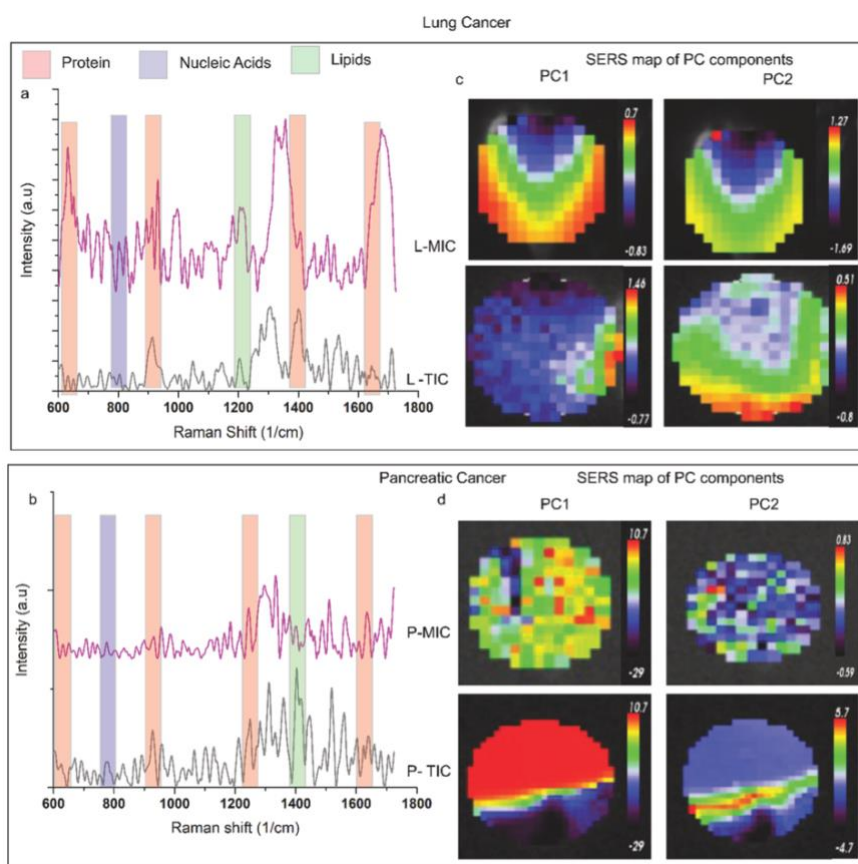


Figura 4. Caratterizzazione spettrale SERS per differenziare tra le cellule TIC e MIC. In figura a) si osservano le cellule del polmone; in figura b) quelle del pancreas. Dai grafici si nota come, sia per le cellule MIC polmonari, sia per quelle del pancreas, si osservano dei picchi con intensità maggiormente elevate per le macromolecole rispetto alle corrispondenti cellule TIC: questo permette di distinguere al meglio le popolazioni cellulari, i cui spettri Raman vengono ulteriormente amplificati grazie all'utilizzo della SERS.

Confrontando gli spettri SERS, è possibile osservare come l'intensità SERS di L-MIC sia maggiore rispetto a quella di L-TIC; una tendenza simile è stata osservata anche confrontando le TIC e le MIC del pancreas. Inoltre, gli spettri SERS mostrano anche i picchi Raman caratteristici associati a diverse biomolecole intracellulari. Le specifiche differenze molecolari tra L-TIC e L-MIC e tra P-TIC e P-MIC sono state ottenute attraverso dei PC di caricamento, analisi statistiche multivariate sugli spettri SERS delle cellule: le principali

differenze di picco sono attribuibili alle variazioni riscontrate a livello delle proteine, dell'RNA e dei lipidi che rispecchiano le differenze macromolecolari tra i due tipi cellulari. Queste variazioni nelle componenti fondamentali dei processi cellulari indicano che TIC e MIC sono popolazioni molto diverse che mostrano un alto grado sia di eterogeneità funzionale che di eterogeneità fenotipica e possono essere utilizzate come marker per investigare le MIC all'interno del microambiente tumorale.

2.6.1 Determinazione dello stato intracellulare delle cellule MIC

Le macromolecole cellulari, in particolare il DNA e le proteine, giocano un ruolo significativo nel determinare il destino cellulare: è possibile quindi utilizzare le differenze nel DNA intracellulare come marcatore per rilevare le cellule TIC in una certa popolazione: in particolare si osserva come le cellule TIC presentano un più elevato contenuto di DNA che permette di distinguerle dalle MIC attraverso il confronto degli spettri SERS. È noto, inoltre, che le cellule TIC sono intrinsecamente resistenti ai danni al DNA dal momento che sono in grado di accelerare i meccanismi di riparazione del DNA stesso alterando il ciclo cellulare che risulta caratterizzato da un maggior numero di *checkpoints*. Tuttavia, gli spettri SERS non mostrano alcuna variazione nei picchi associata con gli eventuali danni alla struttura del DNA, ma dall'analisi raziometrica dei valori di picchi ottenuti, si può dedurre che il DNA intracellulare in TIC subisce dei danni a causa di rotture nel doppio filamento. Studi recenti mostrano un aumento di espressione di c-Myc e dell'iniziazione di EMT in TIC per attivare il meccanismo di riparazione del DNA; quindi, si può suggerire che le TIC sottoposte a questo meccanismo acquisiscano un potenziale migratorio aumentato portando alla colonizzazione di siti remoti. Dopo questo processo, le TIC subiscono la MET e stabiliscono un tumore metastatico: questo avviene con un'attivazione dinamica di geni e proteine che contribuiscono alla riparazione del DNA.

2.6.2 Determinazione dello stato intracellulare delle MIC utilizzando la proliferazione cellulare

Il processo di de-differenziazione e i cambiamenti metabolici associati alle trasformazioni cellulari possono essere utilizzati come marker per distinguere le MIC dalle TIC. Il fenotipo mesenchimale è associato alla proprietà migratoria, mentre il fenotipo epiteliale è associato all'invasione e alla progressione tumorale. Tra le TIC è presente una piccola popolazione che subisce la transizione da mesenchimale a epiteliale, attraversando una serie di cambiamenti biochimici intracellulari che portano a una maggiore proliferazione cellulare. I cambiamenti biochimici intracellulari forniscono informazioni accurate sullo stato di proliferazione delle cellule: in particolare le cellule quiescenti hanno un elevato contenuto lipidico e basse quantità di proteine. In generale è possibile osservare che il rapporto lipidi-proteina è significativamente più alto nelle TIC tumorigeniche. Anche il rapporto tra lipidi e RNA funge da marker per determinare lo stato di proliferazione della cellula: il cancro è generalmente associato a una crescita incontrollata delle cellule, un processo strettamente legato alla sintesi

proteica potenziata. In particolare, si osserva una connessione tra le alterazioni dei componenti del ribosoma e la perdita o l'acquisizione delle proprietà TIC: livello delle cellule MIC il rapporto lipidi-RNA è significativamente più basso e costituisce quindi un marker per la rilevazione di questa sottopopolazione di cellule.

2.6.3 Differenziazione fenotipica cellulare per la rilevazione delle MIC

Le cellule TIC e le cellule MIC possono essere distinte in base al loro stato di differenziazione cellulare: le TIC sono principalmente mesenchimali, mentre le MIC subiscono un'inversione della transizione EMT, diventando epiteliali. La spettroscopia Raman viene utilizzata per distinguere le TIC dalle MIC, concentrandosi sulle differenze nelle proteine e nei lipidi. Inoltre, l'analisi delle immagini SERS rivela che le MIC mostrano una maggiore presenza di acidi nucleici, a indicare una loro capacità di iniziazione di tumori secondari attraverso le metastasi. Infine, le MIC presentano un maggior contenuto di acidi grassi e un minore contenuto di triptofano: entrambi i cambiamenti biochimici consentono di differenziare efficacemente le MIC dalle TIC.

2.6.4 Analisi del microambiente tumorale per la differenziazione delle MIC

L'analisi del microambiente tumorale è un ulteriore mezzo che permette la distinzione tra le cellule MIC e le TIC: le differenze principali tra queste due cellule sono legate al tasso di proliferazione cellulare e ai meccanismi di riparazione del DNA e delle cellule. Dal momento che le TIC sono presenti in quantità minima nei campioni tumorali, la loro caratterizzazione biochimica è particolarmente difficile. Inoltre, il microambiente tumorale svolge un ruolo critico nel mantenimento della plasticità delle TIC: il secretoma del tumore, comprendente proteine e fattori di crescita, viene analizzato utilizzando un mezzo di coltura condizionato. Quello condizionato da TIC presenta una maggiore intensità complessiva rispetto a quello condizionato da MIC, proprio a causa della minore proliferazione delle TIC stesse. Dal punto di vista metabolico, si evidenzia come le TIC consumino l'amminoacido tirosina a un tasso più elevato, indicando uno switch metabolico specifico per la sopravvivenza delle TIC stesse in stato di quiescenza. Queste differenze metaboliche possono essere utilizzate per distinguere i due tipi cellulari.

2.6.5 Applicazione dei nanosensori per differenziare le MIC in una popolazione eterogenea di sferoidi

Una volta che i nanosensori sono internalizzati nella cellula, la posizione della sonda determina la sensibilità e la riproducibilità delle rilevazioni SERS. Gli spettri che si ottengono permettono di confrontare l'intensità della SERS da diverse cellule: si può notare una variazione di intensità con una bassa percentuale di deviazione standard relativa, in particolare inferiore al 3% sia per le MIC sia per le TIC.

Applicando i marcatori intracellulari che distinguono le due tipologie cellulari, è stata ottenuta l'immagine SERS e l'intensità del picco Raman viene codificata a colori: quelli associati alle MIC sono rappresentati in rosso nella **figura 5**, mentre quelli delle TIC sono codificati in blu; l'immagine in luce bianca indica l'area in cui è stata eseguita l'analisi. Si osserva come in una popolazione eterogenea di cellule la maggior parte sono TIC e solo poche sono MIC.

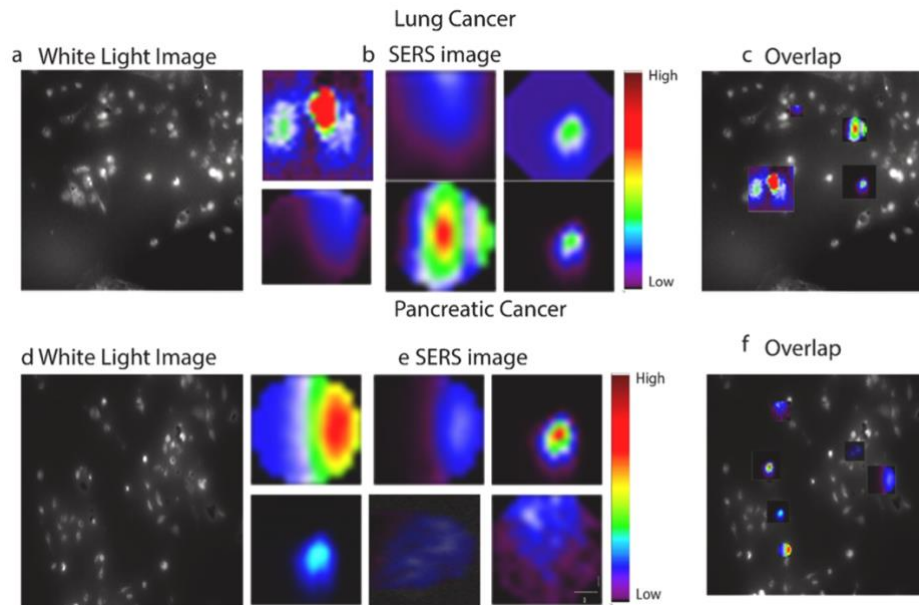


Figura 5. Applicazione dei nanosensori per differenziare le MIC dalle TIC in una popolazione eterogenea di sferoidi. Lo spettro appare rosso per le cellule MIC e blu per le cellule TIC: si possono osservare delle regioni in cui si ha un'overlap dei due colori, ma in generale le cellule MIC sono meno abbondanti delle cellule formanti tumore.

2.7 Previsione delle metastasi attraverso il fenotipo cellulare e il machine learning

Al fine di convalidare ulteriormente l'affidabilità dei nanosensori nella rilevazione e nella differenziazione delle MIC dalle TIC in una popolazione eterogenea, è stato applicato un semplice algoritmo di *machine learning* per analizzare gli spettri SERS ottenuti dalle TIC e dalle MIC: i dati ottenuti dalla popolazione eterogenea di cellule sono stati utilizzati come dati di prova per convalidare gli algoritmi di machine learning per verificare la possibilità di distinguere con accuratezza, sensibilità e specificità le caratteristiche degli spettri, come mostrato in **figura 6**. Il modello PLS-R è stato in grado di distinguere i due gruppi di campioni TIC e MIC con una specificità e sensibilità del 98%: si può notare che la specificità della predizione delle MIC nel cancro pancreatico dipende fortemente dalla posizione dei picchi SERS per le proteine di transizione EMT, dalle vibrazioni delle basi A e G del DNA e

delle vibrazioni di specifici amminoacidi delle proteine; osservazioni simili si riscontrano anche a livello del cancro polmonare. Di conseguenza, è possibile confermare che i processi intracellulari e le differenze macromolecolari sono marciatori affidabili per predire la presenza di cellule iniziatrici di metastasi e quindi anche il potenziale metastatico di un tumore.

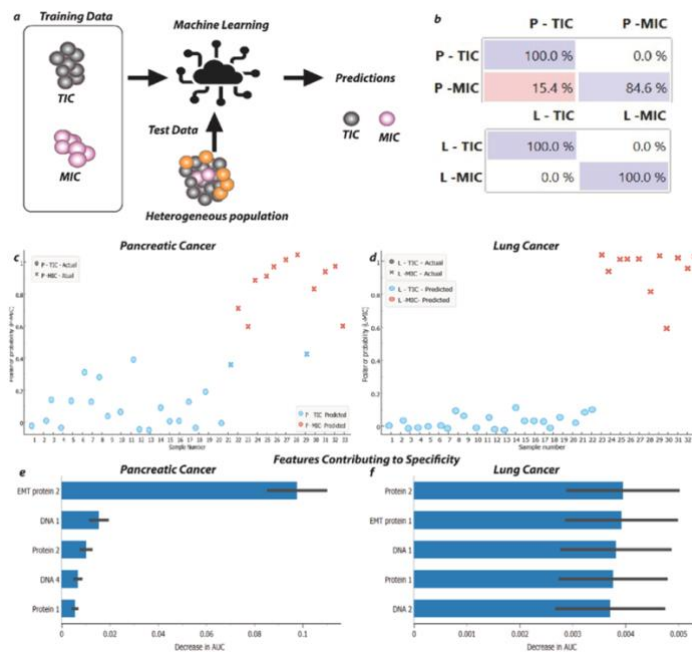


Figura 6. Previsione della metastasi cancerosa attraverso il fenotipo cellulare con l'utilizzo del machine learning. Una popolazione di cellule, suddivise in MIC e TIC, vengono utilizzate per allenare l'algoritmo a distinguere tra le due tipologie (training data). A questo punto l'algoritmo di machine learning è in grado di prevedere se una determinata cellula è una TIC o una MIC con un certo grado di sensibilità. Esistono particolari caratteristiche che contribuiscono ad aumentare la specificità dell'algoritmo e possono influire in modo diverso a seconda del tipo di tumore, come si può osservare negli istogrammi riportati in blu.

2.8 Predizione della metastasi cancerosa da differenti tessuti di origine

L'efficacia dei nanosensori è stata analizzata anche nell'efficacia relativa al differenziamento delle MIC e delle TIC di diverse origini: MIC e TIC del cancro polmonare e del cancro pancreatico sono state utilizzate per addestrare l'algoritmo e i nanosensori erano in grado di individuare l'origine del tessuto canceroso con un'accuratezza e una specificità del 99,62%. La transizione EMT svolge un ruolo importante nella determinazione dell'accuratezza e della specificità della predizione e l'analisi delle caratteristiche fenotipiche ha dimostrato che proprio questa transizione svolge un ruolo predominante nella determinazione del potenziale metastatico delle cellule. Successivamente l'algoritmo è stato utilizzato per differenziare tra i diversi tipi di cellule provenienti da tessuti diversi: questi non si sovrappongono; tuttavia, è stata osservata una classificazione errata tra le cellule pancreatiche, attribuibile allo stato di transizione di queste ultime.

4. DISCUSSIONE

Nell'articolo trattato, è stato identificato un raro sottoinsieme di cellule all'interno del tumore che possiedono il segnale precoce per la metastasi del cancro, note come cellule iniziatrici di metastasi (MIC): si tratta di cellule presenti nella maggior parte dei tipi di cancro metastatizzante, che non vengono influenzate dai cambiamenti epigenetici, e che costituiscono un eccellente marcatore per la metastasi, indipendentemente dal tipo di cancro, dal tessuto di origine e dallo stadio della malattia. Per superare la natura non rilevabile delle MIC, è stato progettato un nanosensore dotato dell'ultra sensibilità necessaria per la rilevazione precoce e la previsione della metastasi: il nanosensore, attivato da SERS, ha consentito la rilevazione delle MIC utilizzando le proprietà dei processi biologici intracellulari e delle loro interazioni nell'ambiente tumorale.

La sonda ha fornito una sensibilità di rilevazione del 98% ed è stata in grado di identificare le MIC in una popolazione eterogenea di cellule tumorali con una specificità del 99,62%. L'applicazione di algoritmi di *machine learning* per la popolazione eterogenea ha prodotto un'accuratezza di previsione dell'89,3%.

Con questo studio, Ganesh e colleghi sono i primi a introdurre una sonda che può fornire sia la diagnosi, sia la previsione della formazione delle cellule MIC fin dalla fase di singola cellula: la capacità di rilevare e prevedere le metastasi del cancro in una fase precoce migliora notevolmente il potenziale diagnostico del cancro. Questo approccio fornisce le basi necessarie per trattamenti che mirano a organi specifici, garantendo una solida base per lo sviluppo di terapie metastatiche efficienti e fornendo una migliore comprensione del meccanismo della metastasi in un contesto diagnostico clinico.

