



OncoPan® Germinale

Test genetico NGS sviluppato dal CGT Laboratory di Cogentech

Introduzione

OncoPan® è il test molecolare innovativo proposto da Cogentech, basato su un processo tecnologico avanzato, denominato Next Generation Sequencing (**NGS**), che consente di rilevare la presenza di variazioni a livello del genoma del paziente. Queste variazioni includono le mutazioni puntiformi (sostituzioni, piccole delezioni/inserzioni o SNV) ed i grandi riarrangiamenti (CNV) dovuti alla delezione/duplicazione di uno o più esoni dei geni in esame. Il test viene offerto **livello germinale** e riguarda i geni coinvolti nella **predisposizione** ai tumori della mammella e dell'ovaio, del colon, dell'endometrio, dello stomaco, della prostata, del pancreas, ai melanomi e a sindromi tumorali note.

Il pannello include anche 313 SNPs per il calcolo del Polygenic Risk Score (PRS), utili a migliorare la definizione del rischio individuale di tumore della mammella e dell'ovaio dei singoli pazienti analizzati. Il valore dello "z score" ottenuto, può essere utilizzato dai clinici nel calcolo sviluppato dal software CanRisk (www.canrisk.org), secondo il modello integrato di predizione del rischio **BOADICEA**, che considera i fattori di rischio personali, anamnesi oncologica familiare, test genetici per la presenza di geni che conferiscono rischio moderato o alto, punteggi di rischio poligenico (PRS), densità mammografica.

I geni compresi nella versione attuale del pannello sono elencati qui di seguito:

ACD, APC, ATM, AXIN2, BAP1, BARD1, BMPR1A, BRCA1, BRCA2, BRIP1, CDH1, CDKN2A (α e β), CDK4 (esone 2), CHEK2, CTNNA1, DICER1, EPCAM, FANCM, GREM1, HOXB13, MBD4, MC1R, MITF (esone 9), MLH1, MSH2, MSH3, MSH6, MUTYH, NBN, NF1, NTHL1, PALB2, PMS2, POLD1, POLE, POT1, PTEN, RAD51C, RAD51D, RNF43, SMAD4, STK11, TERF2IP, TERT, TP53.

Come viene effettuato il test

Il test viene eseguito su DNA genomico estratto da un campione ematico del paziente (**campione germinale**). Il materiale isolato viene arricchito per i geni di interesse mediante la tecnica definita 'a cattura', in cui si usano sonde disegnate tramite un criterio che permette di selezionare specificatamente solo regioni ben definite. Il processo può essere eseguito in manuale (v5.1.0) o in automazione (v5.1.1). Nel dettaglio, le sonde sono state create sfruttando la tecnologia della ditta *Twist Bioscience*, includendo le zone nucleotidiche degli esoni dei geni prescelti ed almeno 21 basi delle aree introniche adiacenti. L'insieme dei frammenti genomici 'catturati' dalle sonde rappresenta le regioni arricchite dei geni d'interesse (**libreria**), che vengono successivamente sequenziate tramite un processo che impiega tecniche NGS: i sequenziatori includono la serie DX di Illumina (Miseq, Nextseq550 o Novaseq 6000).

Le varianti riscontrate vengono confermate con sequenziamento diretto Sanger oppure con Multiplex Ligation-dependent Probe Amplification (MLPA), su seconda aliquota di sangue, qualora si trattasse di VUS (varianti a significato incerto) o di varianti patogeniche. Nel caso di varianti SNV in *PMS2* o *CHEK2* è previsto l'utilizzo di Long Range PCR per selezionare il gene rispetto agli pseudogeni. La piattaforma ABI3500 Dx Genetic Analyzer (Applied Biosystems) viene utilizzata sia per l'identificazione delle sequenze con metodo di Sanger che per i frammenti di MLPA. Le elaborazioni informatiche di questi dati vengono effettuate rispettivamente con software è Mutation Surveyor e Gene Marker, entrambi venduti dalla ditta SoftGenetics.



