

# Preparazione di librerie Agilent SureSelect XT HS RNA: un flusso di lavoro semplificato e migliorato per la costruzione Direct-to-Capture di librerie di sequenziamento di RNA da campioni freschi o FFPE

## Autori

Carsten Carstens, Katherine Felts e Sarah Johns

Agilent Technologies, Inc.

## Abstract

La presente nota applicativa illustra un flusso di lavoro migliorato e più snello per la costruzione di librerie di sequenziamento mirato di RNA. Sono stati apportati quattro significativi miglioramenti al protocollo Agilent SureSelect XT RNA Direct:

1. Sostituzione dell'ibridazione di arricchimento target notturno con un'ibridazione rapida in 90 minuti
2. Eliminazione del trattamento con uracil-deglicosilasi (UDG) (la specificità del filamento di RNA è mantenuta tramite l'uso di un nuovo enzima)
3. Sostituzione degli adattatori SureSelect XT meno recenti con adattatori SureSelect XT HS che consentono l'elaborazione in parallelo di DNA ed RNA dallo stesso campione
4. Introduzione di un codice a barre molecolare (MBC) unico per una migliore de-duplicazione (eliminazione dei duplicati) della PCR e di frammentazione

Tali modifiche hanno permesso di ridurre il tempo di risoluzione da 2-3 giorni a 1 solo giorno. Inoltre, analizzando i dati di fusione con l'uso di vari tipi di campione di partenza (intatto, fresco congelato, FFPE), è stato possibile verificare che questo flusso di lavoro semplificato genera librerie di sequenziamento di RNA di complessità superiore e prestazioni di sequenziamento migliorate con soli 10 ng di RNA.

## Introduzione

L'applicazione alla trascrittomica (sequenziamento di RNA, RNA-seq) del sequenziamento ad alta produttività ha permesso non soltanto la valutazione dell'espressione genica globale, ma ha consentito anche di ottenere informazioni precise su varianti di splicing, trascritti di fusione, eventi di modifica post-traduzionale ed espressione allele-specifica. L'approccio standard alla costruzione di librerie RNA-seq consiste nella frammentazione

del campione di RNA mediante transesterificazione  $Mg^{2+}$ -dipendente, sintesi di cDNA del primo filamento con esamer-primer casuale, seguita da sintesi di cDNA del secondo filamento con marcatura dUTP dei filamenti. Per prevenire la concatenazione, le estremità del cDNA vengono troncate (ottenendo estremità non coesive o "blunt") e viene aggiunta un'adenina al terminale 3' tramite polimerizzazione del DNA senza template. Segue quindi la ligazione di adattatori di sequenziamento specifici per piattaforma. Tutte le fasi successive alla ligazione degli adattatori sono comuni al flusso di lavoro del sequenziamento di DNA. Sebbene le librerie di sequenziamento derivate da DNA e RNA appaiano identiche dopo la ligazione degli adattatori, dal confronto tra sequenziamento di RNA e sequenziamento di DNA emergono differenze specifiche. La prima è la necessità di mantenere la direzionalità dei frammenti di sequenziamento per determinare a quale filamento di DNA genomico corrisponda l'RNA originale. A tale scopo, in genere si include l'uracile nella reazione di sintesi del secondo filamento, si continua con la preparazione della libreria e in un fase successiva si usa l'UDG per rimuovere il secondo filamento. La seconda differenza è l'ampio range dinamico dei trascritti misurati nel sequenziamento di RNA, che può estendersi su oltre cinque ordini di grandezza a causa delle vaste differenze di abbondanza relativa dei trascritti codificanti espressi e dell'abbondanza estremamente elevata di RNA ribosomiale (rRNA) rispetto all'RNA codificante. L'RNA-seq pertanto impone la riduzione della complessità per evitare di sprecare read su trascritti che non codificano informazioni (per esempio: rRNA). Il metodo più comune prevede la rimozione dell'rRNA mediante deplezione mirata (deplezione dell'RNA ribosomiale) o cattura specifica di RNA poliadenilato (mRNA). In un campione tipico, anche dopo la rimozione dell'rRNA, il primo 1% di tutti i geni espressi corrisponde a ≈50% di tutti i trascritti. Ne consegue che se gli obiettivi di ricerca prevedono lo studio di geni a media o bassa espressione, un'ulteriore riduzione della complessità si rivelerà vantaggiosa.

In alternativa agli approcci di deplezione mirati, è possibile ridurre la complessità anche tramite l'arricchimento mirato con l'uso di sonde di arricchimento target biotinilate (note come "esche"), un approccio comunemente adottato nel sequenziamento genomico. Nel sequenziamento dell'RNA, l'arricchimento mirato tramite selezione dell'esca è usato principalmente in abbinamento a campioni derivati da FFPE, per i quali la deplezione dell'rRNA è notoriamente poco riproducibile e non è possibile impiegare l'arricchimento poli(A) a causa della frammentazione del materiale originale<sup>1</sup>. L'arricchimento mirato offre vantaggi anche quando è necessario esaminare solo un sottoinsieme relativamente ridotto del trascrittoma. Un chiaro esempio è l'individuazione di trascritti di fusione che sono indice di un evento di fusione di geni soggiacente. Tuttavia, qualsiasi scenario in cui solo un sottoinsieme di trascritti fornisce informazioni, per esempio fingerprinting trascrizionale o identificazione di rare modifiche post-trascrizionali, trae vantaggio dall'arricchimento mirato<sup>2</sup>.

Agilent offre il kit per la preparazione della libreria SureSelect XT RNA Direct (codice G7564A, G7564B) per consentire la costruzione di librerie di sequenziamento mirato di RNA. Una precedente nota applicativa dimostra la validità di tale kit per l'uso con campioni derivati da FFPE<sup>3</sup>. La presente nota applicativa descrive un flusso di lavoro semplificato e migliorato basato sull'abbinamento tra componenti del kit per la preparazione della libreria SureSelect XT RNA Direct e il kit di arricchimento mirato SureSelect XT HS (codice G9706A) unitamente a un protocollo modificato. Il flusso di lavoro è stato semplificato eliminando la liofilizzazione del campione di RNA, eliminando una fase di purificazione con bead SPRI e sostituendo la fase di ibridazione di 24 ore con esche con una fase di ibridazione rapida di 1,5 ore. È stata altresì eliminata la fase di trattamento con UDG, ottenendo invece la specificità del filamento tramite l'uso di un enzima per PCR che discrimina nei confronti del DNA templato contenente uracile durante la PCR di pre-cattura. Inoltre, l'uso di adattatori di sequenziamento SureSelect XT HS

aggiunge un MBC per l'identificazione di duplicati di frammentazione e diminuisce gli ostacoli pratici all'elaborazione in parallelo di DNA ed RNA dallo stesso campione.

## Condizioni sperimentali

### Fonti di RNA

L'RNA di riferimento umano universale (UHRR) è stato ottenuto come materiale fresco congelato presso Agilent Technologies (Santa Clara, CA, USA, codice 750500-41). Un set corrispondente di tessuto mammario tumorale e tessuto mammario normale adiacente è stato ottenuto sia come materiale fresco congelato sia come materiale FFPE presso CureLine Human Biospecimen CRO (Brisbane, CA, USA, codice personalizzato). Seraseq FFPE Tumor Fusion RNA Reference Material v2 è stato acquistato presso SeraCare (Gaithersburg, MD, USA, n. di cat. 0710-0129).