Il *machine learning* ha un elevato potenziale per quanto riguarda l'analisi dei complessi dataset riguardanti le metastasi: i dati possono essere strutturati, che sono quantitativi e organizzati, o non strutturati, come immagini, e video. Inoltre, ha come obiettivo una diagnosi accurata per arrivare a elaborare un trattamento ottimale e, data l'unicità a livello molecolare dei tessuti metastatizzati, gli algoritmi di *machine learning* possono essere utilizzati per identificare le lesioni primarie dai dati relativi all'espressione genica, impiegate per distinguere tra diversi tipi di tumori.

L'applicazione del *machine learning* nello studio delle metastasi risulta essere un'area emergente della ricerca: può essere utilizzata per diverse analisi che includono il rilevamento delle metastasi stesse, la distinzione tra tipi di tumori, la preferenza riguardo all'utilizzo di un determinato trattamento e la prognosi mirata al miglioramento della qualità della vita (Olutamilayo et al.). Molti studi tendono a utilizzare delle procedure non invasive per la diagnosi e il trattamento del cancro metastatico e, dal momento che la maggior parte dei dati disponibili sono basati sulle immagini, l'obiettivo è quello di sviluppare degli algoritmi che siano efficienti nell'analizzare e interpretare queste risorse. Questo

potrebbe essere fatto attraverso un algoritmo che richieda informazioni ricavabili dalle stesse immagini per la predizione dei tumori.

Un altro approccio è quello di sviluppare degli schemi ibridi che combinino il basso costo del *machine learning* con un'analisi più dettagliata ricavabile dal *deep learning*, al fine di ottenere un notevole potere computazionali per l'analisi di questi dati. Gli ambiti di studio futuri includono il *federated learning*, il *quantum computing*, il *transfer learning* e l'utilizzo di diverse popolazioni di dati per il training degli algoritmi.

In particolare, il *federated learning* include l'allenamento degli algoritmi attraverso diversi server decentralizzati che contengono dati di campioni locali senza necessariamente dividerli, tutelando quindi la privacy dei pazienti. Questo tipo di training è applicato in diversi centri in modo da allenare gli algoritmi su dataset relativamente grandi e su una popolazione varia.

Un'altra tecnica importante è il *transfer learning* che prevede di utilizzare dei modelli principalmente allenati come punto di partenza per un nuovo modello che deve eseguire una nuova task: questo metodo ha dimostrato una buona performance con l'analisi delle immagini mediche, in particolare con i campioni di dimensioni ridotte. Considerando la prevalenza dell'utilizzo delle immagini per raccogliere dati relativi alla ricerca sul cancro metastatico, un graduale shift verso l'utilizzo del transfer learning porta ad avere un modo per allenare gli algoritmi molto più rapido e ottimizzato.

Per concludere, vista la necessità di trattare il cancro con tecniche meno invasive, le possibilità e i prospetti analizzati in questa tesi evidenziano l'utilità e l'importanza del machine learning nella diagnosi e nel trattamento della malattia al fine di garantire una migliore qualità della vita.

BIBLIOGRAFIA

1. Chin VL, Lim CL. Epithelial-mesenchymal plasticity-engaging stemness in an interplay of phenotypes. *Stem Cell Investig.* 2019 Aug 20; 6:25. doi: 10.21037/sci.2019.08.08. PMID: 31559312; PMCID: PMC6737451.
2. Cao MX, Jiang YP, Tang YL, Liang XH. The crosstalk between lncRNA and microRNA in cancer metastasis: orchestrating the epithelial-mesenchymal plasticity. *Oncotarget.* 2017 Feb 14; 8(7):12472-12483. doi: 10.18632/oncotarget.13957. PMID: 27992370; PMCID: PMC5355358.
3. Ember KJI, Hoeve MA, McAughtrie SL, Bergholt MS, Dwyer BJ, Stevens MM, Faulds K, Forbes SJ, Campbell CJ. Raman spectroscopy and regenerative medicine: a review. *NPJ Regen Med.* 2017 May 15; 2:12. doi: 10.1038/s41536-017-0014-3. PMID: 29302348; PMCID: PMC5665621.
4. Eric C. Le Re, Pablo G. Etchegoin “Principles of Surface-Enhanced Raman Spectroscopy and relative plasmonic effects”
5. Tabish, T.A.; Pranjol, M.Z.I.; Horsell, D.W.; Rahat, A.A.M.; Whatmore, J.L.; Winyard, P.G.; Zhang, S. Graphene Oxide-Based Targeting of Extracellular Cathepsin D and Cathepsin L As A Novel Anti-Metastatic Enzyme Cancer Therapy. *Cancers* **2019**, *11*, 319.
6. Petinrin OO, Saeed F, Toseef M, Liu Z, Basurra S, Muyide IO, Li X, Lin Q, Wong KC. Machine learning in metastatic cancer research: Potentials, possibilities, and prospects. *Comput Struct Biotechnol J.* 2023 Mar 29;21:2454-2470. doi: 10.1016/j.csbj.2023.03.046. PMID: 37077177; PMCID: PMC10106342.

APPENDICE

Ganesh, S., Venkatakrisnan, K. & Tan, B. Early detection and prediction of cancer metastasis—Unravelling metastasis initiating cell as a dynamic marker using self-functionalized nanosensors. *Sens. Actuators B Chem.* **361**, 131655 (2022).



Early detection and prediction of cancer metastasis – Unravelling metastasis initiating cell as a dynamic marker using self-functionalized nanosensors

Swarna Ganesh ^{a,b,c,e}, Krishnan Venkatakrishnan ^{a,b,c,e,*}, Bo Tan ^{a,b,d,e}

^a Keenan Research Center for Biomedical Science, Unity Health Toronto, Toronto, Ontario M5B 1W8, Canada

^b Institute for Biomedical Engineering, Science and Technology (IBEST), Partnership between Ryerson University and St. Michael's Hospital, Toronto, Ontario M5B 1W8, Canada

^c Ultrashort Laser Nanomanufacturing Research Facility, Department of Mechanical and Industrial Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada

^d Nanocharacterization Laboratory, Department of Aerospace Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada

^e Nano-Bio Interface facility, Department of Mechanical and Industrial Engineering, Ryerson University, 350 Victoria Street, Toronto, Ontario M5B 2K3, Canada

ARTICLE INFO

Keywords:

Cancer metastasis
Nanosensor
Biosensor
SERS
Self-functionalization

ABSTRACT

The ability of cancer to metastasize to distant organs is an urgent problem to be solved clinically and continue to pose a greater challenge to researchers. Current treatments for cancer are ineffective for metastatic cancer due to inability of conventional imaging techniques to detect at early stages. Additionally, the ability to predict the ability of cancer to metastasize in advance will prove to be a valuable tool to improve patient prognosis. However, current techniques of prediction rely on molecular expression profiles which are highly tumor-specific, patient-specific and does not account for tumor heterogeneity. These factors critically impede the ability to use molecular expression profiles for early diagnosis and prediction of cancer metastasis. Hence, there is a need to concentrate on using tumor cells' phenotypic heterogeneity, which can serve as a diagnostic and predictive marker for cancer metastasis. Here, we have identified a rare subpopulation of cells within the tumor, which can signal metastasis initiation, known as metastasis-initiating cells (MICs) which are cancer non-specific, independent of epigenetic makeup, thereby becoming an excellent, unprejudiced marker for cancer metastasis irrespective of cancer type, the tissue of origin, and cancer stage. However, it is challenging to use MICs as a cellular marker for metastasis from a clinical perspective due to their rare, undetectable nature and the lack of sensitive methods to detect this elusive population of cells. To provide the ultra-sensitivity necessary for the early detection and prediction of cancer metastasis using MICs here we have created a self-functionalised nanosensor which is highly SERS active. The self-functionalized nanosensor enabled the detection of MIC using properties of intracellular biological processes and characters of the tumor microenvironment. The nanosensor enabled a detection sensitivity of 98%. It was able to identify MIC in a heterogeneous population of TICs with an unparalleled specificity of 99.62%. Further, the accuracy of MIC as a predictive marker for metastasis in a heterogeneous population of tumor spheroids was found to be 84.6%. This work contributed to the utilization of cancer cell heterogeneity to identify MIC, which can serve as a universal marker for diagnosis, prediction, and prognosis of tumor metastasis. To the best of our knowledge, this study is the first to design a probe that can provide both the diagnostic signature and predictive signature of cancer metastasis as early as the single cellular stage. This approach holds immense potential in the early diagnosis of metastatic tumors in a clinical setting and considerably improve patient prognosis.

1. Introduction

Cancer metastasis accounts for more than 90% of cancer-related

deaths [1]. The best chance to reduce metastasis-related mortality is through early diagnosis and prediction [2]. However, the understanding of cancer metastasis remains weak [3]. It is critical to develop the ability

* Corresponding author at: Keenan Research Center for Biomedical Science, Unity Health Toronto, Toronto, Ontario M5B 1W8, Canada.
E-mail address: venkat@ryerson.ca (K. Venkatakrishnan).

to detect at early stage and predict cancer metastasis to improve patient survival, disease management and eventually reduce mortality. However, the early detection and prediction of metastasis is critically hampered by the lack of sensitive techniques at a single cellular level, lack of distinguishable markers and the high heterogeneity of metastasis tumors. Hence, there is an urgent need on two fronts, first to search for an unbiased marker can be used for early detection and prediction of cancer metastasis, second to develop an ultrasensitive method with high specificity to perform early detection and prediction.

Metastasis detection is mainly performed through two approaches: non-invasive imaging techniques such as X-ray computed tomography (CT), positron emission tomography [4], and molecular markers such as genes [5], miRNA [6], lncRNA [7], DNA methylation [8]. The non-invasive imaging techniques are not suitable for the detection of early metastases at this resolution. Further, given the high heterogeneity among cancer cells, and the small size of cancer metastasis, it is difficult for the conventional imaging agents to reach these elusive cell population [9]. Also, extensive research has identified multiple epigenetic markers associated with cancer metastasis; however, these markers are highly dependent on the type of tumor and does not accurately provide a universal template for early diagnosis of metastasis. The problems mentioned above impede the use of epigenetic markers to detect cancer metastasis at an early stage.

In addition to the early detection of metastasis, the ability to predict metastasis also plays a critical role in optimizing treatment options and clinical decision making. Conventionally, a primary tumor's ability to metastasize is predicted based on the assumption that the metastatic tumor is molecularly genetically like the primary tumor; however, recent studies show otherwise. The prediction method also assumes that each sample belongs to a specific cancer subtype, based on the tumor's molecular characteristics. This assumption is a highly personalized approach, does not account for tumor heterogeneity, and hinder the repeatability of the prediction model [10]. The high degree of heterogeneity of tumors and the patient-specific nature of molecular/epigenetic markers renders them an unstable marker for metastasis prediction. Hence, there is a need for a universal marker for cancer metastasis that can be used for prediction and diagnosis.

We have identified a rare subpopulation of cells within the tumor that possess an early signal for metastasis known as metastasis-initiating cells (MICs). MICs represent a unique, cancer non-specific, undetectable, distinct subpopulation of cells that possess the fundamental properties that could be used to detect metastasis at a single cellular stage. The dynamic nature of cancer metastasis warrants a prediction marker such as MICs, which remains unaffected by the challenges faced by conventional prediction techniques. MICs represent a unique cell population, representing pro-metastatic events common to most solid tumors. Collectively, MICs have the potential to serve as an unbiased, universal marker for both early detection and prediction of cancer metastasis.

The undetectable nature of MICs warrants for an ultrasensitive, non-destructive method that can provide an all-inclusive representation of MICs. Further, non-invasive molecular detection methods can provide a fresh perspective on cancer metastasis's predictive, diagnostic, and prognostic signature. Hence, probing MICs using a non-invasive technique such as SERS is critical to achieving this goal. Hence, to study the elusive cell population among the heterogeneous tumor cells, there is a need for a sensor, which can efficiently enter the tumor microenvironment and further internalize into the cell and remain benign in the cellular microenvironment. Sub 5 nm graphene oxide presents high biocompatibility [11], remains non-reactive in the tumor microenvironment [12], higher internalization efficiency, thereby making it an ideal candidate to be applied for early detection and prediction of cancer metastasis by probing MICs.

Here, we have attempted to tackle early detection and prediction of cancer metastasis from two fronts: first, to the best of knowledge, we are the first to employ a rare subpopulation of cells within a tumor known as

Metastasis Initiating Cells (MICs) as an unbiased marker for early detection and prediction for metastasis; second, to the best of knowledge, this study is the first to create a self-functionalized nanosensor equipped with ultra-sensitivity to detect MICs.

The accurate diagnosis and prediction of cancer metastasis were achieved by efficiently internalizing the self-functionalized nanosensor into MICs to amplify the signals leading to an accurate diagnosis and prediction of cancer metastasis. The nanosensor was synthesized through ultrashort laser pulse interaction with a carbon substrate in an oxygen-rich atmosphere. The sub-5 nm size (~2 nm) of the self-functionalized nanosensor, along with the presence of abundant oxygen functional groups on the surface of the probe, enabled efficient self-cellular uptake.

Preclinical models of pancreatic cancer and lung cancer cells were used to ascertain the non-cancer-specific nature of MICs and their applicability as a universal marker for cancer metastasis diagnosis and prediction. SERS analysis was performed in a sorted cell population to investigate the significant differences in the signature of TICs and MICs. The predominant features that can aid as a diagnostic and predictive signature were found in the proliferative state, cellular phenotypic plasticity (EMT plasticity), intracellular DNA of MICs, and TICs. The principal component analysis (PCA) of the SERS spectral was able to differentiate them with a statistical difference. The tumor microenvironment was analyzed by tumor secretome, which comprises the metabolites, nutrient consumption, and molecules responsible for cell-cell communication. Results show that the self-functionalized nanosensor distinguished the TIC and MIC in the heterogeneous population without ambiguity.

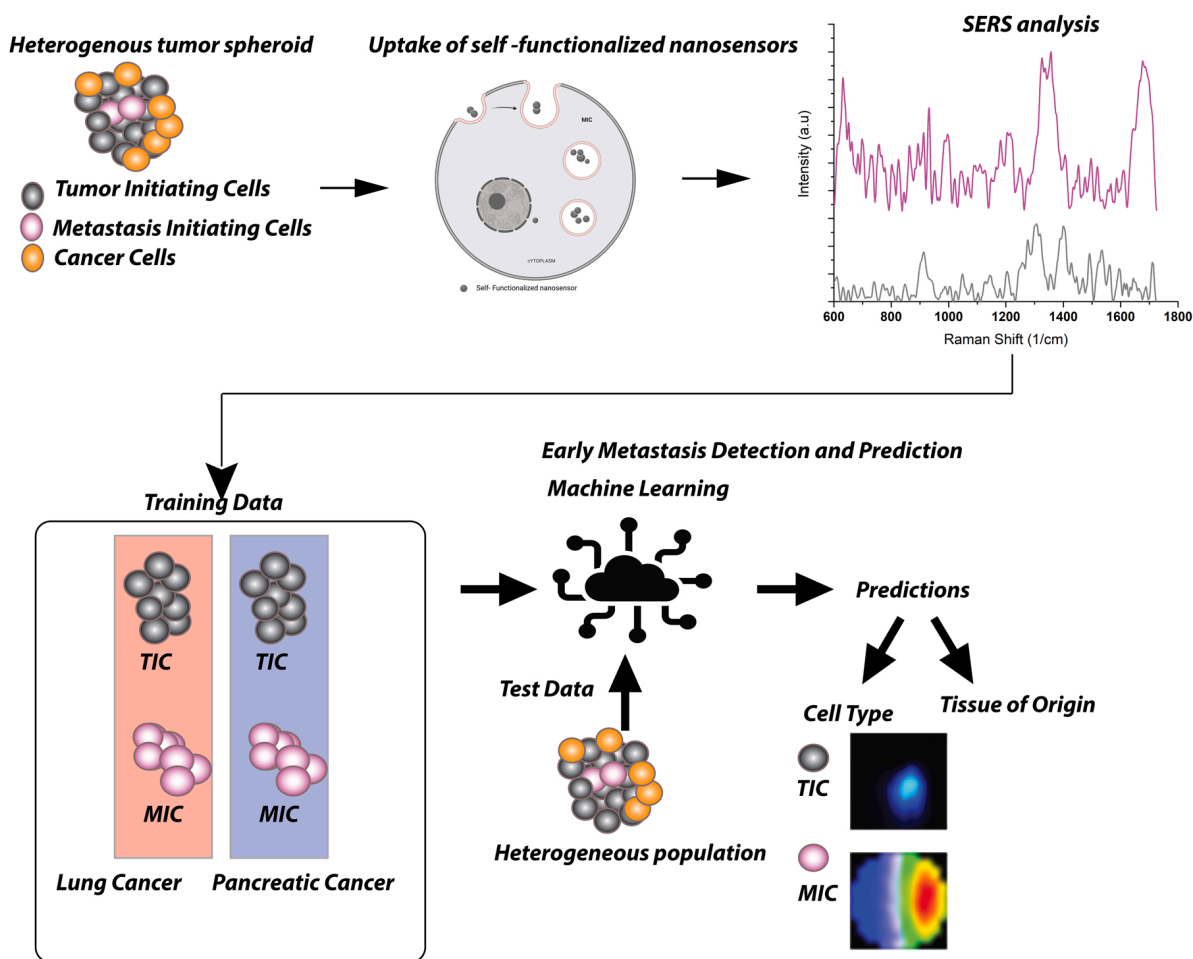
The data obtained from the sorted population was used as a training dataset for predictive modeling using the Support Vector Machine algorithm. The model's prediction accuracy was 99.62% for TIC and MIC's heterogeneous population from different tissue origins. This study is the first to propose a cellular marker for the early detection of cancer metastasis to the best of our knowledge. This proposed new self-functionalized nanosensor enables the pathological identification of metastatic cancer in the early stages and predicts the probability of the onset of metastatic cancer, thereby providing an opportunity to prolong patient survival (Scheme 1).

