

# Analisi delle CNV della linea germinale con flusso di lavoro Illumina DNA Prep with Exome 2.5 Enrichment

Rilevamento efficace di  
varianti del numero di copie  
utilizzando un pannello dei  
parametri normali



## Introduzione

Le variazioni del numero di copie (CNV, copy number variation) sono una delle principali fonti di diversità genetica negli esseri umani.<sup>1</sup> Questa classe di variazione genomica si verifica quando il numero di copie di un gene o di una regione genomica varia da individuo a individuo e include ampie duplicazioni o delezioni genomiche. L'analisi CNV consente ai ricercatori di identificare le varianti associate a tratti complessi o la predisposizione alle malattie, come il cancro, le malattie autoimmuni, i disturbi genetici ereditari e molte altre.<sup>1-4</sup>

Rispetto all'uso di microarray per il rilevamento delle CNV, i metodi di sequenziamento di nuova generazione (NGS, next-generation sequencing) rivelano maggiori dettagli sulle variazioni strutturali. Inoltre, a differenza degli array, il sequenziamento NGS può rilevare alcune varianti minori di 50 kb. Illumina DNA Prep with Exome 2.5 Enrichment fa parte di una soluzione di sequenziamento dell'intero esoma (WES, whole-exome sequencing) umano ad alte prestazioni, veloce e affidabile che genera librerie compatibili con l'analisi CNV NGS. Insieme ai sistemi di sequenziamento e alle pipeline di analisi di Illumina, questo kit di preparazione delle librerie può essere utilizzato per rilevare varianti di singolo nucleotide (SNV, single-nucleotide variant), piccole inserzioni e delezioni (indel) e grandi CNV strutturali.

Questa nota sull'applicazione illustra un flusso di lavoro ottimizzato per il sequenziamento dell'esoma per l'analisi CNV della linea germinale, dalla preparazione delle librerie fino all'ottenimento delle informazioni (Figura 1). La soluzione WES integra Illumina DNA Prep with Exome 2.5 Enrichment, il collaudato NGS di Illumina, con l'analisi secondaria DRAGEN™ estremamente accurata, per l'identificazione delle CNV.

Gli utenti possono anche applicare la piattaforma di ricerca per l'interpretazione delle varianti Emedgene™, basata sull'intelligenza artificiale spiegabile (XAI, explainable artificial intelligence). Il software Emedgene consente l'analisi e la generazione di informazioni sulle varianti, la valutazione e l'automazione delle procedure operative standard (SOP, standard operating procedure) e include opzioni semplificate per la generazione di report di ricerca.

## Metodi

### Campioni

Per consentire il rilevamento delle CNV della linea germinale sull'esoma, gli utenti devono sviluppare un file di riferimento di base da un pannello di campioni presumibilmente normali per l'esecuzione del confronto durante l'analisi. Questo pannello di riferimento dei parametri normali viene utilizzato per normalizzare i conteggi target e consente un rilevamento accurato delle CNV durante la valutazione dei campioni dei casi. I campioni per il pannello dei parametri normali devono essere preparati e sequenziati nelle stesse condizioni dei campioni del caso.

Per creare il pannello dei parametri normali, sono stati selezionati 54 campioni (Coriell Institute for Medical Research) senza CNV patogene note (Tabella 1) per rappresentare la diversità di variazione naturale tra le superpopolazioni umane (in base al 1000 Genomes Project, Tabella 2).<sup>5</sup> Per dimostrare le prestazioni del flusso di lavoro WES per il rilevamento delle CNV, sono stati utilizzati 17 campioni di casi contenenti una CNV precedentemente rilevata all'interno di un gene o di una regione di interesse in un'analisi con il pannello dei parametri normali.



Figura 1: flusso di lavoro WES completo per l'identificazione di CNV. Combinazione di Illumina DNA Prep with Exome 2.5 Enrichment, piattaforme NGS Illumina, analisi rapida dei dati con analisi secondaria DRAGEN e interpretazione delle varianti con il software Emedgene per la creazione di un flusso di lavoro intuitivo e ottimizzato per un sequenziamento accurato dell'esoma e l'interrogazione delle CNV.

