

UNIVERSITÀ DEGLI STUDI DEL PIEMONTE ORIENTALE

“AMEDEO AVOGADRO”



UNIVERSITÀ DEL PIEMONTE ORIENTALE

Dipartimento di Scienze e Innovazione Tecnologica

Corso di Laurea Magistrale in Biologia

Curriculum Biomedico e Biomolecolare

TESI DI LAUREA

*Utilizzo della Digital Pathology per l'analisi morfometrica
quantitativa del carcinoma mammario Triple Negative Breast
Cancer (TNBC) con il software QuPath*

Relatore: Professor Elia Ranzato *Elia Ranzato*

Tutor aziendale: Dottor Alberto Maria Pisacane

Sede di tirocinio: A.S.L. Biella – Anatomia Patologica

Candidata: Giulia Serra

Giulia Serra

Matricola: 20024076

Anno Accademico 2022/2023

Ringraziamenti

Questa tesi magistrale è stata realizzata grazie alla collaborazione tra la S.C. di Anatomia Patologica dell'A.S.L. di Biella e la Fondazione Edo ed Elvo Tempia. Tengo quindi a ringraziare ogni singola persona che ha dato il suo contributo nella realizzazione di questo progetto, in particolare: il mio tutor aziendale Dott. Alberto Maria Pisacane, il Dott. Andrea Vandoni, la Dott.ssa Mariangela D'Andrea, la Dott.ssa Elena Biletta e tutte le tecniche di laboratorio dell'Anatomia Patologica dell'Ospedale degli Infermi di Biella, che è stata la sede del mio progetto formativo.

Inoltre, vorrei fare dei ringraziamenti speciali.

Grazie alla mia famiglia: Mamma Rossana, Papà Pino e mio fratello Alberto, per aver sempre creduto in me e per avermi supportata e sopportata durante tutto il mio percorso universitario e non solo, per tutta la mia vita. Vi voglio tanto bene.

Grazie a Fabio per esserci sempre stato in questi otto anni e per aver sempre saputo dire la cosa giusta nei momenti di difficoltà e sconforto.

Grazie alle mie amiche e ai miei amici, vecchi e nuovi, per tutti i momenti unici vissuti insieme e per tutte le cose belle che verranno.

Grazie alla mia grande passione, l'equitazione, per essere sempre il mio posto felice che, con il mio grande amore per i cavalli, mi permette di essere senza pensieri allontanandomi dalla realtà.

Grazie ad ogni singola persona che fa parte della mia vita. Senza tutti voi non sarei la persona che sono oggi e, forse, non ce l'avrei mai fatta.

Infine, grazie a me per non avere mai mollato nonostante le difficoltà e per aver saputo affrontare ogni sfida a testa alta, raggiungendo i miei obiettivi e ottenendo tante soddisfazioni. Grazie a me per essere sempre la stessa: sensibile, emotiva e sentimentale.

Alla Giulia che verrà.

Grazie, di cuore, a tutti.

Sommario

1. Abstract.....	1
2. Introduzione.....	4
2.1. Carcinoma mammario	4
2.1.1. Diagnostica	12
2.1.2. Trattamento.....	15
2.2. Carcinoma mammario triplo negativo	21
2.2.1. Classificazione molecolare	23
2.2.2. TNBC Luminal Androgen Receptor	25
2.3. Digital Pathology	25
2.3.1. Scanner	29
2.3.2. QuPath	34
3. Materiali e metodi.....	38
3.1. Popolazione di studio: descrizione dei dati	38
3.2. Previsione dei sottotipi TNBC.....	40
3.3. Determinazione del profilo genetico	40
3.4. Ricerca e acquisizione dati: selezione dei casi	41
3.5. Attività istologica del laboratorio di Anatomia Patologica	43
3.5.1. Prelievo.....	43
3.5.2. Accettazione	43

3.5.3.	Fissazione	45
3.5.4.	Riduzione.....	45
3.5.5.	Processazione	47
3.5.6.	Inclusione	48
3.5.7.	Taglio.....	50
3.5.8.	Allestimento vetrini	51
3.5.9.	Colorazione istologici.....	52
3.5.10.	Montaggio.....	55
3.5.11.	Etichettatura.....	56
3.5.12.	Immunoistochimica	57
3.6.	Digitalizzazione dei vetrini.....	59
3.7.	Revisione dei dati	65
3.8.	Analisi dei vetrini con il software QuPath	66
3.8.1.	Creazione di un progetto di analisi.....	67
3.8.2.	Deconvoluzione dei colori: separazione delle regioni e normalizzazione dei colori	67
3.8.3.	Stima dei vettori di colorazione.....	69
3.8.4.	Rilevamento del tessuto: definire la regione di interesse	70
3.8.5.	Aree di misurazione.....	72
3.8.6.	Rilevamento delle cellule	73
3.8.7.	Classificazione delle cellule	75
3.9.	Analisi statistica.....	77

4. Risultati.....	81
5. Discussione.....	85
6. Conclusioni.....	91
7. Bibliografia.....	92

1. Abstract

Scopo dell'attività di ricerca

L'obiettivo di questo lavoro di ricerca sarà quello di trattare come argomento di tesi lo studio del carcinoma mammario attraverso l'intelligenza artificiale della *Digital Pathology* (DP). In particolar modo, si andrà a studiare uno specifico tumore della mammella, ossia il carcinoma mammario triplo negativo, attraverso la raccolta dei dati delle pazienti e la loro digitalizzazione.

Lo scopo di questa tesi di ricerca sarà quello di creare un flusso di lavoro per analizzare le immagini digitali mediante un software su computer, che ci permetterà di effettuare un'analisi morfometrica di istomorfometria quantitativa dei tumori delle pazienti, caratterizzandoli in base ad una serie di parametri che andremo ad indagare da un punto di vista morfologico, tra cui: la quantità di tumore espressa come area in μm^2 , la quantità di stroma espressa come area in μm^2 , il rapporto stroma/tumore, la quantità di necrosi tumorale espressa come area in μm^2 , la purezza tumorale espressa in percentuale (%) e il numero di linfociti peri-tumoral proliferanti presenti in mitosi. Questi parametri determineranno perciò le caratteristiche morfologiche dei tumori in analisi.

Infine, i dati di questi parametri insieme ai dati della componente intraduttale, dell'indice di proliferazione Ki-67 e del diametro espresso come misurazione micrometrica, verranno associati analizzando il *follow-up* delle pazienti per quanto riguarda i dati di sopravvivenza con cui verranno confrontati mediante analisi statistica.

Il software utilizzato è QuPath, piattaforma di analisi *open-source* per la quantificazione automatica di immagini digitali per l'identificazione e sviluppo di biomarcatori prognostici basati sull'imaging nell'applicazione clinica degli algoritmi di Intelligenza Artificiale (IA).

L'analisi computerizzata delle immagini di pazienti affette da TNBC (*Triple Negative Breast Cancer*) può consentire lo studio della morfologia e dell'architettura dei tessuti andando ad indagare sulle caratteristiche del microambiente tumorale, che gioca un ruolo significativo e fondamentale nella progressione della malattia. Infatti, per questo motivo, il microambiente tumorale potrebbe aiutare a scoprire delle differenze biologiche tra diverse popolazioni e porre le basi per una nuova sotto classificazione TNBC. [4]

L'Intelligenza Artificiale ha quindi il potenziale di fare previsioni più ottimizzate e personalizzate sugli esiti della malattia. Gli obiettivi sono una digitalizzazione e una sostenibilità della pratica clinica in Anatomia Patologica, migliorando la qualità e i tempi di formulazione dei processi diagnostici. Gli strumenti digitali, se impiegati correttamente nella pratica clinica, hanno infatti il vantaggio di ridurre notevolmente le tempistiche di formulazione della diagnosi e migliorarne la qualità; ma non solo, anche l'intero flusso di lavoro.

Oggi, vi è la necessità di un significativo potenziamento per rispondere alle esigenze della collettività e l'analisi delle immagini digitali potrebbe essere una risposta alternativa come strumento complementare di supporto che mira a potenziare e rendere l'intero processo diagnostico delle Unità di Anatomia Patologica maggiormente sostenibile e con tempi ridotti. È fondamentale, però, che l'analisi digitale delle immagini istopatologiche, siccome supera la sola analisi visiva, venga

integrata con i dati provenienti dalle informazioni cliniche e demografiche dell'anamnesi, dei test e degli esiti clinici dei pazienti.

Il processo di creazione e diagnosi dei vetrini istologici è un procedimento costituito da diverse fasi in cui la componente manuale è molto presente e significativa, per questo motivo è fondamentale inserire delle tecnologie di intelligenza artificiale come strumenti di supporto e affiancamento siccome le immagini digitali offrono un enorme vantaggio rispetto alle immagini manuali. Si parla di “Transizione Digitale” dall'Anatomia Patologica tradizionale con il microscopio ottico convenzionale alla *Digital Pathology*, una patologia moderna e all'avanguardia. Le ricche risorse di dati sottoforma di immagini tissutali istopatologiche determinano quindi una transizione della patologia da scienza clinica a scienza informatica. [13]

In questo studio, faremo una panoramica sulle caratteristiche generali del carcinoma mammario e del carcinoma mammario triplo negativo, analizzando alcuni approcci basati sull'Intelligenza Artificiale applicata al Deep Learning e, in particolare, uno alla patologia computazionale.

2. Introduzione

2.1. Carcinoma mammario

Il carcinoma mammario, definito come tale ossia “carcinoma”, è un tumore che origina dalle cellule epiteliali. È una malattia eterogenea, per cui i pazienti che apparentemente presentano tumori simili nelle caratteristiche clinico-patologiche, essi non lo sono poi nel decorso clinico.

Questa malattia rappresenta la quinta causa di morte per cancro in tutto il mondo ed è il tumore invasivo più comune nelle donne di tutte le età, nelle quali rappresenta la principale causa di morte. [3] [13]

Come tutti i tumori, il carcinoma della mammella è una malattia multifattoriale siccome le cause non sono del tutto chiare ma sono stati identificati diversi fattori, tra cui: fattori riproduttivi, fattori ormonali, fattori comportamentali, fattori genetici, fattori ambientali, fattori dietetici e metabolici; che sono dei fattori di rischio che possono determinare il suo sviluppo e che quindi devono essere analizzati considerandoli nel loro insieme e non singolarmente. I principali sono: età, abitudini alimentari, obesità, consumo di fumo e/o alcol, terapia ormonale sostitutiva, familiarità con casi di tumore al seno in familiari stretti ed ereditarietà con la presenza di alcune mutazioni predisponenti. [49]

Il rischio di ammalarsi aumenta con il passare dell'età, e questa correlazione è dovuta al continuo stimolo endocrino-proliferativo a cui il tessuto mammario è sottoposto, abbinato al progressivo danneggiamento del DNA e all'accumulo di alterazioni epigenetiche dovute ad uno squilibrio tra oncogeni e geni oncosoppressori.

È uno dei tumori più comuni e mortali, infatti colpisce oltre il 10% delle donne a livello globale. [14] Per questo motivo, la diagnosi e il trattamento precoci possono prevenire significativamente la progressione e l'esito infausto della malattia riducendo i tassi di morbilità e mortalità. Ciò viene attuato con i programmi di prevenzione nella popolazione generale, i quali prevedono che le donne eseguano degli autocontrolli mediante la palpazione come prevenzione primaria e dei controlli di screening regolari effettuati con cadenza annuale o ogni due anni, secondo i programmi nazionali, con l'utilizzo delle strumentazioni ecografiche e mammografiche come prevenzione secondaria periodica. La mammografia rappresenta il metodo di prevenzione attualmente più efficace per la diagnosi in stadio precoce; infatti, lo screening mammografico viene sottoposto a pazienti asintomatiche con il fine di effettuare la diagnosi e in seguito offrire trattamenti meno aggressivi e più efficaci con l'obiettivo di ridurre la mortalità.

Inoltre, in casi selezionati con una certa familiarità è consigliato il counselling genetico con l'esecuzione di test genetici per l'analisi dei geni BRCA-1 e BRCA-2 come strumento di prevenzione aggiuntivo per identificare la presenza di mutazioni genetiche ereditarie. Inoltre, nelle donne ad alto rischio per importante storia familiare di carcinoma mammario o perché portatrici di mutazioni si effettua uno screening di risonanza magnetica mammaria con mezzo di contrasto a cadenza annuale a partire dall'età di 25 anni.

Nel caso in cui venisse riscontrata un'anomalia dagli esami di screening, il protocollo prevede che venga eseguita la procedura bioptica che consiste nell'eseguire una biopsia del tessuto mammario in questione ottenendo un campione, dal quale viene realizzato un vetrino istologico colorato in Ematossilina-

Eosina (EE), che sarà successivamente sottoposto ad osservazione al microscopio ottico da parte di uno specialista anatomo-patologo per eseguire la diagnosi.

Durante la loro valutazione, gli specialisti esaminano al microscopio alcune regioni adiacenti per analizzare il tessuto cercando alcuni elementi rilevanti per l'identificazione del tumore su porzioni microscopiche tissutali, valutandone le proprietà e caratteristiche istologiche relativamente ai nuclei e alla loro disposizione e organizzazione nel tessuto, andando quindi dal particolare al generale identificando le regioni nel contesto più ampio del tessuto. Questa procedura consente di eseguire una prima distinzione tra le regioni maligne del tessuto e le regioni non maligne (benigne) che possono comunque presentare delle variazioni nell'architettura e strutturazione del parenchima mammario, che però non sono correlate alla progressione del tessuto verso la malignità.

Successivamente, secondo la classificazione istologica, le regioni di tessuto individuate come maligne possono essere ulteriormente sotto-classificate suddividendole in due gruppi principali: carcinoma in situ e carcinoma invasivo.

Il carcinoma in situ, per definizione, si identifica come tale quando le cellule cancerose sono delimitate e circoscritte all'interno del sistema duttale-lobulare mammario senza diffondersi in altri tessuti, mentre il carcinoma invasivo viene definito tale in quanto è caratterizzato da una vera e propria invasione del tessuto da parte delle cellule tumorali che si diffondono in tutto il parenchima mammario oltre i dotti, per cui non risultano più delimitate in regioni specifiche ma presenti in tutto il tessuto mammario.

Infatti, i tumori al seno vengono definiti invasivi quando sono caratterizzati dalla capacità del tumore di invadere i tessuti circostanti diffondendosi a partire da un

punto o sito di origine (lobuli o dotti galattofori) e dando origine a metastasi nei tessuti vicini, come linfonodi o organi più lontani.

Secondo la classificazione anatomico-patologica dei tumori della mammella dell'Organizzazione Mondiale della Sanità (OMS), definita come *World Health Organization* (WHO) 2019, i tumori al seno invasivi presentano come forma più comune il carcinoma duttale infiltrante, che ha appunto origine dai dotti galattofori e rappresenta circa il 70-80% di tutte le forme di cancro al seno e viene definito come carcinoma invasivo o infiltrante non di istotipo speciale, dall'inglese "no special type" (NST) o carcinoma duttale di tipo non altrimenti specificato, in quanto l'acronimo NST definisce un gruppo ampio ed eterogeneo di carcinomi, che si identificano come tali in quanto non possiedono caratteristiche sufficientemente particolari per poter essere classificati come tipi istologici speciali.

Il carcinoma lobulare infiltrante, invece, è sempre un tumore invasivo che ha origine nei lobuli, rappresenta circa il 10-15% di tutti i tumori della mammella e viene classificato insieme ad altre tipologie di carcinoma mammario come carcinoma di tipo istologico speciale.

In generale, i carcinomi mammari invasivi rappresentano circa il 70% di tutti i carcinomi mammari ed essendo solitamente diagnosticati in stadi avanzati presentano una prognosi peggiore rispetto ai sottotipi in situ. [12]

Vi sono anche altre forme di carcinoma mammario di tipo speciale, meno frequenti rispetto al carcinoma duttale e al carcinoma lobulare, ovvero: il carcinoma tubulare, il carcinoma papillare, il carcinoma mucinoso e il carcinoma cribriforme; che hanno tutti prognosi favorevole.

In base quindi alla diagnosi, varia il trattamento a cui il paziente viene sottoposto. Infatti, le lesioni benigne possono essere seguite clinicamente nel tempo senza la necessità di un intervento chirurgico, a differenza invece delle lesioni maligne che richiedono quasi sempre un intervento chirurgico seguito oppure no, in base ai casi, da un'ulteriore terapia. [5]

Il carcinoma mammario è suddiviso in diversi sottotipi basati sul profilo molecolare, il quale viene stabilito in base alla variazione di alcuni recettori, in particolare: il recettore per gli estrogeni (ER), il recettore per il progesterone (PR) e il recettore per il fattore di crescita epidermico (HER2). Quando un carcinoma mammario è privo di questi recettori, si parla di carcinoma mammario triplo negativo (TNBC) che, per definizione immunohistochimica, è considerato tale per la triplice negatività. [16]

Ciascun sottotipo del carcinoma mammario mostra una diversa sensibilità agli agenti chemioterapici e alle terapie target, ecco perché è importante la valutazione immunohistochimica dei recettori ormonali, di HER2 e del Ki-67 nella pratica clinica.

Basandosi sulla classificazione molecolare, ottenuta in seguito ad indagini effettuate sui profili di espressione genica con la metodica dei *microarray*, i carcinomi mammari invasivi si suddividono in quattro sottogruppi, molto importanti dal punto di vista prognostico, che sono i seguenti:

- Luminali A, caratterizzati: dall'espressione marcata dei recettori ormonali con una positività che determina una prognosi favorevole, HER2 negativi e bassa attività proliferativa con un ridotto Ki-67. Generalmente si tratta di tumori di basso grado con un'elevata sensibilità alla terapia endocrina o ormonoterapia

adiuvante, la quale può essere completata dalla chemioterapia a cui hanno meno sensibilità;

- Luminali B, spesso identificati come tumori di alto grado e determinati anch'essi da un'elevata espressione dei recettori ormonali rappresentati da una certa positività ma, nonostante ciò, hanno un elevato rischio di sviluppare una recidiva a causa dell'indice di proliferazione Ki-67 alto, che è associato ad un'elevata espressione dei geni di proliferazione. Per questo motivo presentano un fenotipo più aggressivo e una prognosi peggiore rispetto ai Luminali A. Vengono trattati con la chemioterapia in aggiunta alla terapia endocrina e possono essere HER2 negativi oppure HER2 positivi con una sua sovraespressione o amplificazione;
- HER2-positivi, identificati come non luminali e individuati dall'espressione di HER2 con una certa positività sovraespresso o amplificato e da una negatività dei recettori ormonali;
- *Basal-like*, caratterizzati da una più elevata espressione delle citocheratine basali e tipici per non possedere né i recettori ormonali (a causa della loro mancata espressione) né HER2, per questo determinano la prognosi peggiore confrontando l'insieme di tutti i sottotipi. Ne fa parte il carcinoma mammario triplo negativo, il cui trattamento principale è la chemioterapia adiuvante.

Anche all'interno di ciascun sottogruppo, i carcinomi mammari possiedono caratteristiche eterogenee che, per le nuove conoscenze patologiche e molecolari, hanno portato a sottotipizzare ulteriormente questa malattia ridefinendola nuovamente. Un esempio è il caso della classificazione molecolare che riporta la definizione dei sei sottotipi del carcinoma mammario triplo negativo, che andremo ad indagare nel Sottocapitolo 2.2.1.

I tumori e il loro ambiente sono ecosistemi spazialmente organizzati e costituiti da tipi cellulari distinti che possono assumere una varietà di fenotipi, sfociando in sottotipi diversi. La progressione del cancro è un processo che si basa sull'interazione delle cellule all'interno del microambiente tumorale, ma non solo. Infatti, il tumore acquisisce capacità invasive nel momento in cui inizia ad espandersi grazie anche al contatto che si crea con le sole cellule tumorali e il sistema immunitario che interviene, il quale può favorire la crescita tumorale promuovendone lo sviluppo oppure inibirla sopprimendone la crescita stessa. [29]

Questo ha spinto la ricerca ad individuare dei biomarcatori che possono essere identificati come rappresentativi del sistema immunitario, il quale interviene nella malattia di un singolo paziente per dare una risposta immunitaria come difesa immuno-mediata dell'ospite contro il tumore, in modo da andare a predire la risposta alla terapia. Un esempio sono i *Tumor Infiltrating Lymphocytes* (TIL), considerati come fattori prognostici della malattia in quanto la loro elevata presenza nello stroma intratumorale determina una prognosi migliore. Essi sono costituiti da un insieme di diverse tipologie immuno-cellulari (cellule T citotossiche, cellule *natural killer*, cellule dendritiche, cellule T-helper, cellule B e cellule T regolatorie) che sono considerati come biomarcatori della componente adattativa del tumore.

Infatti, con la classificazione WHO del 2019, che ha sostituito la precedente redatta nel 2012, è stato riconosciuto il valore prognostico dei Linfociti Infiltranti il Tumore.

La valutazione dell'infiltrazione linfocitaria viene effettuata visivamente da parte del patologo per effettuare una quantificazione adeguata che, però, è un processo lungo, costoso, soggettivo e di conseguenza può andare incontro ad errori. Ecco perché anche nell'individuazione di questi biomarcatori si è utilizzata la *Digital*

Pathology con l'utilizzo di strumenti computerizzati su campioni istopatologici presenti in immagini EE e IHC di pazienti TNBC, per ottenere una quantificazione più oggettiva dei TIL. [40][41]

Da questi studi si sono estratte diverse misure quantitative dei Linfociti Infiltranti il Tumore che hanno dimostrato di essere associati alla sopravvivenza e, in particolare, la presenza di un'alta densità di TIL è correlata a un miglioramento della sopravvivenza; dimostrando così, in generale, la capacità della Patologia Digitale di utilizzare strumenti computazionali con i quali interpretare i biomarcatori attraverso la loro quantificazione. [31] [32] [33]

Oltre all'analisi sulla densità dei TIL, alcuni studi hanno sfruttato immagini istopatologiche IHC eseguendo su di esse una caratterizzazione spaziale dei TIL nel TNBC con l'utilizzo di strumenti automatici per individuare gruppi cellulari in regioni tissutali differenti, ed estraendo caratteristiche morfometriche da questi gruppi. Sul fronte invasivo si tende ad avere una densità di cellule immunitarie più elevata, evidenziando questa caratteristica come molto importante nell'architettura del tumore. [37]

Inoltre, anche il calcolo del rapporto linfocitario intratumorale in base alle distanze tra i TIL e singole cellule tumorali o gruppi di cellule tumorali, come misura quantitativa. Questa risulta essere statisticamente significativa in associazione con la sopravvivenza malattia-specifica nel TNBC. Inoltre, si sono estratte le caratteristiche spaziali dei TIL calcolandone un punteggio associandolo ai risultati delle pazienti. [34][35]

Inoltre, è noto che pazienti affette da TNBC con malattia invasiva residua hanno una prognosi peggiore rispetto a pazienti affette da carcinoma mammario ER positivo

con malattia invasiva residua. Questi risultati si sono ottenuti da uno studio in cui è stata eseguita la caratterizzazione dell'architettura spaziale dei TIL nel TNBC per identificare le pazienti che sono a maggior rischio di morte o recidiva. [38]

Altri studi hanno invece dimostrato che la componente dei TIL è maggiore nel cancro al seno TNBC e nel carcinoma mammario HER2 positivo rispetto ad altri sottotipi di tumori. [42]

L'insieme di tutti questi studi ha dimostrato l'importanza di indagare l'organizzazione architetturale e strutturale del TNBC come prova dell'analisi sull'architettura spaziale del tumore. Purtroppo, un'importante limitazione sono le dimensioni contenute dei campioni, così come per questo studio, e sono quindi necessarie ulteriori sperimentazioni e analisi su una coorte maggiore di pazienti per determinare se questi approcci possono essere generalizzati su larga scala nel mondo reale. [27] [3]

Prendendo in considerazione le caratteristiche spaziali, quindi, bisogna interrogarsi sulla biologia del tumore e sulla risposta al trattamento.