2. Results

2.1. Fabrication of self-functionalized nanosensors

The Self-functionalized nanosensors was synthesized by using an ultrashort pulsed laser (UPL) ionization process in the presence of oxygen gas. The UPL ionization process produces nanosensors of desired morphology, size, and chemical properties by laser pulse impacting the graphite surface [13]. The femtosecond laser pulse interacts with the graphite surface, and the laser energy is reassigned from the electrons in the valence band to the electrons in conduction band [14], which is later transferred to the graphite lattice through the electron-phonon coupling. This energy transfer causes a total breakdown of graphite into ionic and neutral carbon species in the form of an expanding plume. The introduction of oxygen gas in the plume environment facilitates a slower plume expansion and further confinement of the laser plume. When the laser pulse is discontinued, the plume temperature drastically reduces to room temperature in a period of several nanoseconds. However, in the presence of oxygen gas, the plume temperature is increased due to the shielding effect of oxygen, thereby increasing the ionization potential of the plume [15], leading to smaller particle size.

The morphology and size distribution of the Self-functionalized nanosensors was determined using Transmission electron microscopy (TEM). Fig. 1a shows the TEM images of the nanosensors, which are polygonal in shape. Based on the TEM image, the size distribution of the nanosensors was determined using the ImageJ software. It can be observed from Fig. 1b that the particle size distribution of the probe fits



Scheme 1. Schematic representation of the methodology for early diagnosis and prediction of cancer metastasis.

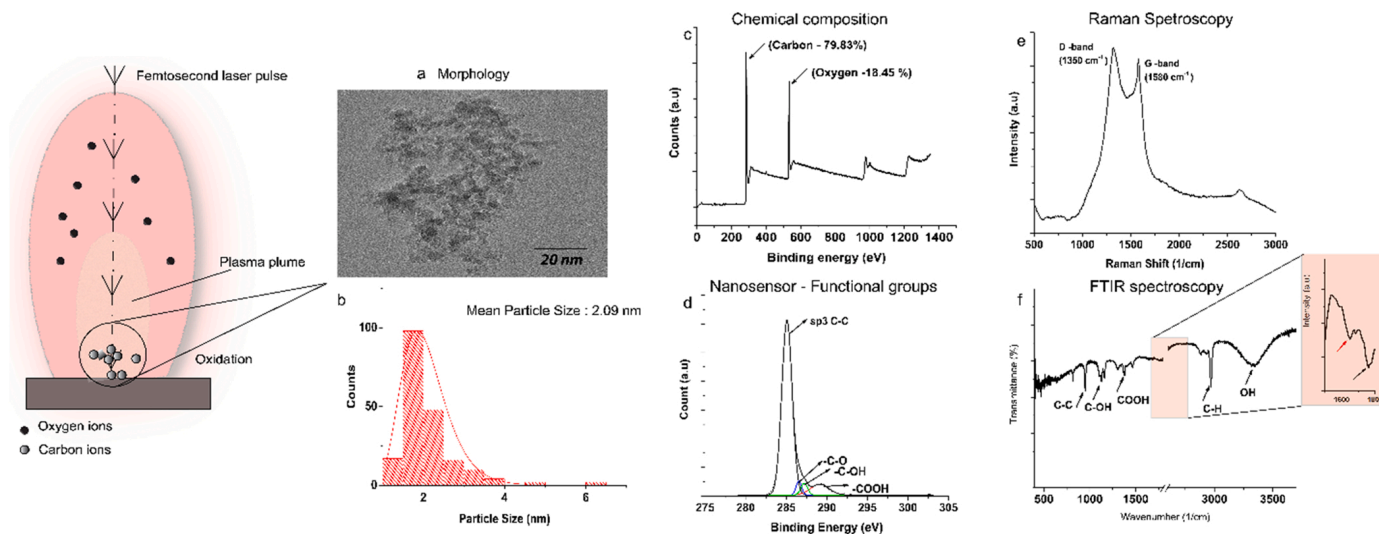


Fig. 1. Physicochemical characterization of Self-functionalized nanosensors using a) TEM imaging showcasing the morphology of Self-functionalized nanosensors, b) Histogram showcasing the size distribution of Self-functionalized nanosensors c) XPS analysis showcasing the chemical composition of Self-functionalized nanosensors d) High-resolution XPS spectra of C1s showcasing the presence of oxygen functional groups in Self-functionalized nanosensors and the predominance of sp³ hybridization e) Raman spectra showcasing the presence of characteristic bands of graphene oxide f) ATR-FTIR spectra confirming the presence of oxygen-containing functional groups in the Self-functionalized nanosensors, Inset – Presence of peaks corresponding to functional groups at vacancies in the carbon lattice.

in the normal distribution curve with a mean particle size of 2.09 nm with a standard deviation of 0.7 nm.

Fig. 1c showcases the XPS survey spectra to determine the chemical composition of the probe. It can be observed that the Self-functionalized

nanosensors primarily of carbon (79.83%) and oxygen (18.45%). Further, the high-resolution C1s spectra reveal the functional groups present on the surface. It can be observed from Fig. 1d that the peak binding energy of C1s is red shifted to 285.3 eV, confirming the

presence of a sp³ hybridized C-C bond. This peak shift is commonly observed when the oxygen functional groups are dominantly present on SERS' surface-activated carbon nanosensors. The deconvolution of C1s spectra yielded oxygen functional groups, including -C=O, COOH, -C-OH corresponding to peak positions at 287 eV, 286.4 eV, and 289 eV, respectively [16]. The predominance of oxygen functional groups was further confirmed using an ATR- FTIR spectroscopy, shown in Fig. 1f. Additionally, we investigated the presence of vacancies and edges in the self- functionalized nanosensor, which can be observed from the inset of Fig. 1f. The peak present at ~1650 cm⁻¹ corresponds to the presence of ketone stretching vibration at vacancies in the carbon lattice [17,18]. Further, the peak at ~1750 cm⁻¹ corresponds the carboxyl stretching and OH vibration at the edges of self- functionalized nanosensor [17, 19]. The Raman spectra of the Self- functionalized nanosensors showed the characteristic peaks of graphene oxide at 1350 cm⁻¹ (D-band) and 1580 cm⁻¹ (G-band), corresponding to the second-order of zone boundary phonons and doubly degenerate zone center mode, respectively [20]. The presence of zone boundary phonons can be predominantly seen only in graphene, with defects caused due to incorporation of oxygen in the graphite lattice [21], which further corroborates the presence of defect related vibrations in the infrared absorption of the self- functionalized nanosensor. The high Raman cross-section of GO, combined with the uniform surface and superior biocompatibility [22], makes the self- functionalized nanosensor a right candidate for SERS based early detection of MICs with a single cellular sensitivity.

2.2. Identification of metastasis initiating cells

A prototypical example of cellular plasticity in tumor progression is the process of Epithelial-Mesenchymal Plasticity (EMP), which encompasses both epithelial-mesenchymal transition (EMT) and the reverse process of mesenchymal to epithelial transition (MET), which is the first step of the metastasis – invasion cascade. The cascade's primary step is the epithelial cells lose the cell-cell adhesion to acquire mesenchymal, a characteristic required to migrate to distant sites through the bloodstream. The EMT is subsequently followed by the reverse process of MET, which enables secondary tumors in distant metastatic sites. The primary mechanism of EMT- MET is through the functional loss and re-expression of proteins related to cellular functions, such as E-cadherin, Vimentin, and Cytokeratin, through multiple genetic mechanisms mutation, hypermethylation, and repression of transcription factors [23–25]. Studies have revealed that only a small percentage of cells that undergo EMT can sustain the reverse process of MET, therefore the ability to seed secondary tumors [26,27]. Multi-omics analysis of these small percentage of cells capable of sustaining MET has proven the presence of an increased number of cancer driver mutations, dysregulated transcriptional levels, chromosomal instability, and heterogeneity [26,28].

TIC and MIC's main difference is the expression of epithelial markers such as E- cadherin, cytokeratin, and EpCAM [29]. Researchers currently use the coexpression of several surface markers to study the MIC subpopulation. The current understanding is that the expression of CD 44 is essential for invasive growth and CD 133 for maintaining the self-renewal property of tumor-initiating cells. Among the numerous epithelial markers associated with tumor initiation, the expression of cytokeratin is associated with poor prognosis in various solid tumors [30]. The up-regulation of cytokeratin is associated with morphological and phenotypical changes in malignant lesions [31]. Besides, cytokeratin is the only epithelial marker that is acquired in cells that have undergone epithelial differentiation [32]. Given the properties mentioned above, we have hypothesized that the use of multiple surface markers may provide an effective mechanism to identify metastasis initiating cells.

To investigate MIC with SERS, first, it has been isolated. Fig. 2 presents a workflow of separation of MIC from TIC. The potential for identifying the tumorigenic TIC is to use CD 44⁺, CD 133⁺, Cytokeratin⁺ cells where CD 44, CD 133 corresponds to the stem cell-like phenotype cytokeratin corresponds to the cells that had undergone the transition from mesenchymal phenotype to epithelial phenotype [33,34]. Fig. 2 showcases the flow cytometric sorting of TIC spheroid cells comprising of a heterogeneous population of Pancreatic TIC spheroids, Lung TIC spheroids. It can be observed from Fig. 2b that the Lung TIC spheroids comprise 31.12% CD 44⁺ cells. These cells were further used to obtain the double-positive cells (CD 44⁺ / CD 133⁺) constituting 15.54% of the population, termed as lung tumor initiating cells (L- TIC). The tumorigenic Lung Cancer Stem Cell (L- MIC) population shown in Fig. 2b comprising of (CD 44⁺ / CD 133⁺ / CK⁺) cells constitutes only 1.53% of the total TIC population. It can be observed from Fig. 2c that the Pancreatic TIC spheroids comprise 26.79% CD 44⁺ cells. These cells were further used to obtain the double-positive cells (CD 44⁺ / CD 133⁺) constituting 29.76% of the population, termed Pancreatic tumor initiating cells (P- TIC). The tumorigenic Pancreatic Cancer Stem Cell (L- MIC) population shown in Fig. 2c comprising of (CD 44⁺ / CD 133⁺ / CK⁺) cells constitutes only 0.82% of the total TIC population. The low ratio of tumorigenic TICs in the cellular population makes it extremely difficult to detect these cells' presence. The cells were sorted using FACS and used for further analysis to characterize the differences between tumor initiating cells and metastasis initiating cells.

2.3. Uptake of self- functionalized nanosensors in TIC and MIC

The SERS analysis of cells requires the uptake of Self- functionalized nanosensors by both TIC and MIC. The representative electron microscopy images shown in Fig. 3a shows the presence of self- functionalized nanosensors in the cytoplasmic region of the MICs from lung cancer and pancreatic cancer. A flow cytometry analysis is conducted to investigate

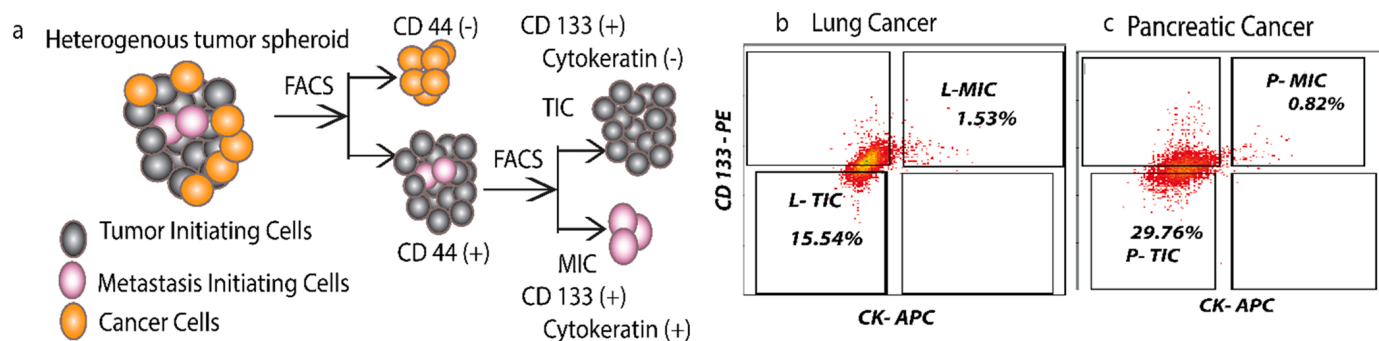


Fig. 2. Identification of metastasis initiating cells in tumor spheroids a) Schematic representation of the workflow associated with the isolation of metastasis initiating cells. Flow cytometric isolation of metastasis initiating cells in b) CD 44 positive lung cancer cells b) L-TIC subpopulation (CD 44 positive / CD 133 positive / CK negative), L-MIC subpopulation (CD 44 positive / CD 133 positive / CK positive) c) P-TIC subpopulation (CD 44 positive / CD 133 positive / CK negative), P-MIC subpopulation (CD 44 positive / CD 133 positive / CK positive).

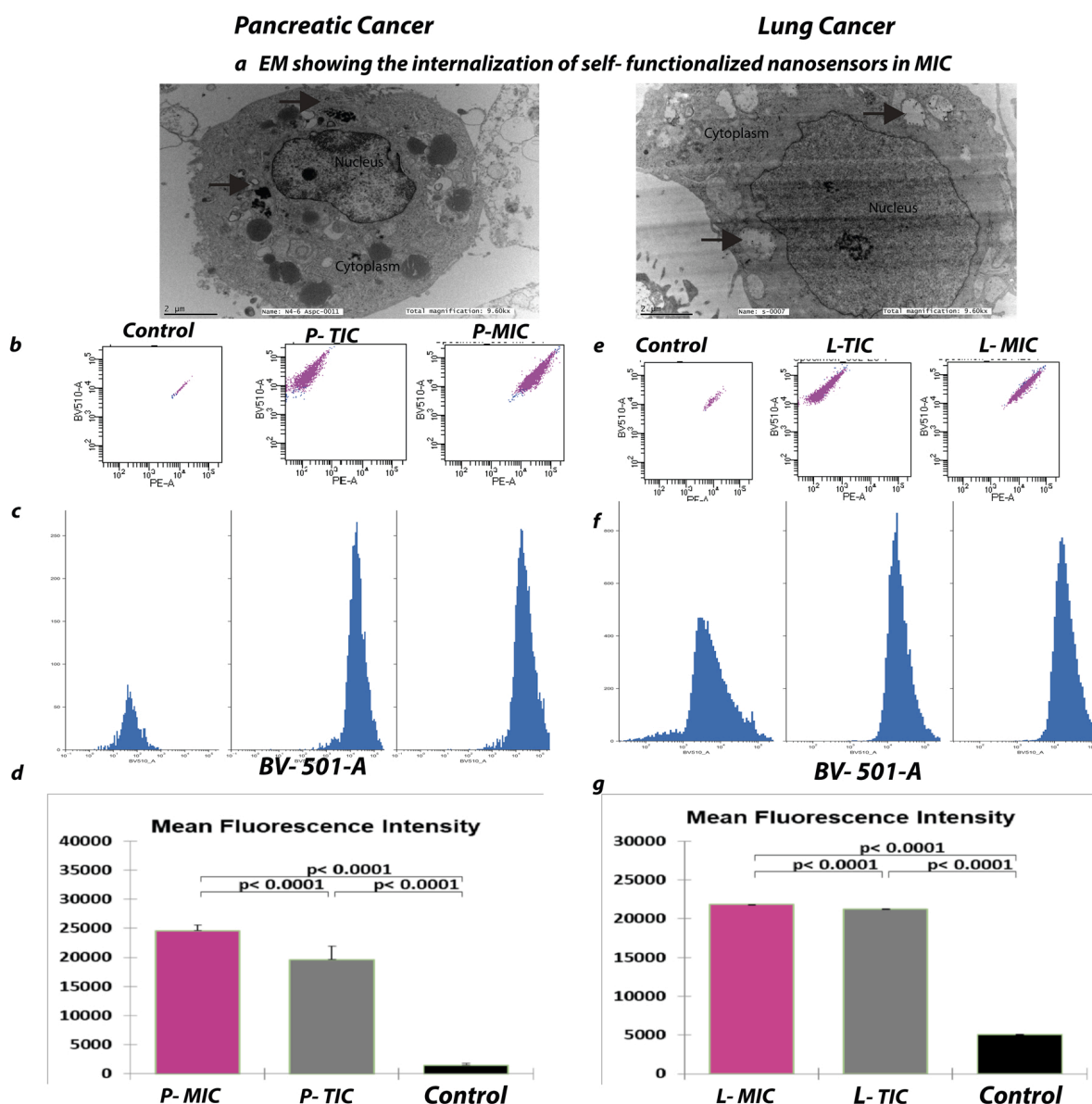


Fig. 3. Self-functionalized nanosensors uptake through flow cytometric analysis and electron microscopy imaging. a) Electron microscopy imaging showing the presence of self-functionalized nanosensors in the cytoplasmic region of MICs. Representative dot plots of Self-functionalized nanosensors uptake in b) Pancreatic Cancer cells e) Lung Cancer Cells, Histogram of fluorescence intensity of control and treated c) Pancreatic tumor initiating cells and pancreatic metastasis initiating cells f) Lung tumor initiating cells and Lung metastasis initiating cells, Comparison of mean fluorescence intensity of d) Pancreatic tumor initiating cells and pancreatic metastasis initiating cells g) Lung tumor initiating cells and Lung metastasis initiating cells.

the internalization of the Self-functionalized nanosensors. Flow cytometry analysis of probe uptake is a highly sensitive technique to quantify the uptake of self-functionalized nanosensors particles with an intrinsic fluorescence property [35]. The result is presented in Fig. 3. The Self-functionalized nanosensors load per cell is characteristically expressed as the number of events of the fluorescent associated with the Self-functionalized nanosensors, showing a characteristic fluorescence emission at 525 nm (BV-510) (Fig. S1). The distinguishing fluorescence of the Self-functionalized nanosensors enables the detection and quantification of probe uptake with a single cellular sensitivity. The internalization of the Self-functionalized nanosensors can be detected from the side scattering, given in Fig. 3b, e. The cells with an internalized Self-functionalized nanosensors show a high intensity of BV-510 in TICs derived from both pancreatic cancer cells and lung cancer cells, as shown in Fig. 3c, f, respectively. Further, it can be observed from Fig. 3d, g that the MIC exhibits an increased fluorescent intensity when compared to TIC and control cells. The high fluorescent intensity could be attributed to increased uptake of Self-functionalized nanosensors by

the MIC.