### Isolamento dell'RNA

Laddove necessario, l'RNA è stato isolato utilizzando il kit RNeasy FFPE o il mini kit RNeasy di Qiagen secondo le istruzioni del produttore (Qiagen USA, Germantown, MD, USA, codice 73504 e 74104 rispettivamente). Per maggiori dettagli sul protocollo, consultare l'Appendice.

### Valutazione della qualità del materiale di partenza e delle librerie di sequenziamento

I campioni di acidi nucleici sono stati valutati sul sistema bioanalizzatore 2100 Agilent (Agilent Technologies, codice G2939B) utilizzando il kit Agilent RNA 6000 Pico (Agilent Technologies, codice 5067-1513) per il punteggio di qualità dell'RNA o il kit Agilent DNA 1000 (Agilent Technologies, codice 5067-1504) per valutare la qualità delle librerie di sequenziamento.

### Altri materiali

L'actinomicina D è stata acquistata presso Sigma (St. Louis, MO, USA, codice A1410) e portata a una concentrazione di 4 µg/µL in una soluzione stock di DMSO. Le purificazioni con bead SPRI sono state effettuate utilizzando bead AMPure XP (Beckman Coulter, Atlanta, GA, USA, codice A63880). La cattura delle sonde biotinilate è stata effettuata utilizzando

bead DynabeadsMyOne Streptavidin T1 (Thermo Fisher Scientific, Waltham, MA, USA, n. di cat. 65601).

### Preparazione di librerie SureSelect XT HS RNA

La costruzione di librerie di sequenziamento di RNA è stata effettuata utilizzando il kit SureSelect XT RNA Direct (Agilent Technologies, codice G7564A) e il sistema di arricchimento target SureSelect XT HS per la libreria di sequenziamento multiplex a paired-end Illumina (Agilent Technologies, codice G9706A). Per una descrizione dettagliata, consultare l'Appendice.

### Preparazione di librerie SureSelect XT RNA Direct

Le librerie RNA Direct sono state generate, arricchite e sequenziate seguendo le istruzioni del protocollo (manuale) di preparazione delle librerie SureSelect XT RNA Direct.

### Arricchimento target

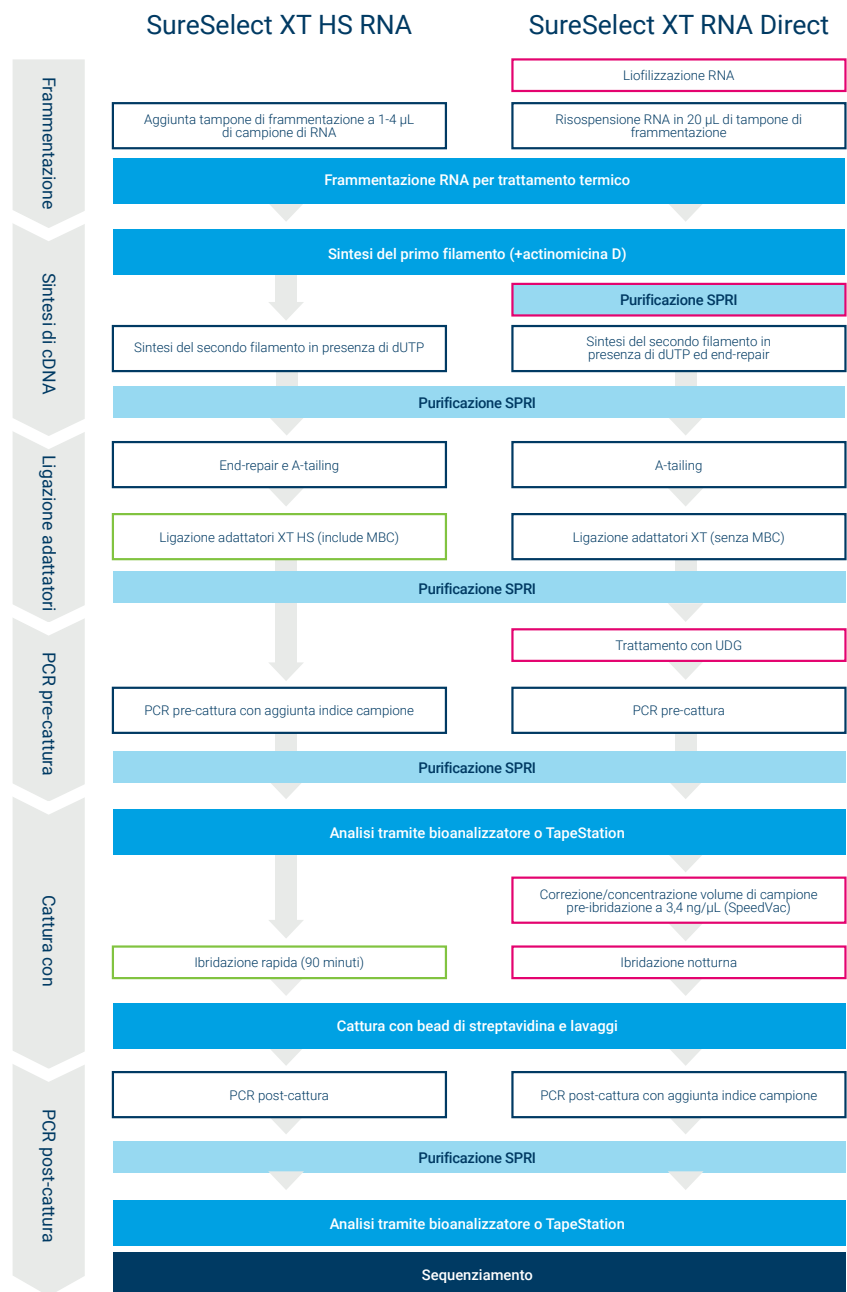
L'arricchimento target delle librerie SureSelect XT HS RNA è stato effettuato utilizzando l'esoma SureSelect Human All Exon V7 (Agilent Technologies, codice 5191-4029) avendo come target il trascrittoma codificante. Un protocollo dettagliato per la cattura delle esche è disponibile nell'appendice.

### Analisi del sequenziamento e dei dati

Le librerie di sequenziamento sono state analizzate su un Illumina HiSeq 4000 tramite sequenziamento paired-end con un formato di lettura 2 x 150. Per l'analisi dell'espressione (dati non mostrati), i file FASTQ sono stati allineati al trascrittoma utilizzando il pacchetto splice-aware STAR versione 2.6.0a utilizzando la struttura genomica hg38 come riferimento. I profili di espressione sono stati quindi generati dal risultato dell'allineamento STAR utilizzando lo strumento RSEM. Le statistiche generali delle librerie (specificità del filamento, bias estremità 5'-3', tassi di duplicazione MBC-blind, stime delle dimensioni delle librerie) sono state generate mediante la procedura di analisi Picard RNA con duplicati marcati utilizzando file .bam sottocampionati a  $2 \times 10^7$  read per generare tassi di duplicazione normalizzati.