2.1.1. Diagnostica

Per tutte le neoplasie mammarie, la diagnosi viene effettuata con esami mammografici ed ecografici, a volte affiancati con la risonanza magnetica.

L'eventuale identificazione di noduli sospetti da parte del medico indirizza ad eseguire un esame bioptico o biopsia, la quale viene effettuata in procedura ambulatoriale presso gli ambulatori di senologia diagnostica con un prelievo mediante ago bioptico. Sul campione della biopsia vengono eseguite delle analisi in istologia per esaminare le caratteristiche del tessuto in esame. Inoltre, ai fini di stabilire la prognosi e il trattamento, sul tessuto prelevato vengono effettuate anche

le indagini molecolari in immunoistochimica per valutare i fattori biologici come: l'espressione dei recettori ormonali, la velocità di crescita o attività proliferativa con l'indice di proliferazione Ki-67 in cui si valuta l'espressione dello stesso come percentuale di nuclei che si colorano dalla reazione antigene (proteina MIB1 codificata dal gene MIB1) e anticorpo, e la sovraespressione del recettore HER2 all'immunoistochimica e la sua amplificazione genica mediante ibridazione in situ in fluorescenza (FISH) che misura l'amplificazione del gene. Questi, analizzati come fattori o indicatori o score prognostici aiutano a prevedere l'*outcome* delle pazienti indipendentemente dal trattamento. Quindi, sono dei validi strumenti di valutazione che permettono di effettuare una panoramica sulla prognosi della malattia e quindi sulla sopravvivenza, essendo correlati ad essa. Essi sono, infatti, molto importanti per stabilire l'aggressività del tumore e per pianificare correttamente la scelta del tipo di terapia o trattamento da somministrare al paziente dopo intervento chirurgico.

Oltre a quelli già citati, altri parametri prognostici sono: la dimensione del tumore espresso come diametro in cm, lo stato dei linfonodi ascellari (con cellule tumorali isolate (ITC) o micro-metastasi nel linfonodo sentinella), il grado o *grading* istologico, il tipo istologico, l'invasione vascolare, l'età della paziente, i profili di espressione genica e i linfociti intratumorali o infiltranti il tumore (TILs).

La stessa procedura di diagnosi, utilizzata per il campione biotico con la realizzazione dell'esame istologico e del suo quadro molecolare, viene applicata nuovamente sul campione prelevato dal pezzo operatorio ottenuto dopo aver eseguito l'intervento chirurgico.

L'Anatomo-Patologo, in seguito all'identificazione dell'area del carcinoma mammario invasivo, effettua ulteriori analisi di differenziazione del tumore

attraverso l'utilizzo degli schemi di classificazione *Bloom-Richardson* e *Nottingham Histologic Grading* stimando l'aggressività del tumore con la valutazione delle seguenti caratteristiche istologiche:

- Formazione di tubuli
- Pleomorfismo
- Mitosi

In base a queste caratteristiche, le quali possono avere valori che variano da 1 a 3 (1=basso, 2=intermedio, 3=elevato), il patologo determina il grado istologico della malattia.

Questo approccio potrebbe rientrare all'interno di una metodologia automatizzata, che consente una riproducibilità per il rilevamento del cancro al seno invasivo, potenziando l'analisi dei vetrini istologici e riducendo così le tempistiche necessarie per la diagnosi e la variabilità.

Come per la maggior parte dei tumori solidi, anche per il carcinoma mammario si utilizza il sistema di stadiazione TNM che, con criteri classificativi tradizionali, definisce l'estensione della malattia. Dal gennaio 2018 è stato rivisto dall'*American Joint Committee on Cancer* (AJCC) come ottava edizione a partire dal AJCC/UICC TNM settima edizione, riconosciuta dalle due importanti organizzazioni scientifiche dell'*American Joint Committee on Cancer* (AJCC) e della *Union for International Cancer Control* (UICC).

In particolare, esso valuta tre parametri o fattori anatomici che pongono le basi per la classificazione anatomica: l'estensione della malattia tumorale (T) che nel caso di multifocalità (ossia la presenza di più focolai tumorali di invasione separati da parenchima sano) definisce le dimensioni del focolaio maggiore oppure dell'intero

carcinoma diffuso che possiede una crescita lobulare e si diffonde in più quadranti, il coinvolgimento dei linfonodi regionali (N) solitamente a livello ascellare e la presenza o assenza di metastasi (M); tenendo conto però anche della classificazione prognostica riportando le informazioni dei fattori prognostici con il grado tumorale o *grading* istologico e il quadro dei recettori ormonali ER e PR insieme all'amplificazione di HER2, che possono influenzare la stadiazione nella prognosi. Questo cambiamento è stato apportato con l'ultima edizione del documento con l'obiettivo di migliorare i criteri del sistema stadiativo, mentre nella versione precedente, AJCC/UICC TNM settima edizione, la classificazione era basata solamente su aspetti anatomici.

La combinazione dei tre fattori TNM insieme permette di assegnare uno stadio, con l'obiettivo di definire in seguito una terapia specifica. Infatti, lo stadio può essere assegnato nelle tre diverse fasi che seguono il paziente durante il suo percorso: fase clinica (cTNM) prima dell'intervento, fase chirurgica (sTNM) all'atto della terapia chirurgica e fase patologica (pTNM) dopo lo studio anatomico-patologico del pezzo operatorio.

La *Digital Pathology* consente uno storage, una visualizzazione e, in seguito, un'analisi della patologia migliorando l'efficienza complessiva del flusso di lavoro della routine diagnostica.

2.1.2. Trattamento

La scelta del percorso terapeutico a cui la paziente viene sottoposta dipende da diversi fattori che influenzano il clinico in questa decisione. Infatti, i fattori predittivi sono quei parametri che sono correlati all'eventuale efficacia di un trattamento antitumorale, e sono i seguenti: HER2, recettori ormonali, PD-L1 con

positività nelle cellule immunitarie infiltranti il tumore > 1% dell'area tumorale, mutazione di PIK3CA su tessuto tumorale e mutazioni germinali dei geni BRCA-1 e BRCA-2.

La terapia per il carcinoma mammario comprende diverse opzioni, le quali possono essere anche abbinate tra loro e che sono: la chirurgia, la radioterapia, la chemioterapia, la terapia ormonale e la terapia biologica.

Generalmente, i tumori al seno in base alle loro caratteristiche molecolari vengono trattati o con la terapia ormonale oppure con farmaci a bersaglio molecolare anti-HER2. Lo stato di HER2 rappresenta, infatti, sia un fattore prognostico sia un fattore predittivo di risposta ai farmaci anti-HER2. Inoltre, i carcinomi mammari con alti livelli dei recettori ormonali si identificano con una maggiore probabilità di trarre beneficio dalla terapia ormonale-endocrina a cui vengono sottoposti.

Il TNBC, rispetto a tutti gli altri, non può beneficiare di questi trattamenti che sono oggi disponibili a causa dell'assenza dei recettori per gli estrogeni, per il progesterone e anche della proteina HER2 (*Human Epidermal Growth Factor Receptor 2*). Quest'ultima caratteristica, quindi, non permette l'applicazione della terapia basata sui farmaci biologici che sfruttano questa proprietà nei tumori HER2-positivi basandosi su un particolare farmaco, Trastuzumab, che è in grado di legarsi alla proteina HER2 bloccando le fasi di crescita e divisione delle cellule tumorali. Questo perché, in condizioni fisiologiche, HER2 regola la crescita e la proliferazione delle cellule che lo possiedono sulla loro superficie, ma in condizioni patologiche determina una crescita cellulare incontrollata. [48]

La terapia ormonale o ormonoterapia, invece, viene applicata in quei tumori del seno che sono stimolati a crescere da ormoni fisiologicamente presenti

nell'organismo, ossia gli estrogeni.

Essa consiste, quindi, in un trattamento in grado di ridurre i livelli di questi ormoni. Esistono diversi farmaci che svolgono questa funzione e la loro scelta viene effettuata sulla base delle caratteristiche del tumore e della paziente affetta dalla malattia.

La chirurgia, definita come terapia locale o trattamento loco-regionale, rappresenta il primo tra i principali step di trattamento che devono essere eseguiti all'interno del protocollo. Per tutti i tumori mammari, essa può essere di due tipologie in base a come viene realizzata: quadrantectomia (di tipo conservativo) o mastectomia (di tipo più invasivo o radicale).

La chirurgia conservativa può essere realizzata portando a risultati diversi, in base: al tipo di tumore e alle sue dimensioni dove è localizzato, dalla quantità di tessuto circostante e quindi anche in base alle dimensioni della mammella.

L'intervento chirurgico di mastectomia, invece, consiste nella rimozione dell'intero tessuto mammario, compreso il capezzolo. Nel caso in cui il tumore si sia diffuso anche ai linfonodi limitrofi detti "sentinella", è necessaria la loro asportazione con la procedura di svuotamento ascellare in quanto la loro positività è un chiaro segnale di migrazione delle cellule tumorali dalla loro sede di origine.

La scelta della tipologia di intervento chirurgico mammario dipende da diversi fattori: localizzazione del tumore, rapporto tra le dimensioni del tumore e quelle della mammella sede della malattia, caratteristiche mammografiche e presenza o assenza di controindicazioni alla radioterapia.

In seguito a intervento chirurgico, le pazienti vengono successivamente sottoposte a chemioterapia, definita come terapia sistemica adiuvante, per eliminare le eventuali

cellule tumorali residue. Questa terapia viene somministrata a sostegno del primo trattamento principale con lo scopo di ridurre al minimo il rischio di recidiva e di migliorare la prognosi del paziente dopo il trattamento chirurgico. A volte, la chemioterapia viene somministrata come terapia neo-adiuvante, ovvero in anticipazione all'intervento chirurgico, per cercare di ridurre da subito le dimensioni del tumore.

Anche la radioterapia, che consiste nell'utilizzo di dosi controllate di radiazioni, può essere utilizzata come terapia adiuvante per eliminare le cellule tumorali rimaste in seguito al trattamento chirurgico o alla chemioterapia.

Poiché il Ki-67 è un marcatore di proliferazione cellulare, un indice più elevato implica che le cellule tumorali proliferano più rapidamente con un'alta velocità. Questo ha ovviamente una relazione con la prognosi del paziente e, in alcuni casi, può indirizzare la scelta del trattamento più appropriato. Per questo motivo, l'intelligenza artificiale può avere un ruolo cruciale come sistema di supporto indirizzando i patologi nella diagnosi e di conseguenza guidando i clinici nel prendere le migliori decisioni terapeutiche per i pazienti. [13]

Infatti, vi è un crescente interesse nell'identificazione e sviluppo di modelli o approcci predittivi basati sui biomarcatori dell'imaging che possano indirizzare verso un trattamento efficace tra la chemioterapia adiuvante o l'immunoterapia, prevedendo la risposta al trattamento. Infatti, i biomarcatori offrono molteplici vantaggi siccome nel caso della patologia computazionale non sono distruttivi dei tessuti e possono fornire delle caratteristiche quantitative in modo da poter caratterizzare l'eterogeneità del tumore. [47]

Data la mancanza di biomarcatori noti utilizzati come target per il trattamento, il TNBC viene spesso trattato con la chemioterapia citotossica convenzionale che rappresenta il trattamento standard abituale di cura più utilizzato ed efficace, sia negli stadi iniziali della malattia sia in quelli tardivi, dopo la quale se i pazienti raggiungono la completa assenza di segni di tumore in situ o invasivo presentano tassi di sopravvivenza favorevoli. Questo perché le pazienti affette da TNBC hanno una risposta più elevata alla chemioterapia, rispetto a pazienti affette da altri sottotipi di cancro. [18]

Sfortunatamente, però, circa l'80% delle pazienti trattate non raggiunge questa condizione per cui esse tendono ad essere sotto alto rischio di recidiva siccome sviluppano una morbilità per aver ricevuto un trattamento altamente citotossico e con benefici limitati; quindi, il trattamento non è completamente efficace per tutte le pazienti e, in particolare per quelle affette da TNBC metastatico. [19][38]

Per questo motivo si ha la necessità di sviluppare biomarcatori per poter identificare e suddividere le pazienti in modo da migliorare la pianificazione del trattamento, rendendolo il più specifico possibile. [20]

Negli ultimi anni, quindi, si sono iniziate ad applicare delle terapie sistemiche innovative (che sono ancora in fase iniziale di sperimentazione a causa dell'elevata tossicità aggiuntiva dei componenti) costituite da combinazioni multi-farmaco in ambito neoadiuvante e adiuvante con l'obiettivo di ottenere ulteriori miglioramenti nei risultati alla terapia. Un esempio di queste nuove terapie è relativo all'immunoterapia antitumorale con inibitori del checkpoint immunitario (ICI) che hanno dato una svolta alle strategie terapeutiche nel trattamento dei carcinomi negli stadi avanzati. Tuttavia, purtroppo, negli studi si è visto che le risposte all'immunoterapia nel TNBC metastatico sono modeste e applicabili solo a una

minoranza di pazienti. [21] Di conseguenza, la selezione delle pazienti affette da TNBC è fondamentale per identificare chi potrà trarre beneficio da questa terapia oppure essere indirizzato direttamente verso un altro trattamento. Infatti, se il tumore presenta la proteina PD-L1 l'immunoterapia viene utilizzata in combinazione alla chemioterapia. [3]

Inoltre, tassi elevati di espressione di PD-L1 nel TNBC sono stati identificati con il software QuPath con un approccio oggettivo, riproducibile e accurato. PD-L1 è una proteina presente sulla superficie delle cellule tumorali, nelle quali viene espressa regolando la risposta immunitaria. Si è visto che un'elevata espressione di PD-L1 è associata a un miglioramento dei risultati clinici nelle pazienti affette da TNBC nel contesto della chemioterapia standard. Infatti, l'espressione di PD-L1 da parte di cellule distinte è correlata all'istologia e alla sopravvivenza del tumore. Inoltre, un numero elevato di macrofagi CD68 con PD-L1 era associato a una migliore prognosi della sopravvivenza globale in quelle pazienti che erano state trattate con chemioterapia, [43] migliorando così il valore predittivo di PD-L1 da solo per la sopravvivenza specifica del cancro al seno. [30] [39]

Alcuni studi clinici e retrospettivi hanno dimostrato che i TIL agiscono, con un ruolo fondamentale, come agenti sinergici essendo associati ad un miglioramento della sopravvivenza delle pazienti affette da TNBC trattati con chemioterapia adiuvante a base di antracicline, pembrolizumab o altri agenti neoadiuvanti. [18] In questo tipo di tumori, spesso, può essere necessario ricorrere anche alla chemioterapia neoadiuvante, ossia somministrata prima del trattamento chirurgico per ridurre le dimensioni del tumore avendo quindi una risposta patologica completa.

2.2. Carcinoma mammario triplo negativo

Il carcinoma mammario triplo negativo, dall'inglese *Triple Negative Breast Cancer* (TNBC), rappresenta circa il 15% - 20% di tutti i tumori mammari. [17]

Diversi studi hanno dimostrato che questa malattia colpisce maggiormente la popolazione africana o quella afroamericana. Infatti, mentre l'incidenza del carcinoma mammario è inferiore nelle donne nere rispetto alle donne bianche, il tasso di mortalità è maggiore e, in particolare, risulta essere il doppio rispetto al rischio di mortalità della popolazione bianca. Infatti, le donne nere affette da TNBC hanno circa il 21% in più di probabilità di morire a causa di questa malattia, considerando al pari le condizioni di vita e i fattori correlati al trattamento. Questo perché si è visto che le pazienti nere sottoposte agli studi presentano una resistenza maggiore alla chemioterapia neoadiuvante rispetto alla popolazione bianca, che potrebbe essere dovuto o per un diverso accesso alla salute per una disuguaglianza sociale oppure perché potrebbero essere presenti delle differenze morfologiche e molecolari nel fenotipo della malattia. [23] Ecco perché è fondamentale una differenziazione delle pazienti in modo da rendere soggettiva la terapia o trattamento per renderla adeguata. [22]

All'esame anatomico-patologico, da un punto di vista architetturale e morfologico, i TNBC sono carcinomi ad alto grado con un elevato indice mitotico Ki-67 e numerose atipie nucleari che, a differenza di tutti gli altri sottotipi, risultano essere più aggressivi in quanto tendono a crescere e diffondersi molto più rapidamente e, per questo motivo, hanno una prognosi negativa e peggiore rispetto agli altri sottotipi di cancro al seno. Questo perché si hanno a disposizione anche meno opzioni terapeutiche a causa della mancanza di biomarcatori target noti che determina quindi una mancata risposta alla terapia ormonale e ai farmaci mirati per

HER2. Esso è quindi un sottotipo distinto di cancro mammario che presenta delle caratteristiche molecolari uniche su base immunohistochimica, proprio a causa dell'assenza di espressione dei recettori ormonali e della negatività di HER2.

Proprio per questo motivo il trattamento delle pazienti affette da TNBC rappresenta una sfida per la ricerca, a causa dell'aggressività e della mancanza di risposta alla terapia ormonale. Infatti, viene applicata la chemioterapia citotossica convenzionale, che rappresenta il trattamento abituale ma che purtroppo non è sempre efficace determinando lo sviluppo di recidive. Per questo motivo si è iniziato ad applicare l'immunoterapia su piccola scala ed essa ha dato dei risultati promettenti, restringendo però il campo ad una minoranza di pazienti.

Solitamente, i TNBC sono associati a pazienti in giovane età di diagnosi che presentano un alto rischio di recidiva e metastasi con elevata probabilità nei primi tre anni dalla diagnosi e di morte entro i successivi cinque anni dalla diagnosi. [3][4]

Da queste problematiche nasce la necessità di studiare questo sottotipo di carcinoma mammario effettuando delle indagini di quantificazione con lo scopo, ad esempio, di sviluppare dei biomarcatori efficaci che facilitino il trattamento e la gestione del paziente in generale.

L'Intelligenza Artificiale (IA), infatti, ha fatto dei passi da gigante nel campo delle applicazioni mediche con l'obiettivo di supportare il processo decisionale clinico andando ad estrarre informazioni specifiche sulla malattia, che non possono essere quantificate con la sola diagnosi mediante l'utilizzo del microscopio ottico.

Queste analisi, nel contesto dei TNBC, hanno la potenzialità di permettere la stratificazione del rischio per identificare le pazienti che hanno una maggior probabilità di sfociare in una recidiva o andare incontro a morte e prevedere la

risposta patologica completa della malattia, con lo scopo di sviluppare un approccio prognostico e predittivo.

L'Intelligenza Artificiale con l'analisi computerizzata delle immagini può essere impiegata per identificare delle eventuali differenze biologiche tra le popolazioni da un punto di vista della morfologia e della architettura del tessuto, e questo può essere molto utile nel caso del carcinoma mammario triplo negativo, in quanto esso colpisce in modo sproporzionato alcune popolazioni rispetto ad altre, come ad esempio le donne nere. Essa ha quindi il potenziale di fornire delle previsioni più ottimizzate e personalizzate sugli esiti della malattia TNBC che potrebbero aiutare tutte le popolazioni, e in particolare quelle più svantaggiate. [24][25]

2.2.1. Classificazione molecolare

Il carcinoma mammario triplo negativo è una malattia molto eterogenea che, secondo la classificazione di Lehmann et al. del 2011, si suddivide in sei diversi sottotipi molecolari:

1. *Basal-like 1* (BL1)
2. *Basal-like 2* (BL2)
3. *Mesenchymal* (M)
4. *Immunomodulatory* (IM)
5. *Mesenchymal stem-like* (MSL)
6. *Luminal androgen receptor* (LAR)

Questi, si contraddistinguono gli uni dagli altri in quanto possiedono caratteristiche in biologia molecolare differenti e sono caratterizzati da un diverso comportamento clinico.

I sei sottogruppi TNBC, riportati nella *Figura 1* insieme alla classe degli *Unstable*, sono stati identificati nello studio riportato da *Akhouayri et al.* attraverso l'utilizzo di uno strumento web di sotto-tipizzazione, basandosi sui profili di espressione genica ad elevato rendimento di campioni TNBC. [4]

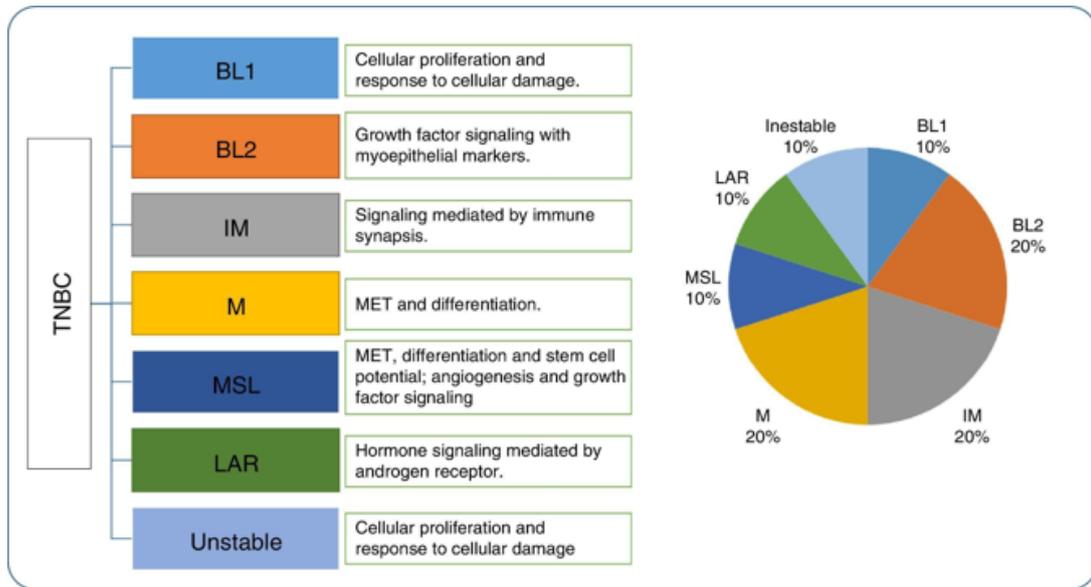


Figura 1: Classificazione molecolare dei sottotipi del carcinoma mammario triplo negativo (TNBC).