2.4. SERS characterization of MIC

To assess the intracellular SERS activity of the self-functionalized nanosensors, we analysed the SERS spectral differences between MIC with nanosensors, TIC with nanosensors and the cells without nanosensors were used as a control. The data is presented in Supplementary Fig. S4. It can be observed that the MICs shows a substantially high signal enhancement when compared to TIC. Additionally, the control cells show no visible discernible signals of intracellular molecules. The high intracellular SERS activity could be attributed to multiple factors including the sub 5 nm size of the nanosensors, the presence of favorable functional groups on the surface of nanosensors and the presence of defects which further amplifies the SERS activity. Studies have shown that the sub 5 nm size of graphene oxide results in the activation of graphene plasmon which provides an high level of signal enhancement of DNA molecules [36]. Additionally, the presence of surface functional

groups results in the interaction of nanosensors with proteins, lipids and RNA molecules intracellularly resulting in charge transfer between the intracellular macromolecules and the nanosensor system [37–39].

Following the uptake of Self- functionalized nanosensors by TIC and MIC, the cells' SERS spectra were obtained. The cells sorted through FACS were further treated with Self- functionalized nanosensors for 24 h. The cells were also fixed using the 4% paraformaldehyde. The results of the SERS analysis are presented in Fig. 4. It is worth stress that the data presented here is obtained at the single-cell mapping of TICs and MIC.

Fig. 4a, showcases the SERS spectra of lung tumor initiating cells (L-TIC) and lung metastasis initiating cells (L- MIC). It can be observed from Fig. 4a that the SERS intensity of L- MIC is higher than that of L-TIC. A similar trend was observed in pancreatic tumor initiating cells (P-TIC) and pancreatic metastasis initiating cells (P- MIC), as shown in Fig. 4b, respectively. This observation is in agreement with MICs' high proliferation rate and enhanced nutrient uptake. Further, the SERS spectra show the characteristic Raman peaks associated with Various intracellular biomolecules.

To obtain adequate information to point out the differences between TIC and MIC, we used multivariate statistical analysis on the cells' SERS spectra. It can be observed from Fig. 4c that the PCs of L- TIC and L- MIC conjugated into separate clusters. Further, it can be seen from the figure that L- TIC and L- MIC (Fig. 4c), P- TIC, and P- MIC (Fig. 4d) form separate clusters with no overlapping, confirming the capability of Self- functionalized nanosensors to differentiate TICs from MICs without ambiguous.

The specific molecular differences between L- TIC and L- MIC, P- TIC and P- MIC, were obtained from PC loading shown in Supplementary Fig. S3 respectively. The peak differences between L- TIC and L- MIC are 635 cm^{-1} , 885 cm^{-1} , 921 cm^{-1} , 1387 cm^{-1} , 1573 cm^{-1} , which can be attributed to C-S vibration (Protein), Ribose – phosphate (RNA), C-C vibration of proline (Protein), CH_3 deformation (proteins, lipids), phenylalanine, tyrosine (Proteins) respectively [40,41]. Similarly, the primary peak differences between P- TIC and P- MIC are 635 cm^{-1} , 882 cm^{-1} , 991 cm^{-1} , 1149 cm^{-1} , 1389 cm^{-1} , 1446 cm^{-1} , 1606 cm^{-1} corresponding to C-S vibration (Protein), Ribose – phosphate (RNA), Ribose phosphate (RNA), C-N vibration of nucleotides (DNA), CH_3

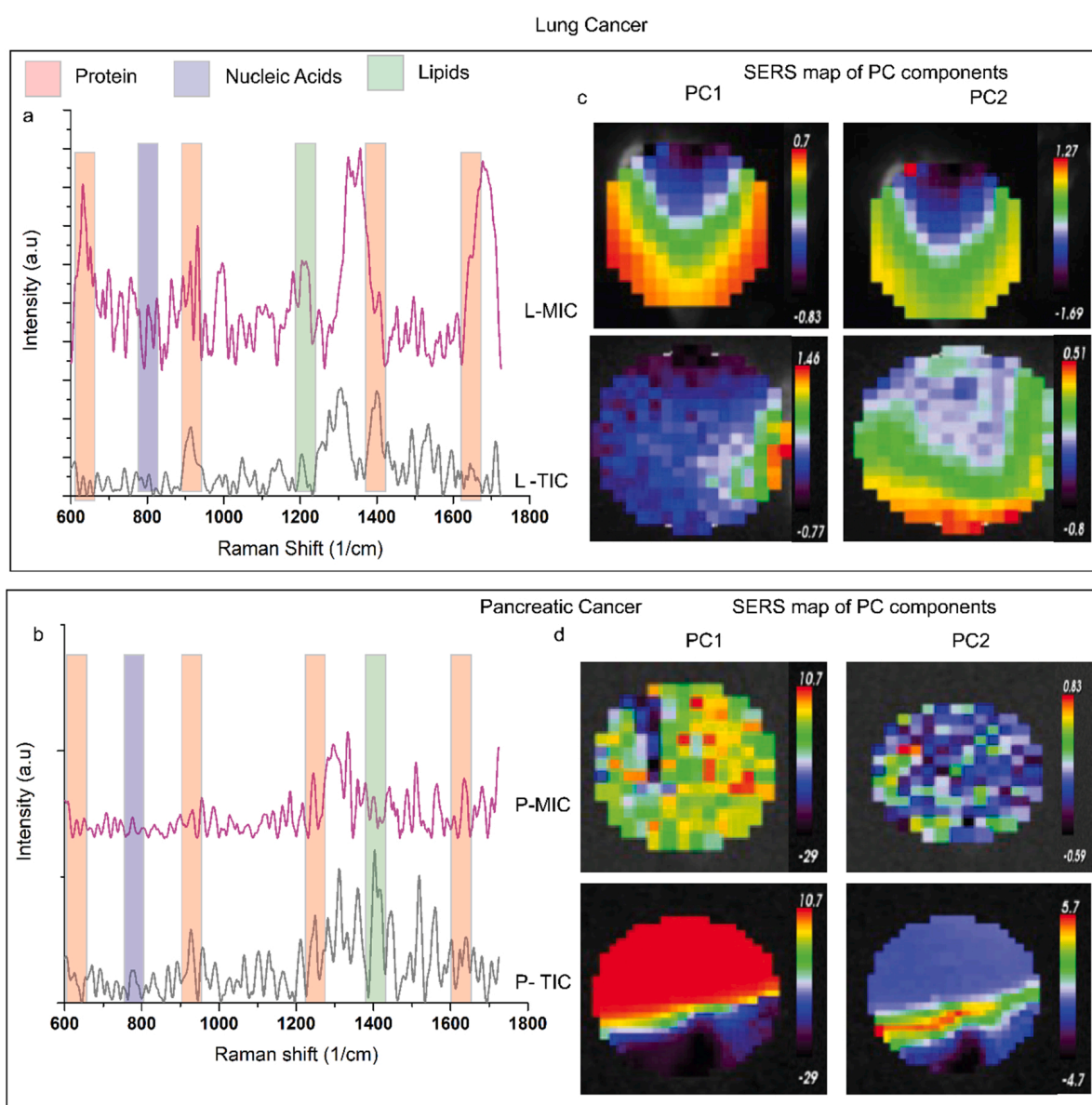


Fig. 4. SERS spectral characterization to differentiate between tumor initiating cells and metastasis initiating cells. SERS spectra of a) lung tumor initiating cells(L-TIC), lung metastasis initiating cells (L-MIC) b) Pancreatic tumor initiating cells (P-TIC), pancreatic metastasis initiating cells(P-MIC). The multivariate statistical analysis SERS mapping points out the differences between tumor initiating cells and metastasis initiating cells derived from c) Lung cancer cells d) Pancreatic cancer cells (Scale 20 μm).

deformation (proteins, lipids), CH₂ deformation of lipids, Phenylalanine, tyrosine (Proteins) respectively [40,41]. These results represent the macromolecular differences between TIC and MIC. The primary differences occur mainly in proteins, lipids, and nucleic acids. These differences in the building blocks of cellular processes indicate that the TIC and MIC are vastly different populations exhibiting a high degree of both functional heterogeneities and phenotypic heterogeneity. The observed difference between MICs and TICs can be used as a marker to investigate MIC within the tumor niche.

2.5. Determining the intracellular state of MIC using intracellular DNA distribution

Growing evidence suggests that TICs are intrinsically resistant to DNA damage by accelerating the DNA repair mechanisms, increasing DNA damage checkpoints, and altering the cell cycle [42]. Moreover, studies have revealed that epigenetic regulation is a crucial player in promoting tumor progression and metastasis by altering critical intracellular signaling pathways [42,43]. TICs have developed the ability to maintain low intracellular reactive oxygen species (ROS) to prevent the ensuing DNA damage and further cell death. The maintenance of low ROS levels was made possible by activating the intracellular scavenger pathways, removing ROS [44]. In certain instances, such as intratumoral hypoxia, transcriptional deregulation, genomic instability, the ROS levels in the tumor increases, leading to ROS induced DNA lesions [45]. The accumulation of ROS induced DNA base lesions is accompanied by translocation of Rac1, thereby prolonging the M phase of cell cycle [46]. Although these intracellular processes help repair the DNA damage caused in TICs, they increase DNA instability, DNA mutations and promote metastatic tumor progression [42]. It can be inferred that intracellular DNA plays a vital role in determining the fate of TIC.

In addition to the differentiation of TIC and MIC based on intracellular processes, the cellular macromolecules, especially DNA and proteins, play a significant role in deciding the cell's fate. Hence, it would be accurate to utilize the differences in intracellular DNA as a marker to detect tumorigenic TIC within a given population, which is presented in Fig. 5.

Fig. 5 showcases the intracellular distribution of DNA in TIC and

MIC. It can be observed from the SERS image of L- TIC and L- MIC (Fig. 5a), P- TIC, and P- MIC (Fig. 5d) the tumorigenic TIC has a higher DNA content. The SERS peak positions corresponding to intracellular DNA structure were compared to determine TIC and MIC differences. The Principal Component Analysis (PCA) scatter plot revealed that SERS peaks associated with intracellular DNA from different cell phenotypes form a distinct cluster on the farthest side of the plot, as seen from Fig. 5c. A similar result can be observed in the PCA plot of intracellular DNA in P- TIC and P- MIC cells (Fig. 5e). Hence, it could be concluded from SERS – PCA analysis that intracellular DNA in a heterogenous TIC population is significantly different, making it one of the crucial molecules in sorting out MIC from TICs.

Fig. 5c and f show the violin plot of SERS peaks associated with DNA structure. It can be observed that the MIC shows a higher intensity owing to the higher DNA content cells. However, the SERS spectra do not show any peak shift associated with DNA damage in the DNA structure. Further, the ratio of 1393 cm⁻¹ / 1334 cm⁻¹, with a value lesser than 1, indicates the ROS induced backbone damage to DNA [47]. The SERS spectra revealed the median ratio of 0.98 and 0.94 for L- TIC and L- MIC. Similarly, the ratio of 1393 cm⁻¹ / 1334 cm⁻¹ for P- TIC and P- MIC were 0.916 and 0.928, respectively. It can be inferred from the SERS ratiometric analysis that intracellular DNA in TIC is damaged due to double-strand breakage. Recent research shows an increased expression of c- Myc and EMT initiation in TIC to activate the DNA repair mechanism [48,49]. Hence, it can be suggested that TICs undergoing EMT acquire an enhanced migratory potential and leads to TIC colonization in remote sites. After TIC colonization, the TIC undergo MET, thereby establishing a metastatic tumor. Hence, it can be deduced, and TICs present a dynamic activation of genes and proteins contributing to DNA repair. Therefore, the detection of a specific DNA repair mechanism and the corresponding intracellular signal transduction pathway essential for TIC survival and maintenance could help determine the treatment course and the corresponding clinical outcome of patients with metastatic cancer.

2.6. Determining the intracellular state of MIC using cellular proliferation

Recent evidence shows an extensive intratumoral heterogeneity

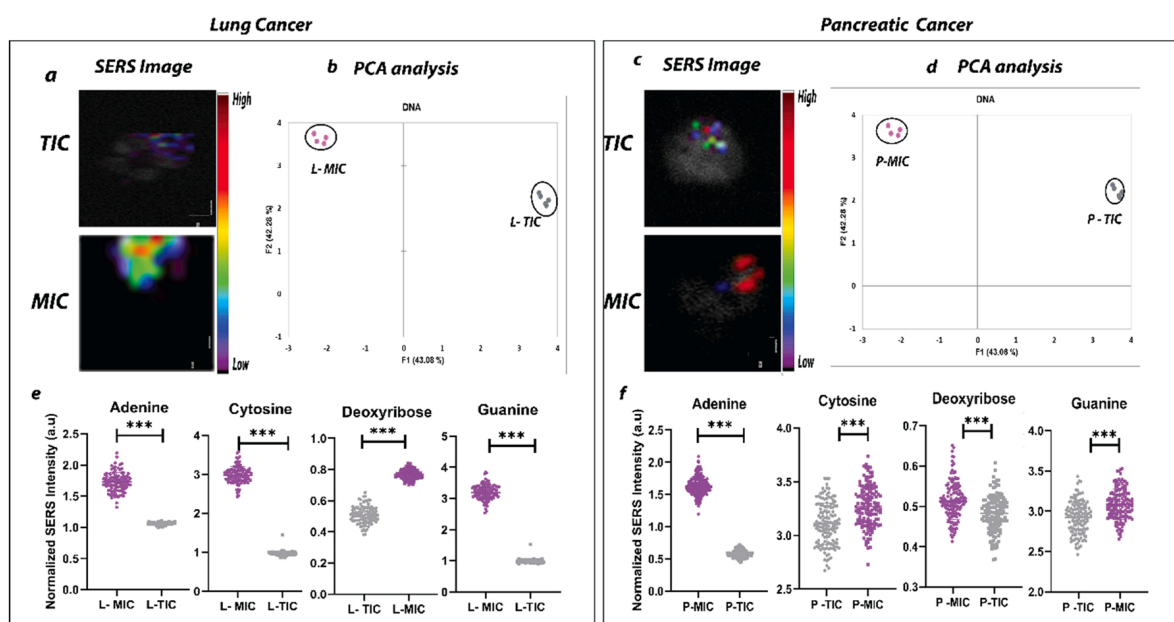


Fig. 5. SERS analysis of intracellular distribution DNA to determine the cellular state of TIC and MIC a) SERS image of TIC and MIC from lung cancer showcasing the distribution of DNA in the cell b) PCA analysis to showcase the difference in SERS peaks associated with DNA in L- TIC and L- MIC c) SERS image of P- TIC and P- MIC showcasing the distribution of DNA in the cell d) PCA analysis to showcase the difference in SERS peaks associated with DNA in P- TIC and P- MIC e, f) Scatter plot comparing the relative SERS intensities of peaks associated with intracellular DNA in lung cancer and pancreatic cancer respectively (***) ($p < 0.001$) (Scale 20 μm).

through the presence of cells with variable states of differentiation [50, 51]. These differences can best be explained by several factors, including epigenetic variation, metabolic variation, microenvironmental differences [52]. Studies have shown that using lineage-tracing experiments shows that a tiny proportion of TIC cells show increased cellular proliferation and comprises the bulk of tumors in a secondary metastatic site [53,54]. These cells' cellular transformation is driven through horizontal gene transfer, mutations, and metabolic shifts, including oxidative phosphorylation [55]. Moreover, the heterogeneous population of TICs displays dynamic changes in epithelial and mesenchymal characteristics. The mesenchymal phenotype of the TICs enables the migratory capacity, whereas the dedifferentiation of mesenchymal to epithelial phenotype is essential for the initiation of metastasis [56,57]. The process of dedifferentiation and the metabolic shifts associated with the cellular transformations can be used as a marker to distinguish the MIC from TIC.

The capacity of epithelial cells to gain a mesenchymal phenotype has been proposed to play a vital role in tumor progression and metastasis. The mesenchymal phenotype is associated with the migratory property, and the epithelial phenotype is associated with invasion and tumor progression [56]. Most TICs are quiescent. However, a small population undergoing mesenchymal to epithelial transition undergoes intracellular biochemical changes, which leads to higher cellular proliferation. Studies have shown that CK-positive cells acquire an epithelial-mesenchymal hybrid state and exhibit high proliferation due to the activation of Smad signalling [58].

The differences in the proliferative state of the cell can be analyzed by understanding the essential biochemical changes. The TIC cells have a very low proliferative index where most of the cells are in the G0 phase of the cell cycle. The MIC cells are cells with a high proliferation index, with most of the cells in the S and G2 phases of cell cycle [59,60]. The representative SERS map of the single TIC cells ($CD44^+$, $CD133^+$, CK⁻) and individual MIC cells ($CD44^+$, $CD133^+$, CK⁺) is presented in Fig. 6.

Fig. 6a, b shows the distribution of RNA, lipid, protein in pancreatic tumor initiating cells (P- TIC), and pancreatic metastasis initiating cells (P- MIC). Fig. 6e, f shows the distribution of intracellular macromolecules RNA, lipid, protein in lung tumor initiating cells (L- TIC) and lung metastasis initiating cells (L- MIC). It can be observed that the relative amounts of nucleic acids and proteins are higher in P- MIC, confirming

the high proliferation status of tumorigenic P- TIC. A similar trend is observed in L- TIC and L- MIC. Hence, it can be generalized that tumorigenic TIC possesses a higher proliferative index compared to TIC [61,62].

The intracellular biochemical changes provide accurate information about the proliferation status of the cell. The quiescent cells possess high lipid content and low protein. However, cells in the exponential growth state consume lipids for cellular processes such as respiration, thereby lowering the total lipid content. Further, the decrease in lipid content could also be attributed to the change in the molecular structure of lipids [63].