Le statistiche di duplicazione con correzione MBC e le stime delle dimensioni delle librerie sono state generate con la stessa procedura ma utilizzando UmiAwareMarkDuplicatesWithMateCigar per la rimozione dei duplicati di

frammentazione. Il punteggio assegnato ai trascritti di fusione è stato ottenuto con STAR-Fusion e i trascritti sono stati visualizzati con FusionInspector Tool, facente parte del kit Trinity Cancer Transcriptome Analysis Tool (CTAT)<sup>4</sup>.



**Figura 1.** Confronto tra il flusso di lavoro Agilent SureSelect XT HS RNA (RNA XT HS) e il kit per la preparazione della libreria Agilent SureSelect XT RNA Direct (RNA Direct). Le fasi eliminate dal flusso di lavoro RNA Direct sono indicate da un riquadro rosa. Le fasi che conferiscono una migliore funzionalità al protocollo RNA XT HS sono indicate da un riquadro verde.

## Risultati e discussione

### Panoramica dei flussi di lavoro RNA XT HS ed RNA Direct

Il presente lavoro mirava allo sviluppo di un flusso di lavoro semplificato e migliorato per generare librerie di sequenziamento di RNA di alta qualità abbinando componenti del kit per la preparazione della libreria SureSelect XT RNA Direct (RNA Direct) con il kit di arricchimento mirato SureSelect XT HS. Il flusso di lavoro risultante è stato denominato SureSelect XT HS RNA (RNA XT HS). Una panoramica di confronto tra i due protocolli è riportata in Figura 1. Il flusso di lavoro RNA XT HS presenta ovvi vantaggi rispetto a quello RNA Direct, tra cui l'eliminazione del trattamento con UDG e la sostituzione dell'ibridazione notturna tradizionale con un'ibridazione rapida in 90 minuti, diminuendo quindi in misura considerevole il tempo della procedura. Gli altri obiettivi del presente lavoro erano determinare se fosse possibile rimuovere la fase di purificazione SPRI tra la sintesi del primo filamento e quella del secondo ed evitare la liofilizzazione del campione di RNA nel flusso di lavoro RNA XT HS (consultare l'Appendice per informazioni dettagliate).

### Confronto delle prestazioni dei flussi di lavoro

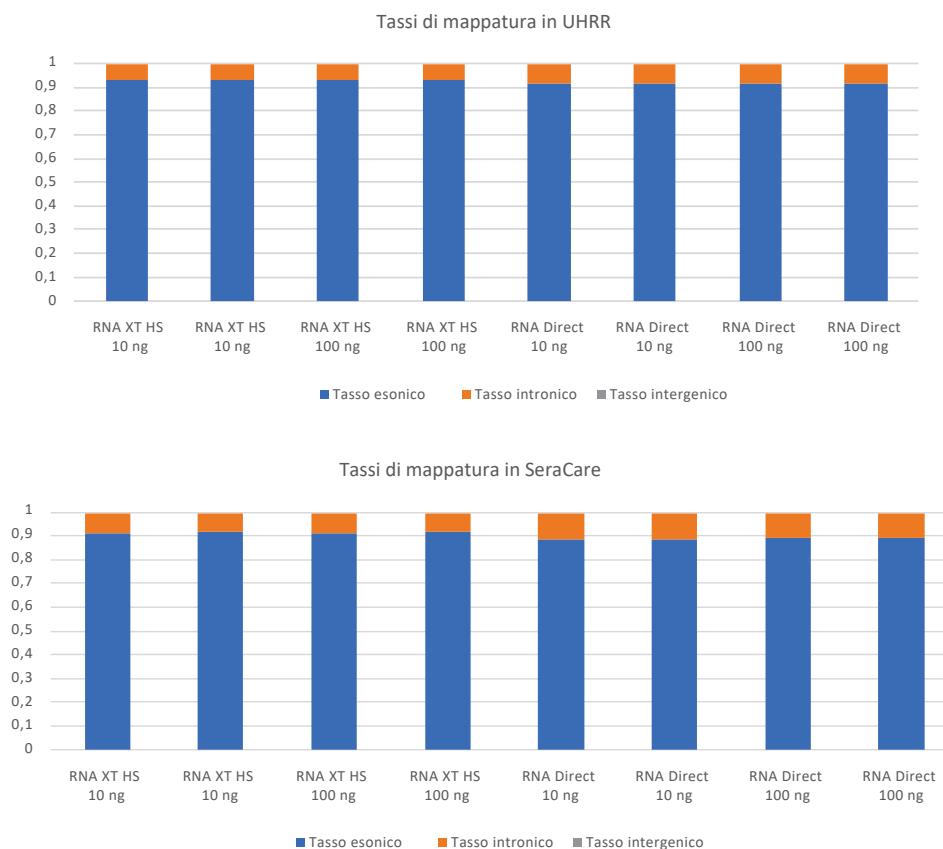
Per confrontare le prestazioni dei due processi di preparazione di librerie, sono state innanzitutto generate librerie impiegando varie quantità di RNA di riferimento umano universale (UHRR) e SeraCare FFPE Tumor Fusion RNA Reference Material v2 (SeraCare) come materiale di partenza, a rappresentazione rispettivamente di un campione intatto e di un campione FFPE ideale. Le librerie RNA XT HS ed RNA Direct sono state arricchite con SureSelect Human All Exon V7. Infine, tali librerie arricchite sono state sequenziate impiegando un sequenziatore Illumina e i loro dati analizzati mediante procedure di analisi dei dati personalizzate (fare riferimento alla sezione Condizioni sperimentali per informazioni dettagliate).

In Tabella 1 è riportato un riepilogo delle statistiche di sequenziamento globali risultanti dall'analisi dei dati di sequenziamento RNA XT ed RNA Direct.

Le librerie sequenziate RNA XT HS sono risultate indistinguibili dalle librerie RNA Direct relativamente a svariate metriche, in particolare per quanto riguarda l'elevata specificità del filamento (>98%) e la bassa contaminazione da rRNA (~0,1%). L'elevata specificità del filamento delle librerie RNA XT HS è prova dell'efficacia del flusso di lavoro RNA XT HS, il cui approccio prevede l'eliminazione del trattamento con UDG in favore dell'utilizzo di un enzima per PCR che non amplifica i template contenenti uracile. La specificità del filamento tende a essere leggermente migliore per i campioni intatti rispetto ai campioni FFPE, come prevedibile in considerazione della minor qualità del materiale di partenza FFPE. Indipendentemente da ciò, le specificità dei filamenti osservate per i campioni FFPE con entrambi i

protocolli rimangono comunque molto elevate (>98% per materiale FFPE). La percentuale di contaminazione da rRNA costantemente bassa per entrambi i metodi di preparazione delle librerie è stata riscontrata in precedenza negli approcci di arricchimento mirato<sup>3</sup>.