- Il sottotipo BL1 presenta due proprietà che lo caratterizzano: la proliferazione cellulare e la risposta al danno cellulare;
- Il sottotipo BL2 caratterizzato dalla segnalazione del fattore di crescita con marcatori mioepiteliali;
- Il sottotipo IM che possiede un *signaling* mediato dalle cellule immunitarie;
- Il sottotipo M specifico per una certa differenziazione;
- Il sottotipo MSL che si caratterizza con una potenziale differenziazione delle cellule staminali, da un'angiogenesi e da un *signaling* di fattori di crescita;
- Il sottotipo LAR definito da un *signaling* ormonale mediato dal recettore per l'androgeno;

- I campioni *Unstable*, classificati in una classe a parte siccome non appartengono a nessuno dei sei precedenti sottotipi e sono raggruppati come tali in quanto possiedono le seguenti caratteristiche comuni: proliferazione cellulare e risposta al danno cellulare.

2.2.2. TNBC Luminal Androgen Receptor

Il sottotipo TNBC Luminal Androgen Receptor (LAR) è un tumore molto differente da tutti gli altri sottotipi TNBC del carcinoma mammario triplo negativo perché nonostante la triplice assenza dei recettori e in particolare la negatività del recettore per gli estrogeni ER, comunque esprime il recettore per gli androgeni (AR) e/o tutti i suoi effettori a valle ed ha, per questa motivazione, un'elevata associazione con i pathways di signaling ormonale come la sintesi degli steroidi e il metabolismo degli estrogeni e degli androgeni. [4]

In questa tesi magistrale ci siamo concentrati sull'indagine del sottotipo LAR, in quanto se considerate nel lungo periodo sono le pazienti TNBC che vanno a peggiorare maggiormente rispetto alle pazienti TNBC caratterizzate dagli altri sottotipi.

2.3. Digital Pathology

La necessità per cui si mira ad avere delle analisi affidabili in un contesto digitalizzato nasce proprio dall'esigenza di creare e sviluppare dei modelli di Intelligenza Artificiale, dall'inglese *Artificial Intelligence* (AI).

L'Intelligenza Artificiale (IA) è una branca dell'informatica nella quale viene utilizzato un dispositivo informatico PC (Personal Computer) per risolvere dei problemi complessi, emulando ciò che farebbe un essere umano che si trova nella stessa situazione. Grazie allo sviluppo dell'intelligenza artificiale e alle tecniche di

apprendimento automatico, negli ultimi dieci anni molti modelli stanno prendendo parte nella nostra vita quotidiana in quanto hanno trasformato notevolmente alcuni settori in funzione delle loro alte prestazioni e di un'elevata efficienza nelle attività di previsione. Ad esempio, il *Deep Learning* (DP) è una classe di tecniche di apprendimento automatico di cui fanno parte diversi approcci attraverso i quali la comunità scientifica sta cercando di elaborare uno strumento di classificazione multi-classe su immagini WSI. [26] [5]

La capacità di acquisire scansioni digitali ad alta risoluzione che rappresentano sezioni di tessuto, denominate immagini a diapositiva intera (dall'inglese *Whole Slide Images* (WSI)) come immagini di dimensioni giga-pixel di interi vetrini, con l'utilizzo di scanners ad alta risoluzione (*High-Resolution Whole Slide Scanners*) sta trasformando la scoperta diagnostica attraverso l'elaborazione digitale posteriore che consiste nella visualizzazione, nell'analisi delle immagini digitali, l'automazione, la quantificazione e lo screening di campioni di tessuto. Quest'area è diventata ampiamente conosciuta come *Digital Pathology*, approccio che nasce come applicazione della tecnologia di imaging su vetrini interi. [28]

La *Digital Pathology* viene vista sempre di più con un crescente interesse siccome è uno strumento di supporto per la scoperta e l'utilità clinica che nasce da una serie di miglioramenti tecnici che nell'arco di 40 anni hanno trasformato la Patologia Digitale in un'applicazione affidabile.

I *Whole Slide Scanners* sono in grado di generare rapidamente delle immagini 2D molto grandi, in cui ciascun piano può contenere fino a 40 GB di dati non compressi.

La *Digital Pathology* nasce e si sviluppa proprio dalla necessità di superare la diagnosi tradizionale eseguita, con una valutazione soggettiva e manuale del medico patologo, su immagini biomediche che contengono dati di processi specifici della malattia che sono difficili da valutare con la sola indagine visiva, con l'obiettivo di supportare maggiormente le sperimentazioni su larga scala dei biomarcatori garantendo un'analisi oggettiva con elevata qualità e riproducibilità della ricerca, essenziale per una correlazione clinica affidabile e per la selezione dei biomarcatori candidati, in particolare nei campi della microscopia e delle immagini ad alto contenuto. In questo modo la valutazione patologica dei tessuti deve essere pratica, accessibile ed affidabile; quindi, funzionale per la scoperta biologica e lo sviluppo di strumenti diagnostici tissutali clinicamente rilevanti.

I progressi della Patologia Computazionale combinati con quelli dell'Intelligenza Artificiale offrono una gestione personalizzata della malattia siccome consentono di andare oltre i limiti visivi umani e quindi di ampliare le conoscenze. [13]

Nessuno strumento software, ad eccezione di QuPath, è in grado di visualizzare e analizzare delle immagini provenienti da diapositive intere e/o dati appartenenti ad immagini 2D molto grandi. Questo perché in precedenza, i software utilizzati nella *Digital Pathology* erano strumenti o piattaforme in grado di gestire i vetrini digitalizzati in alcuni formati come all'interno di una libreria e ritagliare le immagini digitali in riquadri più piccoli e quindi gestibili su cui eseguire l'analisi. [3]

L'analisi microscopica eseguita su una biopsia o un pezzo operatorio rimane comunque uno dei metodi più importanti ed utilizzati per diagnosticare la tipologia di cancro al seno, ma questo tipo di indagine prevede un'analisi specializzata da parte di un anatomo-patologo che richiede un dispendio di energie e denaro piuttosto elevati e che, a volte, portano a dei risultati non consensuali. Una diagnosi

corretta ed accurata è fondamentale per la gestione del paziente siccome apporta una differenziazione nel trattamento, per questo motivo hanno acquisito rilevanza gli algoritmi di classificazione automatica come sistemi precisi e robusti che utilizzano dei set di dati presenti sottoforma di immagini istopatologiche ottenute da vetrini colorati in Ematossilina-Eosina che sono già stati dimostrati, ma i risultati ottenuti per il momento non sono ancora ottimali per l'utilizzo in ambito clinico. Per questo motivo i sistemi di diagnosi assistita da computer (CAD) possono facilitare la procedura fornendo un sistema di supporto per una valutazione complementare e obiettiva allo specialista.

BACH (*BreAst Cancer Histology images*) è un progetto che nasce sulla sfida di *Bioimaging* 2015 con l'obiettivo di identificare una classificazione del tumore mammario differenziandolo nelle quattro classi istopatologiche principali (Normale, Benigno, In situ, Invasivo) con la localizzazione di elementi clinicamente rilevanti, confrontando i risultati tra quelli ottenuti in microscopia e quelli da immagini di vetrini interi WSI, a partire da un ampio set di dati annotati e resi pubblici. Si è dimostrato che gli algoritmi presentati hanno incrementato la classificazione automatica del tumore mammario promuovendola ulteriormente all'interno del campo della Patologia Digitale. In particolare, i partecipanti al progetto hanno proposto principalmente algoritmi basati su uno stesso pensiero che sono gli approcci di rete neurale convoluzionale, dall'inglese *Convolutional Neural Network* (CNN), i quali oggi rappresentano un approccio molto innovativo dimostrando risultati promettenti nel campo dell'analisi di immagini biomediche per tutta una serie di aspetti che rappresentano i punti di forza di questo approccio, siccome essi sono facili da impostare e non richiedono una conoscenza applicata sul campo così ricercata ed approfondita. [5]

I recenti progressi sui sistemi di acquisizione hanno permesso la scannerizzazione dei vetrini digitalizzandoli interamente, evitando la perdita di campione bioptico e fornendo in questo modo un sistema aggiuntivo di supporto per la valutazione patologica da parte degli esperti.

Nel campo dell'analisi delle immagini biomediche, gli approcci di *Deep Learning* si stanno integrando agli approcci standard perché consentono di ottenere dei buoni risultati con elevate prestazioni riducendo significativamente la necessità di conoscenza specifica sul campo. [5]

I *software* che riescono ad utilizzare le immagini WSI rappresentano principalmente delle soluzioni costose che riescono ad elaborare e trattare solo piccole porzioni di quelle immagini WSI andando a definire quindi delle piccole regioni di interesse (ROI), dall'inglese *Small Region Of Interest* (ROIs). Per questo motivo la ricerca è andata avanti progredendo nell'intento di individuare dei validi sistemi alternativi che potessero elaborare dei dati a partire da immagini WSI intere di grandi dimensioni, senza la necessità di selezionare manualmente delle regioni di interesse specifiche o di una potenza di calcolo estremamente elevata garantendo allo stesso tempo, comunque, dei buoni risultati di prestazione.

Questo ha dato vita all'estensione della Patologia Digitale nella Patologia Computazionale, che consiste proprio nell'utilizzo di strumenti di intelligenza artificiale per estrarre informazioni quantitative da immagini di campioni istopatologici.

2.3.1. Scanner

La S.C. di Anatomia Patologica dell'Ospedale degli Infermi di Biella (ASL – BI) è dotata di uno scanner per diapositive digitali ad alta risoluzione e rendimento

elevato che possiede la capacità di elaborare automaticamente un totale di 320 vetrini (40 vetrini x 8 rack) di dimensione pari a 76 mm x 26 mm e uno spessore che può variare dai 0,9 mm ai 1,2 mm. Lo scanner, che presenta un peso approssimativo di circa 209 kg, converte un'area pari a 15 mm x 15 mm di un vetrino in una immagine a colori da 1,1 Giga-Pixel in 30 secondi ad un ingrandimento 20x e 35 secondi ad un ingrandimento 40x e ha una risoluzione di scansione di 0,46 $\mu\text{m}/\text{pixel}$ (ingrandimento 20x) e di 0,23 $\mu\text{m}/\text{pixel}$ (ingrandimento 40x).

Il modello di scanner è il NanoZoomer-XR C12000 dell'azienda produttrice giapponese Hamamatsu Photonics K.K. ed è in grado di ottenere immagini a diapositiva intera, dall'inglese *Whole Slide Imaging* (WSI), riducendo al minimo il carico di lavoro e il tempo di scansione dei vetrini necessario ottenendo una scansione automatica, continua e rapida.

La serie NanoZoomer è una famiglia di scanner per diapositive intere che è stata introdotta a partire dall'anno 2005. Nel tempo, questi scanner sono risultati essere dei dispositivi robusti e stabili, capaci di convertire i vetrini fisici in immagini digitali ad alta risoluzione mediante scansione ad alta velocità.

Essi sono dotati di numerose funzioni di scansione che si differenziano in base alla tipologia di campione che viene inserito al loro interno. In particolare, il NanoZoomer-XR C12000, che risale all'anno 2012, esegue la scansione di vetrini che possono essere sia istologici sia citologici in quanto è dotato di due funzioni di acquisizione d'immagine: quella eseguita in monostrato su un singolo *layer* e la multi-layer Z per una migliore visualizzazione di campioni spessi.

Lo scanner NanoZoomer-XR C12000 è dotato di caratteristiche uniche di nuova concezione, che permettono: una scansione priva di sfocature ed errori, una messa a

fuoco nitida sull'intero campione producendo immagini ben definite e chiare di alta qualità con una messa a fuoco automatica grazie alla tecnologia dinamica di pre-focalizzazione, dall'inglese *Dynamic Pre-Focusing* (DPF), metodo utile per tracciare e mantenere il campione a fuoco durante la scansione soprattutto per campioni che hanno superfici focali ruvide. Esso analizza i fuochi prima di catturare l'immagine e aggiusta l'obiettivo al focus nell'immagine in tempo reale.

È in grado di eseguire tre modalità di scansione:

- *full automatic batch scan*, che corrisponde alla modalità totalmente automatizzata in cui una volta che l'operatore ha caricato i vetrini nello scanner e premuto il pulsante *Start Batch*, esso localizza i campioni nei vetrini e automaticamente setta i parametri di scansione e inizia a scannerizzare i vetrini fino a quando tutte le diapositive sono state completate;
- *semi-automatic batch scan*, modalità che permette all'operatore di settare i parametri di scansione manualmente prima di eseguire la scansione automatica dei vetrini;
- *manually defined scanning parameters single scan*, modalità in cui l'operatore setta manualmente e definisce dettagliatamente i parametri di scansione per ogni singolo vetrino.

Inoltre, lo scanner è dotato di un controllo automatizzato della qualità delle intere immagini che consiste nel generare dei focus per ogni diapositiva che vengono revisionati dall'operatore in tempo reale sul monitor, siccome è lo strumento che valuta immediatamente la qualità con cui viene elaborata e poi prodotta l'immagine scansionata ed esegue una seconda scansione automatica in caso di necessità. In questo modo si riduce al minimo la manualità dell'operatore liberando l'utente dai

lunghe controlli di qualità manuali che, spesso, vengono condotti dopo la scansione per evitare errori, dovuti allo sporco sul vetrino o alla sezione del campione.

Infine, lo scanner possiede un sistema di calibrazione automatica (auto-calibrazione) attraverso il quale si eseguono periodicamente ed automaticamente calibrazioni di routine per l'intensità della luce, il bilanciamento del bianco e l'ombreggiatura, utilizzando uno specifico vetrino di calibrazione con l'obiettivo di mantenere sempre le condizioni di sistema ottimizzate e ottenere le migliori immagini possibili ogni volta che si esegue la scansione. Questo perché l'intensità della luce può variare nel tempo o a causa della temperatura ambiente.



Figura 2: Postazione di Digital Pathology, presente nella S.C. di Anatomia Patologica dell'ASL-BI, costituita dallo scanner per la digitalizzazione dei vetrini istologici e dal pc collegato ad esso e dedicato alla visualizzazione delle immagini WSI.

Lo scanner, presente in *Figura 2*, è collegato alla rete ospedaliera tramite il suo *software* che riceve i dati dei vetrini da scannerizzare direttamente dal sistema informativo WinSap del laboratorio di Anatomia Patologica, dall'inglese *Anatomic Pathology Laboratory Information System* (APLIS), utilizzando un protocollo web standardizzato attraverso il quale avviene lo scambio di dati tra applicazioni e sistemi diversi in modo da tracciare il vetrino che verrà in seguito scannerizzato. WinSap, che è un Sistema Informatizzato di Reparto, permette di avere una tracciabilità del materiale in tutte le fasi di preparazione e analisi del percorso diagnostico che si conclude con la redazione di una diagnosi e la refertazione diagnostica con la stesura del referto.

Una volta ottenute le immagini digitali, esse vengono salvate come tipologia di file *ndpi* in due sistemi: NAS (*Network Attached Storage*) che possiede una capacità di 21 TB di memoria e RAID-6 che protegge i dati degli utenti dagli errori di salvataggio nella memoria.

Esse possono essere visualizzate rapidamente sul monitor di un computer mediante l'utilizzo di un *software* di visualizzazione d'immagine *NDP.view2*, della stessa azienda produttrice, che mediante la tecnica brevettata della mappa di navigazione permette di migliorare l'efficienza e avere quindi un ambiente di visualizzazione nettamente più funzionale rispetto a quello di un microscopio ottico. Infatti, il *software* possiede alcune caratteristiche che il microscopio non potrà mai offrire, ossia: una pulita visualizzazione a schermo intero, uno zoom continuo, una navigazione contemporanea attraverso diverse immagini *ndpi* con una visualizzazione multipla che permette di effettuare dei confronti, delle annotazioni e delle misure.

Oppure, le diapositive scansionate possono essere visualizzate da remoto sulla piattaforma web *Cloud Pathology*, su cui vengono caricate in automatico subito dopo essere state scannerizzate, al fine di poter rendere la loro lettura accessibile ai medici anatomo-patologi in rete in modo da permettere la condivisione dell'interpretazione diagnostica con altri Centri di Anatomia Patologica che dispongono anch'essi delle credenziali di accesso a tale sistema. [6] [7]

2.3.2. QuPath

QuPath, abbreviazione di *Quantitative Pathology*, è un nuovo *software open-source* gratuito per l'analisi di bio-immagini che è stato creato e sviluppato in ambito accademico presso il *Centre for Cancer Research & Cell Biology della Queen's University* di Belfast come una parte di ricerca e sviluppo all'interno di un programma di ricerca in *Digital Pathology*. Esso nasce dalla crescente necessità ed esigenza di visualizzare immagini digitali (bio-immagini) complete di tessuti di vetrini interi, dall'inglese *Whole Slide Imaging (WSI)*, riuscendo ad analizzarne molteplici contemporaneamente e nella loro interezza, nonostante le elevate dimensioni di grandezza in GB dei *file* contenenti miliardi di pixel.

L'obiettivo di questo potente strumento è quindi quello di effettuare un'analisi delle immagini tissutali (*tissue-based image analysis*), in un contesto digitalizzato come applicazione della *Digital Pathology*, rispondendo con un output di patologia quantitativa basata perciò sulla quantificazione di determinati parametri oggetto di analisi, soprattutto nel campo della ricerca e della diagnostica sul cancro con applicazioni cliniche di diagnosi e prognosi, soddisfacendo la richiesta di analisi affidabili e riproducibili ad alto ingrandimento ed incrementando e potenziando sempre di più la quantificazione dei biomarcatori su larga scala e quindi l'analisi dei biomarcatori ad alto rendimento in immunohistochimica.

Il *software* è stato progettato in modo tale da essere dotato di un'interfaccia attraverso la quale gli utenti, sia ricercatori sia diagnostici, possono effettuare delle attività più o meno complesse come ad esempio: il rilevamento, la classificazione e il conteggio automatico di molte migliaia di cellule e oggetti, il tutto garantendo un'alta flessibilità e un'alta produttività nella capacità di elaborazione automatizzata.

Esso è probabilmente il programma di analisi di immagini digitali più utilizzato al mondo, in quanto, considerata la sua capacità di analisi di bio-immagini WSI, QuPath è un *software* multipiattaforma cioè che può essere esteso a diverse discipline nel campo della ricerca biomedica, non solo alla patologia.

Principalmente, viene impiegato nella *Digital Pathology* dai ricercatori per ottenere dei dati quantitativi grezzi che devono essere poi in seguito elaborati, a partire dall'analisi di immagini WSI colorate in Ematossilina-Eosina EE (dall'inglese *Haematoxylin and Eosin* (H&E)) o in immunistoichimica campo chiaro o fluorescente (dall'inglese *Brightfield or Fluorescent*) senza la necessità di doverle ritagliare o sotto-campionare per ridurre le dimensioni dei *file* con l'obiettivo di un'analisi successiva.

QuPath è in grado di aprire e visualizzare simultaneamente più immagini WSI, acquisite attraverso diverse tipologie di scanner per la patologia digitale, e in seguito eseguire delle analisi applicando in un'elaborazione riproducibile dei processi automatizzati, attraverso i quali rilevare i parametri di interesse dell'indagine che si sta eseguendo visualizzando le misurazioni e, infine, esportando i dati ottenuti con l'analisi di un ampio set di immagini in quella che viene definita come elaborazione *batch*.

QuPath è dotato di diverse funzionalità, tra le quali: la localizzazione subcellulare, l'identificazione tumorale automatizzata, la suddivisione delle immagini per eseguire un'analisi più rapida, la stima dell'intensità di colorazione di una specifica zona, la capacità di scambiare i dati con altri *software open-source* e la capacità di leggere e visualizzare molteplici tipologie di immagini. Tutte queste funzioni possono essere raggruppate e organizzate all'interno di uno stesso flusso di lavoro che viene creato con la cronologia dei comandi in maniera automatizzata.

Si crea, quindi, un progetto di analisi all'interno del quale vengono caricate molteplici immagini per facilitare un'elaborazione *batch* ed eseguire l'analisi mediante la creazione di un flusso di lavoro che consente l'automatizzazione dell'analisi, siccome la cronologia dei comandi delle funzioni utilizzate viene visualizzata in uno *script* che contiene tutti i passaggi dei comandi applicati alle immagini.

In particolare, con QuPath è possibile eseguire lo *script* su tutte le immagini facenti parte del progetto con l'obiettivo di analizzarle in maniera riproducibile. Lo *scripting* delle funzionalità risulta essere quindi un modo per velocizzare notevolmente l'analisi delle immagini ed è per questo motivo che QuPath è un programma dotato di grande elasticità e flessibilità siccome è in continua evoluzione e cambiamento. [1]

QuPath è dotato di un'alta capacità di rilevamento per ottenere risultati in modo molto più rapido e riproducibile, spesso con maggiore precisione e meno errori. Viene spesso utilizzato per applicazioni di Patologia Digitale perché offre un potente set di strumenti per lavorare con immagini di diapositive intere, ma può essere applicato anche a molti altri tipi di immagini.

Le caratteristiche includono:

- Potenti strumenti di annotazione e visualizzazione che utilizzano una moderna interfaccia JavaFX;
- Algoritmi integrati per attività comuni, incluso il rilevamento di cellule e tessuti;
- Machine learning interattivo, sia per la classificazione degli oggetti che dei pixel;
- Compatibilità con altri strumenti o tool open-source, tra cui: ImageJ, OpenCV, Java Topology Suite e OMERO;
- Supporto per molti formati di immagine tramite Bio-Formats e OpenSlide, comprese immagini di diapositive intere e dati multiplex;
- Scripting groovy per la personalizzazione e l'interrogazione dei dati più approfondita.