Fig. 6c, f shows the ratiometric analysis of total intracellular protein and lipid in pancreatic cancer and lung cancer, respectively. It can be observed that the protein-lipid ratio is significantly higher in tumorigenic TIC when compared to TIC. Also, the ratio of lipids and RNA acts as a marker for determining the proliferative status of a cell [62]. Cancer is generally associated with the uncontrolled growth of cells. This process is highly related to enhanced protein synthesis, which in turn depends on the biogenesis of ribosome [64]. Recent studies have provided a connection between alterations in ribosome components and the loss or acquisition of TIC properties [65]. Besides, low protein synthesis is essential for TIC metabolism. It can be observed from Fig. 6a, b that the total RNA levels are high in MIC. Further, it can be concluded from Fig. d, h that the lipid - RNA ratio is significantly lower in MIC. Hence, the lipid - RNA ratio can be considered as a marker for the detection of this subpopulation of TIC.

2.7. Cellular phenotypic differentiation for detection of MIC

One of the primary reasons for a high proliferative state of tumorigenic TIC is its ability to regulate the process of EMT differentially. Unlike the majority of TICs which possess a mesenchymal phenotype, the MIC undergo the reverse process of EMT, thereby attaining an epithelial phenotype. The transformation of the mesenchymal phenotype to epithelial phenotype, with the maintenance of stem cell-like features, is a critical property of tumorigenic TIC. Here, we utilize the intrinsic molecular changes associated with both epithelial and mesenchymal phenotype to differentiate between TIC and MIC. The result is presented in Fig. 7.

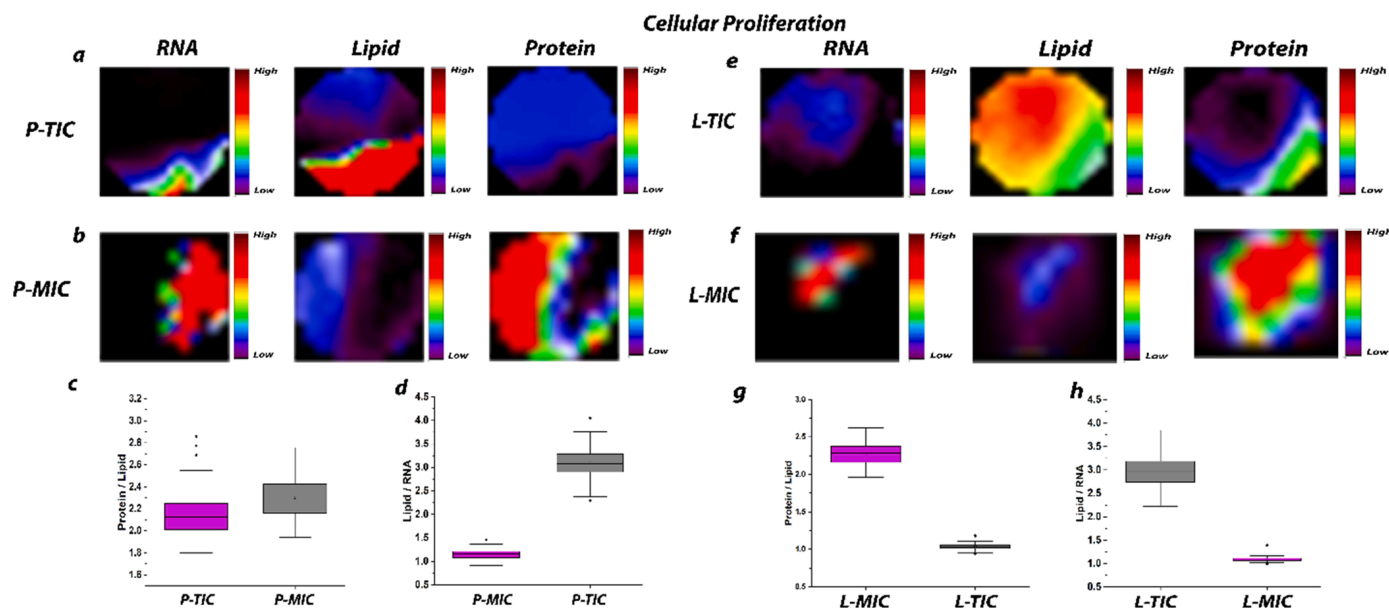


Fig. 6. Cellular proliferation as a marker to detect MIC. a) SERS image showcasing the differences in biochemical signature associated with cellular proliferation in P- TIC and P- MIC b) Box plot comparing the Protein - Lipid ratio in P- TIC and P- MIC ($p < 0.0001$) c) Box plot comparing the Lipid - RNA ratio in P- TIC and P- MIC ($p < 0.0001$) d) SERS image showcasing the differences in biochemical signature associated with cellular proliferation in L- TIC and L- MIC e) Box plot comparing the Protein - Lipid ratio in L- TIC and L- MIC ($p < 0.0001$) f) Box plot comparing the Lipid - RNA ratio in L- TIC and L- MIC ($p < 0.0001$) (Scale 20 μm).

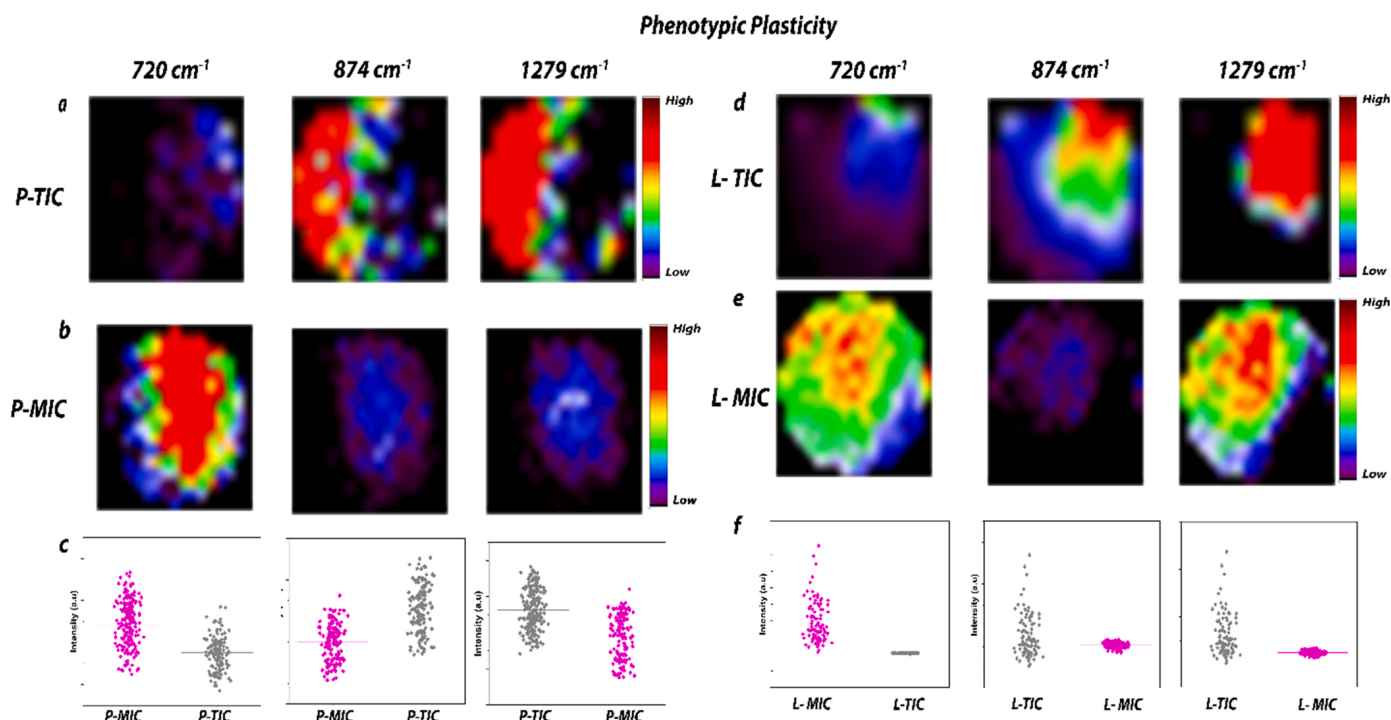


Fig. 7. Cellular phenotypic differentiation as a marker for the detection of MIC. SERS image of characteristic peak differences between epithelial and mesenchymal cells in a) L- TIC, b) L- MIC c) Relative intensity comparison between L- TIC and L- MIC d) P- TIC e) P- MIC f) Relative intensity comparison between P- TIC and P- MIC (Scale 20 μm).

The average spectra of TIC and MIC exhibited minor differences, showing both the cell types possess similar intracellular biochemical components. The primary differences between epithelial and mesenchymal phenotype lie in proteins and lipids [66]. The characterization of epithelial-mesenchymal transition by Raman spectroscopy has revealed that SERS peaks at 1279 cm^{-1} , 874 cm^{-1} , 720 cm^{-1} corresponding to triglycerides (Fatty acids), tryptophan (proteins), Ring breathing of adenine (DNA / RNA) shows a statistically significant difference between epithelial and mesenchymal phenotype [66–68].

The SERS image of single cells corresponding to nucleic acids is shown in Fig. 7a for L- TIC and L- MIC Fig. 6d for P- TIC and P- MIC. It can be observed that L- MIC shows a higher contribution of nucleic acids when compared to L- TIC cells. The SERS intensity comparison presented in Fig. 6b also corroborates the SERS image. Studies have shown a high DNA / RNA content in cells undergoing mesenchymal to epithelial transition, which is consistent with their ability to initiate secondary tumors through metastasis [27,69].

The SERS intensity corresponding to nucleic acids is a combination of genomic DNA, mitochondrial DNA, RNA [70,71]. It can be observed from the SERS image of MIC that the SERS peak corresponding to nucleic acids is distributed throughout the cell, as opposed to the TIC in which the SERS peak is confined to the nucleus. The distribution of DNA peaks could be attributed to the lower mitochondrial DNA in TIC owing to their quiescent nature [72]. Recently, studies have shown that mitochondrial DNA is susceptible to a high mutational rate compared to nuclear DNA due to the lack of repair mechanisms [73]. Moreover, tumorigenic mitochondrial DNA is known to mediate the process of metastatic initiation [74].

The SERS image of single cells corresponding to tryptophan is shown in Fig. 7b for L- TIC and L- MIC Fig. 7e for P- TIC and P- MIC. Tryptophan is one of the crucial amino acids essential for TIC survival in hypoxic conditions [75,76]. Further, intracellular depletion of tryptophan is essential for the maintenance of the TIC phenotype by activation of OCT 4 transcription [76]. It can be observed from the SERS image in Fig. 6b, e that the SERS intensity of the peak corresponding to tryptophan is lower in MIC when compared to TIC. Moreover, studies have proven that high

metabolite concentration is a characteristic element of EMT phenotype [68]. The relative intensity comparison analysis in Fig. 7d confirms the lower content of tryptophan in MIC cells.

Another striking difference between cells of the epithelial and mesenchymal phenotype is their fatty acid content. The SERS image of single cells corresponding to the distribution of fatty acids is shown in Fig. 7c for L- TIC and L- MIC Fig. 7f for P- TIC and P- MIC. Low tryptophan content, along with high fatty acid content, is associated with a highly aggressive metastatic phenotype [68,77]. It can be observed from the SERS image in Figs. 7c and 6f that the fatty acid content is higher tumorigenic TIC when compared to TIC, consistent with the metastatic nature of the cells. Besides, studies have proven that the increased lipid content might act as an alternative fuel source that is needed for the highly proliferative MIC [68]. The SERS image is further supported by the relative intensity comparison provided in Fig. 7. It can be observed that the median SERS intensity of fatty acids in MIC is higher than the median SERS intensity of TIC.

2.8. Analysing the tumor microenvironment for differentiation of MIC

The primary intracellular differences between MIC and TIC correspond to cellular proliferation status and the cell and DNA repair mechanisms' phenotype, presented in Figs. 5, 6, and 7. Their genotypic and phenotypic plasticity initiates the heterogeneity of TIC; minimal presence in the tumor sample makes it extremely difficult to perform biochemical characterization of TIC [78]. The intracellular components play an essential role in the differentiation of heterogenous TIC populations. Moreover, the tumor microenvironment plays a critical role in the determination and maintenance of TIC plasticity [79,80]. Studies have revealed that the secreted macromolecules play a crucial role in studying TIC interaction with its microenvironment [78,81]. The secreted macromolecules in the extracellular space include various proteins, including anti- survival proteins, growth factors, cytokines, cell motility factors [82]. These molecules and nuclear content secreted in the tumor microenvironment comprise the tumor secretome. For in vitro analysis, the conditioned media is used for secretome analysis.

The composition of the tumor secretome depends on de novo synthesis changes, the half-life of the metabolite, and the trafficking processes [83]. Moreover, the signaling molecules in the secretome are believed to modify the phenotype, thereby facilitating metastasis [84]. Studies have determined that deregulation of cellular energetics, invasion, and metastatic potential is regulated by growth factors (EGF, FGF, TGF- β) and metabolites in secretome [83]. Hence, the tumor secretome

can act as a biomarker for the detection of tumorigenic TIC amidst the heterogeneous TIC population. Fig. 8 presents the analysis of extracellular metabolites of TIC and MIC.

The conditioned media of TIC incubated with Self-functionalized nanosensors were used to obtain the extracellular metabolites' SERS spectra. The SERS spectra of cell culture media conditioned by L- TIC, L- MIC, P- TIC, P- MIC are presented in Fig. 8a, b, f, g, respectively. The

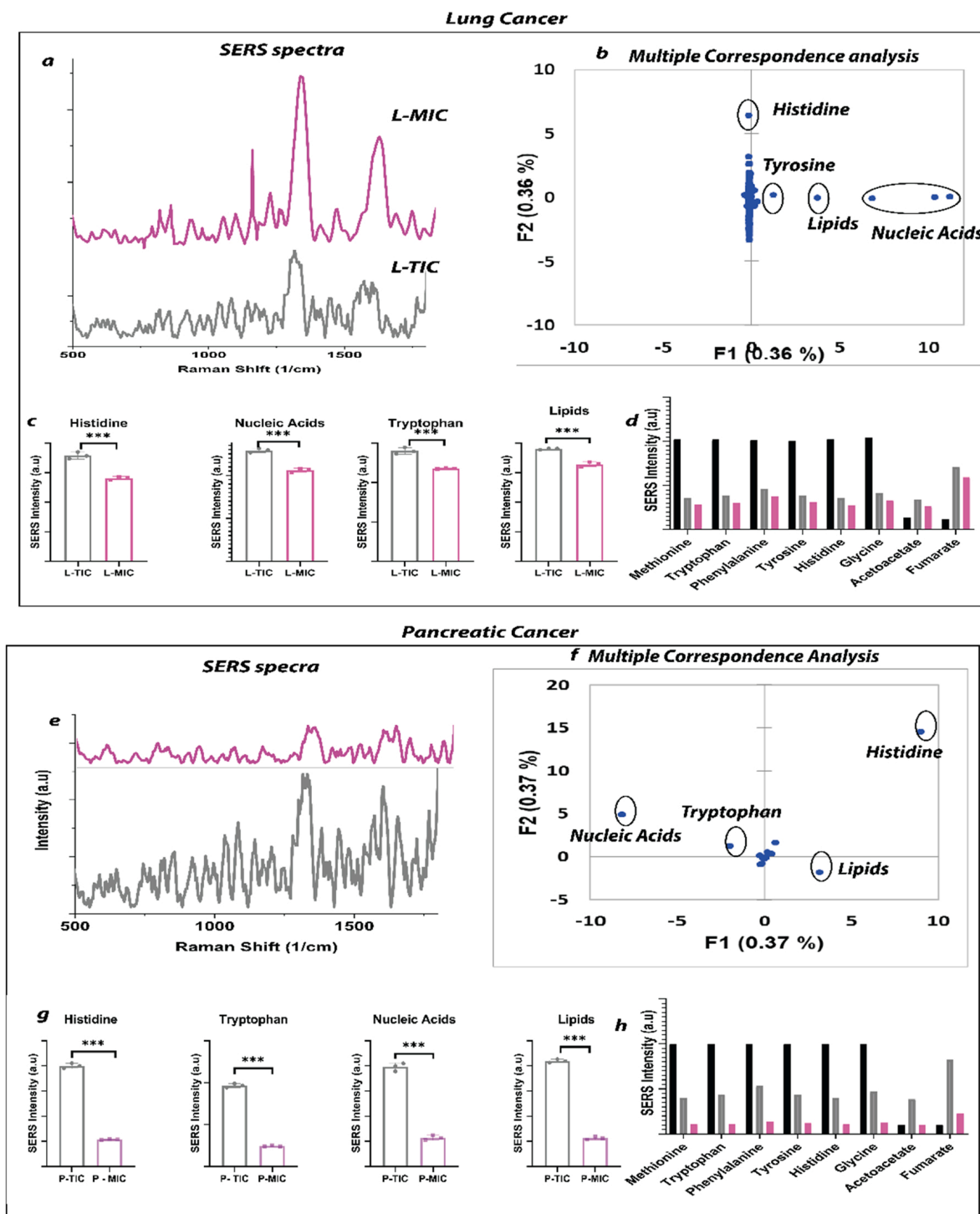


Fig. 8. Extracellular metabolite-based differentiation of MIC. SERS spectra of TIC conditioned media of a) L- TIC, L- MIC, b) Multiple correspondence analysis to pinpoint the difference between extracellular metabolites of L- TIC and L- MIC c) SERS intensity comparison of the characteristic peak differences between L- TIC and L- MIC. d) SERS intensity comparison between L- TIC and L- MIC to determine the nutrient consumption of cells in similar culture conditions. SERS spectra of TIC conditioned media of e) P- TIC, P- MIC f) Multiple correspondence analysis to pinpoint the difference between extracellular metabolites of P- TIC and P- MIC g) SERS intensity comparison of the characteristic peak differences between P- TIC and P- MIC, h) SERS intensity comparison between P- TIC and P- MIC to determine the nutrient consumption of cells in similar culture conditions.

unconditioned cell culture media was used as a control. It can be observed that the overall SERS intensity of L- TIC conditioned media is higher than the overall intensity of the L- MIC. The high SERS intensity can be attributed to the static nature of TIC. Due to the low proliferation rate, nutrient consumption is low, which leads to a high concentration of amino acids in the spent media. However, in MIC, the high proliferative rate leads to high nutrient consumption; hence the concentration is lower in conditioned media.

