

## Strategie di sequenziamento: l'esempio del tumore testicolare in un trio familiare

Marcella Nunziato<sup>1,2</sup>

<sup>1</sup>Ceinge-Biotecnologie Avanzate, Napoli, Italia

<sup>2</sup>Dipartimento di Medicina Molecolare e Biotecnologie Mediche, Università degli Studi di Napoli Federico II, Napoli, Italia

*Questo lavoro è stato in parte presentato al 53° Congresso SIBioC 11-13 Ottobre 2021, Virtual Edition, sotto forma di Poster, ricevendo il premio SIBioC destinato ai 4 migliori poster presentati.*

### ABSTRACT

#### Genome sequencing strategies: the example of testicular cancer in a familial trio.

Testicular cancer is the most common malignancy in men under the age of 50. In 1972 for the first time, the carcinoma *in situ* was mentioned and in particular the tumor of mixed germ cells *in situ*; this type of tumor, in fact, originates in the testicles but can also be found in other parts of the body. The diagnosis of these tumors involves the use of instrumental tests, usually carried out together with some laboratory tests including the determination of tumor markers; these, when increased, can be of help in indicating the presence of the disease. In recent years, there has been a sharp increase in the number of testicular cancer cases diagnosed especially in young males all over the world. Although the understanding of the molecular causes underlying this neoplasm is very limited, Next Generation Sequencing (NGS) strategies can help to shed light on complex and multifactorial pathologies such as in the case of testicular cancer. In addition, a new nucleic acid sequencing strategy is gaining ground: the third generation Oxford Nanopore Technology (ONT), which, allows to obtain a "picture" of the DNA in each individual as complete as possible, by providing very long genome sequences. The aim of this paper is to illustrate how the use of combined nucleic acid sequencing strategies has allowed to analyze the presence of mutations in a family trio where the boy (proband) was affected by testicular neoplasia. The combined sequencing of the genome and the methyloma can be of help in understanding the possible causative or predisposing mutations.

**Parole chiave:** tumore testicolare, sequenziamento di II generazione, sequenziamento di III generazione

### IL TUMORE TESTICOLARE

Il tumore testicolare è la neoplasia più frequente nel sesso maschile sotto i 50 anni d'età. Nel 1972 si è parlato per la prima volta di carcinoma testicolare *in situ* ed in particolare di tumore delle cellule germinali miste *in situ* (germ cell neoplasia *in situ*, GCNIS). Questo tipo di tumore infatti origina nei testicoli ma può trovarsi anche in altre parti del corpo (1). Si stima che circa il 2-3% dei tumori testicolari sia di natura ereditaria e talvolta associato a sindromi come quella di Peutz-Jeghers causata da mutazioni a livello germinale del gene *STK11*. Inoltre è stato notato che maschi con neoplasia testicolare hanno un parente stretto con la stessa condizione oncologica; tra fratelli infatti il rischio è aumentato di circa 8-10 volte e tra padre-figlio fino a 4-6 volte. Un rischio ancora più elevato

è stato osservato tra gemelli monozigoti, suggerendo quindi che cause genetiche svolgano un ruolo rilevante nell'insorgenza di questi tumori (2). Tradizionalmente, i tumori a cellule germinali (GCT) sono classificati in base alle caratteristiche istologiche e vengono suddivisi in due grandi gruppi dall'Organizzazione Mondiale della Sanità (OMS): neoplasie germinali, ossia derivate da cellule della linea germinale, che rappresentano circa il 95% di tutte le neoplasie testicolari, e neoplasie non germinali che invece sono rarissime. Ognuno di questi gruppi presenta al suo interno diversi sottotipi. Il sottotipo istologico più frequente è rappresentato dal seminoma (40-50% dei casi), che è caratterizzato anche da una prognosi migliore. Tra le forme germinali non seminomatose, si annoverano il carcinoma embrionale, che rappresenta il 15-20% dei casi, il teratoma con il 5-10% dei casi (spesso

Corrispondenza a: Marcella Nunziato, Ceinge-Biotecnologie Avanzate, Via Gaetano Salvatore, 486, 80131 Napoli, Email: nunziato@ceinge.unina.it