La versione attualmente presente e che verrà utilizzata per l'analisi è la 0.4.4. Per motivi di riproducibilità, non è necessario mischiare l'analisi delle immagini utilizzando diverse versioni del software all'interno dello stesso progetto o studio di analisi. Questo perché i risultati dei comandi (ad esempio il rilevamento cellulare o la classificazione delle cellule) o degli script potrebbero essere diversi. [2]

3. Materiali e metodi

Nella sezione 3. di “Materiali e Metodi” viene rappresentata la cronologia con la quale si è svolto il flusso di lavoro del progetto di ricerca, andando quindi a descrivere nello specifico tutte le fasi o step appartenenti alla sequenza temporale. In particolare, il progetto è stato caratterizzato da diversi step di lavoro:

1. Preparazione dei dati, fase in cui è avvenuta la scansione dei vetrini ottenendo le corrispondenti immagini WSI (*Whole-Slide Images*), la revisione dei casi e l’archiviazione del set di dati in una memoria esterna;
2. Analisi dei dati, fase nella quale si è utilizzato il *software* QuPath per sviluppare degli algoritmi di analisi sulle immagini WSI;
3. Valutazione dei risultati, fase nella quale vengono analizzati e valutati i risultati ottenuti nella fase 2 mediante l’utilizzo di analisi statistiche.

3.1. Popolazione di studio: descrizione dei dati

I 9 casi su cui questa tesi magistrale è stata svolta fanno parte di uno studio sul carcinoma mammario triplo negativo in cui ha preso parte anche la Fondazione Edo ed Elvo Tempia dell’Ospedale di Biella. [4]

In questo studio sono stati utilizzati due set di dati TNBC: il primo è stato recuperato dal *Gene Expression Omnibus* (GEO) e si riferisce al sequenziamento dell’intero trascrittoma (RNA-seq) eseguito su biopsie di ricerca pretrattamento dal *BrightNess phase III study* (AFT-04). Questo set di dati (GSE164458) consiste in valori di espressione RNA-seq log-normalizzati di tumori allo stadio clinico da II a III e che verrà chiamato GEO-TN. Il secondo set di dati è stato preso dal *Genomic Data Commons* (GDC) *Data Portal of the National Cancer Institute* e si riferisce al progetto sull’atlante del genoma del cancro (TCGA) nel quale sono stati selezionati

63 campioni TNBC basandosi sul loro profilo immunohistochimico (ER-, PR-, HER2-) su un totale di 1093 carcinomi mammari invasivi. Questo set di dati, denominato TCGA-TN, contiene valori di espressione di RNA-seq log-normalizzati e dati clinici.

Il terzo set di dati GSE206912, caricato nell'archivio pubblico GEO, corrisponde al set di dati italiano del quale fanno parte 72 campioni di pazienti italiane TNBC che sono state trattate chirurgicamente presso due strutture sanitarie italiane: l'Ospedale degli Infermi di Biella e il Policlinico Gemelli di Roma, nei quali sono state operate 19 pazienti e 53 pazienti rispettivamente.

Il set di dati, chiamato *Italian-TN*, è stato sottoposto a profilazione dell'espressione genica presso il Laboratorio di Genomica della Fondazione Edo ed Elvo Tempia presente a Biella. La raccolta di campioni è stata approvata dai Comitati Etici di Novara e del Policlinico Gemelli, rispettivamente con Prot. 861 CE 149/19 e Prot. 3559. In seguito, dopo aver reso nota l'area di interesse tumorale evidenziandola tracciandone i confini sui vetrini EE da parte di un Medico Anatomo-Patologo esperto, l'RNA totale (RNAtot) è stato isolato a partire da sezioni macro dissezionate utilizzando il *Kit Agilent Absolutely RNA FFPE*, è stata fatta una trascrizione inversa convertendolo al corrispondente cDNA trascritto in vitro con il Kit di Amplificazione Sigma *TransPlex Whole Transcriptome*. Il cDNA è stato quindi amplificato ed etichettato con il kit di etichettatura del DNA *Agilent SureTag*, ibridato mediante il *Kit Agilent Gene Expression Hybridization* sull'intero genoma *SurePrint G3 Human GE 8 x 60 KV3 microarrays* contenenti sonde per 26803 *coding* RNAs e 30606 lncRNAs. Successivamente i vetrini sono stati lavati con il *Kit Gene Expression Wash Buffer*, scansionati con lo scanner *Agilent version C*. e infine l'analisi dell'immagine *array* è stata eseguita mediante l'utilizzo del *Software*

Agilent Feature Extraction v12.1, da cui i dati di espressione sono stati elaborati e processati con la funzione di sottrazione dello sfondo seguita dalla normalizzazione quantile dell'*array* utilizzando LIMMA (*Linear Models for Microarray Analysis*) all'interno del Software R per Statistica nella versione v.4.1.0. Anche questo terzo set di dati contiene intensità log-normalizzate. [4]

L'uso di set di dati pubblici è ad oggi fortemente consigliato in quanto consente ai ricercatori di essere incoraggiati nel confrontare tra di loro diversi algoritmi o modelli su set di dati standardizzati, e quindi affidabili, sicuri e riproducibili. [3]

3.2. Previsione dei sottotipi TNBC

I dati pre-elaborati presenti nei tre set di dati GEO-TN, TCGA-TN e Italian-TN sono stati caricati sullo strumento online *TNBCtype online tool* che ha, in primo luogo, indagato sulla presenza di qualsiasi campione positivo ai recettori ormonali rimuovendolo in caso affermativo. In seguito, lo strumento ha calcolato il Coefficiente di Correlazione di Spearman tra ogni campione e i sei centroidi dei sottotipi TNBC determinati in precedenza, con lo scopo di assegnare ogni campione al sottotipo più correlato. La denominazione *Unstable* (UNS) è assegnata a quelli che risultano essere instabili, ossia al 26% della coorte su un totale di 72 campioni, in quanto presentano un coefficiente di correlazione molto basso quindi non statisticamente significativo con qualsiasi sottotipo. [4]

3.3. Determinazione del profilo genetico

La determinazione del profilo genetico si è basata sul calcolo differenziale dei geni espressi (DEGs) specifici per ogni sottotipo TNBC. Per fare ciò sono stati selezionati due metodi distinti. In particolare, il primo metodo si è basato sull'utilizzo di LIMMA nel software R e l'espressione differenziale dei geni è stata

fatta applicando un cut-off, mentre il secondo metodo si è basato sul test U di Mann-Whitney (MWU test). Il rilevamento dell'espressione genica differenziale è stato fatto applicando un cut-off alla differenza dell'espressione della mediana tra i vari sottogruppi, rispettivamente per i geni up-regolati e down-regolati.

Tutti e tre i set di dati sono stati sottotipizzati utilizzando il TNBCtype online tool. Dai risultati si evince che: i sottotipi IM e M sono i più prevalenti, a differenza invece dei sottotipi BL2 e LAR che invece sono i meno frequenti. Questo ci dà un'idea sullo squilibrio dei sottogruppi.

Inoltre, con l'applicazione dei due test utilizzati, si è arrivati a determinare un'espressione genica differenziale con due pannelli di geni: il primo con 120 geni up-regolati e il secondo con 81 geni down-regolati. In particolare, 12 componenti dei venti geni LAR up-regolati sono direttamente regolati dall'Androgen Receptor.

Inoltre, delle 19 pazienti operate presso l'Ospedale degli Infermi di Biella, 9 sono state identificate come sottotipo LAR e che sono il nostro oggetto di studio.

3.4. Ricerca e acquisizione dati: selezione dei casi

In questo studio hanno preso parte 9 pazienti a cui è stato diagnosticato il carcinoma mammario triplo negativo presso l'A.S.L. di Biella dove, successivamente, sono state sottoposte a intervento chirurgico nella fascia temporale che va dall'anno 2013 all'anno 2017.

Dopo l'individuazione di quali pazienti presentavano il sottotipo LAR, la ricerca di esse nel sistema informatico è avvenuta mediante l'utilizzo del *software* gestionale WinSap, utilizzato dalla Struttura Complessa (SC) di Anatomia Patologica dell'Azienda Ospedaliera come programma di refertazione. Nello specifico, le pazienti sono state ricercate innanzitutto all'interno del programma attraverso la

selezione dell'anno di diagnosi e come chiave di ricerca alla voce "Tipologia d'esame (descrizione)" è stata inserita la parola "mammella" e si sono utilizzati due codici di ricerca:

- G5 – 017: Mammella – mastectomia o quadrantectomia con linfadenectomia
- G5 – 018: Mammella – mastectomia semplice – quadrantectomia o resezione parziale

Questi, con l'obiettivo di individuare tutte le pazienti che, divise per anno, sono state sottoposte a intervento chirurgico e i cui campioni risultano essere provenienti da Pezzi Operatori (PO) in modo da avere più materiale possibile a disposizione su cui lavorare e ottenere così dei risultati completi e quindi affidabili.

In seguito, sono stati analizzati i referti di tutte le pazienti presenti all'interno degli elenchi risultanti dai codici di ricerca e divisi per anno, facendo una selezione delle sole pazienti triple negative, ovvero coloro che presentavano una triplice negatività (ER- , PR- , HER2-) alla voce dei fattori prognostici e un'ulteriore selezione inserendo i numeri istologici corrispondenti alle nostre pazienti LAR di interesse. Una volta acquisiti e analizzati i referti delle pazienti, si è effettuata la ricerca dei campioni all'interno del locale di archiviazione dei vetrini istologici utilizzando come chiave di ricerca i numeri istologici presenti sui referti, univoci per ogni specifico esame a cui il paziente viene sottoposto e quindi, in questo caso, all'intervento chirurgico.

Successivamente, per ogni paziente sottoposta allo studio, sono state selezionate due sole tipologie di campione, ossia il vetrino istologico in Ematossilina – Eosina (EE) (che solitamente è risultato essere quello su cui era stata fatta in precedenza la

tecnica di immunistoichimica per la diagnosi oncologica siccome si trattava del vetrino con la sezione istologica più adeguata) e il vetrino Immuno Ki-67.

3.5. Attività istologica del laboratorio di Anatomia Patologica

In questo Sottocapitolo 3.5 si intende analizzare tutte le fasi che hanno inizio con l'intervento chirurgico in sala operatoria e terminano con le attività di preparazione dei campioni biologici che portano alla realizzazione dei vetrini da sottoporre ad indagine nella S.C. di Anatomia Patologica. [8] [9]

3.5.1. Prelievo

Le pazienti prese in esame in questo studio sono state tutte sottoposte a intervento chirurgico da parte del chirurgo senologo con una mastectomia, che prevede il prelievo della mammella interessata e i relativi linfonodi sentinella.

Una volta che il materiale viene prelevato, esso viene raccolto all'interno di contenitori specifici e identificato con i dati del paziente e della sede del prelievo con allegati i fogli di lavoro e di richiesta. In seguito, il materiale viene inviato alla S.C. di Anatomia Patologica.

3.5.2. Accettazione

La sala di accettazione è il luogo che riceve e accetta i campioni inviati dall'esterno, i quali possono provenire: dal blocco operatorio, dai reparti, dagli ambulatori, dai distretti esterni all'azienda.

I campioni vengono controllati al momento della ricezione per rilevare eventuali non conformità e, nel caso in cui siano idonei, subiscono un trattamento da parte del personale addetto all'attività di accettazione.

Infatti, bisogna verificare che i campioni soddisfino i seguenti requisiti come verifica di idoneità:

- Documentazione presente e compilata correttamente con timbro e firma del medico richiedente, specifica del materiale e delle notizie cliniche necessarie;
- Integrità dei contenitori e assenza di contaminazione esterna;
- Etichetta di identificazione del paziente presente e leggibile;
- Corrispondenza tra dati anagrafici o numero identificativo, data, tipologia del campione rispetto a ciò che è riportato nella documentazione allegata;
- Congruenza campione del materiale in esame con l'esame richiesto, per quanto riguarda la rispondenza tra il numero di provette o contenitori o vetrini rispetto a ciò che è riportato nella scheda di accompagnamento del materiale prelevato;
- Temperatura e tempistiche idonee dal prelievo per la tipologia di campione e dal tipo di esame richiesto;
- Assenza di alterazioni della qualità del campione;
- Quantità sufficiente di materiale campionato per l'esame richiesto.

Al termine di questa verifica, i campioni idonei vengono accettati, assegnati ad un numero istologico con QR-code univoco sul programma WinSap con il quale si assegna ad ogni numero identificativo il materiale in esame (Citologici Cervico-Vaginali (CCV), Citologici Extra-Vaginali (CEV), istologici, autopsie) e tutte le notizie cliniche riportate sul foglio di accettazione, creando un protocollo specifico per ogni campione. In seguito, i campioni vengono avviati al processo analitico di analisi del materiale secondo le specifiche procedure.

3.5.3. Fissazione

Il processo analitico prosegue con la fase di fissazione, durante la quale il campione viene messo a fissare con un liquido fissativo nel quale viene immerso all'interno del proprio contenitore. Questo liquido fissativo d'elezione è la formalina tamponata neutra (formaldeide 4% e tampone acetato 0,05 M).

Il tempo minimo richiesto per la fase di fissazione è pari a 24 ore ma esso può variare a seconda della grandezza del campione. Questo perché più il campione è grande più dovrà essere lasciato a fissare affinché il fissativo agisca completamente in modo da fissare interamente il campione e consentire ad esso di sopportare tutta una serie di stress fisico-chimici a cui verrà sottoposto nelle successive fasi di disidratazione, inclusione e sezionamento.

3.5.4. Riduzione

Tra le operazioni preliminari, il tecnico prepara per i campioni istologici le relative cassette che sono di due colori:

- Rosa, se si tratta di biopsie gastriche, ossee o di colorazioni aggiuntive rispetto alla colorazione EE e richieste dal dirigente medico;
- Bianco, per tutti gli altri tipi di prelievo.

Le cassette vengono numerate riportando: le ultime due cifre dell'anno, il numero istologico corrispondente al paziente che gli è stato assegnato automaticamente dal computer in fase di registrazione e il numero aggiuntivo in caso di prelievo multiplo.

Durante la fase di riduzione, il medico anatomo-patologo effettua la descrizione macroscopica (che verrà in seguito riportata sul referto insieme alla diagnosi) di ogni campione registrando i dati del paziente sul sistema informatico WinSap

associandoli al numero istologico corrispondente con data, cognome e nome del paziente, struttura di provenienza, sede anatomica, tipo di materiale; mediante l'aiuto di un tecnico di laboratorio che controlla la corrispondenza tra i dati scritti sulla richiesta e quelli del contenitore.

Durante la descrizione macroscopica, le caratteristiche riportate sono le seguenti:

- Organo o parte di esso con strutture anatomiche riconoscibili;
- Dimensioni e peso dell'organo e delle lesioni;
- Numero delle lesioni;
- Rapporti anatomici delle lesioni, in particolare la distanza dai margini;
- Sede dei prelievi, assegnando ad ognuno un codice numerico;
- Segnalazione di eventuali prelievi per tecniche speciali.

Successivamente, in base alla tipologia e alle dimensioni del materiale in esame, l'anatomo-patologo decide se procedere al campionamento di esso oppure no in corrispondenza della stazione *ProPath* di riduzione dei pezzi chirurgici in preparazione all'inclusione.

I campioni, in seguito, vengono trasferiti in bio-cassette adibite che sono posizionate in cestelli immersi nel fissativo della formalina (in cui rimangono sotto cappa per una notte) siccome necessitano di ulteriore fissazione prima della fase di processazione di routine, e sulle quali viene riportato il numero istologico univoco e il QR-Code corrispondente. Solitamente alla formalina viene aggiunta qualche goccia di eosina in modo che sia visibile il contrasto di colore dopo la fase di processazione, con l'obiettivo di facilitare la fase di inclusione successiva.

Al termine del campionamento nella fase di riduzione, il materiale residuo viene mantenuto a fissare sigillandolo con lo strumento della macchina sottovuoto e in

attesa di diagnosi viene riposto negli armadi aspiranti. Inoltre, viene effettuata la decontaminazione, pulizia e disinfezione di tutti gli strumenti metallici impiegati per la riduzione utilizzando una soluzione acquosa decontaminante di polifenoli al 3% per una tempistica di almeno 10 minuti.

3.5.5. Processazione

La fase di processazione avviene in seguito alla fase di riduzione e nella quale le bio-cassette vengono inserite nella camera di processazione del processatore, dopo aver effettuato il lavaggio della camera, e vengono caricate con l'utilizzo dei cestelli coprendoli con appositi coperchi. Il processatore effettuerà quindi la preparazione dei campioni per la fase di inclusione.

Questa fase si basa sull'utilizzo di uno specifico strumento che è il Processatore Leica APS 6025, posto all'interno di un box aspirante. Con questo strumento si possono utilizzare diversi programmi di analisi: giornaliero, week-end e festivo. Normalmente, il programma maggiormente utilizzato è il programma giornaliero che ha una tempistica di circa 24 ore siccome il carico del processatore avviene al mattino e lo strumento lavora fino al mattino successivo, quando viene scaricato il cestello contenente i campioni istologici processati.

1	Formalina	1 ora	37°
2	Acqua	2 sec	
3	Etanolo 70°	40 min	45°
4	Etanolo 80°	40 min	45°
5	Etanolo 95°	40 min	45°
6	Etanolo Assoluto Denaturato	1 ora	45°
7	Etanolo Assoluto Denaturato	1 ora	45°

8	Etanolo Assoluto Denaturato	1 ora	45°
9	Xilolo	1 ora	45°
10	Xilolo	1 ora	45°
11	Xilolo	1 ora	45°
12	Paraffina	1 ora	65°
13	Paraffina	1 ora	65°
14	Paraffina	1 ora	65°

Tabella 1: Schema dei passaggi della fase di processazione del Processatore Leica APS 6025 presente nella S.C. di Anatomia Patologica presso l'ASL-BI.

Secondo lo schema del programma di routine chirurgica riportato in *Tabella 1*, le bio-cassette inizialmente immerse in formalina subiscono un lavaggio con acqua in modo che le molecole d'acqua raggiungano tutti gli spazi del campione occupandoli. Dopo questa fase iniziale, il processatore esegue una prima fase di disidratazione con l'utilizzo di un particolare alcol che è l'Etanolo, presente in tre fasi a concentrazione crescente, che progressivamente sostituisce gli spazi occupati precedentemente dalle molecole d'acqua. Successivamente, il processatore effettua una fase di chiarificazione o diafanizzazione durante la quale i campioni disidratati vengono immersi nell'agente diafanizzante Xilolo per essere chiarificati. Infine, il processatore termina il programma con la fase di impregnazione utilizzando la paraffina liquida.

3.5.6. Inclusione

In questa fase i campioni fissati e impregnati di paraffina vengono inclusi all'interno di blocchetti di paraffina per creare un mezzo solido e resistente che permette il successivo taglio dei campioni in sezioni sottili.

Sul bancone di inclusione Asem, l'apparecchiatura utilizzata è la Centralina di Inclusione o Inclusore Leica EG 1150 H costituito da tre parti: a sinistra una piastra refrigerante ad una temperatura di -15° (Leica EG 1150 C), al centro una stazione di piastra riscaldata costituita da un dispensatore di paraffina liquida mantenuta ad una temperatura di $+58^{\circ}$ (*Dispenser Unit*), a destra uno strumento suddiviso in due parti (*Thermal Unit*) e utilizzato per mantenere la camera in cui sono tenute le formelle pronte per l'uso a 40° (parte posteriore dello strumento) e per tenere liquida la paraffina a 58° nella camera di stazionamento (parte anteriore dello strumento).

In particolare, i cestelli contenenti le bio-cassette dei campioni istologici provenienti dal processore vengono disposti nelle camere di stazionamento dell'Inclusore. Alcune, successivamente, vengono messe nell'unità termica in corrispondenza della piastra riscaldata *Dispenser Unit*, da cui una alla volta vengono aperte e con l'utilizzo di una pinzetta viene prelevato il campione istologico, messo e pressato per garantire la corretta adesione dell'intera superficie sul fondo della formella di metallo di dimensioni idonee, nella quale in precedenza viene dispensata la paraffina liquida dal dispensatore dell'Inclusore che la mantiene liquida a 58° .

In seguito, le formelle di metallo vengono chiuse con le bio-cassette posizionate al di sopra con aggiunta ulteriore di paraffina liquida e si lasciano raffreddare sul piano freddo della centralina, che rappresenta la *Cool Unit* dell'Inclusore, per una tempistica di almeno 5 minuti e che determinerà la formazione dei blocchetti. Una volta raffreddati, gli inclusi vengono tolti dalle formelle metalliche e posizionati in ordine numerico crescente su vassoi raffreddati presi dal congelatore a -20° , in cui vengono riposti nuovamente per il tempo necessario e sufficiente affinché acquisiscano le caratteristiche di robustezza e solidità idonee per la fase di taglio.

3.5.7. Taglio

In questa fase l'operatore esegue il taglio del materiale in esame, incluso precedentemente nei cubetti di paraffina, mediante l'utilizzo del microtomo con lo scopo di ottenere delle sezioni istologiche sottili dello spessore di circa 4 μm che, in seguito, vengono colorate nella fase successiva di colorazione per la valutazione morfologica.

L'apparecchiatura utilizzata per effettuare la fase di taglio può variare tra due diverse tipologie di strumenti:

- Microtomo a slitta (Jung SM 2000 R e Leica SM 2010 R)
- Microtomo rotativo (Leica RM 2145)

In particolare, il microtomo a slitta è caratterizzato da un movimento completamente manuale secondo cui la lama viene fatta scorrere in senso orizzontale lungo una robusta guida metallica mentre il blocchetto di paraffina è fissato. Il microtomo rotativo, invece, è semimotorizzato con un movimento circolare manuale della manovella per cui la lama è fissa e il blocchetto viene fatto scorrere verticalmente su e giù contro la lama.

L'operazione di taglio prevede i seguenti step di procedura:

1. Accensione del dispositivo di essiccamento *Dryer* 3F ad una temperatura regolata di 50°;
2. Preparazione dei bagni stendifette termostatati Medite TFB35 o Leica HI2125 con acqua distillata ad una temperatura regolata di circa 40° (siccome temperature superiori scioglierebbero la paraffina disperdendo il campione in acqua, mentre temperature inferiori non permetterebbero una distensione corretta ed ottimale delle sezioni);

3. Disposizione della lama nell'apposito alloggiamento del microtomo, regolandone l'angolatura;
4. Posizionamento e fissaggio del cubetto di paraffina con il morsetto nella posizione di lavoro ottimale;
5. Regolazione dell'avanzamento della lama fisso a 3-4 μm ;
6. Realizzazione della prima fase di taglio grossolana, detta sgrossamento, della paraffina in eccesso;
7. Realizzazione della seconda fase di taglio vera e propria, quando tutto il materiale è sullo stesso piano, prelevando le sezioni sottili ottenute complete dell'intero campione con delle pinzette.

