



**SARDEGNA
RICERCHE**

Sardegna FESR 2014/2020 - ASSE PRIORITARIO I

“RICERCA SCIENTIFICA, SVILUPPO TECNOLOGICO E INNOVAZIONE”

Azione 1.1.4 Sostegno alle attività collaborative di R&S per lo sviluppo di nuove tecnologie sostenibili, di nuovi prodotti e servizi

**PROGETTO CLUSTER TOP DOWN
“NIASMIC - Not Invasive Analysis of Somatic
Mutations In Cancer”**

RAPPORTO TECNICO

**Messa a punto e ottimizzazione del protocollo
sperimentale**

1/12



**SARDEGNA
RICERCHE**

Indice

Premessa	3
Purificazione del DNA	3
Dosaggio degli acidi nucleici	4
Messa a punto delle condizioni ottimali per la preparazione delle librerie	5
Primo esperimento	7
Secondo esperimento	7
Terzo esperimento	8
Quarto esperimento	8
Conclusioni	9
Bibliografia	12



**SARDEGNA
RICERCHE**

Premessa

Dalla preparazione degli acidi nucleici fino alla generazione dei dati di sequenziamento il protocollo consiste in numerosi passaggi. Dal punto di vista della messa a punto del protocollo di laboratorio, l'attività del progetto è iniziata con l'esplorazione della letteratura e dei protocolli di kit commerciali per ognuna delle fasi di trattamento dei campioni. Inoltre, la scelta dei reagenti e della strategia di analisi si è avvalsa dello scambio di informazioni con colleghi di altri laboratori e delle informazioni acquisite dall'esperienza pluriennale del laboratorio del CRS4 su svariate applicazioni della tecnologia NGS per migliaia di campioni biologici di molteplici origini, sia per progetti di ricerca e collaborazioni, che per servizi svolti in qualità di facility.

Purificazione del DNA

L'estrazione degli acidi nucleici dai diversi tessuti è stata eseguita manualmente con diversi kit.

Per il cfDNA nella maggior parte della letteratura e dei protocolli commerciali è indicato l'utilizzo del QIAamp MinElute ccfDNA Mini Kit (QIAGEN) [1]. Per avere la possibilità di una replica, un'aliquota di 2,6ml di plasma è stata utilizzata per ogni estrazione seguendo le indicazioni del protocollo, mentre una seconda aliquota è stata conservata come backup.

Per l'estrazione del DNA genomico costituzionale (gDNA) abbiamo utilizzato il QIAamp Blood DNA Mini Kit (QIAGEN) e il relativo protocollo per buffy coat partendo da un volume di 200ul di campione.

L'estrazione del DNA dal tessuto tumorale, ricevuto dalla chirurgia del reparto di Otorinolaringoiatria del Policlinico Universitario di Monserrato (CA) in una soluzione di formalina e tampone fosfato e successivamente conservato a -80°C in una soluzione di etanolo al 70% in acqua, è stata eseguita con il kit QIAamp Blood DNA Mini Kit (QIAGEN) e il relativo protocollo. Brevemente: una porzione di circa 25-30mg di tessuto è stato frantumato meccanicamente su una piccola piastra petri con l'aiuto di un pestello, risospeso in 200ul di Buffer ATL (QIAGEN) e dopo aggiunta di proteinasi K la lisi è stata eseguita per 3 ore a 56°C in un ThermoMixer (Eppendorf) in agitazione a 500rpm. In seguito si è proceduto alla purificazione del DNA seguendo le indicazioni del protocollo.

Il DNA proveniente dal CNR di Li Punti (SS) e dall'Ospedale Gaslini di Genova è stato estratto da tessuto fissato in formalina e incluso in paraffina (FFPE) mediante il kit GeneRead DNA FFPE (QIAGEN) e seguendo la procedura indicata dal manuale.