L'analisi bioinformatica

Alla fine della fase di sequenziamento NGS, i dati vengono sottoposti a una serie di **analisi bioinformatiche** avanzate che includono:

- l'**analisi primaria**, ovvero la generazione delle sequenze (*reads*) e la valutazione della loro qualità;
- l'**analisi secondaria**, che consiste nell'appaiamento dei dati ottenuti con le corrispondenti regioni del genoma scelto come riferimento (GRCh37-hg19): si definisce la presenza di eventuali varianti SNV o di CNV;
- l'**analisi terziaria**, che permette l'interpretazione delle varianti, a cui segue eventualmente l'emissione di un referto di diagnosi molecolare.

Mentre l'analisi primaria viene effettuata dal sequenziatore NGS, le altre sono eseguite da una procedura automatizzata (*pipeline*) sviluppata in collaborazione con una ditta esperta del settore (enGenome). La *pipeline* è stata creata in modo da:

- evidenziare eventuali regioni di interesse con bassa copertura di sequenziamento, ovvero che hanno una profondità di lettura (numero di sequenze) minore di 30X. Queste zone verranno rianalizzate tramite il sequenziamento col metodo di Sanger, per garantire sempre una copertura della regione di interesse del 100% nell'analisi;
- generare risultati relativi ai soli geni richiesti.

Verifica del pannello

Il pannello OncoPan® è stato aggiornato negli anni come numero di regioni analizzate, ma si è mantenuta costante la scelta della metodologia per la creazione della *libreria* e il tipo di strumentazioni usate.

Per verificare la robustezza del pannello sono stati confrontati i dati ottenuti con l'attuale pannello con quelli ricavati dallo stesso campione ma con la versione precedente del pannello, validato in origine confrontando dati generati tramite il metodo Sanger, MLPA (Multiplex Ligation-dependent Probe Amplification) o NGS (le cui librerie erano state create con un approccio differente). Sono stati assegnati i valori di *specificità* e *valore predittivo positivo (PPV)*, indagando alcuni geni presenti nel pannello, relativamente agli esoni codificanti.

I parametri analitici sono definiti come:

Sensibilità (SE): capacità di un test diagnostico di identificare uno specifico evento (tasso di veri positivi)

$$SE = \frac{\text{Veri Positivi (TP)}}{\text{Veri Positivi (TP)} + \text{Falsi Negativi (FN)}} \times 100$$

Valore Predittivo Positivo (PPV): probabilità che un evento rilevato sia effettivamente reale

$$PPV = \frac{\text{Veri Positivi (TP)}}{\text{Veri Positivi (TP)} + \text{Falsi Positivi (FP)}} \times 100$$

dove:

TP (Veri Positivi) = varianti realmente presenti e rilevate dall'esperimento da validare

FP (Falsi Positivi) = varianti rilevate dall'esperimento da validare ma non realmente presenti.

FN (Falsi Negativi) = varianti realmente presenti ma non rilevate dall'esperimento da validare



In **Tabella 1** sono elencati gli esiti di questa analisi in OncoPan®, utilizzando il protocollo su Miseq (in automazione):

Parametro analitico	SNV	CNV
Sensibilità	99.99%	99.99%
Positive predicted value (PPV)	99.99%	96.00%

Limiti del pannello

Quest'analisi non è in grado di evidenziare varianti introniche profonde, riarrangiamenti genomici, espansioni di triplette, mosaicismi della linea germinale o mosaicismi sotto il 15% di frequenza dell'allele meno rappresentato.

Conclusioni

Il pannello OncoPan® si è dimostrato affidabile nei limiti dei valori riportati. Per migliorare sempre più l'accuratezza dell'analisi CNV, saranno aggiunti ulteriori dati per irrobustirne la 'baseline', ovvero l'insieme di controlli necessari al processo per valutare la variazione della profondità di lettura delle regioni indagate, con particolare attenzione a quei geni che nel genoma umano hanno dei corrispondenti pseudogeni (*PMS2* e *CHEK2*). La **profondità di lettura** media ottimale (calcolata rispetto a tutte le regioni di ogni campione) è stata stimata intorno a 200x per i campioni germinali, con un minimo di 30X. Infine ogni regione sequenziata con tecnica NGS viene visualizzata tramite il programma IGV (Integrative Genomics Viewer del Broad Institute), per ridurre ulteriormente la possibilità di avere falsi negativi.