The comparative amino acid consumption rate of L- TIC, L- MIC, P- TIC, and P- MIC is presented in Fig. 8c, h, respectively. It should be observed that the SERS intensity of peaks corresponding to amino acids shows a decrease in intensity. The prominent peaks observed in TIC and MIC's conditioned media are 727 cm^{-1} , 828 cm^{-1} , 1004 cm^{-1} , 1180 cm^{-1} , 1244 cm^{-1} , 1488 cm^{-1} Methionine, tyrosine, phenylalanine, tyrosine, amide III (protein), histidine, respectively [85]. The peaks increasing in intensity are 1158 cm^{-1} , 1291 cm^{-1} , corresponding to acetoacetate, fumarate, respectively [86].

It can be observed from Fig. 8c, h that the decrease in intensity of tyrosine peak is accompanied by an increase in metabolite peaks, including acetoacetate and fumarate, by-products of the tyrosine degradation pathway. The presence of tyrosine by-products indicates that tyrosine was consumed at a higher rate by TICs. In contrast, MICs show lower levels of tyrosine degradation. Recent studies have proven that quiescent TICs use a unique metabolic switch from glycolysis to tyrosine degradation for survival [87,88]. When TIC is active, lower consumption of tyrosine, indicating the switch to glycolysis when the active TIC is the MIC. so we can use the degradation of tyrosine to distinguish MICs from TIC.

We used Multiple Correspondence Analysis (MCA) to unscramble the TIC conditioned media's molecular information. The SERS spectra of the TIC conditioned media after baseline correction was used to perform MCA analysis, which differentiated the metabolites that differed within the cell groups. The score plots corresponding to the components are shown in Fig. 8b, f for L- TIC and L- MIC, P- TIC, and P- MIC, respectively. It can be observed that most extracellular metabolites form a cluster around the center. However, the components with a statistically significant difference ($p < 0.0001$) were found separately from the cluster. These components formed specific differences between TIC and MIC. It can further be observed from Fig. 8b and f that both L- TIC and L- MIC, P- TIC, and P- MIC show common differences in secreted metabolites corresponding to nucleic acids, histidine, lipids. However, the difference between TICs of pancreatic origin and lung origin can be attributed to tyrosine degradation and tryptophan degradation.

2.9. Application of self- functionalized nanosensors to differentiate MIC in a heterogeneous spheroid population

Following the characterization and detection of MIC using intracellular biological process and secretome characterization, the quantum probe's ability to provide reproducible cellular detection was investigated. The results are presented in Fig. S2. Once the Self- functionalized nanosensors is internalized in the cell, the probe's location determines the SERS sensitivity and reproducibility [89]. Fig. S3a, S3c, S3e, S3g shows the SERS spectra of 5 different cells from L- TIC, L- MIC, P- TIC, and P- MIC cells. The spectra show a comparable and repeatable spectrum. Further, the SERS intensity comparison from 10 different single

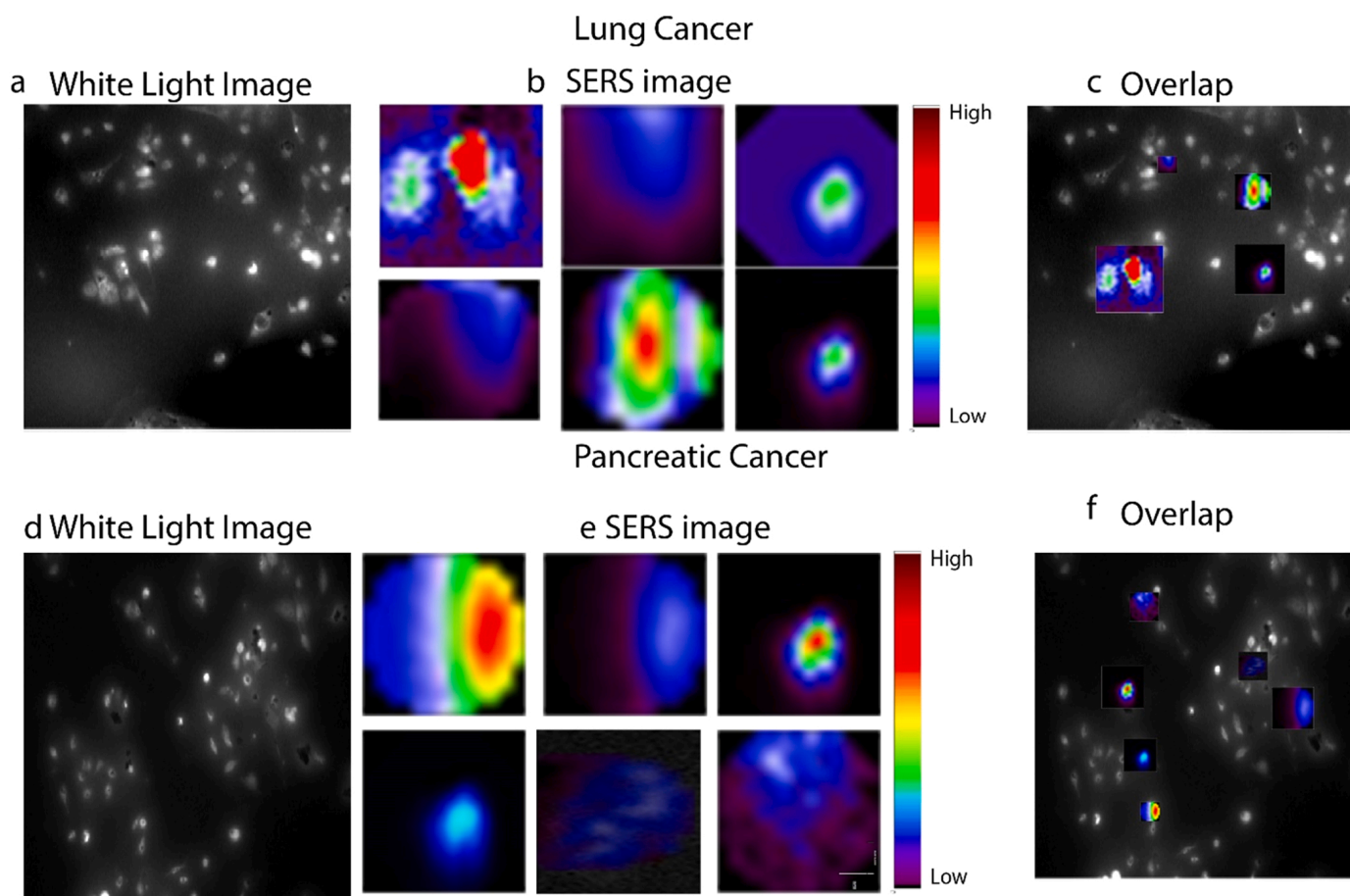


Fig. 9. Application of Self- functionalized nanosensors to differentiate TIC and MIC in a heterogeneous TIC spheroid population. White light image of heterogenous TIC spheroid derived from a) Lung cancer d) Pancreatic cancer. SERS image of cells from heterogenous TIC spheroid derived from b) Lung cancer e) Pancreatic cancer. Overlap of SERS and white light image to show differentiation of TIC and MIC in a heterogeneous population of cells in c) lung cancer f) Pancreatic Cancer. (Scale 20 μm).

cells shows intensity variation with a low percentage of relative standard deviation. It can be observed that the relative change in intensity is below 3% for both TIC and MIC.

After confirming the molecular and microenvironment differences between TIC and MIC, we applied the Self-functionalized nanosensors to differentiate TIC and MIC in an unsorted heterogeneous population of cells.

The SERS images of the representative cells are provided in Fig. 9. The white light image indicates the area where SERS mapping analysis was performed (Fig. 9a, e). By applying the intracellular markers discerning TIC from MIC, presented in section 2.4.2- 2.4.3, the SERS image was obtained. The Raman peak intensity is color-coded. The ones associated with MIC are coded with red and TIC with blue. Fig. 9b, f shows the SERS map of a selected few cells from the unsorted heterogeneous population. Most of the cells are TIC, and only a few are MIC.

2.10. Prediction of cancer metastasis through cellular phenotype using machine learning

To further validate the reliability of the Self-functionalized nanosensors to detect and differentiate MIC from TIC in a heterogeneous population a simple machine learning algorithm was applied to analyse the SERS spectra obtained from the TIC and MIC. The SERS spectral data obtained from the heterogeneous cell population was used as the test data to validate the machine learning algorithms if it was possible to discern the spectral features with high accuracy, sensitivity, specificity [90]. The spectral signature sorted population of TIC and MIC was used as the training data. The spectral signature of the unsorted heterogeneous population was used as the test data. Based on the established PLS-R model, the two-sample groups (TIC and MIC) were discernable with 98% specificity and sensitivity. The posterior probability of the cell to be classified as MIC are presented in 10 c, 10 d corresponding to pancreatic

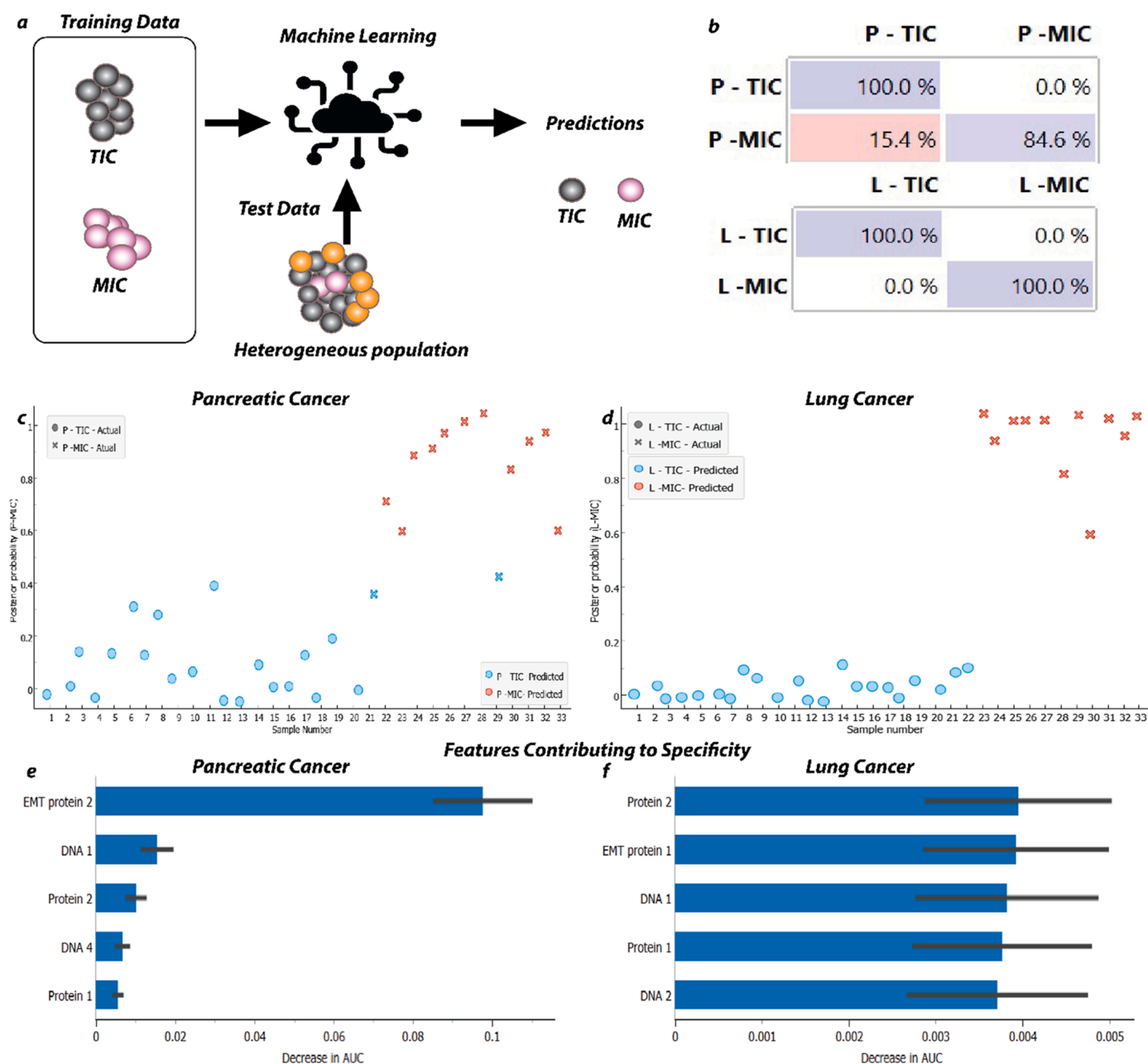


Fig. 10. Prediction of cancer metastasis through cellular phenotype using machine learning a) Schematic of prediction of metastatic potential using metastasis initiating cell as a marker b) Confusion matrix showing the specificity of prediction, Prediction of metastasis initiating cell in heterogeneous population c) Pancreatic cancer d) lung cancer, Effect of phenotypic features on prediction specificity in e) Pancreatic cancer f) lung cancer.

cancer and lung cancer, respectively. The threshold for MIC was determined to be 0.5. It can be observed from graph in Fig. 10e that the prediction specificity of MIC in pancreatic cancer is highly dependent on the SERS peak positions to EMT proteins, adenine, guanine vibrations of DNA and tryptophan vibrations of proteins. Similar observations were made in lung cancer (Fig. 10f) as well. Hence, it can be confirmed that the intracellular processes and macromolecular differences are reliable markers to predict the presence of metastasis initiating cells, hence can be used to predict the metastatic potential of a tumor.

2.11. Prediction of cancer metastasis from different tissue of origin

We also analyzed the efficiency of Self- functionalized nanosensors in differentiating TIC and MIC of different origins. Fig. 11a showcases the schematic representation of the prediction algorithm. The TIC and MIC from lung cancer and pancreatic cancer were used to train the algorithm.

The SERS signature of the heterogenous population were used as the test data. It can be observed from Fig. 11b that the self- functionalized nanosensors were able to find the tissue of origin of the cancer with a high accuracy and specificity 99.62%. Further, the EMT transition plays a major role in determining the prediction accuracy and specificity.

Next, we aimed to use the algorithm to differentiate between cell types from different tissue of origin. It can be observed from Fig. 11c that the cell types from different tissue of origin do not overlap, however we observed one misclassification between pancreatic TIC and MIC, which could be attributed to transition state of the cells. The analysis of phenotypic features contributing most to the classification showed that EMT transition plays a predominant role in determining the metastatic potential of the cells. Taken together, here we have demonstrated a new intracellular biomarker for early detection and prediction of cancer metastasis with high sensitivity, specificity and accuracy.

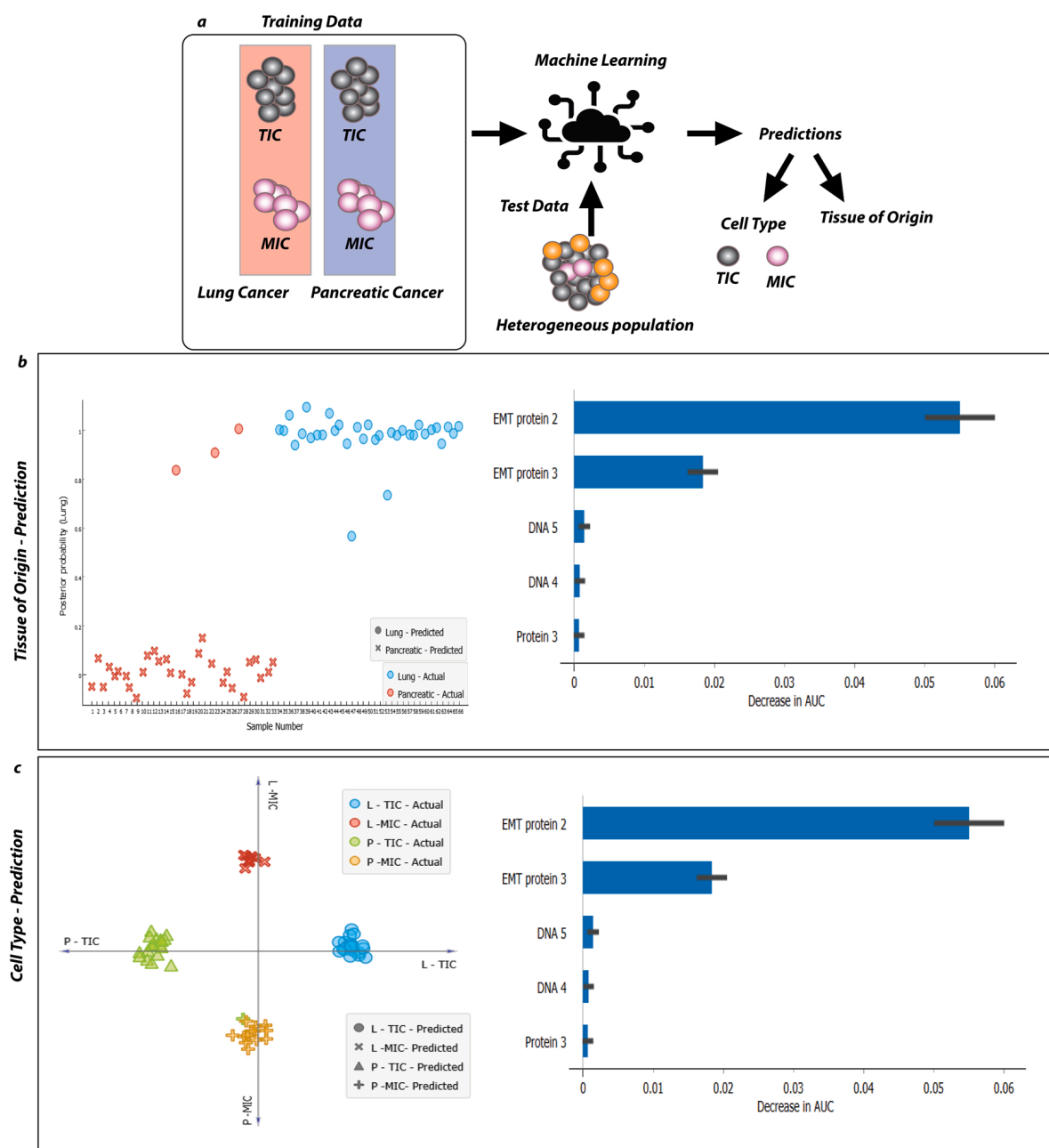


Fig. 11. Prediction of cancer metastasis from different tissue of origin a) schematic representation of cancer tissue of origin through MIC as a marker b) Prediction of cancer tissue of origin using MIC as a marker, phenotypic features contributing most to the prediction accuracy c) Prediction of cell type in a heterogenous population, using the phenotypic and intracellular features of TIC and MIC, phenotypic features contributing most to the prediction accuracy.