**SARDEGNA
RICERCHE**

Dosaggio degli acidi nucleici

Il dosaggio del DNA è stato eseguito con l'utilizzo di due diversi metodi: la lettura spettrofotometrica mediante Nanodrop (ThermoFisher) rileva la concentrazione di tutti gli acidi nucleici (dsDNA e ssDNA) oltre alla presenza di contaminanti, evidenziabile da valori inferiori a 1,8 per il rapporto 260/280nm (proteine) e valori inferiori a 1,6 per il rapporto 260/230nm (sali e solventi); il dosaggio fluorimetrico mediante Qubit BroadRange reagent (ThermoFisher) e lettore di micropiastre Infinite F200Pro (TECAN) rileva la concentrazione del solo DNA a doppia elica.

Per il gDNA i valori di concentrazione risultano sempre sovrapponibili.

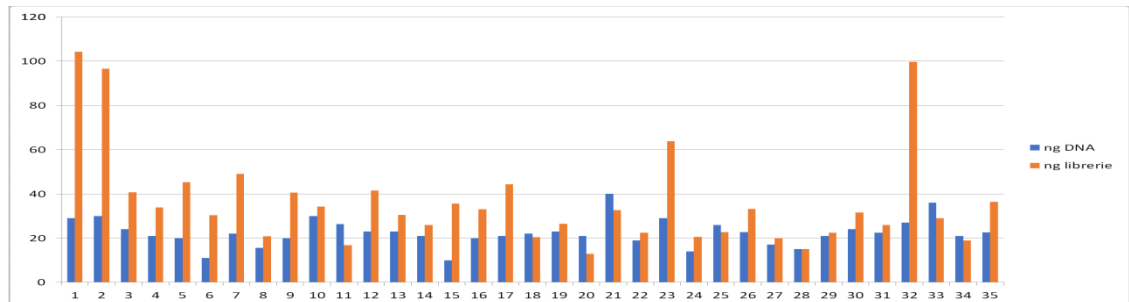
Il dosaggio del cfDNA spesso dà valori poco riproducibili (duplicati discordanti oltre il 30%) e non correlati alla reale riuscita e quantità di librerie che si ottengono (figura 1). Inoltre, data la scarsità del materiale ottenuto dall'estrazione (range: 10-40ng/ul), si è preferito non consumarne per averne a disposizione una maggiore quantità.

Il dosaggio del tDNA risulta quasi sempre discordante tra le due metodiche. La concentrazione data dal Nanodrop risulta sempre molto maggiore rispetto a quella rilevata in fluorescenza (nell'ordine di 10 volte). Questo risultato indica che il DNA estratto da tessuto è principalmente sotto forma di singola elica. Tale fenomeno è esclusivo del DNA da tessuto tumorale e si riscontra sia per i campioni estratti nel nostro laboratorio che per quelli ricevuti dai collaboratori. Inoltre, in passato è stato evidenziato su numerosi campioni, provenienti da diversi laboratori e ricevuti dal CRS4 per servizi di NGS, e non riguarda solo il kit di estrazione della QIAGEN ma anche altri sistemi (automatizzati o manuali), indifferentemente che sia da tessuto fresco o da FFPE. Per tentare di risolvere il problema sono stati interpellati gli specialist di prodotto della Qiagen (molto poco utili) e sono state eseguite delle prove modificando le temperature di incubazione (una seconda incubazione prevista dal protocollo è stata portata da 70°C a 56°C) ma senza ottenere risultati soddisfacenti e riproducibili. Come vedremo successivamente, la presenza di DNA denaturato, non solo influisce sull'efficienza del protocollo di library prep, richiedendo un maggiore numero di cicli di amplificazione, ma porta anche a ottenere sequenze con una maggiore quantità di GC rispetto a frammenti ricchi in AT; fenomeno già osservato eseguendo l'analisi dell'esoma su campioni di DNA estratto da tessuto tumorale.