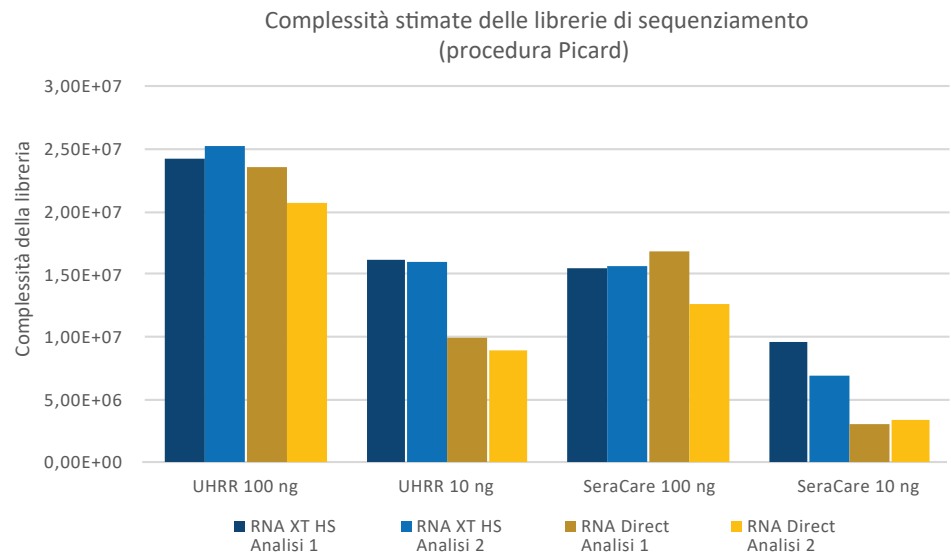
Anche il confronto dei tassi di mappatura esonica tra le librerie RNA XT HS ed RNA Direct ha evidenziato prestazioni confrontabili (Figura 2). Un passaggio critico per ridurre i tempi del nuovo flusso di lavoro RNA XT HS è la sostituzione dell'ibridazione notturna tradizionale con una fase di ibridazione rapida. Come mostrato in Figura 2, la cattura accelerata non incide sui tassi di mappatura; inoltre, i tassi esonici osservati sono costantemente del 90% con pochissime read mappate a regioni intergeniche, indipendentemente



**Figura 2.** Confronto dei tassi di mappatura delle librerie sequenziate RNA XT HS ed RNA Direct. Le librerie RNA-seq sono state preparate da campioni UHRR (A) e SeraCare (B) utilizzando il flusso di lavoro RNA XT HS o RNA Direct. Sono mostrati i tassi di mappatura delle librerie di sequenziamento arricchite dell'esoma Agilent SureSelect Human All Exon V7 alle sequenze esoniche, introniche e intergeniche. Si sospetta che il tasso intronico rifletta il sequenziamento di mRNA non elaborato.

dal materiale di partenza. I trascritti target della libreria di cattura V7 costituivano il 92,9-94,1% dell'intera espressione osservata, indipendentemente dalla quantità di partenza o dal tipo di campione. Il restante 7% circa è principalmente dovuto a trascritti contaminanti ad alta espressione, quali geni mitocondriali, oppure a errori di annotazione.

Un indicatore cruciale dell'efficienza del processo complessivo di preparazione delle librerie è la complessità stimata della libreria iniziale. Librerie più grandi e quindi più complesse sono preferibili, in quanto sequenziandole si ottengono informazioni più significative e una quantificazione più affidabile. La complessità di una libreria si calcola a partire dalla profondità di sequenziamento (in coppie di read) e dal numero di varianti uniche osservate, ottenuto dal tasso di duplicazione. Per calcolare le complessità stimate delle librerie in questo lavoro si è utilizzata la procedura Picard, la cui ipotesi è che tutti i duplicati sono duplicati della PCR. Questo approccio sottostima la vera complessità delle librerie e sarà discusso in maggior dettaglio nella sezione relativa ai risultati. Le complessità proiettate risultanti delle librerie sono elencate in Tabella 1 e mostrate in Figura 3. Il confronto tra i flussi di lavoro RNA XT HS ed RNA Direct non ha evidenziato differenze significative tra le librerie costruite da alte quantità (100 ng) di campioni di RNA fresco e FFPE. Tuttavia, riducendo la quantità di campione a 10 ng sia per UHRR sia per SeraCare, sono state riscontrate notevoli differenze nella complessità delle librerie. In primo luogo, le librerie con meno materiale di partenza sono più piccole e meno complesse di quelle con una maggior quantità di materiale di partenza. Ciò non sorprende, considerato che quanto più bassa è la quantità di campione di RNA da convertire in una libreria di sequenziamento, tanto più piccola dovrebbe essere la libreria finale. Con una quantità di campione più bassa, inoltre, il flusso di lavoro semplificato RNA XT HS produce librerie aventi un'efficienza 1,5-2 volte maggiore rispetto al flusso di lavoro RNA Direct. Al momento sono oggetto di studio svariati fattori del flusso di lavoro RNA XT HS (dati non mostrati) che hanno dato luogo a questo guadagno di efficienza.



**Figura 3.** Differenze di complessità tra le librerie RNA XT HS ed RNA Direct con una bassa quantità di campione di partenza. Le librerie di sequenziamento preparate usando le fonti e quantità indicate sono state sequenziate su una piattaforma Illumina HiSeq 4000 tramite sequenziamento paired-end. Le complessità stimate delle librerie sono state determinate in base ai tassi di duplicazione osservati e al rispettivo numero di coppie di read determinate dallo strumento di analisi Picard RNA. Questa procedura non discrimina tra duplicati della PCR e di frammentazione.

**Tabella 1.** Statistiche di sequenziamento globali delle librerie SureSelect XT HS RNA e SureSelect XT RNA Direct: le librerie di sequenziamento di RNA generate dallo stesso materiale di partenza utilizzando il protocollo RNA XT HS o RNA Direct sono state sequenziate sulla piattaforma Illumina HiSeq 4000. Dopo il sottodimensionamento a  $2 \times 10^7$  read, le statistiche delle librerie sono state generate con lo strumento di analisi Picard RNA. Le statistiche con correzione MBC sono state generate con la stessa procedura dopo la marcatura degli MBC con UmiAwareMarkDuplicatesWithMateCigar. (A) Campione di RNA fresco congelato ("intatto"); (B) campione FFPE.

#### A. RNA di riferimento umano universale (UHRR)

Protocollo	SureSelect XT HS RNA				SureSelect XT RNA Direct			
	100	100	10	10	100	100	10	10
Campione di partenza (ng)	100	100	10	10	100	100	10	10
Read analizzate (milioni)	1,9	1,9	1,9	1,9	1,9	1,9	1,9	1,9
Tasso rRNA (%)	0,1%	0,1%	0,1%	0,1%	0,1%	0,1%	0,1%	0,1%
Specificità del filamento (%)	99,1%	99,1%	99,0%	99,0%	99,1%	99,2%	99,1%	99,1%
Tasso di duplicazione (%)	17,3%	16,7%	24,3%	24,7%	17,9%	20,0%	35,8%	38,7%
Dimensioni stimate della libreria ( $\times 10^6$ )	24,2	25,2	16,2	15,9	27,5	23,6	10,3	9,2
Tasso di duplicazione con correzione MBC (%)	7,1%	6,4%	17,1%	17,8%	NA	NA	NA	NA
Dimensioni stimate della libreria con correzione MBC ( $\times 10^6$ )	65,9	72,8	24,6	23,6	NA	NA	NA	NA