Ricevuto: 01.06.2022

Revisionato: 16.06.2022

Accettato: 05.07.2022

Publicato on-line: 03.08.2022

DOI: 10.19186/BC\_2022.051

ben differenziato, ma potenzialmente maligno), il tumore del sacco vitellino che è più frequente in età infantile e infine il corion-carcinoma (<1% in forma pura, con spiccata tendenza all'invasione di altri organi e tessuti) (3). Il fattore di rischio più conosciuto è il criptorchidismo, un difetto già presente alla nascita in cui uno o entrambi i testicoli non sono nello scroto e talvolta restano presenti in addome; questo sembra aumentare il rischio di sviluppare un tumore testicolare di circa 5 volte rispetto ai soggetti sani. Inoltre anche alcuni fattori ambientali sembrano avere un ruolo importante nella patogenesi del tumore testicolare ed in particolare lo stile di vita e l'esposizione ad alcuni agenti chimici che potrebbero alterare il normale funzionamento del sistema endocrino agendo quindi come "innesco" per lo sviluppo della neoplasia (4,5). La diagnosi di questi tumori prevede alcuni esami strumentali, quali ecografia e TAC/PET, per determinare l'eventuale stadiazione, unitamente ad alcuni esami di laboratorio, tra cui la determinazione di alcuni marcatori tumorali, che, se aumentati, possono contribuire ad indicare la presenza di malattia neoplastica. In particolare, si valuta l'alfa-fetoproteina (AFP), una proteina che normalmente è sintetizzata durante la vita fetale dal sacco vitellino e dal fegato del feto, e che, elevata alla nascita subisce una riduzione drastica nei periodi di vita successivi. E' prodotta nei tumori germinali non-seminomatosi ma non nei seminomi puri; pertanto l'aumento dei livelli di AFP potrebbe consentire di effettuare una diagnosi differenziale tra seminoma e non seminoma. Di interesse inoltre la gonadotropina corionica umana (hCG), una glicoproteina con attività ormonale, associata alla gravidanza, con la funzione di favorire un ambiente ormonale e tissutale adeguato allo sviluppo dell'embrione. La proteina è presente in concentrazione elevata sia nei tumori germinali di tipo seminoma che non-seminoma (6). Tra i tumori stromali, che includono neoplasia delle cellule di Leydig e di Sertoli, non si riscontra produzione di AFP o hCG. Un altro marcatore tumorale usato per effettuare diagnosi di tumore testicolare è la lattato deidrogenasi (LDH) che si riscontra fortemente incrementata in caso di malattia oncologica. L'enzima LDH è presente in quasi tutti i tessuti, e viene rilasciata in caso di danno; non è dunque specifica per il tumore testicolare e può aumentare nel circolo ematico in seguito a molteplici altre patologie o condizioni cliniche.

La sopravvivenza a 5 anni per il tumore testicolare si attesta intorno al 91%, percentuale che aumenta se il tumore viene identificato nelle primissime fasi.

### **DIVERSE STRATEGIE DI ANALISI: DAL SEQUenziAMENTO TARGET AL SEQUenziAMENTO DELL'INTERO GENOMA**

Negli ultimi anni è stato osservato un forte incremento del tumore del testicolo in tutto il mondo, soprattutto nei giovani con età compresa tra i 18 e i 44 anni. Nell'anno 2020 sono stati riportati 74 458 nuovi casi e 9 334 morti. L'incidenza maggiore è in Europa (che conta il 33,7%), seguita dall'Asia (con il 27,7%). La mortalità più alta