**SARDEGNA
RICERCHE**

Figura 1. Librerie con 10 cicli di PCR. Concentrazione ng/ul di cfDNA estratto (blu), librerie ottenute (arancio)



Messa a punto delle condizioni ottimali per la preparazione delle librerie

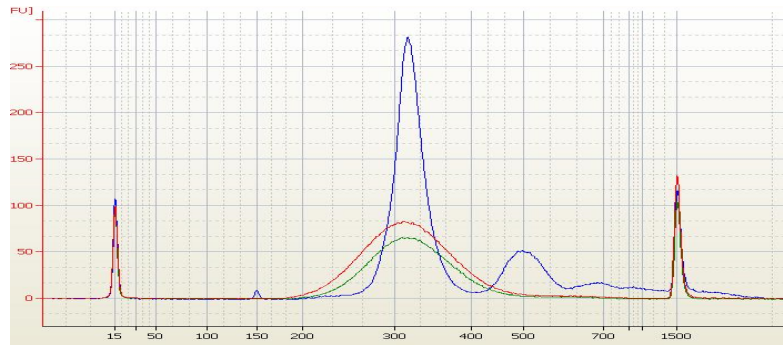
Esistono sul mercato diversi kit per la preparazione di librerie, molti dei quali volti ad ottenere il prodotto finale con la minore quantità di DNA di partenza e con il minor numero di step al fine di ridurre, sia i tempi di lavoro manuale, sia le perdite di DNA che si hanno quando per ogni reazione il prodotto deve essere purificato. Per lo sviluppo del protocollo NIASMIC sono stati selezionati reagenti che, sulla base della letteratura e dell'esperienza pregressa, possono essere considerati i più idonei.

Per la preparazione delle librerie è stato utilizzato il kit KAPA HyperPlus (Roche), basato sulla frammentazione enzimatica, quando necessario, e l'esecuzione delle diverse reazioni in una singola provetta senza purificazioni intermedie. Protocollo che in letteratura è stato descritto come il più efficiente nella preparazione di librerie a partire da basse quantità di DNA [2]. Per ottenere gli inserti della grandezza desiderata è stato necessario eseguire alcune reazioni di messa a punto per identificare il tempo di incubazione ideale nella reazione di frammentazione. In Figura 2 le grandezze ottenute per diverse librerie.



**SARDEGNA
RICERCHE**

Figura 2. Profili elettroforetici delle librerie ottenute da cfDNA (blu), gDNA (rosso), tDNA (verde)



Gli adattatori sono stati acquistati dalla IDT (Integrated DNA Technologies) e hanno la caratteristica di portare al loro interno due diverse sequenze index di 8nt uniche per le due estremità dei frammenti (Unique Dual-Index); il doppio index unico è necessario per mitigare il problema dell'index hopping. Una sequenza random di 9nt è inserita subito al 3' dell'index 1 ed è detta UMI (Unique Molecular Identifiers). Come è noto la frammentazione nucleosomica produce frammenti uguali o molto simili tra loro, l'introduzione di una sequenza UMI negli adattatori permette di associare ad ogni singolo frammento uno dei 262.144 tag possibili in maniera casuale. Nella successiva analisi bioinformatica tali sequenze consentono di riconoscere, sia i duplicati di PCR dalle sequenze uniche, sia i falsi positivi dovuti a errori di incorporazione durante la PCR.

La ligazione può essere eseguita con diverse condizioni. Per ottenere la maggiore efficienza possibile sono state sperimentate due diverse incubazioni: a 20°C per 4 ore e a 4°C per 16 ore. In questa reazione la concentrazione degli adattatori da utilizzare dipende dalla quantità di DNA iniziale; solo per i campioni di cfDNA vengono diluiti alla concentrazione 3uM in Duplex Buffer (IDT) per minimizzare la formazione di dimeri ed evitare di trattenerli nelle successive reazioni. La ligazione viene quindi purificata mediante biglie paramagnetiche (AMPure XP beads, Beckman Coulter) alla concentrazione 0,8X allo scopo di eliminare i reagenti residui della reazione e i dimeri di adattatori. Per uniformare la grandezza dei prodotti di ligazione da tDNA e gDNA a quella che si ottiene da cfDNA, allo scopo di portare nella reazione di ibridazione frammenti omogenei, è necessario eseguire una size-selection che si ottiene variando la concentrazione delle biglie magnetiche.