#### B. Seraseq FFPE Tumor Fusion RNA Reference Material v2 (SeraCare)

Protocollo	SureSelect XT HS RNA				SureSelect XT RNA Direct			
	100	100	10	10	100	100	10	10
Campione di partenza (ng)	100	100	10	10	100	100	10	10
Read analizzate (milioni)	18,6	18,6	18,7	18,6	19	18,8	18,9	18,9
Tasso rRNA (%)	0,1%	0,2%	0,1%	0,1%	0,1%	0,2%	0,1%	0,1%
Specificità del filamento (%)	98,6%	98,6%	98,6%	98,7%	98,9%	98,9%	98,9%	98,9%
Tasso di duplicazione (%)	24,8%	24,7%	36,0%	45,0%	23,5%	29,4%	69,1%	67,2%
Dimensioni stimate della libreria ( $\times 10^6$ )	15,5	15,5	9,6	6,9	18,6	13,3	3	3,3
Tasso di duplicazione con correzione MBC (%)	13,2%	13,7%	29,3%	38,8%	NA	NA	NA	NA
Dimensioni stimate della libreria con correzione MBC ( $\times 10^6$ )	32,5	31,2	12,7	8,66	NA	NA	NA	NA

## Effetto degli identificatori molecolari unici (MBC) sulle librerie di sequenziamento di RNA

Come ricordato in precedenza, la misura accurata delle complessità delle librerie è cruciale per la valutazione dell'efficienza del processo di preparazione delle librerie. Le stime della complessità delle librerie di cui sopra si basavano sull'ipotesi che le coppie di read con gli stessi siti di inizio e fine fossero duplicati della PCR derivati dalla stessa molecola della libreria originale. I duplicati, tuttavia, possono avere origine anche da una frammentazione casuale, che genera due frammenti indipendenti con le stesse estremità. Poiché i duplicati di frammentazione sono veri membri indipendenti delle librerie di sequenziamento, la determinazione delle complessità delle librerie andrebbe condotta solo in base ai veri duplicati della PCR.

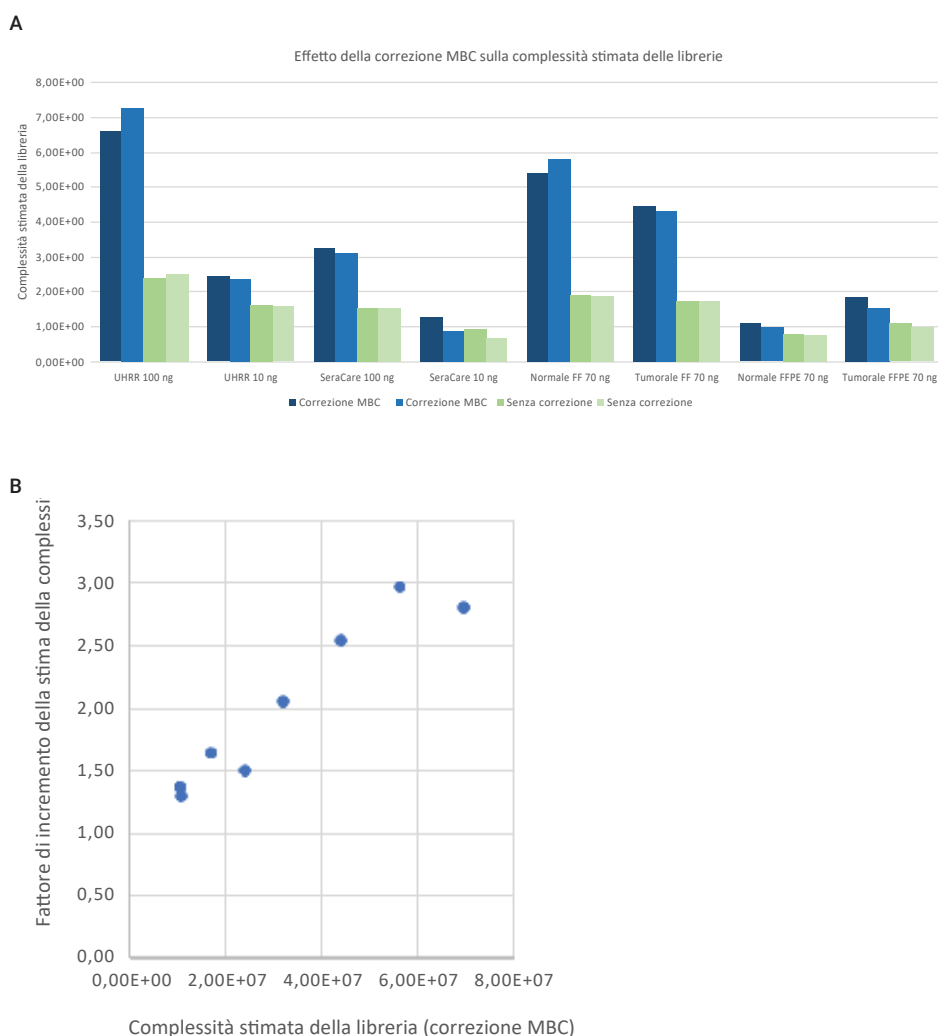
Rispetto al sequenziamento del DNA, il sequenziamento dell'RNA può generare un maggior numero di duplicati di frammentazione, poiché l'alta espressione di alcuni geni incrementa la possibilità casuale di duplicati di frammentazione. Il risultato sono stime meno accurate delle complessità delle librerie. Un potenziale vantaggio del flusso di lavoro RNA XT HS è che le librerie sono costruite con adattatori XT HS contenenti singoli MBC da 10 bp. È possibile ipotizzare che questo MBC possa essere utilizzato per distinguere i duplicati della PCR e di frammentazione. I dati di sequenziamento RNA XT HS di cui sopra sono stati nuovamente analizzati impiegando una procedura di analisi dei dati modificata che sfrutta i dati MBC (nota: le librerie RNA Direct non includono un MBC e pertanto non sono state incluse nella nuova analisi). Il risultato di questa analisi è riportato in Tabella 1 e in Figura 4.

Come mostrato in Figura 4, la correzione che tiene conto dei duplicati di frammentazione determina stime più elevate delle complessità delle librerie corrispondenti, in particolare per le quantità di partenza più alte, caso in cui la complessità della libreria viene sottostimata di circa tre volte se non si considerano i duplicati di frammentazione. Il bias più pronunciato osservato per le librerie più grandi e le corrispondenti

quantità di partenza più alte è prevedibile in considerazione della maggior probabilità di osservare duplicati di frammentazione quando più molecole derivate dalla stessa sequenza di codifica vengono elaborate nelle librerie.

Il presente lavoro ha quindi preso in esame l'effetto della correzione MBC su un set di campioni "reali". A tale scopo, sono state generate librerie RNA XT HS impiegando un set corrispondente di tessuto mammario tumorale e tessuto mammario normale adiacente, entrambi conservati e trattati

come campioni freschi congelati e campioni FFPE. Queste librerie sono state arricchite, sequenziate e analizzate come sopra con le librerie RNA XT HS UHRR e SeraCare. Le statistiche di riepilogo di tali librerie sono disponibili in Tabella 2 e in Figura 4. Sono stati osservati specificità del filamento e tasso rRNA eccellenti in dati di sequenziamento tumorale/normale fresco congelato ed FFPE, confrontabili con le librerie UHRR e SeraCare. È stato inoltre osservato che le librerie FFPE sono significativamente più piccole delle librerie di tessuto fresco



**Figura 4.** Effetto della correzione con identificatori molecolari unici (MBC) sulle stime delle dimensioni delle librerie. Le dimensioni delle librerie RNA XT HS sono state calcolate o per de-duplicazione Picard (senza correzione) o dopo l'utilizzo di MBC (correzione MBC). La de-duplicazione Picard ipotizza che tutti i duplicati siano duplicati della PCR, mentre la correzione MBC distingue tra duplicati della PCR e di frammentazione. **A)** Complessità stimate delle librerie per varie librerie di sequenziamento RNA XT HS derivate da svariati tipi e quantità iniziali di campione. **B)** Le differenze tra complessità delle librerie con e senza correzione MBC sono più pronunciate per le librerie più grandi e, pertanto, per le corrispondenti quantità di partenza più alte.

congelato, un fenomeno ben noto.