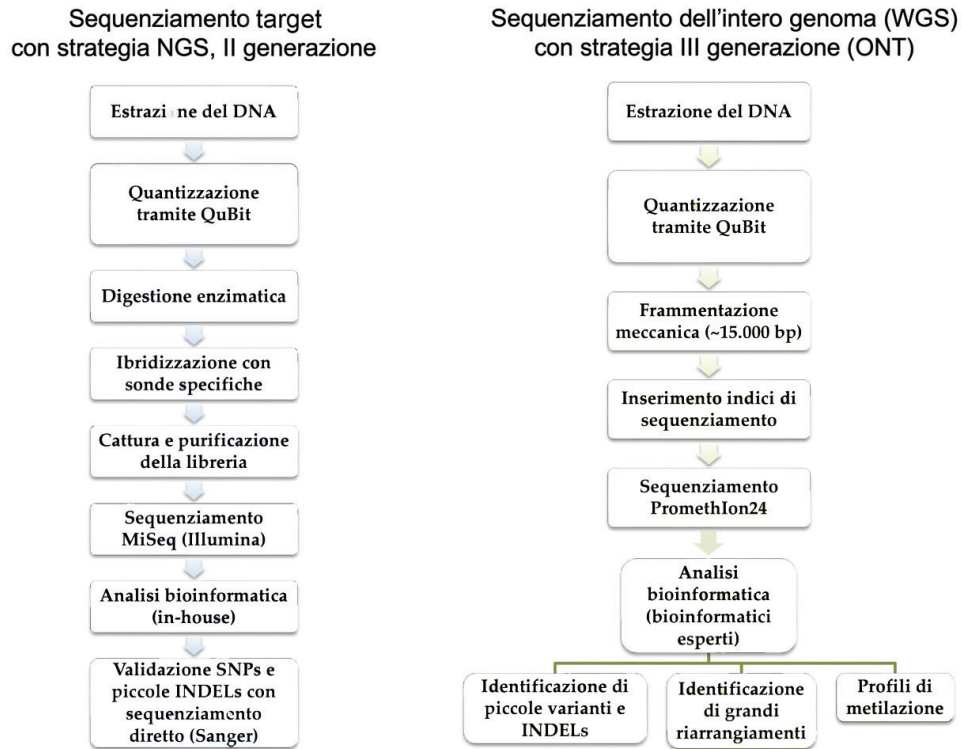
invece riguarda l'Asia (42,8%), seguita dall'America Latina e Caraibi, mentre l'Europa segue con una più bassa percentuale che si attesta al 16,8% (7).

Nel contesto di una patologia particolarmente complessa, multi-genica e multi-fattoriale, l'analisi di interi trio (composti da madre, padre e figlio) potrebbe aiutare e facilitare la comprensione della malattia e la corrispondenza tra genotipo e fenotipo.

Le strategie di Next Generation Sequencing (NGS) applicate allo studio di pannelli multi-genici, di interi esomi (WES) o interi genomi (WGS) aiutano e soprattutto rendono possibili queste tipologie di analisi grazie anche a costi sempre più ridotti e alla possibilità di avere risultati in tempi abbastanza rapidi (8).

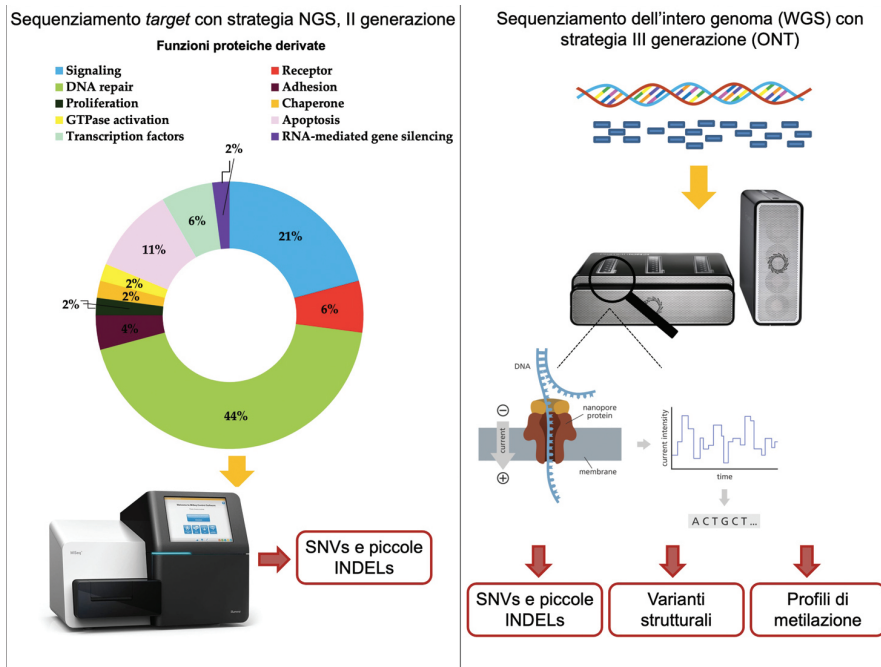
Le strategie più usate sono quelle di sequenziamento "target", con sonde specifiche per pannelli di geni selezionati. In secondo luogo si può utilizzare il sequenziamento dell'esoma, in cui si analizza l'esoma clinico, ovvero tutti i cosiddetti geni-malattia, o di tutta la regione codificante, fino al sequenziamento dell'intero genoma (Figure 1 e 2). La scelta della migliore strategia di sequenziamento da usare dipende da diversi fattori. L'approccio "target" infatti consente di ottenere un sequenziamento veloce di specifiche regioni di interesse, già note per la presenza di varianti causative. Con questa strategia è possibile sequenziare più pazienti insieme permettendo anche l'abbattimento dei costi; questa modalità può inoltre rappresentare una buona soluzione per una analisi di primo livello grazie anche ai tempi di analisi ridotti. I pannelli target sono inoltre facilmente personalizzabili in base ad esigenze specifiche sia in termini di malattie da indagare ai fini di diagnostica differenziale che in termini di regioni da analizzare. Il sequenziamento dell'esoma clinico consente invece di analizzare tutte le regioni codificanti dei cosiddetti geni-malattia; questa è una strategia che può essere considerata quando si ha il sospetto di una patologia genetica ma i segni e/o i sintomi sono sovrapponibili ad altre condizioni che non permettono al clinico di effettuare una diagnosi certa. In questo caso i costi sono più alti, la mole di dati ottenuti è maggiore con tempi di analisi più lunghi ed è possibile sequenziare un minor numero di campioni contemporaneamente. Infine, il sequenziamento dell'intero genoma permette di ottenere certamente un quadro pressoché completo, ma la gestione bioinformatica dei dati con l'analisi delle varianti putative a maggior rischio patogenetico e i tempi di analisi restano ancora di difficile gestione nei laboratori di Biologia Molecolare Clinica.