La Polymerase Chain Reaction (PCR) con primer complementari alle estremità p5 e p7 degli adattatori deve essere eseguita con il minor numero di cicli possibile per ridurre la produzione di duplicati di PCR. L'analisi della letteratura e dei protocolli commerciali ha suggerito l'utilizzo della 2X KAPA HighFidelity HotStart Ready Mix [3] che, oltre a minimizzare gli errori di incorporazione



**SARDEGNA
RICERCHE**

dei nucleotidi, diminuisce i bias di amplificazione tipici di altre polimerasi che risultano in una sottorappresentazione di sequenze ricche in GC o in AT rispetto a sequenze neutre. La correzione di questi bias porta alla riduzione dei falsi positivi e una maggiore uniformità di copertura delle sequenze target.

Primo esperimento

Inizialmente si era pensato di eseguire le reazioni di ibridazione producendo dei pool di librerie con i campioni di DNA di 3 pazienti per svolgere in un'unica reazione l'analisi di cfDNA, tDNA e gDNA, in modo da ottenere i dati corrispondenti da una lane della flow cell di sequenziamento. Per raggiungere le rispettive coperture delle regioni geniche era necessario sbilanciare le quantità in nanogrammi di librerie proporzionalmente al coverage. Per un coverage 5000X del cfDNA servono 1000ng di libreria, per il tDNA a 1000X servono 200ng e per il gDNA a 200X 40ng e allestire quindi la reazione di ibridazione con 2500-3000ng complessivi di librerie, come indicato in diversi protocolli. Per il cfDNA sono stati necessari un minimo di 12 cicli di PCR, rispetto ai 5 cicli richiesti da tDNA e gDNA. I risultati dell'analisi bioinformatica di questi primi 12 campioni ha evidenziato un'elevata percentuale di duplicati di PCR (media=80,6%) per il cfDNA, rispetto a valori più accettabili per il tDNA (60,6%) e per il gDNA (31%).

Per questo primo esperimento l'ibridazione delle sonde con le librerie è stata eseguita per 18 ore a 65°C e alla stessa temperatura sono stati eseguiti i lavaggi di stringenza per eliminare i frammenti aspecifici. Queste condizioni sono il frutto di una scelta ponderata sulla base di informazioni ricavate dallo studio di diversi protocolli. Sia il manuale Agilent (fornitore del pool di sonde a RNA), che quello della IDT (fornitore del kit di reagenti contenente i buffer di ibridazione e di lavaggio), nei rispettivi protocolli di ibridazione e dei successivi lavaggi utilizzano una temperatura di 65°C. In seguito alla selezione dei target la PCR di arricchimento finale è stata eseguita con 12 cicli.

Dall'analisi bioinformatica dei dati è emerso che le sequenze ricche in GC erano over-rappresentate rispetto ai frammenti ricchi in AT (60%GC).

Negli esperimenti successivi sono state apportate diverse modifiche al fine di diminuire i duplicati di PCR e correggere lo sbilanciamento nel contenuto di GC.

Secondo esperimento

Nelle sessioni di analisi successive è stata adottata una strategia di multiplexing diversa per diminuire la quantità di librerie di cfDNA necessaria per l'ibridazione. Si è deciso di produrre dei pool uniformi, composti quindi da 12 campioni di cfDNA, 12 di tDNA e 12 di gDNA. In questo modo per ogni campione sono necessari 250ng di libreria e si è potuto portare il numero di cicli a 10 per cercare di minimizzare i duplicati di PCR che vengono sequenziati. Anche la seconda amplificazione post-cattura è stata portata a 10 cicli. Inoltre, le temperature di ibridazione e dei



**SARDEGNA
RICERCHE**

lavaggi post-cattura sono state impostate a 62°C per correggere lo sbilanciamento a favore delle sequenze con maggiore contenuto di GC.

Risultati ottenuti: inaspettatamente la quantità di duplicati di PCR per il cfDNA è risultata essere più elevata (media 88%), mentre per i campioni di gDNA e di tDNA è stata ottenuta una diminuzione, rispettivamente 16,2% e 55,1%. Il bias di contenuto delle basi sequenziate è stato corretto (51,3% GC).