Analizzando i dati di sequenziamento FFPE con l'approccio basato sulla correzione MBC, sono stati osservati miglioramenti di un fattore 0,3-3,5 nella complessità stimata delle librerie e, di conseguenza, nei dati recuperati che sarebbero andati persi nell'ipotesi che tutti i duplicati fossero duplicati della PCR anziché duplicati della PCR e di frammentazione. Ciò evidenzia che allo scopo di incrementare al massimo le informazioni di sequenziamento fornite da campioni di RNA FFPE, è essenziale identificare accuratamente in tali campioni i duplicati della PCR e di frammentazione.

### Rilevamento di fusioni di geni

È stato dimostrato che il sequenziamento mirato di RNA è utile nell'individuare eventi di fusione di geni, in particolare per i campioni più complicati (FFPE). Il presente lavoro si proponeva di determinare se l'approccio RNA XT HS potesse funzionare per questo caso di utilizzo nelle condizioni di arricchimento con sonde esoma (rispetto a un pannello specifico per fusione) e massimizzazione della lunghezza della sequenza. Come ricordato in precedenza, le librerie RNA XT HS da tutti i campioni (UHRR, SeraCare, fresco congelato ed FFPE) sono state sottoposte ad arricchimento mirato con SureSelect Human All Exon V7 e le librerie arricchite sono state sequenziate come libreria paired-end a lunghezza di lettura  $2 \times 150$ . I dati di sequenziamento tumorale/normale fresco congelato ed FFPE sono stati analizzati con STAR-Fusion e visualizzati con FusionInspector.

FusionInspector ha rilevato un range di fusioni putative inclusi falsi positivi noti quali eventi di ricombinazione VDJ, a indicare la presenza di cellule immunitarie in alcuni campioni. I ricombinanti VDJ sono stati rimossi da Tabella 3. L'analisi dei dati mostra 15 potenziali eventi di fusione di geni nei campioni in esame; 12 dei 15 partner di fusione sono stati precedentemente segnalati in quanto coinvolti in fusioni di geni correlate a tumori, sebbene non nella combinazione osservata. Una delle fusioni, FCHSD2-FAM168A (evidenziata in grigio in Tabella 3), è stata precedentemente

**Tabella 2.** Statistiche di sequenziamento globali delle librerie Agilent SureSelect RNA XT HS di tessuto fresco congelato ed FFPE. Librerie di sequenziamento di RNA generate da un set corrispondente di tessuto tumorale e normale adiacente sia come fresco congelato sia FFPE utilizzando il protocollo RNA XT HS. Le librerie sono state sequenziate sulla piattaforma Illumina HiSeq 4000. Dopo il sottodimensionamento a  $2 \times 10^7$  read, le statistiche delle librerie sono state generate con lo strumento di analisi Picard RNA con o senza marcatura MBC.

Tipo di campione	Fresco congelato				FFPE			
	Normale	Normale	Tumorale	Tumorale	Normale	Normale	Tumorale	Tumorale
Fonte								
Campione di partenza (ng)	70	70	70	70	70	70	70	70
Read analizzate (milioni)	18,8	18,8	18,7	18,8	18,6	18,6	18,7	18,7
Tasso rRNA (%)	0,2%	0,2%	0,2%	0,2%	0,2%	0,2%	0,1%	0,1%
Specificità del filamento (%)	99,0%	99,0%	99,1%	99,2%	98,4%	98,4%	98,8%	98,8%
Tasso di duplicazione (%)	21,0%	21,0%	23,0%	23,0%	41,0%	43,0%	33,0%	35,0%
Dimensioni stimate della libreria ( $\times 10^6$ )	19	18,8	17,3	17,3	7,9	7,4	10,9	9,8
Tasso di duplicazione con correzione MBC (%)	8,0%	8,0%	10,0%	10,0%	33,0%	35,0%	22,0%	25,0%
Dimensioni stimate della libreria con correzione MBC ( $\times 10^6$ )	53,9	58,2	44,7	43,1	11	9,93	18,6	15,4

**Tabella 3.** Rilevamento di fusioni di geni in campioni di tessuto normale/tumorale fresco congelato ed FFPE. Le librerie di sequenziamento RNA XT HS sono state sequenziate e successivamente analizzate per rilevare eventi di fusione di geni mediante una procedura STAR-Fusion. Le fusioni putative rilevate in uno o più campioni sono elencate in ciascuna riga con le evidenze a supporto del rilevamento della fusione in questione. Read di giunzione e read di supporto sono espresse come read per milione di coppie di read uniche della libreria. Ciascuna colonna rappresenta la media di due replicati tecnici. La fusione evidenziata in grigio è identificata nel database TCGA come una fusione associata al cancro della mammella e potrebbe essere una mutazione driver per questo tumore. Gli eventi di ricombinazione VDJ osservati indicativi della presenza nei campioni di cellule immunitarie sono stati rimossi dall'analisi. Il rilevamento si basa sull'analisi di  $0,7-1,4 \times 10^7$  coppie di read uniche per ciascun campione.

Fusione putativa	Fresco congelato				FFPE			
	Normale		Tumorale		Normale		Tumorale	
	Giunzione	Supporto	Giunzione	Supporto	Giunzione	Supporto	Giunzione	Supporto
MYLK-LPAR6	0,00	0,00	8,86	0,86	0,00	0,00	1,68	0,00
FCHSD2-FAM168A	0,00	0,04	5,60	0,21	0,00	0,00	2,79	0,04
FAM157A-RB1	0,00	0,00	2,59	0,04	0,00	0,00	0,91	0,00
RP4-565E6.1-HYDIN	0,00	0,00	1,47	0,00	0,00	0,00	1,33	0,00
CDR2-FRG1	0,00	0,00	0,42	0,00	0,00	0,00	0,48	0,00
SLC7A5-RP11-645C24.2	0,00	0,04	0,33	0,20	0,00	0,00	0,09	0,04
PHRF1-TXNDC5	0,00	0,00	2,78	0,17	0,00	0,00	0,00	0,00
NUDT1-AC004840.8	0,00	0,00	1,71	0,00	0,00	0,00	0,00	0,00
MPZL1-RCS1	0,00	0,00	1,34	0,06	0,00	0,00	0,00	0,00
CCDC66-SLMAP	0,00	0,00	0,96	0,00	0,00	0,00	0,00	0,00
UBE2Q2-C15orf27	0,00	0,00	0,77	0,06	0,00	0,00	0,00	0,00
POLR2J-AC004980.9	0,00	0,00	0,00	0,00	11,53	0,00	0,00	0,00
POLR2J-UPK3B	0,00	0,00	0,00	0,00	3,96	0,00	0,00	0,00
RP11-634B7.4-TRIM58	0,36	0,00	0,00	0,00	0,00	0,00	0,00	0,00
RB1-MYLK-AS1	0,00	0,00	0,16	0,03	0,00	0,00	0,00	0,00