3. Conclusion

Here, we have identified a rare subset of cells within the tumor, which possess the early signal for cancer metastasis, known as metastasis-initiating cells (MICs). MICs are present in most metastatic cancer types, unaffected by epigenetic makeup changes, thereby becoming an excellent, rational marker for cancer metastasis irrespective of cancer type, the tissue of origin, and cancer stage.

To overcome the undetectable nature of MIC, here we have designed a self-functionalized nanosensor with the ultra-sensitivity necessary for the early detection and prediction of cancer metastasis using MICs. The SERS-activated nanosensor enabled MIC detection using properties of intracellular biological processes and their interaction in the tumor microenvironment. The probe enabled a detection sensitivity of 98%. It was able to identify MIC in a heterogeneous population of TICs with a high specificity of 99.62%. Further, applying machine learning algorithms for the heterogeneous population yielded a prediction accuracy of 89.3%.

To the best of our knowledge, this study is the first to introduce a probe that can provide both the diagnostic signature and predictive signature of cancer metastasis as early as the single cellular stage. The ability to detect and predict cancer metastasis at an early stage dramatically improves metastatic cancer's diagnostic potential. Targeting specific organs for treatment will provide a strong foundation for the development of effective metastatic therapies. This approach holds immense potential in the early diagnosis of metastatic tumors and provides a better understanding of metastasis's mechanism by providing unique insights in a clinical diagnostic setting.

4. Materials and methods

4.1. Synthesis of Self-functionalized nanosensors

The Self-functionalized nanosensors was synthesized using a Clark-MXR IMPULSE pulsed Yb-doped fiber-amplified ultrashort pulsed laser to ionize a graphite substrate. The laser conditions were maintained constant at a wavelength of 1030 nm with a circular polarization. The average laser power of 17 W and the laser pulse width of 214 fs was used for the fabrication of Self-functionalized nanosensors. The graphite substrate was secured on a piezo-driven system with a repetition of lines designed using EzCAD software. The pulse repetition rate of 4 MHz was used for generating the Self-functionalized nanosensors. To efficiently manipulate the laser plume dynamics low pressure oxygen gas was introduced in the environment. The introduced oxygen gas was evenly contained using separate nozzles in the ionization zone at a constant flow rate.

4.2. Characterization of self-functionalized nanosensors

The TEM image showcasing the morphology of Self-functionalized nanosensors was obtained using a Hitachi H-7000 HRTEM on copper mesh grids. The images acquired were used to determine the size distribution of Self-functionalized nanosensors. The size distribution of Self-functionalized nanosensors was obtained by employing ImageJ software. The XPS measurement was acquired using a Thermo fisher K alpha XPS system with an Al K α X-ray source. The quantification analysis was implemented using Avantage software.

The optical characterization Raman spectra was attained using Renishaw InVia Confocal Raman spectrometer equipped with a Leica DMI6000 epifluorescence microscope. The wavelength of 785 nm was used to obtain Raman spectra for characterizing Self-functionalized nanosensors. A Perkin-Elmer spectrometer fitted with a thermo scientific iS50 ATR accessory was used to perform ATR-FTIR analysis.

4.3. Cell culture

ASPC-1 (Pancreatic Cancer), H69-AR (Lung Cancer) cells obtained from ATCC were cultured separately using RPMI containing 10% FBS, 1% pen/strep at 37 °C in 5% CO₂ atmosphere. The cells were passaged in the ratio of 1:3 on reaching confluence. The cells were collected using Trypsin-EDTA solution, centrifuged at 1000 rpm for 5 min to pellet the cells. The sphere formation media from MammoCult medium (Stem Cell Technologies) was used to generate CSC spheroids from the individual cells.

4.4. Flow cytometric cell sorting

Antibodies for cell surface markers CD 44, CD 133, Cytokeratin were obtained from Miltenyi biotec. Cell surface marker staining was performed in DMEM / 5% FBS. The cells stained with the cell surface markers were kept on ice for 15 min. The stained cells were then washed to remove unbound cell surface markers, and then resuspended in cell sorting buffer. Flow cytometric sorting was performed using a MoFlo – Astrios Cell sorter. The analysis was performed using FCS express 7 software. Forward Scattering – height vs forward scatter width and side scatter height vs side scatter width were applied to eliminate cell aggregates. Dead cells were excluded by DAPI staining, which enabled efficient single live cells to be employed for further analysis.

4.5. Cell uptake study

The CSC spheroids were incubated with Self-functionalized nanosensors in a Petri dish with 3 mL of media for 24 h. The cells sorted using flow cytometry were separately incubated with Self-functionalized nanosensors, which was then subjected to flow cytometry (BD Biosciences) to characterize the uptake percentage of Self-functionalized nanosensors based on the characteristic fluorescence of the probe.

4.6. Cellular SERS analysis

The cells were incubated with Self-functionalized nanosensors for 24 h. The spent media was aspirated. The cells washed with 1% PBS three times to remove any trace media present and incubated 4 mL of 4% paraformaldehyde solution for 20 min at room temperature. After incubation, the paraformaldehyde solution is aspirated, and the cells were washed with PBS and then air-dried before SERS analysis. The Raman spectra of cells were collected using a Renishaw inVia Confocal Microscope with a 20X magnifying lens. The wavelength was maintained at 785 nm. The acquisition time was kept constant at 10 s; 3 acquisitions were made for each spectrum to ensure the reproducibility and uniformity of signals. Single-cell Raman mapping was done on cells fixed for Raman analysis. The mapping was performed in the Renishaw inVia confocal microscope with a 20X magnification. The map obtained for a single cell with a step size of 0.1 × 0.1.

4.7. Statistical analysis

All experiments were carried out in triplicates. The data points were averaged unless and otherwise stated. The error bars show standard deviation. WiRE 3.3 software was used to process the raw spectral data. The processed spectral data was then employed to determine the differences between TIC and MIC. All the statistical analysis was performed with the in-built package in the XLSTAT software[91]. The training dataset was obtained from the SERS spectra of the FACS sorted cells. SERS spectra from 20 cells were used to perform machine learning analysis. Here we utilized the Support vector machine model to test the accuracy of detection. The spectral signature sorted population of TIC and MIC was used as the training data. The spectral signature of the unsorted heterogeneous population was used as the test data. test the differences between the TIC and MIC, a student t-test was conducted.

The sample size ($n = 20$) was maintained throughout unless otherwise specified. The p -values < 0.05 was statistically significant.

CRedit authorship contribution statement

Swarna Ganesh: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft, Validation, Formal analysis, Visualization, Software, **Krishnan Venkatakrishnan:** Conceptualization, Methodology, Writing – review & editing, Resources, Supervision, Project Management, Funding acquisition, **Bo Tan:** Conceptualization, Methodology, Writing – review & editing, Resources, Supervision, Project Management, Funding acquisition.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Funding: This research was funded by NSERC Discovery Grant 132950, 134361.

Author contributions: S.G, K.V, B.T worked together in designing the project, S.G performed the experiments and wrote the manuscript, K.V. and B.T assisted in results, discussion and edited the manuscript.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.snb.2022.131655.

References

- [1] H. Dillekås, M.S. Rogers, O. Straume, Are 90% of deaths from cancer caused by metastases? *Cancer Med* 8 (2019) 5574–5576, <https://doi.org/10.1002/cam4.2474>.
- [2] L.W, Q.Z. Linlin Yuan, Fei Guo, Prediction of tumor metastasis from sequencing data in the era of genome sequencing, *Brief. Funct. Genom.* 18 (2019) 412–418.
- [3] J. Fares, M.Y. Fares, H.H. Khachfe, H.A. Salhab, Y. Fares, Molecular principles of metastasis: a hallmark of cancer revisited, *Signal Transduct. Target. Ther.* 5 (2020), <https://doi.org/10.1038/s41392-020-0134-x>.
- [4] M. Zhang, Early detection of multi-organ metastases, *Nat. Biomed. Eng.* 1 (2017) 934–936, <https://doi.org/10.1038/s41551-017-0173-y>.
- [5] B.V.S.K. Chakravarthi, S. Nepal, S. Varambally, Genomic and epigenomic alterations in cancer, *Am. J. Pathol.* 186 (2016) 1724–1735, <https://doi.org/10.1016/j.ajpath.2016.02.023>.
- [6] Y. Shi, C. Liu, X. Liu, D.G. Tang, J. Wang, The microRNA miR-34a inhibits non-small cell lung cancer (NSCLC) growth and the CD44hi stem-like NSCLC cells, *PLoS One* 9 (2014) 1–8, <https://doi.org/10.1371/journal.pone.0090022>.
- [7] Q. Huang, J. Yan, R. Agami, Long non-coding RNAs in metastasis, *Cancer Metastasis Rev.* 37 (2018) 75–81, <https://doi.org/10.1007/s10555-017-9713-x>.
- [8] M. Kim, J. Costello, DNA methylation: an epigenetic mark of cellular memory, *Exp. Mol. Med.* 49 (2017) e322–e328, <https://doi.org/10.1038/emm.2017.10>.
- [9] D.S. Micalizzi, S. Maheswaran, D.A. Haber, A conduit to metastasis: circulating tumor cell biology, *Genes Dev.* 31 (2017) 1827–1840, <https://doi.org/10.1101/gad.305805.117>.
- [10] M.J. Jahid, T.H. Huang, J. Ruan, A personalized committee classification approach to improving prediction of breast cancer metastasis, *Bioinformatics* 30 (2014) 1858–1866, <https://doi.org/10.1093/bioinformatics/btu128>.
- [11] N.L. Teradal, R. Jelinek, Carbon nanomaterials in biological studies and biomedicine, *Adv. Healthc. Mater.* 6 (2017) 1–36, <https://doi.org/10.1002/adhm.201700574>.
- [12] T.A. Tabish, M.Z.I. Pranjol, D.W. Horsell, A.A.M. Rahat, J.L. Whatmore, P. G. Winyard, S. Zhang, Graphene oxide-based targeting of extracellular cathepsin D and Cathepsin L as a Novel anti-metastatic enzyme cancer therapy, *Cancers* 11 (2019), <https://doi.org/10.3390/cancers11030319>.
- [13] J.A. Powell, K. Venkatakrishnan, B. Tan, Towards universal SERS detection of disease signalling bioanalytes using 3D self-assembled non-plasmonic near-quantum scale silicon probe, *ACS Appl. Mater. Interfaces* (2017), <https://doi.org/10.1021/acsami.7b15393>.
- [14] N.G. Semaltianos, Nanoparticles by laser ablation, *Crit. Rev. Solid State Mater. Sci.* 35 (2010) 105–124, <https://doi.org/10.1080/10408431003788233>.
- [15] A. Bogaerts, Z. Chen, D. Bleiner, Laser ablation of copper in different background gases: comparative study by numerical modeling and experiments, *J. Anal. Spectrom.* 21 (2006) 384–395, <https://doi.org/10.1039/b514313f>.
- [16] A. Bratt, A.R. Barron, XPS of carbon nanomaterials, *Measurement* (2011) 1–16.
- [17] C. Zhang, D.M. Dabbs, L.M. Liu, I.A. Aksay, R. Car, A. Selloni, Combined effects of functional groups, lattice defects, and edges in the infrared spectra of graphene oxide, *J. Phys. Chem. C* 119 (2015) 18167–18176, <https://doi.org/10.1021/acs.jpcc.5b02727>.
- [18] M. Acik, G. Lee, C. Mattevi, M. Chhowalla, K. Cho, Y.J. Chabal, Unusual infrared-absorption mechanism in thermally reduced graphene oxide, *Nat. Mater.* 9 (2010) 840–845, <https://doi.org/10.1038/nmat2858>.
- [19] E. Fuente, J.A. Menéndez, M.A. Díez, D. Suárez, M.A. Montes-Morán, Infrared spectroscopy of carbon materials: a quantum chemical study of model compounds, *J. Phys. Chem. B* 107 (2003) 6350–6359, <https://doi.org/10.1021/jp027482g>.
- [20] A.C. Ferrari, J.C. Meyer, V. Scardaci, C. Casiraghi, M. Lazzeri, F. Mauri, S. Piscanec, D. Jiang, K.S. Novoselov, S. Roth, A.K. Geim, Raman spectrum of graphene and graphene layers, *Phys. Rev. Lett.* 97 (2006) 1–4, <https://doi.org/10.1103/PhysRevLett.97.187401>.
- [21] C. Castiglioni, C. Mapelli, F. Negri, G. Zerbi, Origin of the D line in the Raman spectrum of graphite: a study based on Raman frequencies and intensities of polycyclic aromatic hydrocarbon molecules, *J. Chem. Phys.* 114 (2001) 963–974, <https://doi.org/10.1063/1.1329670>.
- [22] J. Lin, X. Chen, P. Huang, Graphene-based nanomaterials for bioimaging, *Adv. Drug Deliv. Rev.* 105 (2016) 242–254, <https://doi.org/10.1016/j.addr.2016.05.013>.
- [23] H. Zheng, Y. Kang, Multilayer control of the EMT master regulators, *Oncogene* 33 (2014) 1755–1763, <https://doi.org/10.1038/onc.2013.128>.
- [24] B. De Craene, G. Berx, Regulatory networks defining EMT during cancer initiation and progression, *Nat. Rev. Cancer* 13 (2013) 97–110, <https://doi.org/10.1038/nrc3447>.
- [25] J.H. Tsai, J. Yang, Epithelial-mesenchymal plasticity in carcinoma metastasis, *Genes Dev.* 27 (2013) 2192–2206, <https://doi.org/10.1101/gad.225334.113>.
- [26] M. Takaishi, M. Tarutani, J. Takeda, S. Sano, Mesenchymal to epithelial transition induced by reprogramming factors attenuates the malignancy of cancer cells, *PLoS One* 11 (2016) 1–15, <https://doi.org/10.1371/journal.pone.0156904>.
- [27] N.P.A.D. Gunasinghe, A. Wells, E.W. Thompson, H.J. Hugo, Mesenchymal-epithelial transition (MET) as a mechanism for metastatic colonisation in breast cancer, *Cancer Metastasis Rev.* 31 (2012) 469–478, <https://doi.org/10.1007/s10555-012-9377-5>.
- [28] S. Bhatia, J. Monkman, T. Blick, P.H. Duijff, S.H. Nagaraj, E.W. Thompson, Multi-omics characterization of the spontaneous mesenchymal–epithelial transition in the PMC42 breast cancer cell lines, *J. Clin. Med.* 8 (2019) 1253, <https://doi.org/10.3390/jcm8081253>.
- [29] A. Koren, M. Rijavec, I. Kern, E. Sodja, P. Korosec, T. Cufer, BMI1, ALDH1A1, and CD133 transcripts connect epithelial-mesenchymal transition to cancer stem cells in lung carcinoma, *Stem Cells Int.* 2016 (2016), <https://doi.org/10.1155/2016/9714315>.
- [30] V. Neumeister, S. Agarwal, J. Bordeaux, R.L. Camp, D.L. Rimm, In situ identification of putative cancer stem cells by multiplexing ALDH1, CD44, and cytokeratin identifies breast cancer patients with poor prognosis, *Am. J. Pathol.* 176 (2010) 2131–2138, <https://doi.org/10.2353/ajpath.2010.090712>.
- [31] N.P. de Andrade, M.F.S.D. Rodrigues, C.O. Rodini, F.D. Nunes, Cancer stem cell, cytokeratins and epithelial to mesenchymal transition markers expression in oral squamous cell carcinoma derived from orthotopic xenoinplantation of CD44high cells, *Pathol. Res. Pract.* 213 (2017) 235–244, <https://doi.org/10.1016/j.prp.2016.12.009>.
- [32] A. Eramo, F. Lotti, G. Sette, E. Pilozi, M. Biffoni, A. Di Virgilio, C. Conticello, L. Ruco, C. Peschle, R. De Maria, Identification and expansion of the tumorigenic lung cancer stem cell population, *Cell Death Differ.* 15 (2008) 504–514, <https://doi.org/10.1038/sj.cdd.4402283>.
- [33] M.K. Jolly, M. Boareto, B. Huang, D. Jia, M. Lu, H. Levine, E. Ben-jacob, C.S. Cells, C. Systems, Implications of the hybrid epithelial / mesenchymal phenotype in metastasis Running title: clusters of hybrid E/M cells: primary ‘bad actors’ of metastasis, *Front. Oncol.* (2015).
- [34] U.D. Kahlert, J.V. Joseph, F.A.E. Kruyt, EMT- and MET-related processes in nonepithelial tumors: importance for disease progression, prognosis, and therapeutic opportunities, *Mol. Oncol.* 11 (2017) 860–877, <https://doi.org/10.1002/1878-0261.12085>.
- [35] B. Drasler, D. Vanhecke, L. Rodriguez-Lorenzo, A. Petri-Fink, B. Rothen-Rutishauser, Quantifying nanoparticle cellular uptake: which method is best? *Nanomedicine* 12 (2017) 1095–1099, <https://doi.org/10.2217/nmm-2017-0071>.
- [36] L. Ju, B. Geng, J. Horng, C. Girit, M. Martin, Z. Hao, H.A. Bechtel, X. Liang, A. Zettl, Y.R. Shen, F. Wang, Graphene plasmonics for tunable terahertz metamaterials, *Nat. Nanotechnol.* 6 (2011) 630–634, <https://doi.org/10.1038/nnano.2011.146>.
- [37] H. Medina, Y.C. Lin, D. Obergfell, P.W. Chiu, Tuning of charge densities in graphene by molecule doping, *Adv. Funct. Mater.* 21 (2011) 2687–2692, <https://doi.org/10.1002/adfm.201100401>.
- [38] Kenry, A. Geldert, Y. Liu, K.P. Loh, C. Teck Lim, Nano-bio interactions between carbon nanomaterials and blood plasma proteins: why oxygen functionality matters, *NPG Asia Mater.* 9 (2017), <https://doi.org/10.1038/am.2017.129> (e422–e422).
- [39] V.C. Sanchez, A. Jachak, R.H. Hurt, A.B. Kane, Biological interactions of graphene-family nanomaterials: an interdisciplinary review, *Chem. Res. Toxicol.* 25 (2012) 15–34, <https://doi.org/10.1021/tx200339h>.
- [40] W.A. El-Said, T.H. Kim, H. Kim, J.W. Choi, Analysis of intracellular state based on controlled 3D nanostructures mediated surface enhanced Raman scattering, *PLoS One* 6 (2011), <https://doi.org/10.1371/journal.pone.0015836>.