Gli approcci di sequenziamento sono molto cambiati negli ultimi anni: i sequenziatori che utilizzano la chimica di "sequencing-by-synthesis" (piattaforme Illumina, NGS) hanno di ottime caratteristiche e sono facili da usare nella routine diagnostica. Sicuramente, la riduzione dei tempi di analisi e dei costi ha permesso un forte sviluppo di queste tecnologie che oggi sono essenziali per l'analisi simultanea di diversi geni, implicati non solo in patologie di natura oncologica. D'altro canto il costante aumento della richiesta e l'avanzare della ricerca volta ad effettuare analisi sempre più complesse, quali il sequenziamento dell'intero genoma, ha evidenziato le problematiche



**Figura 1**

Nella figura è mostrato il flusso di lavoro per la preparazione di una libreria con pannello multi-genico e sequenziamento con strumenti Illumina (MiSeq) – NGS, II generazione e il flusso di lavoro per il sequenziamento dell'intero genoma (in questo caso il sequenziamento è effettuato con PromethION24 di Oxford Nanopore (ONT, III generazione). NGS, Next Generation Sequencing; ONT, Oxford Nanopore Technologies. SNVs, polimorfismo a singolo nucleotide, INDEL, inserzioni/delezioni;



**Figura 2**

Rappresentazione di sintesi dei due tipi di sequenziamento (II e III generazione) e le informazioni ottenibili dalle due strategie (sequenziamento target con strumenti di NGS, II generazione vs. sequenziamento dell'intero genoma con strategia ONT, III generazione). NGS, Next Generation Sequencing; SNVs, Variazioni di un singolo nucleotide; INDELS, inserzioni/delezioni; WGS, Whole Genome Sequencing; ONT, Oxford Nanopore Technologies.

legate alle cosiddette “short reads”, aprendo altresì la strada ad approcci che invece sfruttano le “long reads” come PacBio e Oxford Nanopore Technologies (ONT) (sequenziamento di III generazione) (9,10). Le “short reads” infatti non permettono di sequenziare lunghi tratti continui di DNA; i filamenti devono essere prima frammentati e amplificati e dopo il sequenziamento molto spesso si ricorre all'uso di programmi bioinformatici per l'analisi delle sequenze ottenute. I passaggi di amplificazione, però, possono introdurre numerosi errori. Inoltre, questo tipo di sequenziamento potrebbe non riuscire a generare una sovrapposizione sufficiente tra i frammenti di DNA. Dunque, il sequenziamento di un genoma altamente complesso e con regioni altamente ripetitive, come quello umano, può essere molto difficile.