Tabella 1: ID campione del Coriell Institute inclusi nel pannello dei parametri normali

HG00096	HG01393	HG01950	HG03006	NA10830	NA12877	NA18502	NA18970	NA21112
HG00190	HG01441	HG01985	HG03870	NA10831	NA12878	NA18508	NA19681	NA24143
HG00262	HG01551	HG01990	HG03882	NA10835	NA12889	NA18622	NA19720	NA24149
HG00626	HG01599	HG02013	HG03898	NA10838	NA12890	NA18637	NA20509	NA24631
HG00628	HG01896	HG02348	HG04090	NA10839	NA12891	NA18942	NA20875	NA24694
HG01392	HG01914	HG02521	HG04214	NA12249	NA12892	NA18957	NA21098	NA24695

## Preparazione delle librerie

Le librerie di sequenziamento dell'esoma sono state preparate in duplicato seguendo il protocollo Illumina DNA Prep with Exome 2.5 Enrichment (Documento n. 1000000048041 v07) utilizzando 50 ng di DNA genomico (gDNA, genomic DNA) da campioni del Coriell Institute, se non diversamente indicato. Sono stati utilizzati IDT for Illumina DNA/RNA UD Indexes Sets A-D, Tagmentation (Illumina, n. di catalogo 20027213, 20027214, 20042666 e 20042667) per consentire l'indicizzazione univoca su tutte le librerie preparate. Le librerie pre-aricchimento sono state raggruppate in pool per massa (target 250 ng per campione) a 12 plex per pool e ibridate durante la notte utilizzando Twist Bioscience for Illumina Exome 2.5 Panel, disponibile all'interno dei kit Illumina DNA Prep with Exome 2.5 Enrichment (Illumina, n. di catalogo 20077595 e 20077596). Le librerie sono state preparate con e senza aggiunta di Twist Bioscience for Illumina Mitochondrial Panel (Illumina, n. di catalogo 20093180). L'aggiunta del pannello mitocondriale è avvenuta dopo la diluizione 1:100 descritta nella Guida di consultazione di Illumina DNA Prep with Exome 2.5 Enrichment (Documento n. 1000000157112).

## Sequenziamento

Le librerie arricchite sono state sequenziate con NovaSeq™ 6000 System e NextSeq™ 2000 System utilizzando reagenti standard per il sequenziamento mediante sintesi (SBS, sequencing by synthesis) e corse paired-end da 150 bp. Sono stati utilizzati gli output di dati più elevati disponibili per ottenere un'elevata profondità di copertura per campione e consentire il sottocampionamento come parte dell'analisi. Il pannello di campioni dei parametri normali e i campioni del caso sono stati sequenziati utilizzando NovaSeq 6000 S4 Reagent Kit v1.5 (300 cycles) (Illumina, n. di catalogo 20028312) e NextSeq 2000 P3 Reagents (300 cycles) (Illumina, n. di catalogo 20040561). I kit di reagenti per NovaSeq 6000 System sono progettati per facilitare l'uso con tre cartucce pronte all'uso preriempite con tutti i reagenti necessari.

Tabella 2: attributi riepilogativi del pannello dei parametri normali

Sesso	N. di campioni
Maschio	27
Femmina	27
Superpopolazione <sup>5</sup>	
Europea (EUR)	18
Americana mista (AMR)	8
Africana (AFR)	7
Asiatica meridionale (SAS)	9
Asiatica orientale (EAS)	12

Tabella 3: processività e copertura del sequenziamento su tutte le piattaforme per il pannello dei parametri normali e i campioni CNV

Piattaforma	Campioni per corsa	Copertura media per campione
NovaSeq 6000 System	72	Circa 540×
NextSeq 2000 System	12	Circa 280×

NextSeq 2000 System è più flessibile grazie a molteplici configurazioni di celle a flusso e offre l'analisi DRAGEN integrata. Il numero di campioni per corsa e la copertura media ottenuta per campione in tutte le corse sono indicati nella [Tabella 3](#).