Tutti i campioni usati in questa sessione sono stati estratti nel laboratorio del CRS4.

Terzo esperimento

Per capire se la scarsa qualità dei risultati è legata a problemi tecnici nella purificazione dei DNA, quindi alla qualità del materiale di partenza, la sessione sperimentale successiva è stata eseguita su campioni provenienti dal laboratorio dell'IRGB-CNR di Sassari, senza modificare le condizioni di preparazione delle librerie. L'efficienza della preparazione delle librerie è risultata essere più che sufficiente al fine di produrre dei pool con 250ng per ciascun campione. Sono stati prodotti 3 diversi pool per la successiva cattura dei target: il pool A conteneva cinque gDNA, il pool B con undici cfDNA, il pool C con le cinque librerie di cfDNA che, grazie alla maggiore quantità ottenuta, permettevano una replica.

Al termine della preparazione i pool sono stati dosati con TECAN-Qubit e controllati con Agilent Bioanalyzer. Il pool A di gDNA aveva una concentrazione di 3,5ng/ul e una grandezza media di 355bp (15,2nM); il pool B risultava il più concentrato con 5,4ng/ul e all'Agilent presenta i soliti 3 picchi caratteristici della frammentazione nucleosomica (320bp come picco principale, 500bp e 700bp secondari con rapporti della molarità di 1:10 e 1:20 circa); il pool C si presenta con gli stessi picchi ma, come atteso, è meno concentrato (2ng/ul o 9,6nM)

I risultati hanno dato indicazioni riguardo alla migliore strategia di multiplexing, in quanto i campioni di cfDNA che fanno parte del pool B hanno dato una maggiore quantità di duplicati (85,8%, range 74,6%-91,2%) rispetto alle stesse librerie raccolte con minore livello di plexing nel pool C (77%, range 66,2%-85,4%).

Come atteso dalla modifica del numero di cicli di PCR, i duplicati per le librerie di gDNA sono diminuiti a 8,7%. Le percentuali di off-target e contenuto di GC sono risultati uniformi tra i diversi pool e con valori ottimali, rispettivamente 11% e 54%.

Quarto esperimento

In questa sessione, che ha coinvolto 12 campioni di gDNA e 24 campioni di cfDNA, è stato ulteriormente diminuito il numero di cicli di PCR, sia per l'amplificazione dei prodotti di ligazione (9



**SARDEGNA
RICERCHE**

cicli di PCR), sia per la PCR successiva all'arricchimento (9 cicli). La media delle percentuali dei duplicati di PCR per il cfDNA risulta ancora elevata (84%), mentre per il gDNA la media è del 10%.

Tabella 1. Tabella riassuntiva dei valori ottenuti dagli esperimenti descritti

	Out of target	% duplicati	% GC	Coverage pre-dedup
Esp 1_cfDNA	8,5 - 24,3	70,6-91,2	56-60	5000-7000X
Esp 1_tDNA*	8,8 - 41,3	34,3-87,1*	56-64	300-800X
Esp 1_gDNA	10,3 - 24,6	27,5-33,1	61-64	100-200X
Esp 2_cfDNA	12,8 - 15	77,3-90,8	50-52	4500-6500X
Esp 2_tDNA°	11,6 - 14,6	43,8-71°	51-55	200°-1800X
Esp 2_gDNA	10,6 - 12,3	15,2-16,6	53-55	100-200X
Esp 3_cfDNA	10,9 - 13,1	74,6-91,1	53-56	5000-6500X
Esp 3_gDNA	9,3 - 9,7	8,1-9,3	56	100-200X
Esp 3bis_cfDNA	10,5 - 12,5	66,2-85,3	53-55	4500-5500X
Esp 4_cfDNA	11,2 - 12,9	63 - 91,4	51 - 55	5500-7000X
Esp 4_gDNA	9,4-11	9,3-12,2	54 - 58	100-200X