Il sequenziamento di pannelli di circa 50 geni identifica solitamente dalle 450 alle 750 varianti per campione; il numero di dati generati è facilmente gestibile in laboratorio tramite l'utilizzo di piattaforme *ad hoc* come per esempio il software Alissa (Agilent Technologies) che combina due differenti strumenti: il primo che permette di assemblare e chiamare le varianti mentre l'altro risulta essere utilissimo per l'interpretazione clinica delle stesse. Il sequenziamento degli esomi produce dalle 70 000 alle 90 000 varianti per campione, i dati ottenuti possono essere coordinati dal laboratorio, sempre utilizzando software come il precedente, purché ci si avvalga di operatori competenti in bioinformatica. Gli approcci di sequenziamento che coinvolgono l'intero genoma necessitano al contrario di un gruppo di bioinformatici esperti in grado di gestire e analizzare l'enorme mole di dati ottenuti che non possono essere analizzati con i software di cui sopra.

## LA BIOLOGIA MOLECOLARE NEL TUMORE TESTICOLARE

L'associazione di geni al tumore testicolare è un dato ancora molto limitato in letteratura. Nel nostro laboratorio, abbiamo dapprima effettuato un sequenziamento target di un pannello di 48 geni che codificano per proteine coinvolte nei processi di riparo del danno del DNA (44%). Il pannello è stato personalizzato nel nostro laboratorio: comprende 50 paia di basi introniche a monte e a valle di ciascun esone di ogni gene, nonché le regioni non tradotte (3' e 5' UTRs, promotore e post-COOH non tradotte). Questa scelta lo differenzia dai pannelli in commercio che solitamente indagano molte meno basi nelle giunzioni introne-esone e non contengono le UTRs. Il pannello così disegnato, per ora utilizzato al solo scopo di ricerca, è composto da alcuni geni e *locus* genici già noti in letteratura per l'associazione con alcune delle più frequenti neoplasie. Successivamente, abbiamo altresì optato per l'analisi dell'intero genoma in tutto il trio familiare in questione. Pertanto, data la grande quantità di dati generati, l'analisi, è stata iniziata con l'uso dei cosiddetti “pannelli virtuali” andando cioè a verificare la presenza di varianti patogeniche o possibilmente patogeniche in poco più di altri 100 geni correlati ai tumori testicolari e già noti in letteratura (circa 23 000-25 000 varianti identificate).

Come accennato in precedenza, una nuova strategia si sta facendo strada: il sequenziamento basato su nanopori (ONT); questa rappresenta un approccio completamente diverso, in cui la sequenza degli acidi nucleici è identificata grazie ai cambiamenti nella corrente ionica, quando una singola molecola di DNA passa attraverso un nanoporo proteico. Fino a qualche anno fa il rendimento, relativamente basso e caratterizzato dall'inserimento di molti errori di sequenza (circa il 20% di errori), ha limitato il suo utilizzo. Tuttavia, i recenti miglioramenti, che ne aumentano l'accuratezza con cui è possibile determinare la sequenza del DNA che passa attraverso il nanoporo proteico e la velocità con cui il DNA può attraversarlo per poter effettuare più corse con la copertura desiderata, hanno notevolmente aumentato la capacità della strumentazione, rendendo questa strategia pronta per effettuare sequenziamenti più complessi come quello del genoma umano (10). Negli ultimi anni, la ricerca sull'eterogeneità genomica e trascrittomico a livello della singola cellula hanno acquisito molta importanza soprattutto in oncologia ed immunologia. Gli approcci di II generazione non permettono pienamente di identificare l'abbondanza dei trascritti a livello delle varie isoforme, la III generazione risolve questo problema andando a sequenziare il trascritto interamente, in singole sequenze.

La strategia Nanopore inoltre permette di ottenere diversi tipi di informazioni con un solo esperimento, attraverso l'identificazione di varianti che comportano la sostituzione di un solo nucleotide (SNVs), l'identificazione di varianti strutturali (SVs) tra cui grossi riarrangiamenti genici (LGRs) e variazioni del numero di copie di sequenze geniche (CNVs), e infine l'identificazione dei profili di metilazione (10,11). Il duplice approccio, dapprima il sequenziamento del pannello multigenico con strategia di II generazione e poi quello dell'intero genoma con strategia di III generazione, ha avuto, nella nostra esperienza, i seguenti vantaggi:

- indagare la presenza di ulteriori varianti di interesse in altri geni non indagati con il pannello;
- analizzare i profili di metilazione (non possibile con il sequenziamento target) identificando, al momento, 260 finestre differenzialmente metilate nel trio;
- validare un approccio così complesso quale il sequenziamento dell'intero genoma andando a verificare la presenza delle varianti già identificate con il pannello.