3.5.8. Allestimento vetrini

La fase di taglio è immediatamente seguita dalla fase di allestimento dei vetrini. In particolare, una volta che le sezioni sottili vengono prelevate singolarmente con la pinzetta dal microtomo, esse vengono disposte con l'aiuto di un pennello sulla superficie dell'acqua contenuta nel bagno stendifette termostato, con lo scopo di stenderle in modo ottimale. Successivamente, con l'utilizzo di un vetrino portaoggetti precedentemente numerato, avviene l'allestimento vero e proprio inserendo il vetrino in acqua sotto di esse, con il quale si effettua il recupero della sezione sottile, facendola aderire al piano del vetrino stesso con l'ausilio del pennellino. Infine, si dispone il vetrino sul bordo del bagno stendifette per eliminare l'eccesso di acqua e sul quale si appone, sulla banda satinata, la corrispondente etichetta riportante il numero istologico, il numero del blocchetto tagliato e la specifica colorazione da eseguire.

Come preparazione alla fase di colorazione EE, i vetrini vengono poi disposti nel cestello che verrà direttamente utilizzato in seguito nella Colora-Vetrini Leica ST

5020, provvista anche di postazione asciuga-vevtrini. Altrimenti, i vevtrini allestiti per eseguire altre colorazioni vengono fatti asciugare sulla piastra termica del bagno riscaldato. Nel caso di sezioni sottili per l'immunoistochimica, l'allestimento prevede che vengano utilizzati dei particolari vevtrini carichi elettrostaticamente: i *Leica Bond Plus Slides*.

3.5.9. Colorazione istologici

Nella sezione "Colorazioni" della S.C. di Anatomia Patologica dell'ASL-BI sono presenti due coloratori: il Coloratore Papanicolaou (Leica ST 5020) che effettua l'omonima colorazione PAP per CCV e CEV, e il Coloratore Ematossilina-Eosina (Leica ST 5020) che automaticamente esegue la colorazione EE come protocollo di routine.

Su richiesta, possono essere effettuate ulteriori colorazioni specifiche istochimiche, immunoistochimiche e di Immunofluorescenza (DIF). Le colorazioni istochimiche, definite come speciali, sono: Giemsa (G) eseguita come colorazione manuale, Warthin-Starry (WS), Peiriodic Acid Shiff (PAS), Peiriodic Acid Shiff con Diastasi (PASD), Alcian Blu (AB), Alcian Blu Pas (ABP), Rosso Congo (RC), Perls (PERLS), Ziehl-Neelsen (ZN), Impregnazione Argentica Reticolo (RET), Grocott (GROCOTT), Tricromia di Masson (MASSON); tutte effettuate dal Coloratore automatico *Artisan Link Pro Special Staining System, Agilent Dako*.

In particolare, ci soffermeremo sulla colorazione EE siccome è quella che è stata utilizzata sui vevtrini istologici facenti parte del nostro studio.

La colorazione Ematossilina-Eosina (EE), dall'inglese *Hematoxylin and Eosin* (H&E), è una delle colorazioni d'elezione delle sezioni istologiche più comune ed importante in istologia per la diagnosi medica negli esami istopatologici. Infatti,

rappresenta il *gold standard* per i medici anatomo-patologi essendo una colorazione di routine.

Serve per evidenziare le componenti più importanti di un tessuto biologico. In particolare, essa è data dalla combinazione di due coloranti:

- l'ematossilina, colorante basico che colora di blu violetto i componenti cellulari carichi negativamente come: acidi nucleici, proteine di membrana e tutte le componenti basofile che si trovano prevalentemente a livello nucleare;
- l'eosina, colorante acido che invece determina la colorazione rosso rosata dei componenti carichi positivamente come molte proteine cellulari, mitocondriali e fibre di collagene, che vengono detti acidofili o eosinofili e sono localizzati nella matrice extracellulare e nel citoplasma.

Il seguente schema riporta i passaggi della colorazione nella durata e successione dei vari trattamenti singoli.

1	Forno	40 minuti
2	Xilene	20 minuti
3	Etanolo Assoluto	4 minuti
4	Etanolo 80°	2 minuti
5	Etanolo 70°	2 minuti
6	Etanolo 50°	2 minuti
7	Acqua di fonte	5 minuti
8	Ematossilina di Gill	9 minuti
9	Acqua di fonte	30 secondi
10	Acqua acidulata	30 secondi
11	Acqua di fonte	1 minuto
12	Soluzione di Scott	6 minuti

13	Acqua di fonte	1 minuto
14	Eosina alcolica	4 minuti
15	Etanolo 70°	30 secondi
	Etanolo Assoluto	13 minuti
	Xilene	13 minuti

Tabella 2: Schema dei passaggi di colorazione del Coloratore EE Leica ST-5020, presente nella S.C. di Anatomia Patologica dell'ASL-BI, riportante i singoli step con le rispettive tempistiche.

L'Etanolo Assoluto, l'Etanolo 95° e lo Xilolo si trovano in commercio già pronti all'uso, mentre l'Etanolo 80°, l'Etanolo 70° e l'Etanolo 50° devono essere preparati secondo i rapporti di concentrazione tra acqua distillata ed Etanolo 95°.

L'Acqua Acidulata si ottiene versando 10 mL di Acido Acetico Glaciale in un cilindro graduato, aggiungendo 490 mL di acqua distillata. In questo modo si ottiene una soluzione al 2% di Acido Acetico.

La soluzione di Scott, presente normalmente al 10%, viene diluita per ottenere la soluzione di lavoro aggiungendo 9 volumi di acqua distillata ad 1 volume della soluzione di Scott al 10%.

L'Eosina, presente in polvere, viene preparata sciogliendo 2 g di essa in 400 cc di Alcool 95°. All'utilizzo se ne prendono 500 mL che vengono acidificati con 4 cc di Acido Acetico Glaciale.

L'Ematossilina si trova in commercio già pronta all'uso.

La colorazione EE con il Coloratore automatico prevede sia l'asciugatura dei vetrini per una durata di 40 minuti in forno sia la sparaffinatura, a differenza invece dei CEV e CCV che invece non necessitano né di asciugatura né di sparaffinatura ed

iniziano direttamente con la colorazione in Alcool Assoluto. Inoltre, il vantaggio nell'utilizzare questo strumento è che permette una colorazione "continua", ovvero i cestelli possono essere caricati in qualsiasi momento siccome il computer calcola autonomamente le tempistiche di ogni cestello in modo che non ci siano sovrapposizioni.

Al termine della colorazione i cestelli si trovano immersi nello Xilene, in attesa di essere spostati nel montavetrini automatico che determinerà il montaggio in autonomia dei vetrini.

In seguito alla colorazione istochimica dei vetrini si ha come operazione quella del montaggio, cui segue l'etichettatura.

3.5.10. Montaggio

La fase di montaggio dei vetrini può avvenire sia manualmente con l'operatore sia automaticamente con l'utilizzo di uno specifico strumento che è il Montavetrini automatico Leica CV 5030, posto nella sezione "Strumenti" del laboratorio. In particolare, durante il passaggio manuale l'operatore versa una goccia di balsamo su ogni vetrino copri-oggetto. Successivamente, con l'utilizzo di pinzette estrae uno alla volta i vetrini porta-oggetto contenuti nel cestello che, in seguito, vengono adagiati sul piano della cappa venendo montati con il vetrino copri-oggetto posizionato al di sopra di essi con il balsamo, facendo attenzione che non si creino delle bolle d'aria e che tutta la sezione venga coperta.

Per i vetrini istologici e citologici si utilizzano dei vetrini copri-oggetto di dimensioni 24 mm x 60 mm, mentre per quanto riguarda i vetrini copri-oggetto dell'immunoistochimica essi hanno dimensioni di 24 mm x 50 mm.

Il passaggio automatico, invece, prevede l'uso del Montavetrini automatizzato che è unito e direttamente collegato con il Coloratore EE automatico, il quale è esso stesso ad avviare il montaggio una volta terminata la fase di colorazione dei vetrini EE. Esso è costituito da due componenti principali: una camera contenente una vaschetta con Xilolo, nella quale vengono immersi i vetrini porta-oggetto per subire poi la fase di montaggio premendo il pulsante *start*; e un ago che ha la funzione di erogare una goccia di balsamo per ogni vetrino in modo da far aderire in seguito il vetrino copri-oggetto al vetrino porta-oggetto.

3.5.11. Etichettatura

La fase di etichettatura viene effettuata in seguito al processo di montaggio come passaggio finale per concludere l'attività svolta dal laboratorio prima della fase diagnostica, eseguita dal Medico Anatomico-Patologo. I vetrini vengono quindi disposti in ordine numerico progressivo ed etichettati con le etichette stampate dall'etichettatrice ZEBRA direttamente collegata alla postazione con il computer.

Su tutte le etichette dei vetrini vengono riportati i seguenti dati:

- Nome della Struttura Ospedaliera (ad esempio Ospedale di Biella);
- Tipologia di campione (A Istologico, B Citologico extra-vaginale, C Citologico vaginale);
- Anno in corso;
- Numero istologico progressivo;
- Sigla delle colorazioni.

3.5.12. Immunoistochimica

Questa fase di colorazione in immunoistochimica viene effettuata con l'Immunocoloratore automatico Ventana PD-L1 (SP263) Assay, la cui procedura può essere applicata sia su campioni istologici sia a campioni citologici.

Il principio fondamentale si basa sulla reazione specifica tra antigene e anticorpo, la quale viene evidenziata con l'utilizzo di particolari enzimi, come ad esempio la perossidasi e la fosfatasi, che agiscono su un substrato quando avviene la reazione. Per l'evidenziazione della reazione intervengono quindi degli enzimi che vengono rivelati con dei reattivi specifici, come la Dab e il Fast Red.

Nei casi specifici di questo studio l'antigene di interesse è Ki-67, ossia una proteina nucleare strettamente associata con la proliferazione cellulare e che, nel tumore al seno, è considerato un marcatore tumorale fondamentale in quanto è l'indice di proliferazione essenziale per comprendere la crescita tumorale e determinare il grado.

Come per tutte le fasi precedentemente trattate, si deve prestare molta attenzione sia nei passaggi di prelievo che nella preparazione dei campioni da sottoporre ad esame, in modo che non si creino degli artefatti o false positività o false negatività.

Questa procedura prevede che vengano utilizzati dei particolari vetrini carichi elettrostaticamente, ossia i *Leica Bond Plus Slides*. Essi sono richiesti sia nella fase di riduzione quando viene creato il protocollo o in seguito alla colorazione EE su indicazione del Medico Refertante che, in base alle prime considerazioni e valutazioni diagnostiche sul vetrino EE, richiederà ulteriori approfondimenti tramite il sistema WinSap per avere un pannello di anticorpi più chiaro e dettagliato.

Perciò, il cubetto di paraffina contenente il campione viene messo in congelatore per una tempistica minima di 5 minuti, che gli permetterà di essere freddo a sufficienza per poter essere tagliato. Le sezioni ottenute vengono quindi allestite con la fase di allestimento che termina con la formazione dei vetrini, i quali vengono inseriti negli appositi alloggiamenti del porta-vetrini che viene posto in essiccamento ad una temperatura di 50° per 40 minuti all'interno dell'essiccatoio e, in seguito i vetrini sono lasciati raffreddare per una notte intera a temperatura ambiente.

Al mattino, si accende lo strumento e si inseriscono gli anticorpi specifici da utilizzare in quella corsa e i relativi reagenti dei kit di rivelazione. Subito dopo, vengono stampate le etichette per i vetrini dei casi in esame in modo da poter poi caricarli nei trenta scompartimenti nell'Immunocoloratore.

Al termine della fase di immunoistochimica, dopo aver eseguito sparaffinatura e idratazione, si estraggono i vetrini dai vari scompartimenti dello strumento in cui si trovano alloggiati e si inseriscono all'interno di un porta-vetrini. Si esegue poi un lavaggio con acqua di fonte e sapone per una tempistica di 15 minuti con l'obiettivo che la soluzione oleosa adesa al vetrino venga eliminata e, in seguito, viene effettuata l'operazione di differenziazione-disidratazione sotto cappa aspirante per eliminare l'eccesso di colorante dal tessuto biologico. Questa procedura prevede, nel dettaglio, le seguenti fasi e tempistiche:

1. Lavaggio in acqua di fonte con sapone, per una durata di 15 minuti;
2. Lavaggio in Etanolo 50°, per una durata di 2 minuti;
3. Lavaggio in Etanolo 70°, per una durata di 2 minuti;
4. Lavaggio in Etanolo 80°, per una durata di 2 minuti;
5. Lavaggio in Etanolo Assoluto, per una durata di 2 minuti;
6. Lavaggio in Xilolo, per una durata di 5 minuti.

Infine, i vetrini vengono montati nella fase di montaggio e consegnati ai medici refertanti.

3.6. Digitalizzazione dei vetrini

Ogni vetrino istologico EE ed Immuno Ki-67 è stato sottoposto a digitalizzazione ad un ingrandimento 40x utilizzando lo scanner NanoZoomer XR Digital Pathology dell'azienda Hamamatsu presente presso la S.C. di Anatomia Patologica dell'Ospedale degli Infermi di Biella. La digitalizzazione è stata eseguita durante il periodo del progetto formativo, a differenza invece della realizzazione dei vetrini istologici che si sono ottenuti dopo il campionamento dei pezzi operatori in seguito alle operazioni chirurgiche eseguite negli anni in cui le pazienti sono state operate.

Prima di essere caricati all'interno dello scanner, i vetrini devono essere portati alle condizioni standard quindi: il vetrino copri-oggetto e l'etichetta non devono sporgere all'esterno per evitare che il vetrino si incastri all'interno dello strumento; perciò, è fondamentale che l'operatore utilizzi una lima per eliminare le sporgenze (se presenti); inoltre, i vetrini devono essere puliti con alcol e garza e successivamente caricati nei rack con l'etichetta rivolta verso l'operatore. I rack sono dei supporti di plastica che presentano nella loro struttura degli scomparti appositi dove i vetrini vengono inseriti verticalmente. Infine, i rack vengono caricati all'interno di uno specifico spazio dello scanner e messi in posizione bilanciata.

Successivamente si esegue la procedura attraverso l'utilizzo di un computer collegato direttamente allo scanner e dal quale si utilizza il programma dello scanner, ossia DP SCAN, dal quale si possono eseguire due funzionalità diverse per la scansione dei vetrini:

- *Batch of slide*: funzione che permette la scansione di più vetrini contemporaneamente;
- *Single batch*: funzione che permette la scansione di un singolo vetrino per volta.

In seguito, si esegue il *batch control* dal quale con il pulsante *Start Batch* si inizia la scansione preliminare e la selezione dell'area di interesse. L'immagine deve essere compresa nel riquadro blu e le varie sezioni nei riquadri gialli, al cui interno sono presenti dei focus che vengono messi in automatico dallo scanner e che possono essere spostati e/o eliminati e/o aggiunti in base alle regioni di interesse dell'operatore. Dopo aver sistemato tutti i focus, la scansione viene avviata. Una volta terminata, per poter aprire lo sportello dello scomparto è necessario cliccare sul tasto *Exit Job* per poter permettere l'apertura e lo scarico dei rack.

Per i nostri vetrini EE si è utilizzata la modalità *Brightfield* e si è eseguita la scannerizzazione all'ingrandimento 40x; mentre ad esempio per scannerizzare i vetrini su cui è stata eseguita l'immunoistochimica (HER2, ER, PR, Ki67) si utilizza la modalità *Immuno* e li si scannerizza sempre ad un ingrandimento 40x.

Il set di dati è composto in totale da 18 immagini WSI di carcinomi mammari tripli negativi, in particolare 9 immagini relative ai vetrini in Ematossilina-Eosina e 9 immagini Immuno su Ki-67.

Le immagini così ottenute sono immagini a vetrino intero (WSI) siccome contengono l'intero tessuto campionato e, per questo motivo, sono immagini ad alta risoluzione che possono contenere più di una regione patologica al loro interno. Per la grande dimensione degli input, i modelli ottenuti con gli algoritmi consentono di avere caratteristiche sia microscopiche che di organizzazione dei tessuti.