* comprende campioni outlier

° outlier per campioni da cattura con pool misto

Esp1: 12 + 10 cicli PCR; ligazione 20°C x 4h; ibridazione 65°C; pool di cattura misti

Esp2: 10 + 10 cicli PCR; ligazione 4°C O/N; ibridazione 62°C; pool per tipo DNA + pool misto

Esp3: 10 + 10 cicli PCR; ligazione 20°C 4h; ibridazione 62°C; pool per tipo DNA (11cf in pool)

Esp3bis: stessi campioni di esp3 (5cf in cattura)

Esp4: 9 + 9 cicli PCR; ligazione 20°C 4h; ibridazione 62°C; pool per tipo DNA (11cf in pool)

Conclusioni

Il confronto tra i risultati ottenuti dai diversi esperimenti svolti, rispetto anche ai dati estrapolati dalla letteratura, suggeriscono che il protocollo può essere ulteriormente ottimizzato.

In particolare:

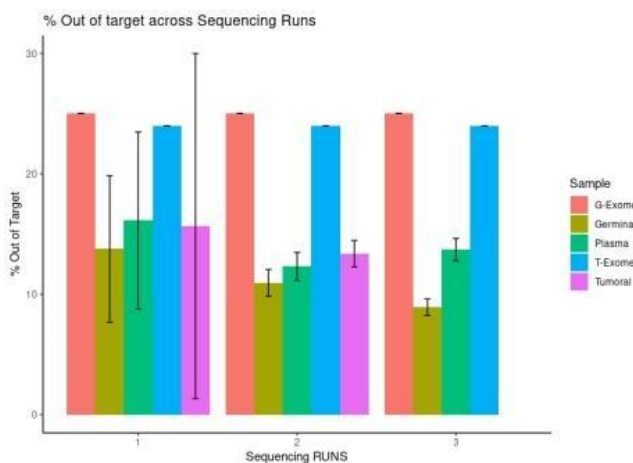
- 1- Per quanto riguarda l'efficienza e la specificità della cattura, la percentuale di sequenze mappate sul target rispetto alle sequenze off-target è molto buona. Come parametro di riferimento abbiamo preso la cattura di un esoma (Agilent v7), che consiste in un target di circa 45Mbp, e che dà generalmente una percentuale di off-target del 25%. Il pannello NIASMIC (sonde Agilent), con un target di 2,06Mb, presenta una media del 12% di sequenze off-target alla condizione identificata come più performante (Figura 3).



**SARDEGNA
RICERCHE**

- 2- La stessa modifica della temperatura di ibridazione ha portato a uniformare il contenuto di GC rispetto ad AT, raggiungendo valori ottimali e paragonabili a quelli che si ottengono da un kit commerciale per l'esoma (Figura 4); le piccole differenze potrebbero anche essere attribuite all'effettivo contenuto in GC dei geni che costituiscono il pannello NIASMIC rispetto a un intero esoma.
- 3- La percentuale di duplicati di PCR rimane un fattore critico per il quale gli esperimenti svolti hanno dato indicazioni sulle possibili correzioni da apportare al protocollo. Su questo tema la diminuzione del numero di cicli di PCR, dai complessivi 24 del primo esperimento ai 18 cicli dell'ultimo, non ha dato i risultati attesi; mentre è apprezzabile una diminuzione dei duplicati per le librerie di gDNA e di tDNA, per il cfDNA sembrano essere addirittura aumentati.