Negli ultimi anni, numerosissimi studi hanno cercato di comprendere meglio il ruolo delle variazioni epigenetiche nelle malattie umane, ma ci sono ancora tante domande che non hanno ancora trovato risposta. La metilazione è un processo mediante il quale i gruppi metilici vengono aggiunti alle molecole di DNA: si tratta di una modifica epigenetica fondamentale nella regolazione della trascrizione genica, con la quasi totalità delle modificazioni del DNA eucariotico che si verificano essenzialmente a livello della citosina (la 5-metil-citosina). È stata definita da Kulis et al. (12) *“un cambiamento soft e potenzialmente reversibile del genoma che può definire o adattarsi alla biologia del tumore ed è funzionalmente equivalente a cambiamenti genetici come le mutazioni o le delezioni”*.

È noto che l'ipometilazione promuove l'instabilità genomica, causando una erronea segregazione cromosomica durante le divisioni cellulari e l'attivazione indesiderata di elementi trasponibili all'interno del genoma stesso, facilitando l'instaurarsi di ulteriori danni genetici. L'ipermetilazione, d'altro canto, può guidare il silenziamento dei principali geni onco-soppressori o delle regioni regolatrici all'interno del genoma portando a una dis-regolazione della crescita cellulare o ad una risposta alterata alle terapie contro il cancro. L'insieme di tutti questi meccanismi epigenetici può, quindi, contribuire allo sviluppo e/o alla progressione tumorale (13).

In questo contesto, l'uso di strategie combinate di sequenziamento di acidi nucleici ha permesso, ad esempio, di analizzare la presenza di mutazioni in un trio familiare in cui il figlio (probando) era affetto da una neoplasia testicolare. Abbiamo inoltre ottenuto il sequenziamento del genoma quasi completo nonchè del metiloma dando anche interpretazioni per comprendere le possibili alterazioni geniche causative o predisponenti.

#### CONFLITTO DI INTERESSI

Nessuno

#### BIBLIOGRAFIA

1. King J, Adra N, Einhorn, L. H. Testicular cancer: biology to bedside. *Cancer Res* 2021;81:5369-76.
2. Kratz C P, Mai PL, Greene M. H. Familial testicular germ cell tumours. *Best Pract Res Clin Endocrinol Metab* 2010;24:503-13.
3. Al-Obaidy KI, Idrees M. T. Testicular tumors: a contemporary update on morphologic, immunohistochemical and molecular features. *Adv Anat Pathol* 2021;28:258-75.
4. Leão R, Ahmad A E, Hamilton R J. Testicular cancer biomarkers: a role for precision medicine in testicular cancer. *Clin Genitourin Cancer* 2019;17:e176-e83.
5. Szucs M, Riesz P, Mavrogenis S, et al. Testicular cancer - diagnosis and treatment. *Orv Hetil* 2008;149:894-6.
6. Damjanov I. Testicular germ cell tumors: serological and immunohistochemical diagnosis. *Acta Med Acad* 2021;50:58-70.
7. De la Rosa AH., Manoharan M, Goolam A S. Current concepts of epigenetics in testicular cancer. *Indian J Surg Oncol* 2017;8:169-74.
8. McInerney-Leo A M, Duncan E L. Massively parallel sequencing for rare genetic disorders: potential and pitfalls. *Front Endocrinol (Lausanne)* 2021;11:1-10.
9. Da Costa E Silva Carvalho S, Cury N M, Brotto D B, et al. Germline variants in DNA repair genes associated with hereditary breast and ovarian cancer syndrome: Analysis of a 21 gene panel in the Brazilian population. *BMC Med Genomics* 2020;13:21.
10. Petersen LM, Martin I W, Moschetti W, et al. Third-generation sequencing in the clinical laboratory: sequencing. *J Clin Microbiol* 2019;58:1-10.
11. Oikonomopoulos S, Bayega A, Fahiminiya S, et al. Methodologies for transcript profiling using long-read technologies. *Front Genet* 2020;11:1-20.
12. Kulis M, Esteller M. DNA Methylation and Cancer. *Adv Genet* 2010;70:27-56.
13. Klutstein M, Nejman D, Greenfield R, et al. DNA methylation in cancer and aging. *Cancer Res* 2016;76:3446-50.