identificata in associazione al cancro della mammella e potrebbe essere potenzialmente una mutazione driver per questo tumore<sup>5</sup>. Per tutte le altre fusioni tumore-specifiche, almeno un partner di fusione (e in genere entrambi) è elencato nel database TCGA ed è associato alla formazione di tumori. Per il trascritto di fusione A:B individuato, sia A che B sono presenti nel database dei tumori. Non è invece presente la voce A:B. Sono tuttavia presenti esempi di trascritti di fusione A:C e D:B. Questi dati suggeriscono che le librerie RNA XT HS arricchite con un esoma permettono ai ricercatori di identificare le fusioni a livello dell'intero trascrittoma, anche nei campioni più complicati.

## Conclusione

Per migliorare l'efficienza della costruzione di librerie di sequenziamento mirato di RNA, questo lavoro si proponeva di migliorare l'attuale flusso di lavoro SureSelect XT RNA Direct utilizzando componenti del kit per la preparazione della libreria SureSelect XT RNA Direct con il kit di arricchimento mirato SureSelect XT HS. Il nuovo flusso di lavoro SureSelect XT HS RNA integra un insieme di miglioramenti significativi:

- La liofilizzazione del campione di RNA è stata sostituita dall'aggiunta diretta del tampone di frammentazione al campione di RNA.
- La fase di purificazione con bead SPRI dopo la sintesi di cDNA del primo filamento è stata eliminata.
- È stato utilizzato un enzima per PCR alternativo che non amplifica i template contenenti uracile, eliminando la necessità di trattamento con UDG per mantenere la specificità del filamento.
- La procedura di cattura con sonde ("esche") è stata modificata da un'ibridazione di 24 ore a un protocollo di ibridazione rapida di 1,5 ore.
- Gli adattatori di sequenziamento XT sono stati sostituiti dagli adattatori XT HS, che permettono l'uso di un MBC per una stima più accurata della complessità della libreria e "recupero" delle read di sequenziamento.

I risultati ottenuti dimostrano che il flusso di lavoro RNA XT HS riduce il tempo di risoluzione per la creazione della libreria e l'arricchimento mirato da 2-3 giorni a 1-2 giorni. Il nuovo flusso di lavoro semplificato genera librerie RNA-seq che sono indistinguibili dalle librerie generate con il protocollo SureSelect XT RNA Direct in termini di specificità del filamento, tassi rRNA e tassi di mappatura con quantità di partenza alte. Con quantità di partenza più basse (10 ng), i risultati ottenuti dimostrano che il flusso di lavoro abbreviato non incide sulle prestazioni complessive e, in effetti, sembra essere più efficiente. I risultati dimostrano anche che l'inclusione di MBC nel design migliorato (XT HS) degli adattatori consente ora l'identificazione di duplicati di frammentazione. L'identificazione più efficace dei duplicati di frammentazione migliora il risultato del sequenziamento evitando la perdita di read che si riscontra nel caso dei metodi di duplicazione start stop standard.

Le prove effettuate su campioni "reali" hanno dimostrato che il flusso di lavoro RNA XT HS continuava a produrre dati di alta qualità anche nel caso dei campioni FFPE. L'analisi preliminare delle fusioni di geni ha rilevato potenziali fusioni geniche nei campioni di tessuto tumorale fresco congelato ed FFPE, suggerendo un potenziale caso di utilizzo nel mondo reale del sequenziamento di RNA FFPE con il protocollo RNA XT HS. Sebbene questo lavoro non riporti dati di espressione genica o splicing, l'analisi preliminare condotta indica che il sequenziamento mirato di RNA può essere impiegato nell'analisi dell'espressione genica globale, nel rilevamento delle varianti di splicing, nel rilevamento dell'espressione delle varianti e nell'analisi dell'espressione allele-specifica. Per finire, lo sviluppo del flusso di lavoro SureSelect RNA XT HS lo "allinea" al kit di arricchimento mirato DNA SureSelect XT HS, semplificando il sequenziamento in parallelo di DNA ed RNA dallo stesso campione. Ciò potrebbe apportare vantaggi in diverse applicazioni, inclusa la ricerca multi-omica generica e l'elaborazione parallela di campioni per l'analisi TMB-MSI e il rilevamento della fusione.

## Abbreviazioni

FFPE, fissato in formalina, incluso in paraffina; TPM, trascritti per milione di chilobasi; nt, nucleotide; UHRR, RNA di riferimento umano universale; MBC, codice a barre molecolare; UDG, uracil-deglicosilasi;

## Bibliografia

- 1) Cieslik, M., *et al.* The Use of Exome Capture RNA-Seq for Highly Degraded RNA with Application to Clinical Cancer Sequencing. *Genome Res.* **2015**, 25, 1372–1381.
- 2) Mittempergher, L., *et al.*, MammaPrint and Blueprint Molecular Diagnostics Using Targeted RNA Next-Generation Sequencing Technology. *J. Mol. Diagn.* **2019**, 21, 808–823.
- 3) Jones, J. C.; Alex Siebold, A.; Lucas, A. B. SureSelect XT RNA Direct Protocol Provides Simultaneous Transcriptome Enrichment and Ribosomal Depletion of FFPE RNA. Nota applicativa Agilent Technologies, codice pubblicazione 5991-8119EN, **2017**.
- 4) Haas, B., *et al.* STAR-Fusion: Fast and Accurate Fusion Transcript Detection from RNA-Seq. bioRxiv. 120295 (**2017**).
- 5) Hu, X., *et al.* TumorFusions: An Integrative Resource for Cancer-Associated Transcript Fusions. *Nucleic Acids Res.* **2018**, 4, 46(D1), D1144-D1149.

## Appendice

### 1. Protocollo dettagliato per la costruzione di librerie di sequenziamento di RNA

#### Preparazione del campione di RNA

L'RNA totale da sezioni FFPE è stato isolato utilizzando un kit Qiagen RNeasy FFPE secondo le istruzioni del produttore. L'RNA da tessuto congelato è stato isolato utilizzando un mini kit Qiagen RNeasy. Tutti i campioni di RNA totale sono stati analizzati su un sistema bioanalizzatore 2100 Agilent utilizzando un kit RNA 6000 Pico. I campioni sono stati riscaldati a 80 °C per 2 minuti prima del caricamento sul chip. I valori RIN e DV200 sono stati calcolati dal software Bioanalyzer. Tali metriche di qualità dei campioni per tutti i campioni di RNA totale testati sono riportate in Tabella S1.