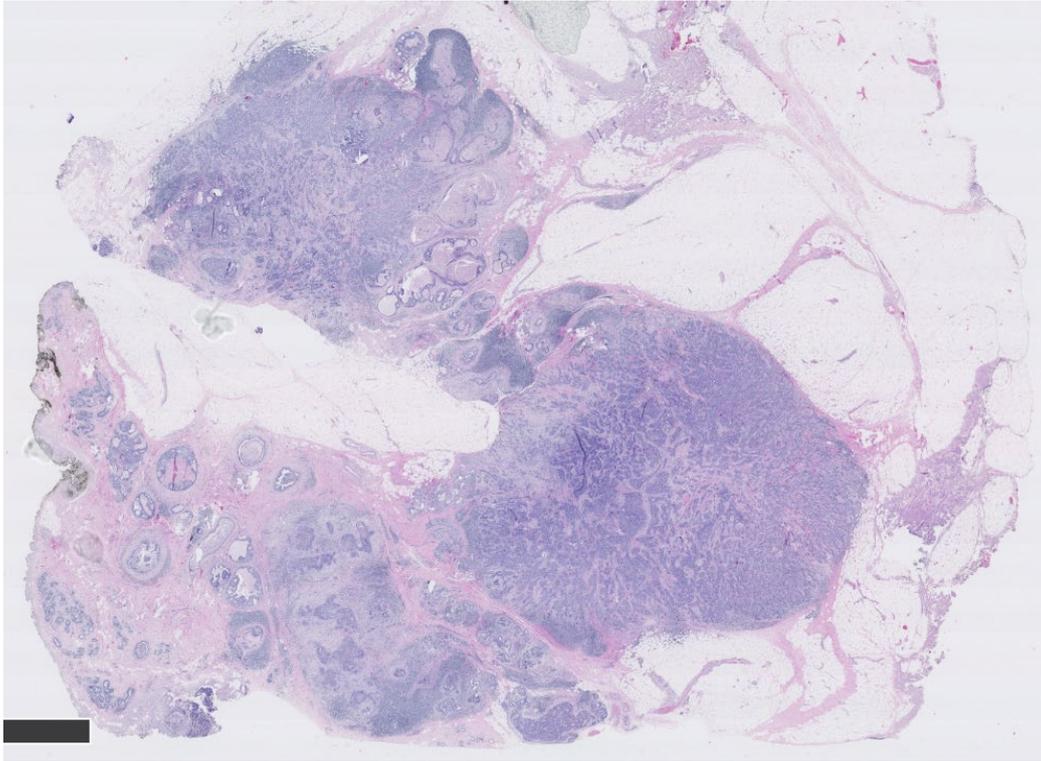


Figura 3: Immagine WSI del vetrino istologico EE del campione 2017.5257

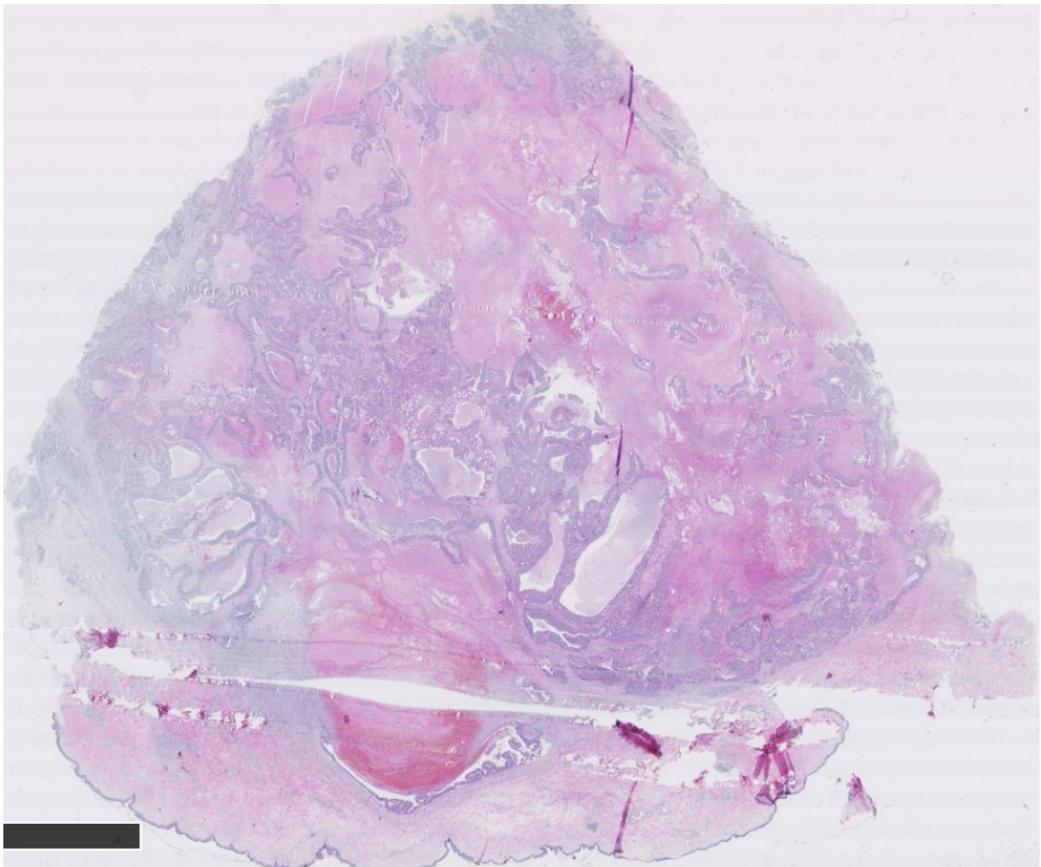


Figura 4: Immagine WSI del vetrino istologico EE del campione 2016.8698

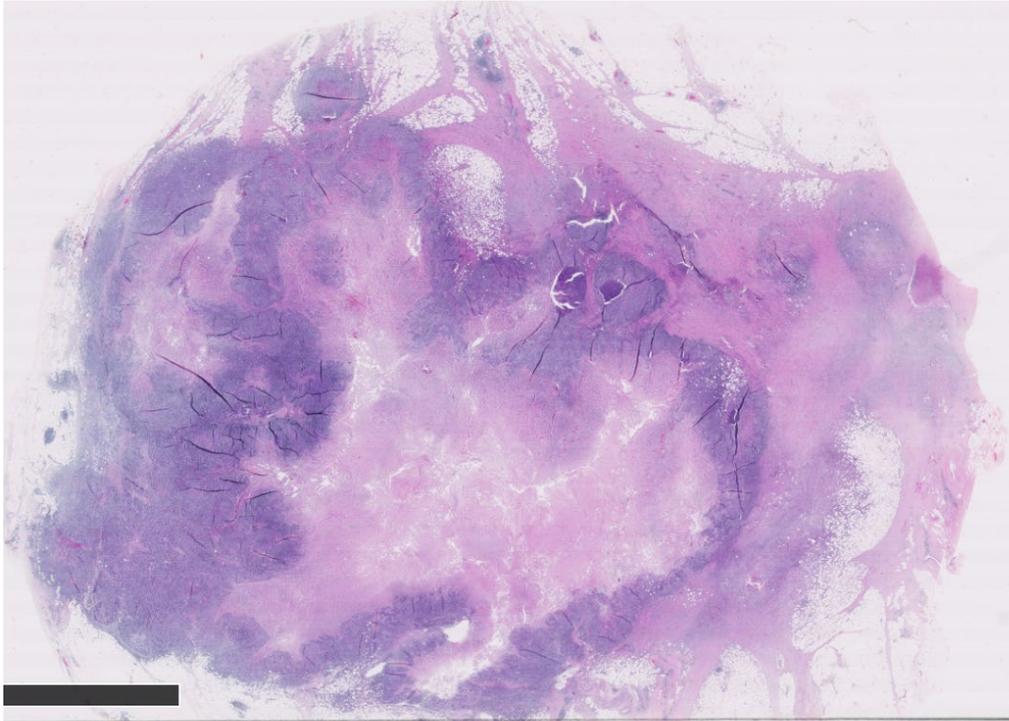


Figura 5: Immagine WSI del vetrino istologico EE del campione 2013.7011

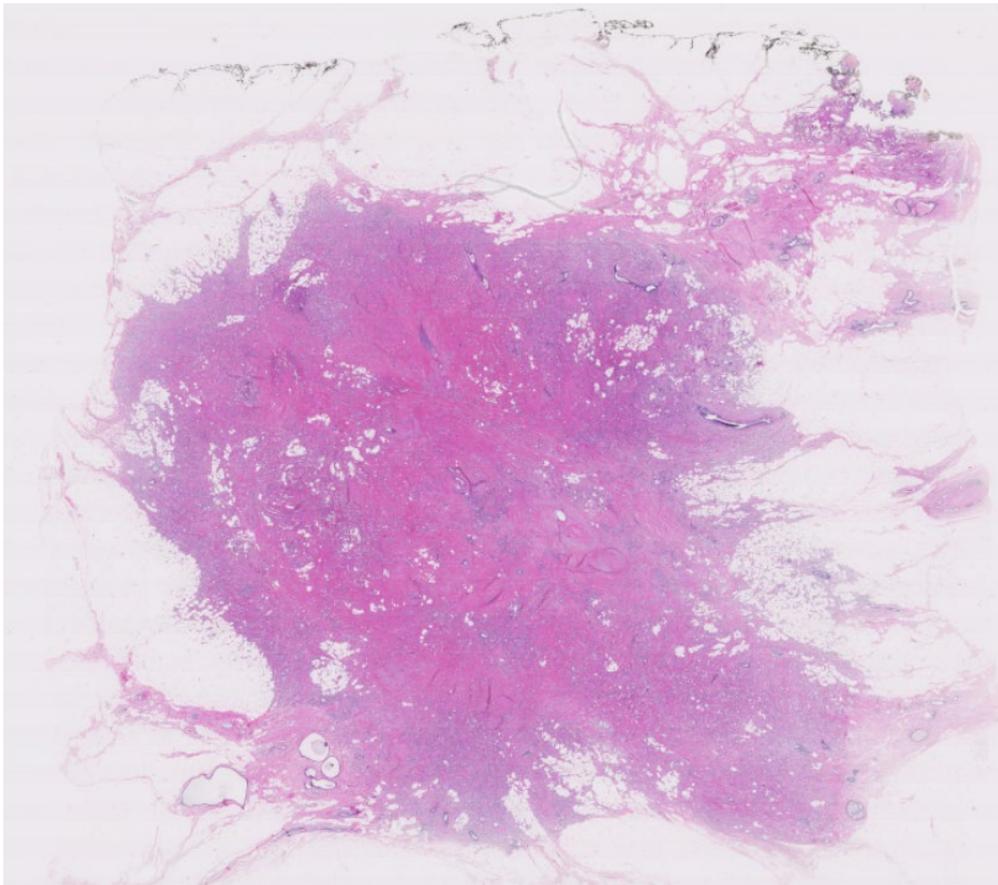


Figura 6: Immagine WSI del vetrino istologico EE del campione 2014.4762

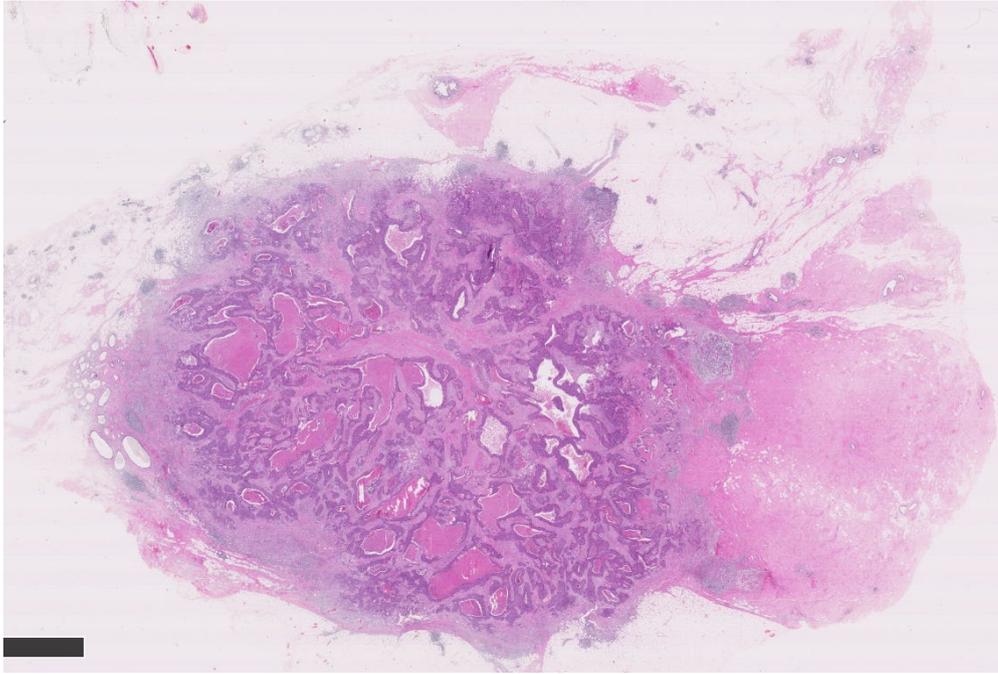


Figura 7: Immagine WSI del vetrino istologico EE del campione 2015.2474

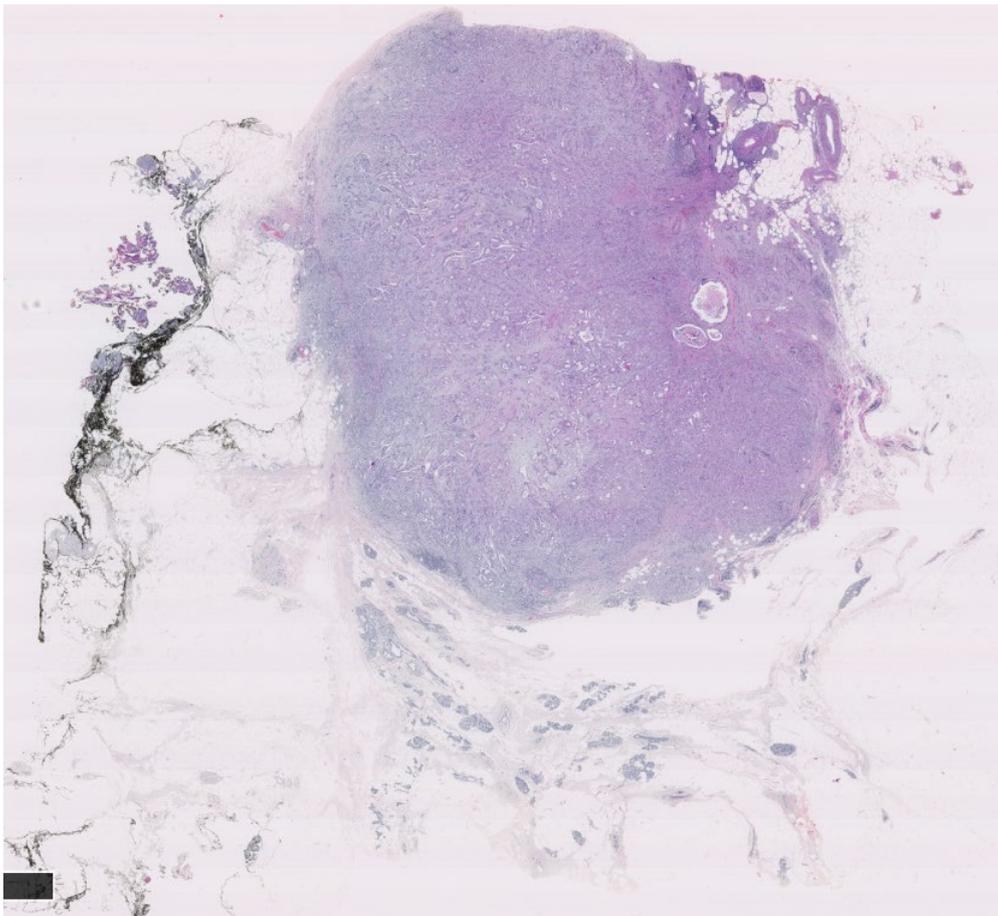


Figura 8: Immagine WSI del vetrino istologico EE del campione 2015.1022

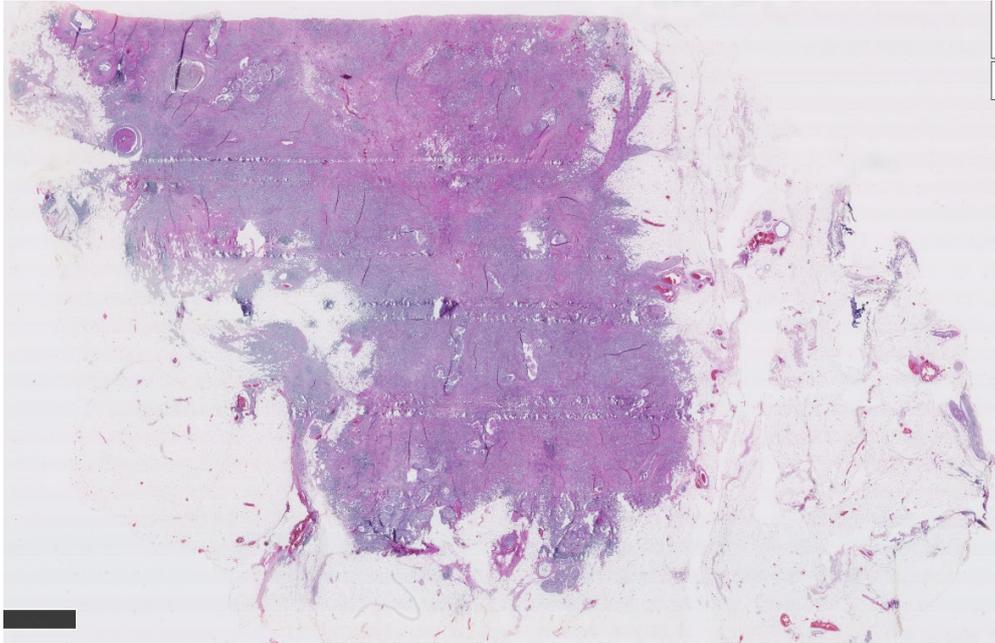


Figura 9: Immagine WSI del vetrino istologico EE del campione 2013.6402



Figura 10: Immagine WSI del vetrino istologico EE del campione 2015.4683

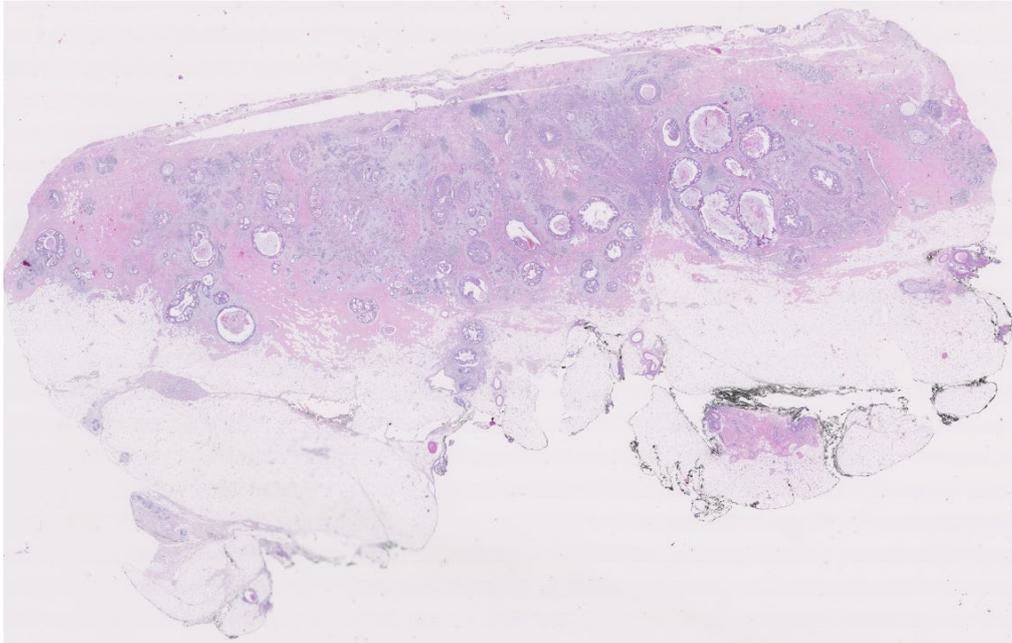


Figura 11: Immagine WSI del vetrino istologico EE del campione 2013.699

3.7. Revisione dei dati

In seguito alla ricerca e acquisizione dei dati delle pazienti triple negative, si è effettuata una revisione completa da parte di un Medico Patologo esperto attraverso un'analisi in cieco di tutti i campioni istologici delle pazienti in modo da confermare le corrette diagnosi precedentemente effettuate. I tumori delle pazienti sono stati quindi riclassificati dividendoli per grado (basso grado e alto grado) facendo una valutazione del grado istologico basata su tre parametri, secondo Nottingham/Elston-Ellis:

1. Formazione di tubuli
2. Pleomorfismo nucleare
3. Indice mitotico

Considerando il pTNM secondo i criteri del AJCC Staging Manual, 8^a Edizione.

3.8. Analisi dei vetrini con il software QuPath

L'analisi digitalizzata è avvenuta sui nove casi presi in considerazione sotto indicazione del Fondo Edo Tempia, i quali avevano tutti la caratteristica in comune di essere dei carcinomi tripli negativi, e in particolare del sottotipo LAR (*Luminal Androgen Receptor*).

Questi 18 campioni in tutto (due vetrini per ogni caso), una volta digitalizzati mediante lo scanner, sono stati salvati su una memoria esterna di grandi dimensioni per poter utilizzare le immagini durante l'analisi.

I 9 campioni EE sono stati quindi processati uno per volta mediante l'utilizzo del *software* QuPath analizzando, durante la prima fase analitica, tre parametri fondamentali:

1. Tumore (μm^2)
2. Stroma (μm^2)
3. Necrosi (μm^2)

Nella seconda fase di analisi, una volta ottenuti i tre parametri principali nella fase precedente, si sono calcolati:

4. Rapporto Stroma/Tumore
5. Purezza tumorale

Infine, nella terza e ultima fase, si è effettuata un'analisi sui 9 campioni Immuno Ki-67 indagando sul seguente parametro di indagine:

6. Linfociti infiltranti lo stroma (linfociti peri-tumorali proliferanti in mitosi)

Per eseguire l'analisi digitalizzata delle immagini istologiche si è seguita una specifica procedura, caratterizzata da diversi step che di seguito andremo ad approfondire.

In sintesi, in ordine vediamo gli scopi per i quali viene applicata questa procedura:

- Creare oggetti sottoforma di annotazioni, in modo da identificare delle regioni di interesse specifiche;
- Aggiungere misurazioni alle annotazioni identificate precedentemente;
- Classificare gli oggetti identificati.

3.8.1. Creazione di un progetto di analisi

La creazione di un progetto permette di raggruppare più immagini correlate all'interno dello stesso progetto di analisi in modo da passare facilmente da un'immagine all'altra e favorire il più possibile l'indagine. Inoltre, permette anche di organizzare i file di dati insieme allo *script*.

Per impostare un progetto da zero è fondamentale creare una cartella vuota sul computer che viene identificata come cartella destinataria del progetto con i seguenti comandi: *File* → *Project* → *Create new project*.

Dopo aver creato un nuovo progetto, le immagini vengono caricate trascinandole sulla finestra del *software* QuPath oppure con il seguente percorso: *File* → *Project* → *Add images*.

3.8.2. Deconvoluzione dei colori: separazione delle regioni e normalizzazione dei colori

La prima azione da svolgere quando viene caricata un'immagine all'interno di un progetto in QuPath è quella di scegliere la tipologia di immagine, che tipicamente

può essere o il Campo Chiaro (*Brightfield*) o la Fluorescenza (*Fluorescence*). Questo viene fatto in quanto la scelta della tipologia di immagine può avere un impatto su come le colorazioni possono essere separate; infatti, lo *stain separation* nelle immagini *Brightfield* è più difficile rispetto alle immagini in fluorescenza siccome QuPath utilizza il metodo *color deconvolution* introdotto da Ruifrok e Johnston che, fondamentalmente, separa digitalmente la deconvoluzione dei colori in tre colorazioni a partire da un'immagine RGB (Red-Green-Blue) in campo chiaro in modo significativo. [10]

In particolare, con la funzione *Brightness/Contrast* è possibile visualizzare questa separazione di colori in tre tipologie: *Hematoxylin*, *Eosin*, *Residual*.

Questo metodo viene utilizzato per individuare strutture e determinare aree con colorazioni simili. Infatti, lo *stain vectors* è la chiave per la deconvoluzione del colore in quanto i vettori di colorazione danno una rappresentazione normalizzata dei colori per ogni colorazione pura dell'immagine caratterizzandola, indipendentemente dalla sua colorazione.

Quando viene scelta la tipologia di immagine, che nel nostro caso specifico per i primi 9 campioni è *Brightfield* (H&E), il *software* QuPath determina in automatico i vettori di colorazione che caratterizzeranno l'ematossilina, l'eosina e il colore residuale automaticamente generato, siccome si tratta di un'immagine con diverse colorazioni cromogeniche. [11]

In questa prima fase, quindi, avviene la standardizzazione e normalizzazione delle immagini istologiche con l'obiettivo di migliorarne la qualità, che presenta comunque delle problematiche importanti, andando ad ottimizzare digitalmente la

colorazione del vetrino digitale senza alterare le informazioni cliniche, in modo da facilitarne l'analisi mediante il *software*.

3.8.3. Stima dei vettori di colorazione

Prima di effettuare l'analisi per la determinazione dei parametri e la loro misurazione, su ogni immagine-campione si è svolto il seguente comando: *Analyze* → *Preprocessing* → *Estimate stain vectors*; il quale consente la stima dei vettori di colorazione e la loro separazione (*staining vectors*) in modo da normalizzare i canali RGB (Red – Green – Blu) su ogni immagine digitale utilizzando la deconvoluzione del colore. Questo comando rappresenta la prima fase di analisi su cui poi tutte le misurazioni successive si basano. Prima di effettuare il comando *Estimate stain vectors*, bisogna prima identificare una regione di interesse ROI rappresentativa che contiene anche una porzione di *background* e che viene individuata con una annotazione.

In seguito, QuPath costruisce dei grafici a dispersione (*scatterplots*) per visualizzare le relazioni tra i pixel di colore rosso, verde e blu; e siccome è difficile visualizzare un grafico basato su tre variabili (3D *scatterplots*) allora il programma costruisce tre grafici separati (2D *scatterplots*) nei quali mostra rispettivamente: la relazione tra il colore rosso e il colore verde, la relazione tra il colore rosso e il colore blu e la relazione tra il colore verde e il colore blu.

In questi grafici, il *software* mostra la distribuzione di colori dei pixel sparsi disegnando anche le linee colorate che indicano i vettori di colore esistenti che, idealmente, dovrebbero circondare la maggior parte dei punti facenti parte della dispersione. Se i vettori risultano essere inaccurati, allora QuPath, mediante la modalità *Auto*, effettua una miglior scelta dei vettori di colorazione applicando un

aggiustamento automatico che viene mostrato nei grafici di distribuzione. Inoltre, per aggiustare ulteriormente i parametri viene selezionata la funzione *Exclude unrecognised colors (HE only)* che effettua una auto-stima dei vettori di colorazione escludendo e rimuovendo i colori che non vengono riconosciuti come Ematossilina o Eosina. In questo modo i vettori risultanti tendono ad essere maggiormente affidabili e completi.

Infine, dando l'ok si inserisce il nome per identificare i nuovi vettori di colorazione.

La funzione *Estimate stain vectors* è molto importante perché è la funzione che utilizza e si basa sulla deconvoluzione dei colori. Essa deve essere applicata come uno dei primi step nelle primissime fasi di analisi, da eseguire subito dopo il caricamento di un'immagine digitale all'interno del *software* prima di eseguire qualsiasi altra funzione di rilevamento o misurazione.

3.8.4. Rilevamento del tessuto: definire la regione di interesse

Questa funzione viene eseguita con l'impostazione di una soglia, che è solitamente la scelta migliore per individuare delle regioni e annotarle. Essa viene utilizzata per definire le regioni di interesse, senza che esse vengano identificate manualmente dall'operatore.

Identificare un valore di soglia significa determinare la distinzione tra due classi di pixel appartenenti a due regioni differenti: una prima regione che presenta valori superiori alla soglia specificata e una seconda regione che possiede valori inferiori rispetto alla soglia individuata.

La creazione di una soglia avviene sfruttando i valori di colorazione che i pixel possiedono nell'immagine digitale che si sta analizzando e questo avviene attraverso il comando: *Classify* → *Pixel classification* → *Create threshold*. Questo ci

permette di creare oggetti, che è la funzione più semplice e più comune nell'utilizzo dei *software* di immagini digitali.