- [41] K. Bando, N.I. Smith, J. Ando, K. Fujita, S. Kawata, Analysis of dynamic SERS spectra measured with a nanoparticle during intracellular transportation in 3D, *J. Opt.* 17 (2015), 114023, <https://doi.org/10.1088/2040-8978/17/11/114023>.
- [42] S. Skvortsov, P. Debbage, P. Lukas, I. Skvortsova, Crosstalk between DNA repair and cancer stem cell (CSC) associated intracellular pathways, *Semin. Cancer Biol.* 31 (2015) 36–42, <https://doi.org/10.1016/j.semcancer.2014.06.002>.
- [43] T.B. Toh, J.J. Lim, E.K.H. Chow, Epigenetics in cancer stem cells, *Mol. Cancer* 16 (2017) 1–20, <https://doi.org/10.1186/s12943-017-0596-9>.
- [44] C. Blanpain, M. Mohrin, P.A. Sotiropoulou, E. Passegué, DNA-damage response in tissue-specific and cancer stem cells, *Cell Stem Cell* 8 (2011) 16–29, <https://doi.org/10.1016/j.stem.2010.12.012>.
- [45] L. Liu, A.V. Salnikov, N. Bauer, E. Aleksandrowicz, S. Labsch, C. Nwaeburu, J. Mattem, J. Gladkich, P. Schemmer, J. Werner, I. Herr, Triptolide reverses hypoxia-induced epithelial-mesenchymal transition and stem-like features in pancreatic cancer by NF- κ B downregulation, *Int. J. Cancer* 134 (2014) 2489–2503, <https://doi.org/10.1002/ijc.28583>.
- [46] Y. Yan, P.M. Greer, P.T. Cao, R.H. Kolb, K.H. Cowan, RAC1 GTPase plays an important role in γ -irradiation induced G2/M checkpoint activation, *Breast Cancer Res.* 14 (2012) 1–14, <https://doi.org/10.1186/bcr3164>.
- [47] S.R. Panikkanvalappil, M.A. Mahmoud, M.A. MacKey, M.A. El-Sayed, Surface-enhanced raman spectroscopy for real-time monitoring of reactive oxygen species-induced DNA damage and its prevention by platinum nanoparticles, *ACS Nano* 7 (2013) 7524–7533, <https://doi.org/10.1021/nn403722x>.
- [48] I. Skvortsova, S. Skvortsov, U. Raju, T. Stasyk, O. Riesterer, E.M. Schottdorf, B. A. Popper, B. Schiestl, P. Eichberger, P. Debbage, A. Neher, G.K. Bonn, L.A. Huber, L. Milas, P. Lukas, Epithelial-to-mesenchymal transition and c-myc expression are the determinants of cetuximab-induced enhancement of squamous cell carcinoma radioresponse, *Radiother. Oncol.* 96 (2010) 108–115, <https://doi.org/10.1016/j.radonc.2010.04.017>.
- [49] M.H. Yang, S.Y. Chang, S.H. Chiou, C.J. Liu, C.W. Chi, P.M. Chen, S.C. Teng, K. J. Wu, Overexpression of NBS1 induces epithelial-mesenchymal transition and co-expression of NBS1 and Snail predicts metastasis of head and neck cancer, *Oncogene* 26 (2007) 1459–1467, <https://doi.org/10.1038/sj.onc.1209929>.
- [50] A. Turdo, V. Veschi, M. Gaggiani, A. Chinnici, P. Bianca, M. Todaro, G. Stassi, Meeting the challenge of targeting cancer stem cells, *Front. Cell Dev. Biol.* 7 (2019) 1–16, <https://doi.org/10.3389/fcell.2019.00016>.
- [51] E. Lagasse, Cancer stem cells with genetic instability: the best vehicle with the best engine for cancer, *Gene Ther.* 15 (2008) 136–142, <https://doi.org/10.1038/sj.gt.3303068>.
- [52] B.C. Prager, Q. Xie, S. Bao, J.N. Rich, Cancer stem cells: the architects of the tumor ecosystem, *Cell Stem Cell* 24 (2019) 41–53, <https://doi.org/10.1016/j.stem.2018.12.009>.
- [53] G. Driessens, B. Beck, A. Caauwe, B.D. Simons, C. Blanpain, Defining the mode of tumour growth by clonal analysis, *Nature* 488 (2012) 527–530, <https://doi.org/10.1038/nature11344>.
- [54] H.C. Arnout G. Schepers, * Hugo J. Snippert, **† Daniel E. Stange, Maaïke van den Born, Johan H. van Es, Marc van de Wetering, Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas, 337, (2012), 730–736.
- [55] R.K. Nimmakayala, S.K. Batra, M.P. Ponnusamy, Unraveling the journey of cancer stem cells from origin to metastasis, *Biochim. Biophys. Acta Rev. Cancer* 1871 (2019) 50–63, <https://doi.org/10.1016/j.bbcan.2018.10.006>.
- [56] H. Wang, J.J. Unternaehrer, Epithelial-mesenchymal transition and cancer stem cells: at the crossroads of differentiation and dedifferentiation, *Dev. Dyn.* 248 (2019) 10–20, <https://doi.org/10.1002/dvdy.24678>.
- [57] T. Brabletz, EMT and MET in metastasis: where are the cancer stem cells? *Cancer Cell* 22 (2012) 699–701, <https://doi.org/10.1016/j.ccr.2012.11.009>.
- [58] T. Kawai, K. Yasuchika, T. Ishii, H. Katayama, E.Y. Yoshitoshi, S. Ogiso, S. Kita, K. Yasuda, K. Fukumitsu, M. Mizumoto, E. Hatano, S. Uemoto, Biology of human tumors Keratin 19, a cancer stem cell marker in human hepatocellular carcinoma, *Clin. Cancer Res.* 21 (2015), <https://doi.org/10.1158/1078-0432.CCR-14-1936>.
- [59] T. Celià-Terrassa, Y. Kang, Distinctive properties of metastasis-initiating cells, *Genes Dev.* 30 (2016) 892–908, <https://doi.org/10.1101/gad.277681.116>.
- [60] E. Aguilar, I.D.E. Marin Mas, E. Zodda, S. Marin, F. Morrish, V. Selivanov, S. S. Meca-corté, H. Delowar, N. Pons, I.S. Izquierdo, T. Celià, P.D.E. Atauri, J. J. Centelles, D. Hockenbery, T.M. Thomson, M. Cascante, Metabolic reprogramming and dependencies associated with epithelial cancer stem cells independent of the epithelial-mesenchymal transition program, *Stem Cells* 34 (2016) 1163–1176, <https://doi.org/10.1002/stem.2286>.
- [61] Y. Chen, J. Dai, X. Zhou, Y. Liu, W. Zhang, G. Peng, Raman spectroscopy analysis of the biochemical characteristics of molecules associated with the malignant transformation of gastric mucosa, *PLoS One* 9 (2014), <https://doi.org/10.1371/journal.pone.0093906>.
- [62] K.W. Short, S. Carpenter, J.P. Freyer, J.R. Mourant, Raman spectroscopy detects biochemical changes due to proliferation in mammalian cell cultures, *Biophys. J.* 88 (2005) 4274–4288, <https://doi.org/10.1529/biophysj.103.038604>.
- [63] A. Echard, D. Burgess, The changing lipidome during cell division, *Cell* 156 (2014) 394–395, <https://doi.org/10.1016/j.cell.2014.01.018>.
- [64] A. Bastide, A. David, The ribosome, (slow) beating heart of cancer (stem) cell, *Oncogenesis* 7 (2018), <https://doi.org/10.1038/s41389-018-0044-8>.
- [65] W.H. Yong, M. Shahbikhani, D. Telesca, S. Yang, J.L. Tso, J.C. Menjívar, B. Wei, G. M. Lucey, S. Mareninova, Z. Chen, L.M. Liao, A. Lai, S.F. Nelson, T.F. Cloughesy, C. L. Tso, Ribosomal proteins RPS11 and RPS20, two stress-response markers of glioblastoma stem cells, are novel predictors of poor prognosis in glioblastoma patients, *PLoS One* 10 (2015) 1–19, <https://doi.org/10.1371/journal.pone.0141334>.
- [66] X. Cao, X. Chen, C. Shi, M. Zhang, W. Lu, L. Li, J. Dong, X. Han, W. Qian, Process characterization of epithelial-mesenchymal transition in alveolar epithelial type II cells using surface-enhanced Raman scattering spectroscopy, *RSC Adv.* 6 (2016) 14321–14328, <https://doi.org/10.1039/c5ra17022b>.
- [67] S. Chen, X. Chen, W. Li, T. Shan, W.R. Lin, J. Ma, X. Cui, W. Yang, G. Cao, Y. Li, L. Wang, Y. Kang, Conversion of epithelial-to-mesenchymal transition to mesenchymal-to-epithelial transition is mediated by oxygen concentration in pancreatic cancer cells, *Oncol. Lett.* 15 (2018) 7144–7152, <https://doi.org/10.3892/ol.2018.8219>.
- [68] M. Marro, C. Nieva, R. Sanz-Pamplona, A. Sierra, Molecular monitoring of epithelial-to-mesenchymal transition in breast cancer cells by means of Raman spectroscopy, *Biochim. Biophys. Acta Mol. Cell Res.* 1843 (2014) 1785–1795, <https://doi.org/10.1016/j.bbamcr.2014.04.012>.
- [69] P.H. Wu, D.M. Gilkes, J.M. Phillip, A. Narkar, T.W.T. Cheng, J. Marchand, M. H. Lee, R. Li, D. Wirtz, Single-cell morphology encodes metastatic potential, *Sci. Adv.* 6 (2020) 1–8, <https://doi.org/10.1126/sciadv.aaw6938>.
- [70] N.A. Brazhe, A.B. Evlyukhin, E.A. Goodilin, A.A. Semenova, S.M. Novikov, S. I. Bozhevolnyi, B.N. Chichkov, A.S. Sarycheva, A.A. Baizhumanov, E. I. Nikelshparg, L.I. Deev, E.G. Maksimov, G.V. Maksimov, O. Sosnovtseva, Probing cytochrome c in living mitochondria with surface-enhanced Raman spectroscopy, *Sci. Rep.* 5 (2015) 1–13, <https://doi.org/10.1038/srep13793>.
- [71] J. Taylor, A. Milton, M. Willett, J. Wingfield, S. Mahajan, J. Milton, M. Willett, J. Wingfield, S. Mahajan, What do we actually see in intracellular SERS? Investigating nanosensor-induced variation, *Faraday Discuss.* (2017) 1–3, <https://doi.org/10.1039/c7fd00156h>.
- [72] I.S. Song, Mitochondria as therapeutic targets for cancer stem cells, *World J. Stem Cells* 7 (2015) 418, <https://doi.org/10.4252/wjsc.v7.i2.418>.
- [73] M.V. Berridge, L. Dong, J. Neuzil, Mitochondrial DNA in tumor initiation, progression, and metastasis: role of horizontal mtDNA transfer, *Cancer Res.* 75 (2015) 3203–3208, <https://doi.org/10.1158/0008-5472.CAN-15-0859>.
- [74] H. Imanishi, K. Hattori, R. Wada, K. Ishikawa, S. Fukuda, K. Takenaga, K. Nakada, J.I. Hayashi, Mitochondrial DNA mutations regulate metastasis of human breast cancer cells, *PLoS One* 6 (2011) 1–7, <https://doi.org/10.1371/journal.pone.0023401>.
- [75] D.M. Ferns, I.P. Kema, M.R. Buist, H.W. Nijman, G.G. Kenter, E.S. Jordanova, Indoleamine-2,3-dioxygenase (IDO) metabolic activity is detrimental for cervical cancer patient survival, *Oncoimmunology* 4 (2015) 1–7, <https://doi.org/10.4161/2162402X.2014.981457>.
- [76] J. Liu, X. Qin, D. Pan, B. Zhang, F. Jin, Amino acid-mediated metabolism: a new power to influence properties of stem cells, *Stem Cells Int* 2019 (2019), <https://doi.org/10.1155/2019/6919463>.
- [77] N. Santana, C. Nieva, A. Sierra, M. Marro, S. Rao, D. Petrov, Raman microscopy is a tool to identify the metastatic ability of breast tumors, 2011, *Int. Work. Biophotonics 2011* (2011) 1–3, <https://doi.org/10.1109/IWBP.2011.5954814>.
- [78] J. Brandi, E.D. Pozza, I. Dando, G. Biondani, E. Robotti, R. Jenkins, V. Elliott, K. Park, E. Marengo, E. Costello, A. Scarpa, M. Palmieri, D. Ceconi, Secretome protein signature of human pancreatic cancer stem-like cells, *J. Proteom.* 136 (2016) 1–12, <https://doi.org/10.1016/j.jprot.2016.01.017>.
- [79] K. Chen, Y.H. Huang, J.L. Chen, Understanding and targeting cancer stem cells: Therapeutic implications and challenges, *Acta Pharmacol. Sin.* 34 (2013) 732–740, <https://doi.org/10.1038/aps.2013.27>.
- [80] J. Frimodig, D. Kumar, V. Snyder, T.C. Reed-Newman, L. Arnold, S. Mary Thomas, S. Anant, Cancer stem cell metabolism and potential therapeutic targets, *Cancer Stem Cell Metab. Potential Ther. Targets Front. Oncol.* 8 (2018) 203, <https://doi.org/10.3389/fonc.2018.00203>.
- [81] Y.C. Lee, M.S. Gajdosik, D. Josic, J.G. Clifton, C. Logothetis, L.Y. Yu-Lee, G. E. Gallucci, S.N. Maity, S.H. Lin, Secretome analysis of an oncogenic prostate tumor identifies complex signaling networks mediating cross-talk of cancer and stromal cells within the tumor microenvironment, *Mol. Cell. Proteom.* 14 (2015) 471–483, <https://doi.org/10.1074/mcp.M114.039909>.
- [82] A.G. Woods, Advancements of mass spectrometry in biomedical research, (2014), <https://doi.org/10.1007/978-3-319-06068-2>.
- [83] J.L. Paltridge, L. Belle, Y. Khew-Goodall, The secretome in cancer progression, *Biochim. Biophys. Acta Proteins Proteom.* 1834 (2013) 2233–2241, <https://doi.org/10.1016/j.bbapap.2013.03.014>.
- [84] H. Peinado, S. Lavotshkin, D. Lyden, The secreted factors responsible for pre-metastatic niche formation: old sayings and new thoughts, *Semin. Cancer Biol.* 21 (2011) 139–146, <https://doi.org/10.1016/j.semcancer.2011.01.002>.
- [85] V. Shalabaeva, L. Lovato, R. La Rocca, G.C. Messina, M. Dipalo, E. Miele, M. Perrone, F. Gentile, F. De Angelis, Time resolved and label free monitoring of extracellular metabolites by surface enhanced Raman spectroscopy, (2017), <https://doi.org/10.1371/journal.pone.0175581>.
- [86] J. De Gelder, K. De Gussem, P. Vandenaabee, L. Moens, P.V. L.M.D. Joke De Gelder, Kris De Gussem, Reference database of Raman spectra of biological molecules, *J. Raman Spectrosc.* 38 (2007) 1133–1147, <https://doi.org/10.1002/jrs>.
- [87] L. Sun, L. Zhang, J. Chen, C. Li, H. Sun, J. Wang, H. Xiao, Activation of tyrosine metabolism in CD13⁺ cancer stem cells drives relapse in hepatocellular carcinoma, *Cancer Res. Treat.* (2019) 1–18, <https://doi.org/10.4143/crt.2019.444>.
- [88] G. Gueron, N. Anselmino, P. Chiarella, E.G. Ortiz, S. Lage Vickers, A.V. Paez, J. Giudice, M.D. Contin, D. Leonardi, F. Jaworski, V. Manzano, A. Strazza, D. R. Montagna, E. Labanca, J. Cotignola, N. Daccorso, A. Woloszynska-Read, N. Navone, R.P. Meiss, R. Ruggiero, E. Vazquez, Game-changing restraint of Ros-damaged phenylalanine, upon tumor metastasis article, *Cell Death Dis.* 9 (2018), <https://doi.org/10.1038/s41419-017-0147-8>.

- [89] G. Kuku, M. Altunbek, M. Culha, Surface-enhanced Raman scattering for label-free living single cell analysis, *Anal. Chem.* 89 (2017) 11160–11166, <https://doi.org/10.1021/acs.analchem.7b03211>.
- [90] F. Lussier, V. Thibault, B. Charron, G.Q. Wallace, J.F. Masson, Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering, *TrAC Trends Anal. Chem.* 124 (2020), <https://doi.org/10.1016/j.trac.2019.115796>.
- [91] AddinSoft, Addinsoft (2019). XLSTAT statistical and data analysis solution, (n.d.).

Dr. Venkatakrisnan has been working on fundamental and applied research in the areas of pulsed laser micro and non-fabrication and laser micro/nano measurement systems. His

research during these years has since generated over 150 international journal publications

Dr. Tan's primary areas of research includes laser synthesis of nanomaterials, applications of nanomaterials for biomedical applications. Her research in these areas has generated various highly cited publications in high impact journals.

Dr. Swarna's primary areas of research focusses on using carbon-based nanostructures for biomedical applications.