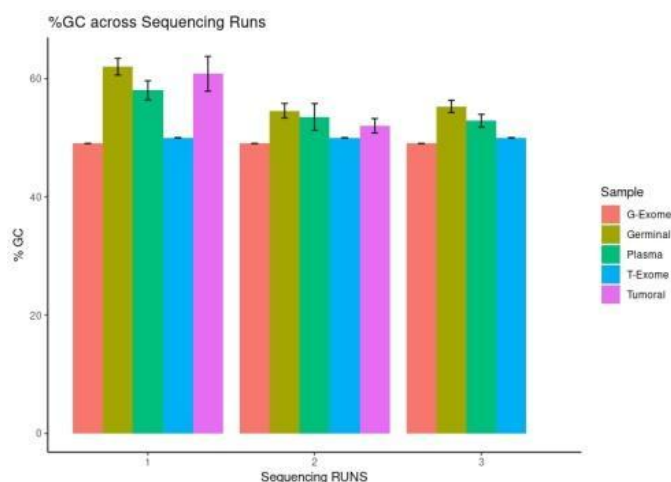
Figura 3. Confronto out of target





**SARDEGNA
RICERCHE**

Figura 4. Confronto GC content



Nel proseguire la sperimentazione verranno valutati pool di librerie con livelli di plexing scalari, anche a costo di dover riportare il numero di cicli di PCR a valori più alti, così come suggerito dallo studio del protocollo di un kit commerciale della Illumina, il TruSight Oncology 500 ctDNA, recentemente messo sul mercato. Nel kit Illumina la cattura viene fatta a singoli campioni e vengono eseguiti complessivamente 33 cicli di PCR, 15 per la prima amplificazione e 18 per quella post cattura. Purtroppo non sono riportati dati quantitativi, sia per quanto riguarda la quantità di ng di libreria necessari per l'ibridazione, in quanto viene messo in cattura tutto il prodotto di amplificazione a prescindere dalle concentrazioni, sia per quanto riguarda la percentuale di duplicati di PCR che vengono rimossi.

Un altro protocollo che può essere preso come riferimento è quello del kit AVENIA della Roche. Sul sito dell'azienda è consultabile un *white paper* nel quale vengono descritti i risultati che si ottengono dalla cattura di tre diversi pannelli per il ctDNA. Questi pannelli hanno target più ristretti rispetto al NIASMIC: 81Kbp pannello Target, 192Kbp pannello Expanded, 198Kbp pannello Surveillance e i valori percentuali dei duplicati di PCR sono rispettivamente 83%, 54% e 54%. Purtroppo non è stato possibile consultare il manuale del kit per poter estrapolare le condizioni e le strategie di analisi utilizzate.

I protocolli descritti in letteratura raramente riportano le quantità di duplicati rimossi ma si limitano a elencare i tool bioinformatici utilizzati. Un paper che può essere utile perché descrive i risultati ottenuti valutando la modifica di alcuni parametri, in particolare per quanto riguarda la quantità di DNA di partenza e il livello di multiplexing [4], evidenzia come la quantità di duplicati aumenta al diminuire della quantità del DNA input per la generazione delle librerie ed è minore per catture con un basso livello di multiplexing pre-cattura.



**SARDEGNA
RICERCHE**

Un altro articolo [5] descrive la messa a punto di un protocollo simile al NIASMIC nel quale il target è rappresentato da 68 geni per un totale di 0,5Mb. In questo lavoro vengono riportati i valori di duplicati ottenuti con una variabilità molto ampia (18%-55%) ed una media di 35,1%. Di conseguenza il coverage grezzo, a monte della rimozione dei duplicati, risulta essere di 16700X e permette di ottenere un coverage medio pulito di 462X.

Bibliografia

1. **Laure Sorber et al.** A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma. J Mol Diagn, 2017 Jan; 19(1):162-168.
2. **Louise Aigrain et al.** Quantitation of next generation sequencing library preparation protocol efficiencies using droplet digital PCR assays - a systematic comparison of DNA library preparation kits for Illumina sequencing. BMC Genomics volume 17, Article number: 458 (2016)
3. **Erwin L van Dijk et al.** Library preparation methods for next-generation sequencing: tone down the bias. Exp Cell Res. 2014 Mar 10;322(1):12-20
4. **Chung J et al.** The minimal amount of starting DNA for Agilent's hybrid capture-based targeted massively parallel sequencing. Scientific Reports 2016; 6: 26732
5. **Cimmino F et al.** A Targeted Gene Panel for Circulating Tumor DNA Sequencing in Neuroblastoma. Front. Oncol. 2020; 10: 596191