Quando si esegue il comando, ci si trova davanti una finestra di dialogo nella quale sono presenti i seguenti parametri che possono essere modificati e regolati in base alla nostra immagine digitale:

- Risoluzione, definisce la risoluzione dell'immagine a cui viene applicato il valore di soglia. Essa sarà sempre inferiore rispetto alla risoluzione completa dell'immagine in quanto si deve scegliere la risoluzione più bassa che si ritenga il più precisa possibile nell'individuazione dell'area di interesse. Per le analisi delle nostre immagini digitali si è utilizzata la risoluzione "Moderata";
- Canale, identifica il canale dell'immagine a cui verrà applicata la soglia in modo che si abbia un netto contrasto per ciò che si vuole rilevare. Per visualizzare tutti i canali disponibili si utilizza il riquadro "Luminosità/Contrasto" che include tutti i canali, compresi quelli che sono generati automaticamente dal software;
- Prefiltro e smoothing sigma, definiscono il modo in cui viene smussata l'immagine digitale dal valore di soglia. Solitamente si utilizza il prefiltro "Gaussiano";
- Valore di soglia, inserita dall'operatore per identificare il valore da inserire affinché si abbia una distinzione tra i valori sopra la soglia e quelli sotto;
- Sopra la soglia, definisce la classificazione dei pixel con valori al di sopra della soglia identificata;
- Sotto la soglia, definisce la classificazione dei pixel con valori al di sotto della soglia;

- Regione, controlla la quantità di soglie dell'immagine digitale quando vengono variati i parametri di impostazione e si visualizzano in anteprima i risultati. Permette quindi di impostare il valore di soglia non solo a livello dell'immagine intera ma anche a singole annotazioni o oggetti.

Una volta impostato il valore di soglia corretto, esso viene salvato all'interno del progetto come classificatore di pixel. In questo modo si può ricaricare un valore di soglia salvato in precedenza seguendo i comandi: *Classify* → *Pixel classification* → *Load pixel classifier*.

In seguito al rilevamento del tessuto con l'impostazione di una soglia, mediante la funzione *Create objects* si possono individuare annotazioni che potranno poi essere misurate e classificate come oggetti. Con questa funzione, anche in questo caso, si può decidere se creare oggetti considerando l'intera immagine oppure in tutte le regioni annotate o in una sola annotazione selezionata.

Questa funzione porta come risultato finale a regioni di tessuto rilevate ed annotate, e che successivamente verranno, infine, classificate.

3.8.5. Aree di misurazione

Questa funzione è una delle più importanti per l'analisi digitale in *Digital Pathology*, in quanto permette di quantificare una regione di interesse ROI, precedentemente identificata e definita, conteggiando i pixel positivi.

Essa permette quindi la visualizzazione delle misurazioni selezionando la funzione *Annotations*, attraverso cui il software QuPath genera in automatico le misurazioni in tempo reale delle aree di interesse in funzione dei parametri di soglia. Le aree riportate dalla classificazione dei pixel si basano sul conteggio dei pixel

nell'immagine classificata che, a seconda della risoluzione della classificazione, determinerà quindi la somma delle diverse aree identificate.

Nel pratico, la funzione di misurazione delle aree avviene con la voce *Measure* che, dopo aver identificato un valore di soglia e creato gli oggetti, può identificare le misure degli oggetti creati oppure dell'intera immagine. In seguito, è possibile visualizzare le misure accedendo dal comando *Measure* nel pannello delle funzioni principali che consente di visualizzare i valori di misurazione sottoforma di tabelle, che nel nostro progetto sono state salvate ed esportate su di una memoria esterna per il loro salvataggio e utilizzo nell'analisi statistica.

3.8.6. Rilevamento delle cellule

Questa funzione di QuPath corrisponde alla capacità di rilevamento delle cellule con l'obiettivo di ottenere dei risultati in modo molto più rapido e riproducibile, spesso con maggiore precisione.

Il programma è in grado di rilevare rapidamente in autonomia le cellule e calcolare la percentuale di positive all'interno di una specifica annotazione, dopo che l'operatore manualmente ha identificato alcune cellule tumorali segnandole per creare delle annotazioni.

Il primo step di questa procedura è rappresentato dall'annotare una specifica regione di interesse ROI che contiene cellule, le quali devono essere rilevate e contate. Successivamente, si esegue il rilevamento delle cellule positive seguendo questi comandi: *Analyze* → *Cell detection* → *Positive cell detection*; secondo i quali le cellule verranno rilevate e in seguito classificate come positive oppure negative.

Da questi step verrà visualizzata una finestra di dialogo che riporta i valori predefiniti, sufficienti alle nostre indagini, e che sono i seguenti:

- *Setup parameters: detection image* (tipologia di immagine), *requested pixel size* (0,5 μm);
- *Nucleus parameters: background radius* (8 μm), *median filter radius* (0 μm), *sigma* (1,5 μm), *minimum area* (10 μm^2), *maximum area* (400 μm^2);
- *Intensity parameters: Threshold* (0,1), *max background intensity* (2);
- *Cell parameters: cell expansion, include cell nucleus*;
- *General parameters: smooth boundaries, make measurements*;
- *Intensity threshold parameters: Single threshold*.

Se si è interessati a classificare le cellule in base all'intensità della colorazione, allora è possibile impostare fino a tre diverse soglie. In questo caso non si è interessati a classificazioni multiple, bensì ad una sola classificazione binaria tra positive e negative. Per questo motivo viene selezionato il comando *Single threshold* e infine il tasto *Run*. In questo modo, come risultato si avrà che una singola grande annotazione conterrà tanti oggetti di rilevamento, che si possono visualizzare selezionando l'annotazione stessa. Quindi si possono visualizzare i dati sul numero totale di cellule detectate, il numero di quelle positive e il numero di quelle negative. Questo viene effettuato con i comandi: *Measure* → *Show detection measurements*, con i quali si ha accesso alle tabelle riportanti i risultati con le misurazioni che sono poi state esportate su una memoria esterna per poter essere salvate ed analizzate.

La funzione di rilevamento delle cellule non deve per forza essere solamente applicata su una singola annotazione, ma vi è la possibilità che si possa estendere a tutte le annotazioni presenti contemporaneamente nell'immagine digitale. Quindi, anche in presenza di più annotazioni, sarà l'operatore a decidere se effettuare il

rilevamento su una singola annotazione selezionandola oppure su tutte quelle presenti con un'elaborazione in parallelo deselezionandole.

Quando si hanno molteplici annotazioni, in automatico si crea una tabella che riporta i risultati delle singole annotazioni come regioni di interesse ROI. Per visualizzare questa tabella si devono seguire i seguenti comandi: *Measure* → *Show annotation measurements*.

3.8.7. Classificazione delle cellule

Questa funzione di QuPath è relativa a come il software può essere addestrato a distinguere diverse tipologie di cellule in base alle loro caratteristiche morfologiche. In questo modo, si fornisce un metodo di analisi alternativo che permette di rilevare tutte le cellule presenti in una regione di interesse identificata con un'annotazione e poi calcolare i punteggi in base alle diverse tipologie cellulari presenti in essa.

Per prima cosa, è necessario annotare una regione di interesse ROI sottoforma di annotazione all'interno della quale dovrebbero essere identificate e rilevate le cellule. Questo step viene svolto molto rapidamente ed è fondamentale che la scelta della ROI avvenga in una porzione di tessuto istologico in cui vi è una miscela di cellule tumorali e cellule non tumorali affinché la classificazione sia significativa.

Successivamente, una volta identificata la regione di interesse dedicata all'indagine, viene eseguito il comando per detectare le cellule: *Analyze* → *Cell detection*. Questo comando esegue esattamente la stessa funzione della *Positive Cell Detection*, quindi di rilevamento e identificazione delle cellule, con la differenza sostanziale che quest'ultimo comando ha il vantaggio extra di detectare le cellule e suddividerle classificandole come positive o negative in accordo con le intensità delle regioni.

In seguito, il passo successivo è creare delle annotazioni più piccole in regioni di interesse differenti per classificare le cellule in esse contenute. In questo modo vengono quindi create delle annotazioni più piccole e modeste di dimensioni all'interno delle ROI più grandi ed esse vengono associate alla classificazione appropriata dall'elenco, cliccando il tasto *Set class*. Da questo elenco vengono visualizzati i valori delle annotazioni create per ogni classe.

È fondamentale seguire correttamente la procedura in cui prima si esegue il rilevamento delle cellule presenti nelle ROI con il *Cell Detection*, e poi successivamente si annotano regioni per la distinzione di cellule differenti.

L'obiettivo di questa procedura è arrivare ad addestrare un classificatore di cellule che distingue le tipologie cellulari in base alle annotazioni con classi diverse precedentemente create. Essa viene svolta secondo i seguenti passaggi: *Classify* → *Object classification* → *Train object classifier*; e con la funzione *Live Update* il classificatore viene applicato su tutte le cellule all'interno delle ROI facendo vedere gli aggiornamenti in tempo reale mentre si disegnano sempre più annotazioni, impostando le relative classi con la funzione *Set Class* per rendere il classificatore il più preciso possibile. Questo avviene se il computer è sufficientemente veloce o se il numero di annotazioni è sufficientemente piccolo.

Un valore molto utile per l'identificazione di cellule tumorali è il rapporto tra l'area del nucleo e l'area della cellula. Questo perché nelle cellule tumorali questa misurazione tende ad essere più elevata siccome i nuclei sono più densamente concentrati e più grandi in questa tipologia cellulare rispetto ai nuclei di cellule non tumorali. Questo rapporto racchiude e incorpora quindi entrambe le caratteristiche.

Il classificatore predefinito è il *Random Trees* siccome tende ad ottenere una buona combinazione di velocità e precisione.

La scelta delle annotazioni è importante per le prestazioni del classificatore e una volta che esso è sufficientemente accurato, si salva fornendogli un nome in quanto può essere applicato ad un'immagine simile in futuro. Questo perché in seguito potrà essere caricato nuovamente grazie ai seguenti passaggi: *Classify* → *Object classification* → *Load object classifier*.

3.9. Analisi statistica

Sui dati delle pazienti LAR che hanno preso parte in questo studio, sono stati applicati due test per poterli confrontare con il parametro della sopravvivenza: il t-test di Student e il test di Mann-Withney; questo perché si sono confrontate due variabili differenti, ovvero una variabile quantitativa (valori dei dati) e una variabile qualitativa (sopravvivenza).

Il test di Mann-Withney risulta essere più appropriato perché è un test non parametrico per campioni statisticamente indipendenti, a differenza invece del t-test di Student che necessita di avere una distribuzione normale dei dati ma siccome lo studio è stato effettuato sulle sole pazienti LAR, che sono 9, ci sono pochi dati e quindi è probabile che la distribuzione non sia normale e, per questo motivo, il t-test non sia appropriato.

Il t-test di Student è un test statistico di tipo parametrico che ha lo scopo di verificare se il valore medio di una distribuzione di dati si discosta significativamente da un certo valore di riferimento. Questo test è stato applicato utilizzando il linguaggio di programmazione R ed il software RStudio con lo scopo di confrontare tra di loro i valori di due gruppi e, in particolare, le medie delle

pazienti in vita e le medie delle pazienti decedute per ogni parametro di indagine e, per capire se la differenza osservata tra le medie è una differenza statisticamente significativa oppure è dovuta al caso.

Il test di Mann-Whitney è l'alternativa non parametrica al più conosciuto test t per il confronto delle medie di due gruppi tra loro indipendenti. Esso è stato calcolato utilizzando R Commander.

Il test t per campioni indipendenti è un test parametrico che permette di determinare se ci sono differenze statisticamente significative tra le medie di due gruppi tra loro indipendenti. Essendo quest'ultimo un test che utilizza le medie, si può applicare solo nei casi in cui è preferibile utilizzare come indice di tendenza centrale la media rispetto alla mediana.

Il test U di Mann-Whitney, invece, si basa sui ranghi e proprio per questo è un test non parametrico che può essere utilizzato in tutti quei casi in cui non è consigliabile effettuare un confronto tra medie.

Questa trasformazione dei valori osservati in ranghi permette infatti di avere degli indici più robusti rispetto ai valori anomali. Nel caso di dati per i quali siano verificate tutte le ipotesi del test parametrico t, si dimostra che utilizzare questa alternativa non parametrica porta però ad una riduzione della potenza del test. Questo significa che il test t, se le assunzioni su cui si basa sono verificate, può individuare come statisticamente significative differenze più piccole a parità di numerosità campionaria. E più il campione è di piccole dimensioni, più la differenza tra i due test può diventare rilevante.

Inoltre, il test parametrico permette di calcolare anche gli intervalli di confidenza delle medie, mentre il test non parametrico restituisce come output solo il p-value.

Nel pratico, si considerano tutte le coppie di osservazioni formate considerando un'osservazione da un gruppo e l'altra osservazione dall'altro gruppo ed il test è basato sul numero di coppie per cui l'osservazione del primo gruppo è maggiore.

Come abbiamo visto, è possibile ottenere il risultato del test utilizzando due diversi metodi:

- Il metodo asintotico, che attribuisce una maggiore potenza statistica rispetto al metodo esatto. Nel metodo asintotico, infatti, la significatività statistica si valuta infatti a partire dall'ipotesi di grandi campioni distribuiti normalmente. Tuttavia, in caso di dati asimmetrici, non è possibile affidarsi a tale metodo, ma è necessario fare affidamento sui risultati esatti;
- nel metodo esatto, infatti, la significatività statistica non è basata su una distribuzione teorica (come quella Normale) ma sulla distribuzione del campione osservato, ossia sulla "esatta" distribuzione dei dati.

Il test U di Mann Whitney restituisce il seguente p-value:

- superiore ad α (in genere si considera $\alpha=0.05$): non c'è evidenza di una differenza statisticamente significativa tra le due distribuzioni;
- inferiore ad α : la sua interpretazione cambia a seconda della forma che la distribuzione della variabile dipendente ha nei due gruppi:
 - se ha forma diversa nei due gruppi: c'è evidenza a favore di una differenza statisticamente significativa tra le medie dei ranghi dei due gruppi;
 - se la forma della distribuzione è la stessa: c'è anche evidenza a favore di una differenza statisticamente significativa tra le mediane dei due gruppi.

Il test permette quindi di capire se, quando nel campione che si sta analizzando si trovano differenze tra le medie dei ranghi, questa è abbastanza grande da essere considerata statisticamente significativa.

Nel caso la forma della distribuzione sia poi simile nei due gruppi, allora questo test permette di verificare anche se ci sia evidenza di una differenza statisticamente significativa tra le mediane dei due gruppi.

Ciò che quindi possiamo dire è che: solo quando la forma della distribuzione può essere considerata la stessa, se si ottiene un p-value con un valore piccolo inferiore ad α , si può concludere che vi è evidenza di una differenza statisticamente significativa tra le mediane dei due gruppi.

4. Risultati

Nel capitolo 4. sono stati riportati i risultati delle analisi statistiche effettuate sui dati delle pazienti LAR ottenuti in *Digital Pathology* con il *software* QuPath, in relazione alla sopravvivenza.

I parametri analizzati con il *software* sui vetrini EE sono i seguenti:

1. Quantità di stroma (area μm^2)
2. Quantità di tumore (area μm^2)
3. Rapporto stroma/tumore
4. Necrosi tumorale (area μm^2)
5. Purezza tumorale (%)

Il parametro analizzato sui vetrini immuno Ki-67:

6. Linfociti infiltranti lo stroma (linfociti peritumorali proliferanti in mitosi)

Inoltre, in aggiunta ai dati dei parametri analizzati con il *software*, sono state fatte indagini statistiche anche sui seguenti dati riportati nei referti diagnostici:

- Componente intraduttale;
- Indice di proliferazione (Ki-67);
- Diametro espresso come misurazione micrometrica.

I parametri che differiscono in maniera significativa tra le pazienti in vita e le pazienti decedute sono:

- area stroma, che è inferiore nelle decedute (p-value che varia da 0,01 a 0,03);
- purezza tumorale, che è inferiore nelle decedute (p-value = 0,04);
- rapporto stroma/tumore, ancora inferiore nelle decedute (p-value = 0,03);

- componente intraduttale, superiore nelle decedute (qui ho messo 0 e 1 a seconda che fosse presente o assente) con un p-value che varia da 0,01 a 0,03.

Ki-67 risulta essere mediamente più alto nelle decedute, ma la differenza non è statisticamente significativa.

I linfociti peritumorali proliferanti (%) sono inferiori nelle decedute, ma la differenza non è significativa.

Questi risultati vengono confermati con entrambi i test applicati, solo per la purezza tumorale la differenza risulta essere leggermente meno significativa con l'applicazione del primo test.

I parametri dell'area tumorale e della necrosi non sono significativi, anche se comunque l'area tumorale risulta essere superiore nelle decedute.

Per quanto riguarda il diametro espresso come misurazione micrometrica, è stato applicato solamente il t-test di Student ma i risultati non sono significativi per la dimensione tumorale.

Nella *Tabella 3* sono riportati i dati ottenuti dalle analisi con il *software* QuPath per quanto riguarda tutti i parametri di indagine, riportando per ogni paziente il numero istologico univoco.

Paziente	Area tumorale	Area stroma	Area necrosi	Linfociti peritumorali		
				proliferanti (%)	Purezza tumorale	Rapporto stroma/tumore
2013.699	14695088.28	396606295	4689024.7	17.24690792	88.06062481	26.9890379
2013.6402	31332186.11	138621073	1826080.8	7.271820449	63.89358757	4.42423878
2013.7011	86237194.67	127700743	146385031	10.58646449	55.32759762	1.48080818
2014.4762	27183158.72	184090906	NA	10.03307607	70.37697402	6.77224115
2015.1022	37961157.2	55282475.1	555417.37	16.98886936	58.55958549	1.45629057
2015.2474	58471945.2	190549352	17842864	12.5687956	71.75320204	3.25881671
2015.4683	67078482.9	58205604.2	578645.31	7.706592386	79.84723854	0.86772392
2016.8698	57147162.22	176820029	138592602	7.81650283	82.18874989	3.0941174
2017.5257	32733481.45	99099019.9	5762746.8	7.626208378	72.57000178	3.02745127

Tabella 3: Dati delle 9 pazienti LAR analizzati in Digital Pathology su immagini WSI dei vetrini istologici EE e Immuno Ki-67 e ottenuti con il software QuPath. In tabella sono riportati tutti i risultati per ogni singolo parametro: area tumorale, area stromale, area necrotica, linfociti peritumorali proliferanti, purezza tumorale e rapporto stroma/tumore.

Paziente	Sopravvivenza	Ki-67	Mammella	Dimensione micrometrica	Componente intraduttale
2013.699	viva	20%	dx	1.2 cm	presente, peritumorale, 80%
2013.6402	morta	35%	sx	2.2 cm	presente, peritumorale, 10%
2013.7011	morta	70%	dx	1.7 cm	presente, peritumorale, 10%
2014.4762	viva	7%	dx	4.8 cm	assente
2015.1022	morta	20%	dx	3.8 cm	presente, peritumorale, 5%
2015.2474	viva	30%	dx	2.5 cm	assente
2015.4683	morta	15%	dx	1.1 cm	presente, peritumorale, 4%
2016.8698	viva	38%	sx	1.4 cm	NA
2017.5257	morta	30%	sx	1.4 cm	presente, peritumorale, 40%

Tabella 4: Dati clinici di rilevanza delle 9 pazienti LAR utilizzati per ulteriori indagini statistiche. In particolare, sono riportati in ordine i dati dei seguenti parametri: sopravvivenza, indice di proliferazione Ki-67, sede mammaria del tumore, dimensione micrometrica e componente intraduttale espressa in percentuale.

Nella *Tabella 4*, invece, sono riportati i dati relativamente ai parametri che sono stati presi dalle diagnosi delle pazienti e confrontati poi in analisi statistica con il parametro della sopravvivenza, anch'esso riportato in tabella. Le 9 pazienti TNBC - LAR sono tutte pazienti che sono state operate tra l'anno 2013 e l'anno 2017 con un anno di nascita che varia dal 1930 al 1980. Essendo TNBC presentano una triplice negatività ai fattori prognostici e un Ki-67 che varia in valori percentuali dal 7% al 70%. Tutte hanno in comune la diagnosi, stabilita come carcinoma duttale infiltrante, e una dimensione micrometrica sulla sezione istologica che identifica la dimensione tumorale espressa sottoforma di diametro della porzione del tumore in valori che si discostano da 1,1 cm a 4,8 cm. Inoltre, la componente intraduttale per alcune pazienti risulta essere assente mentre per altre è presente in valori percentuali che variano dal 4% all'80%.

5. Discussione

L'analisi digitale dei parametri di indagine sui casi LAR consente di integrare sia caratteristiche locali sia caratteristiche globali dell'organizzazione tissutale, con l'obiettivo di confrontare questi dati tra loro e trovare delle caratteristiche comuni per poterli identificare e classificare.

Il calcolo con il software QuPath dei tre parametri analizzati durante la prima fase analitica è stato eseguito considerando le WSI nella loro interezza, quindi analizzando le immagini digitali intere senza dover restringere il campo e individuare specifiche regioni di interesse (ROI). L'analisi delle immagini WSI intere del carcinoma mammario ha riscontrato alcune difficoltà relativamente all'utilizzo del programma su computer con una RAM non così grande ed avanzata, che quindi non permetteva un'analisi ad alta risoluzione per la grandissima quantità di dati da visualizzare e processare. Inoltre, a seconda delle dimensioni delle regioni di interesse, della potenza del computer e della risoluzione dell'analisi, è necessario più tempo affinché vengano visualizzati i valori di misurazione delle aree. Questo perché, in presenza di annotazioni piuttosto grandi, il software QuPath suddivide le ROI in regioni più piccole sottoforma di tanti quadrati che vengono elaborati in parallelo. Questo approccio migliora la velocità di analisi e riduce i requisiti di memoria. Per questa ragione è stato creato un nuovo progetto per ogni immagine digitale analizzata, per evitare che si sovraccaricasse un'unica cartella con un contenuto veramente troppo pesante per il computer da supportare.

In particolare, per quanto riguarda i parametri della prima fase di analisi, c'è stata una segmentazione automatica dei compartimenti tissutali da parte del software abilitato all'Intelligenza Artificiale.

L'area tumorale è quella che è stata identificata per prima siccome questo passaggio rappresenta il passo più importante per il patologo nella diagnosi del cancro al seno. Infatti, la delimitazione del tumore dal resto del tessuto di fondo è uno step fondamentale e necessario per la successiva stadiazione e grading, [12] ed è avvenuto impostando uno specifico valore di soglia che delimitasse tutte le aree tumorali presenti nelle immagini WSI.