#### Preparazione di cDNA utilizzando i reagenti del kit Agilent SureSelect XT RNA Direct

*Nota: una soluzione stock di actinomicina D a 4 µg/µL in DMSO è stata preparata preventivamente e conservata congelata in aliquote monouso (3 µL) a -20 °C.*

*Nota: la miscela di frammentazione, il master mix per il primo filamento, l'enzima per il secondo filamento e le miscele oligo del kit SureSelect XT RNA Direct sono state scongelate in ghiaccio e agitate in vortex per 5 secondi ad alta velocità, quindi brevemente centrifugate prima dell'uso.*

1. I campioni di RNA totale sono stati preparati in un volume di 4 µL di acqua senza nucleasi. Le quantità di partenza differivano tra gli esperimenti e sono indicate nella presentazione dei risultati.

*Nota: è possibile usare volumi di partenza minori (<4 µL) di campioni di RNA ma si sconsiglia l'utilizzo di volumi di partenza maggiori (>4 µL) di campioni di RNA.*

2. La miscela di frammentazione è stata aggiunta al campione di RNA per ottenere un volume di campione finale pari a 20 µL.
3. Il campione di RNA è stato frammentato per riscaldamento in un'unità SureCycler 8800 (o

Tabella S1. Metriche di qualità e condizioni di frammentazione dei campioni.

Descrizione del campione	RIN	DV200	Frammentazione
RNA di riferimento umano universale (UHRR)	9,2	94%	94 °C, 8 min
Seraseq FFPE Tumor Fusion RNA Reference Material v2	2	54%	94 °C, 3 min, 65 °C, 2 min
Tessuto mammario normale congelato	6,3	94%	94 °C, 8 min
Tessuto mammario tumorale congelato	5	88%	94 °C, 8 min
Tessuto mammario normale (FFPE)	2,1	48%	65 °C, 5 min
Tessuto mammario tumorale (FFPE)	2	47%	65 °C, 5 min

Tabella S2. Miscela di reazione del primo filamento.

Reagente	Volume per una reazione	Volume per otto reazioni + 10% in eccesso
Master mix RNA-seq primo filamento	8 µL	70,4 µL
Actinomicina D (120 ng/µL)	0,5 µL	4,4 µL
<b>Totale</b>	<b>8,5 µL</b>	<b>74,8 µL</b>

Tabella S3. Parametri del protocollo di lavaggio SPRI.

Volume bead AMPure	105 µL (1,8x il volume di campione)
Tempo di incubazione bead	5 minuti
Lavaggio in etanolo 70% (eseguire due volte)	200 µL
Tempo di asciugatura a 37 °C	1-2 min o meno
Volume d'eluizione	50 µL di acqua senza nucleasi

- termociclatore equivalente) in presenza di condizioni dipendenti dalle metriche di qualità del singolo campione di RNA, così come consigliato nel protocollo del kit SureSelect XT RNA Direct (G9691). I parametri di frammentazione usati per ciascun campione sono indicati in Tabella S1. Dopo la frammentazione, i campioni sono stati conservati in ghiaccio in attesa del passaggio alla fase di sintesi del primo filamento.
4. Una soluzione stock di actinomicina D a 4 µg/µL in DMSO è stata diluita in acqua fino a 120 ng/µL (3 µL di actinomicina D + 97 µL di acqua).
5. Una miscela di reazione sfusa per otto campioni è stata preparata per la sintesi del primo filamento (Tabella S2). La miscela di reazione è stata agitata in vortex e tenuta in ghiaccio fino all'uso.
6. 8,5 µL della miscela di reazione per il primo filamento sono stati aggiunti a ciascuna aliquota di 20 µL di campione

- frammentato conservato in ghiaccio. I campioni sono stati miscelati con il vortex e brevemente centrifugati.
7. Le miscele di reazione da 28,5 µL sono state incubate in un'unità SureCycler 8800 pre-programmata per 10 minuti a 25 °C, quindi per 40 minuti a 37 °C e infine conservate a 4 °C (o in ghiaccio) fino al passaggio alla sintesi del secondo filamento.
8. Sia la provetta con la miscela enzimatica secondo filamento + end-repair sia la provetta con la miscela oligo RNA-seq secondo filamento + end-repair sono state agitate in vortex prima dell'uso.
9. 25 µL della miscela enzimatica secondo filamento + end-repair (tappo blu) sono stati aggiunti ai 28,5 µL della reazione del primo filamento in ghiaccio.
10. Sono stati immediatamente aggiunti 5 µL della miscela oligo secondo filamento + end-repair (tappo giallo).

11. I campioni sono stati tappati, miscelati con il vortex, brevemente centrifugati e quindi rimessi in ghiaccio.
12. Le miscele di reazione da 58,5 µL sono state incubate in un'unità SureCycler 8800 pre-programmata per 60 minuti a 16 °C e quindi conservate a 4 °C (o in ghiaccio) fino al passaggio alla purificazione SPRI.
13. Dopo la sintesi del secondo filamento, il cDNA è stato sottoposto a purificazione SPRI in base al protocollo descritto in Tabella S3.

**Preparazione di librerie di cDNA SureSelect XT HS utilizzando i reagenti del kit per la preparazione della libreria SureSelect XT HS**

14. End-repair, dA-tailing e ligazione degli adattatori XT HS del cDNA sono stati effettuati utilizzando i reagenti e le istruzioni del sistema di arricchimento mirato SureSelect XT HS per il kit e protocollo di sequenziamento multiplex a paired-end Illumina (G9702).

*Nota: il protocollo SureSelect XT HS è stato seguito a partire dalla Fase 3, pagina 27, con le eccezioni e modifiche descritte nelle fasi dalla 15 alla 20 del presente protocollo.*

15. Fase 6, pagina 34, del protocollo SureSelect XT HS. L'amplificazione delle librerie con PCR pre-cattura è stata eseguita secondo le istruzioni di tale protocollo, modificando il numero di cicli PCR come segue:
  - a. UHRR, Seraseq™ v2 per 12 cicli PCR (RNA di alta qualità)
  - b. Campioni di quadranti della mammella per 14 cicli PCR (RNA di bassa qualità)
16. Le librerie PreCap sono state valutate mediante bionalizzatore o TapeStation per determinare la resa e la distribuzione delle dimensioni delle molecole delle librerie.

**Arricchimento target e sequenziamento delle librerie di cDNA SureSelect XT HS**

17. Fase 1, pagina 46, del protocollo SureSelect XT HS. L'arricchimento mirato è stato effettuato con 200 ng di libreria PreCap come materiale di partenza, utilizzando i reagenti di ibridazione rapida SureSelect XT HS. Per l'ibridazione sono stati utilizzati 5 µL di sonde esoma SureSelect Human All Exon V7.
18. La cattura con bead di streptavidina e i successivi lavaggi sono stati effettuati secondo il protocollo.
19. L'amplificazione con PCR post-cattura è stata effettuata secondo il protocollo utilizzando 12 cicli PCR per tutti i campioni. Le librerie PostCap sono state valutate mediante bionalizzatore o TapeStation per determinare la resa e la distribuzione delle dimensioni delle molecole delle librerie.
20. Tutte le librerie sono state sequenziate su una piattaforma Illumina HiSeq 4000 a una lunghezza di lettura 2 × 150.

[www.agilent.com](http://www.agilent.com)

**Solo per scopi di ricerca. Non utilizzabile per procedure diagnostiche.**

Le informazioni fornite possono variare senza preavviso.

PR7000-2381  
© Agilent Technologies, Inc. 2019, 2020  
Stampato negli Stati Uniti, 29 gennaio 2020  
5994-1644ITE