In seguito, si è riscontrata una criticità per quanto riguarda l'identificazione dell'area stromale e dell'area di necrosi e questo perché entrambi i parametri possiedono valori di colorazione simili, quindi non era possibile identificare un valore di soglia specifico per lo stroma e un altro valore di soglia per la necrosi. A causa di questa problematica, si è quindi valutato di identificare le regioni necrotiche manualmente andando a creare con il software QuPath delle ROI specifiche che contornavano i confini delle aree di necrosi permettendo di calcolarne l'area. Successivamente, per poter identificare il terzo parametro della prima fase di analisi, ossia lo stroma, è stato necessario individuare le regioni del bianco (rappresentato dallo sfondo e dal tessuto adiposo) con uno specifico valore soglia che delimitasse queste regioni calcolandone in seguito l'area. Questa procedura è stata fondamentale in quanto il valore di soglia per identificare le regioni stromali identificava non solo lo stroma ma anche il bianco e la necrosi, di conseguenza si è agito per sottrazione andando a sottrarre al primo valore calcolato l'area della necrosi e l'area del bianco, ottenendo così la sola area dello stroma.

Per questa prima fase di analisi digitale, in cui è avvenuto lo studio dei tre parametri di interesse, per ogni campione si è creato un progetto di analisi e tutte le immagini sono state sottoposte alla normalizzazione dei colori con la deconvoluzione stimando i vettori di colorazione. In seguito, si è eseguito lo step del rilevamento del

tessuto che, come abbiamo detto in precedenza, è avvenuto manualmente per l'area necrotica mentre per l'area tumorale e l'area stromale sono stati utilizzati dei valori di soglia per definire le regioni di interesse (ROI). Infine, i risultati delle misurazioni sono stati esportati sottoforma di tabelle Excel in una memoria esterna per poter sommare tra di loro le aree delle singole annotazioni individuate, con lo scopo di ottenere per ogni caso un'area tumorale totale, un'area stromale totale e un'area necrotica totale.

Una volta terminata la prima fase di analisi digitale, con i valori dei parametri ottenuti dall'area stromale e dall'area tumorale caratterizzati da importanti informazioni spaziali e architettoniche, si è calcolato il rapporto stroma/tumore dividendo l'area stromale per l'area tumorale. Di questo parametro può essere calcolato anche il reciproco, ossia il rapporto tumore/stroma, di cui è stato scoperto che un rapporto basso con valori di stroma elevati in casi TNBC, è associato a una prognosi sfavorevole. Diversi studi hanno infatti dimostrato che esiste quindi una specifica correlazione tra il rapporto stroma-tumore e la prognosi del paziente in diversi tipi di cancro al seno, incluso appunto il TNBC. [3][45]

Inoltre, sempre nella seconda fase di analisi, con un esperto Anatomo-Patologo si è eseguita l'indagine per il calcolo della purezza tumorale (%) andando a identificare e selezionare per ogni campione EE cinque ROI di interesse, caratterizzate da componenti cellulari e morfologia relativamente omogenee, nelle quali effettuare l'analisi. Anche in questo caso, per ogni campione EE è stato creato un nuovo progetto di analisi, si è eseguita la normalizzazione dei colori con la stima dei vettori di colorazione, è stato effettuato prima il rilevamento delle cellule presenti all'interno delle ROI con la funzione *Cell detection* riportando il numero totale di cellule detectate e poi la loro classificazione individuando le sole cellule tumorali,

esprimendo infine la purezza tumorale come il rapporto tra le cellule tumorali e il totale delle cellule detectate, in percentuale.

Lo stesso approccio si è mantenuto anche per la terza e ultima fase analitica, in cui si è sfruttato lo stesso ragionamento di base andando sempre a identificare delle regioni di interesse ROI più piccole in ogni campione Immuno Ki-67 per verificare e calcolare la presenza di linfociti peritumorali proliferanti in mitosi con la funzione *Positive cell detection*, che ha dato come risultati: il numero totale delle cellule detectate, il numero di quelle positive e negative, e la percentuale delle positive.

Tutti questi parametri indagati con un approccio di Patologia Computazionale possono essere considerati dei biomarcatori predittivi e prognostici basati sull'imaging, in particolare: l'identificazione dell'area tumorale distinguendo perciò il tumore, lo stroma e il tessuto adiposo. Infatti, la componente tumorale e quella stromale acquisiscono un ruolo primario e centrale nella crescita e quindi anche nello sviluppo e diffusione dei tumori solidi, come il TNBC, e nella risposta alla terapia.

L'analisi istopatologica, quindi, viene applicata per ottenere dei dati sulle informazioni visive correlate al tumore che possono essere utilizzate anche per stratificare il rischio dei pazienti. [44]

Estraendo prima le caratteristiche visive relativamente alla intensità del colore, alla struttura, alla forma e alla densità cellulare, ciò che si nota dall'analisi visiva di questi 9 campioni LAR in EE è che sono caratterizzati da un'eterogeneità morfologica per quanto concerne lo sviluppo del tumore stesso, che viene rilevata anche con una eterogeneità intra-tumorale per la distribuzione delle cellule tumorali stesse e un'eterogeneità peri-tumorale. Infatti, alcuni casi presentano un'estesa

necrosi centrale circondata esternamente da una abbondante flogosi periferica, mentre altri casi presentano una crescita tumorale distesa e diffusa, poca flogosi e minima o assente necrosi.

L'eterogeneità del tumore è un indicatore vitale delle distinte sottopopolazioni molecolari e cellulari presenti all'interno dell'area tumorale, che può quindi implicare diverse conseguenze sulla prognosi, sulla risposta al trattamento e sulla progressione della malattia. [3]

Questo è dato dal fatto che le immagini sono rappresentative delle intere sezioni del tessuto e forniscono quindi una rappresentazione completa dei campioni analizzati. Con la Patologia Digitale è possibile anche integrare queste caratteristiche visive con degli aspetti di tipo quantitativo, andando a caratterizzarle correlandole e trovando un'associazione con il parametro di sopravvivenza globale. [35]

Una delle conclusioni a cui si è arrivati al termine di questo studio è che negli ultimi anni si sta dando tanta rilevanza all'analisi digitale delle immagini perché si sostiene che essa sia obiettiva ed oggettiva rispetto alla valutazione visiva-manuale che effettua il Medico Anatomico-Patologo. La verità è che non è così, in quanto l'analisi digitale con l'utilizzo di un software deve essere svolta da un operatore che, durante le fasi analitiche, dovrà prendere delle decisioni in merito a come muoversi e agire. Di conseguenza, persone diverse possono prendere decisioni soggettive differenti che possono influenzare i risultati dell'analisi determinando delle discordanze.

Nonostante la colorazione Ematossilina-Eosina sia aspecifica, in quanto non è in grado di distinguere specifiche tipologie cellulari identificando dei veri e propri fenotipi cellulari come accade invece in quegli approcci che utilizzano tecniche ausiliarie di quantificazione come IHC o MIBI-TOF per l'espressione di diverse

proteine come antigeni tumorali o immunitari e proteine immunoregolatrici che sono però in realtà più costosi, richiedono molto più tempo e non sono utilizzati abitualmente nella routine ospedaliera, essa viene utilizzata come tecnica di riferimento. [46] Per questo motivo, siccome la colorazione EE è considerata come il *gold standard* della diagnosi patologica essa è molto diffusa e applicata quotidianamente. Ecco perché è fondamentale rivolgere l'attenzione su approcci o modelli che si basano su questa tecnica, in quanto hanno la possibilità di essere attuati regolarmente nella realtà diagnostica-clinica di tutti i giorni all'interno del contesto ospedaliero. [3]

Tuttavia, sebbene sia molto importante essere consapevoli che l'analisi delle immagini digitali non è obiettiva e che un limite importante di questo studio è il numero molto limitato dei campioni TNBC, ciò la rende comunque un importante e utile strumento di supporto che, usato dal Medico Anatomo-Patologo con cautela e con un'attenta validazione, può essere più veloce, più riproducibile e sicuramente più quantitativa della valutazione visiva; così come abbiamo dimostrato in questa tesi magistrale con un approccio innovativo che può essere uno spunto per studi futuri in *Digital Pathology*.

6. Conclusioni

In questa tesi magistrale si è svolto un lavoro sperimentale che ha avuto come oggetto di indagine casi di pazienti affette da carcinoma mammario, in particolare il sottotipo triplo negativo, che presenta come fattori prognostici una triplice negatività a livello immunostochimico: negatività dei recettori per gli estrogeni, negatività dei recettori per il progesterone e una negatività per HER-2.

Il nostro lavoro si è basato sullo studio di 9 casi di pazienti triple negative LAR con numeri istologici risalenti agli anni: 2013, 2014, 2015, 2016, 2017; periodi in cui sono state sottoposte a intervento chirurgico.

I vetrini istologici utilizzati nello studio sono stati presi da pazienti operate nella Azienda Sanitaria di Biella a partire dall'anno 2013, quindi si tratta totalmente di pezzi operatori e non biopsie o agoaspirati in modo da avere più materiale possibile a disposizione su cui lavorare e ottenere dei risultati completi e affidabili.

Questa indagine è stata svolta grazie all'utilizzo dello scanner Hamamatzu presente nella S.C. di Anatomia Patologica dell'Ospedale di Biella, il quale ha permesso la digitalizzazione dei vetrini istologici delle pazienti e quindi la loro semplice manipolazione come file digitali, facendo un'indagine attraverso uno specifico software chiamato QuPath. Mediante l'utilizzo del software si è effettuata un'analisi morfometrica dei vetrini come immagini digitalizzate attraverso quella che viene chiamata tecnologia WSI (*Whole Slide Imaging*).

L'analisi ha permesso lo studio di alcuni parametri, considerati come potenziali fattori prognostici per studi futuri in cui potrebbero essere oggetto di indagine con analisi più approfondite estese ad una coorte maggiore di pazienti, coinvolgendo anche gli altri sottotipi TNBC.

7. Bibliografia

[1] Humphries M.P., Maxwell P., Salto-Tellez M., **QuPath: The global impact of an open source digital pathology system**, Computational and Structural Biotechnology Journal, Volume 19, 2021, Pages 852-859, ISSN 2001-0370, <https://doi.org/10.1016/j.csbj.2021.01.022>.

(<https://www.sciencedirect.com/science/article/pii/S200103702100026X>)

[2] Bankhead, P. et al. **QuPath: Open source software for digital pathology image analysis**. Scientific Reports (2017). <https://doi.org/10.1038/s41598-017-17204-5>

[3] Germán Corredor et al, **A Review of AI-Based Radiomics and Computational Pathology Approaches in Triple-Negative Breast Cancer: Current Applications and Perspectives**, Clinical Breast Cancer, <https://doi.org/10.1016/j.clbc.2023.06.004>

[4] Akhouayri L, Ostano P, Mello-Grand M, Gregnanin I, Crivelli F, Laurora S, Liscia D, Leone F, Santoro A, Mulè A, Guarino D, Maggiore C, Carlino A, Magno S, Scatolini M, Di Leone A, Masetti R, Chiorino G. **Identification of a minimum number of genes to predict triple-negative breast cancer subgroups from gene expression profiles**. Hum Genomics. 2022 Dec 20;16(1):70. doi: 10.1186/s40246-022-00436-6. PMID: 36536459; PMCID: PMC9764480.

[5] Guilherme Aresta, Teresa Araújo, Scotty Kwok, Sai Saketh Chennamsetty, Mohammed Safwan, Varghese Alex, Bahram Marami, Marcel Prastawa, Monica Chan, Michael Donovan, Gerardo Fernandez, Jack Zeineh, Matthias Kohl, Christoph Walz, Florian Ludwig, Stefan Braunewell, Maximilian Baust, Quoc Dang Vu, Minh Nguyen Nhat To, Eal Kim, Jin Tae Kwak, Sameh Galal, Veronica Sanchez-Freire, Nadia Brancati, Maria Frucci, Daniel Riccio, Yaqi Wang, Lingling Sun, Kaiqiang Ma, Jiannan Fang, Ismael Kone, Lahsen Boulmane, Aurélio Campilho, Catarina Eloy, António Polónia, Paulo Aguiar, **BACH: Grand challenge on breast cancer histology images**,

Medical Image Analysis, Volume 56, 2019, Pages 122-139, ISSN 1361-8415,
<https://doi.org/10.1016/j.media.2019.05.010>.

(<https://www.sciencedirect.com/science/article/pii/S1361841518307941>)

[6] Hamamatsu Photonics K.K., **HAMAMATSU photon is our business**,
<https://www.hamamatsu.com/eu/en/product/life-science-and-medical-systems/digital-slide-scanner.html>

[7] Hamamatsu Photonics K.K. 2012, **High-throughput and high-resolution Digital Slide Scanner NanoZoomer-XR “Redefining the Art of Whole-slide Imaging”**,

Systems Division 812 Joko-cho, Higashi-ku, Hamamatsu City, 431-3196, Japan

<http://www.hamamatsu.com/>

[8] Azienda Sanitaria Locale di Biella (ASL-BI) Dipartimento dei Servizi Struttura Complessa Anatomia Patologica, **Istruzione operativa per l'attività cito-istologica agoaspirativa** , IO.DG.ANP.001

[9] Sergio Daniel, Tiziano Zanin, **Manuale di Tecnica Cito-Istologica- Documentazione scientifica**, Editrice Bologna Manuali

[10] Ruifrok, A C, and D A Johnston, **Quantification of Histochemical Staining by Color Deconvolution**. *Analytical and Quantitative Cytology and Histology*, 23 (4): 291–9.

[11] Landini G, Martinelli G, Piccinini F. **Colour Deconvolution – stain unmixing in histological imaging**. *Bioinformatics* 2020, <https://doi.org/10.1093/bioinformatics/btaa847>

[12] Cruz-Roa A, Gilmore H, Basavanhally A, Feldman M, Ganesan S, Shih NNC, Tomaszewski J, González FA, Madabhushi A. **Accurate and reproducible invasive breast cancer detection in whole-slide images: A Deep Learning approach for quantifying tumor extent**. *Sci Rep*. 2017 Apr 18;7:46450. doi: 10.1038/srep46450. PMID: 28418027; PMCID: PMC5394452.

[13] Wang CW, Muzakky H. **Computational Pathology for Breast Cancer and Gynecologic Cancer**. *Cancers* (Basel). 2023 Feb 2;15(3):942. doi: 10.3390/cancers15030942. PMID: 36765900; PMCID: PMC9913809.

[14] Sung H, Ferlay J, Siegel RL, et al. **Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries**. *CA A Cancer J Clin* . 2021;71(3):209–249 .

- [15] Giaquinto AN, Sung H, Miller KD, et al. **Breast Cancer Statistics, 2022.** *CA A Cancer J Clin* . 2022;72(6):524–541 .
- [16] Abramson VG, Lehmann BD, Ballinger TJ, Pietenpol JA. **Subtyping of triple-negative breast cancer: Implications for therapy: subtyping triple-negative breast cancer.** *Cancer* . 2015;121(1): 8–16 .
- [17] Perou CM, Sørlie T, Eisen MB, et al. **Molecular portraits of human breast tumours.** *Nature* . 2000;406(6797): 747–752 .
- [18] Brown LC, Salgado R, Luen SJ, Savas P, Loi S. **Tumor-infiltrating lymphocytes in triple-negative breast cancer: update for 2020.** *Cancer J* . 2021;27(1):25–31 .
- [19] Harris LN, Ismaila N, McShane LM, et al. **Use of biomarkers to guide decisions on adjuvant systemic therapy for women with early-stage invasive breast cancer:** American Society of Clinical Oncology Clinical Practice Guideline. *J Clin Oncol*. 2016;34(10):1134–50.
- [20] Bergin ART, Loi S. **Triple-negative breast cancer: recent treatment advances.** *F1000Res* . 2019;8. F1000 Faculty Rev–1342. doi: 10.12688/f1000research.18888.

- [21] Wu M, Zhang Y, Zhang Y, Liu Y, Wu M, Ye Z. **Imaging-based biomarkers for predicting and evaluating cancer immunotherapy response.** *Radiol: Imaging Cancer* . 2019;1(2) .
- [22] Jatoi I, Sung H, Jemal A. **The emergence of the racial disparity in U.S. breast-cancer mortality.** *N Engl J Med* . 2022;386(25):2349–2352 .
- [23] Sturtz LA, Melley J, Mamula K, Shriver CD, Ellsworth RE. **Outcome disparities in African American women with triple negative breast cancer: a comparison of epidemiological and molecular factors between African American and Caucasian women with triple negative breast cancer.** *BMC Cancer* . 2014;14(1):62 .
- [24] Bera K, Schalper KA, Rimm DL, Velcheti V, Madabhushi A. **Artificial intelligence in digital pathology – new tools for diagnosis and precision oncology.** *Nat Rev Clin Oncol* . 2019;16(11):703–715 .
- [25] Hong R, Fenyö D. **Deep learning and its applications in computational pathology.** *BioMedInformatics* . 2022;2(1):159–168 .
- [26] LeCun Y, Bengio Y, Hinton G. **Deep learning.** *Nature* . 2015;521(7553):436–444 .

[27] Iwamoto T, Kajiwara Y, Zhu Y, Iha S. **Biomarkers of neoadjuvant/adjvant chemotherapy for breast cancer.** *Chin Clin Oncol* . 2020;9(3):27. doi: 10.21037/cco.2020.01.06

[28] Mukhopadhyay S, Feldman MD, Abels E, et al. **Whole slide imaging versus microscopy for primary diagnosis in surgical pathology: a multicenter Blinded Randomized Noninferiority Study of 1992 Cases (Pivotal Study).** *Am J Surg Pathol* . 2018;42(1):39–52 .

[29] Keren L, Bosse M, Marquez D, et al. **A structured tumor-immune microenvironment in triple negative breast cancer revealed by multiplexed ion beam imaging.** *Cell* . 2018;174(6):1373–1387.e19

[30] Wimberly H, Brown JR, Schalper K, et al. **PD-L1 expression correlates with tumor-infiltrating lymphocytes and response to neoadjuvant chemotherapy in breast cancer.** *Cancer Immunol Res* . 2015;3(4):326–332 .

[31] Gao G, Wang Z, Qu X, Zhang Z. **Prognostic value of tumor-infiltrating lymphocytes in patients with triple-negative breast cancer: a systematic review and meta-analysis.** *BMC Cancer* . 2020;20(1):179 .

[32] Kilmartin D, O'Loughlin M, Andreu X, et al. **Intra-Tumour heterogeneity is one of the main sources of inter-observer variation in scoring stromal tumour infiltrating lymphocytes in triple negative breast cancer.** *Cancers* 2021;13(17):4410 .

- [33] García-Tejido P, Cabal ML, Fernández IP, Pérez YF. **Tumor-infiltrating lympho-cytes in triple negative breast cancer: the future of immune targeting.** *Clin Med Insights Oncol* . 2016;10:CMO.S34540. doi: 10.4137/CMO.S34540
- [34] Lu Z, Xu S, Shao W, et al. **Deep-learning–based characterization of tumor-infil-trating lymphocytes in breast cancers from histopathology images and multiomics data.** *JCO Clin Cancer Inform* . 2020(4):480–490 .
- [35] Wang C, Pécot T, Zynger DL, Machiraju R, Shapiro CL, Huang K. **Identifying survival associated morphological features of triple negative breast cancer using multiple datasets.** *J Am Med Inform Assoc* . 2013;20(4):680–687 .
- [36] Yu G, Li X, He TF, et al. **Predicting relapse in patients with triple negative breast cancer (TNBC) using a deep-learning approach.** *Front Physiol* . 2020;11 .
- [37] Mi H, Gong C, Sulam J, et al. **Digital Pathology Analysis Quantifies Spatial Heterogeneity of CD3, CD4, CD8, CD20, and FoxP3 Immune Markers in Triple-Negative Breast Cancer.** *Front Physiol* . 2020;11 .
- [38] Hatzis C, Symmans WF, Zhang Y, et al. **Relationship between complete pathologic response to neoadjuvant chemotherapy and survival in triple-negative breast cancer.** *Clin Cancer Res* . 2016;22(1):26–33 .

[39] Humphries MP, Hynes S, Bingham V, et al. **Automated tumour recognition and digital pathology scoring unravels new role for PD-L1 in predicting good outcome in ER-/HER2 + breast cancer.** *J Oncol* . 2018;2018:1–14 .

[40] Balkenhol MCA, Ciompi F, 'Swiderska-Chadaj ' Z, et al. **Optimized tumour infil-trating lymphocyte assessment for triple negative breast cancer prognostics.** *The Breast* . 2021;56:78–87 .

[41] Sun P, He J, Chao X, et al. **A computational tumor-infiltrating lymphocyte assess-ment method comparable with visual reporting guidelines for triple-negative breast cancer.** *EBioMedicine* . 2021;70 .

[42] Yuan Y. **Modelling the spatial heterogeneity and molecular correlates of lympho-cytic infiltration in triple-negative breast cancer.** *J R Soc Interface* . 2015;12(103). doi: [10.1098/rsif.2014.1153](https://doi.org/10.1098/rsif.2014.1153)

[43] Wang J, Browne L, Slapetova I, et al. **Multiplexed immunofluorescence identifies high stromal CD68 + PD-L1 + macrophages as a predictor of improved survival in triple negative breast cancer.** *Sci Rep* . 2021;11(1):21608

[44] Millar EK, Browne LH, Beretov J, et al. **Tumour stroma ratio assessment using digital image analysis predicts survival in triple negative and luminal breast cancer.** *Cancers* . 2020;12(12):3749

[45] Kramer CJH, Vangangelt KMH, van Pelt GW, Dekker TJA, Tollenaar RAEM, Mesker WE. **The prognostic value of tumour–stroma ratio in primary breast cancer with special attention to triple-negative tumours: a review.** *Breast Cancer Res Treat* . 2019;173(1):55–64 .

[46] Zhao S, Yan CY, Lv H, et al. **Deep learning framework for comprehensive molec-ular and prognostic stratifications of triple-negative breast cancer.** *Fundamental Research* . Published online June 2022:S2667325822002771. doi: [10.1016/j.fmre.2022.06.008](https://doi.org/10.1016/j.fmre.2022.06.008) .

[47] Fournier MV, Goodwin EC, Chen J, Obenauer JC, Tannenbaum SH, Brufsky AM. **A predictor of pathological complete response to neoadjuvant chemotherapy stratifies triple negative breast cancer patients with high risk of recur-rence.** *Sci Rep* . 2019;9(1):14863 .

[48] Liedtke C, Mazouni C, Hess KR, et al. **Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer.** *JCO* . . 2008;26(8):1275–1281 .

[49] Sun YS, Zhao Z, Yang ZN, Xu F, Lu HJ, Zhu ZY, Shi W, Jiang J, Yao PP, Zhu HP. **Risk Factors and Preventions of Breast Cancer.** *Int J Biol Sci*. 2017 Nov 1;13(11):1387-1397. doi: 10.7150/ijbs.21635. PMID: 29209143; PMCID: PMC5715522.