## Analisi dei dati

È stata utilizzata l'app DRAGEN Baseline Builder v4.2, accessibile in BaseSpace™ Sequence Hub, per creare il pannello dei singoli file di conteggio target dei parametri normali con il genoma di riferimento non grafico hg38. I dati di sequenziamento per i campioni di casi CNV sono stati sottocampionati a 80 milioni di letture (profondità di copertura media su target di circa 120-140 volte) e 50 milioni di letture (profondità di copertura media su target di circa 70-90 volte) e analizzati utilizzando l'app DRAGEN Enrichment v4.2, accessibile anche in BaseSpace Sequence Hub, con rilevamento CNV abilitato. Per ciascuna piattaforma, la normalizzazione dei campioni è stata ottenuta utilizzando il pannello corrispondente dei parametri normali. L'app DRAGEN Enrichment esegue un'analisi secondaria accurata ed efficiente per l'identificazione di varianti completa, tra cui SNV, indel, CNV e varianti strutturali (SV, structural variant), tra le altre applicazioni. Anche l'identificazione di varianti DRAGEN per Illumina DNA Prep with Exome 2.5 Enrichment può essere eseguita su NextSeq 2000 System o completamente integrata nel flusso di lavoro di ricerca per l'interpretazione delle varianti del software Emedgene.

## Risultati

### Rilevamento di CNV in un intervallo di variazione

L'analisi con la pipeline DRAGEN CNV nell'app DRAGEN Enrichment ha rivelato un efficace rilevamento delle CNV per le varianti valutate in questo studio a soli 50 milioni di letture, ovvero a una profondità di copertura media di circa 70-90 volte, su tutte le piattaforme. Il rilevamento delle CNV è determinato dalla formula: (numero di regioni identificate)/(regioni che si sovrappongono al pannello dell'esoma e all'attendibilità). Le CNV sono state considerate rilevate se questa metrica era superiore al 98% (Tabella 4).

Molte delle varianti testate si trovano nel gene *DMD*, che è un target importante nella ricerca sulle malattie genetiche (Tabella 4).<sup>6</sup> Utilizzando Integrated Genomics Viewer, gli eventi previsti di *DMD* CNV rilevati possono essere visualizzati per i sette campioni interessati testati (Figura 2). Questo dimostra la capacità del flusso di lavoro Illumina Exome 2.5 nel rilevare le CNV su questo gene. La traccia del file BED del pannello Exome 2.5 mostra le regioni coperte dal pannello (Figura 2, in alto). I risultati dipenderanno da diversi fattori ed è consigliabile sequenziare fino a una copertura media superiore a 150× per l'analisi iniziale di un flusso di lavoro di laboratorio, al fine di garantire una copertura sufficiente per la maggior parte delle varianti.

## Ulteriori analisi possibili con il software Emedgene

Il software Emedgene semplifica e scala l'interpretazione delle varianti, consentendo di risparmiare il 50-75% delle tempistiche per caso di ricerca.<sup>7</sup> Le diverse caratteristiche potenziano l'interpretazione definita dall'utente, includendo la XAI per classificazioni automatizzate trasparenti e supportate da evidenze di potenziali varianti causative per i campioni, un grafico delle annotazioni e delle evidenze sempre aggiornato, la visualizzazione e la valutazione delle varianti, l'automazione definita dall'utente e molte altre funzionalità che favoriscono l'interpretazione informata delle varianti. Il software Emedgene è stato progettato per offrire un'esperienza utente efficiente e intuitiva.

Per semplificare la revisione dei casi, la funzionalità di prioritizzazione delle varianti IA o "selezione" compila le varianti, comprese le CNV, che hanno maggiori probabilità di risolvere un caso. Questa funzionalità include evidenze di supporto e consente all'analista un risparmio di tempo significativo. In uno studio separato<sup>8</sup> di 51 singoletti precedentemente risolti da una variante CNV, la variante di risoluzione è stata identificata in una selezione di 10 varianti nel 92% dei casi. Nel 6% (n = 3), la variante di risoluzione era presente nell'elenco dei candidati. Durante la revisione dei casi, il software Emedgene consente di rimuovere la marcatura delle varianti individuate dalla selezione o di identificare manualmente le varianti non selezionate dalla selezione. La classificazione ACMG automatizzata\* copre anche le CNV.

### Pannello di accesso ai file dei parametri normali

Il pannello dei file dei parametri normali, generati utilizzando campioni del Coriell Institute, fornisce ai ricercatori un set di dati noti per il confronto con i campioni del caso, senza che sia necessario investire tempo e risorse per l'acquisizione, il sequenziamento e l'analisi di tale set di campioni. Questi file predefiniti vengono utilizzati al meglio per il test iniziale del flusso di lavoro Illumina DNA Prep with Exome 2.5 Enrichment per l'analisi CNV; pertanto, si consiglia ai laboratori di sviluppare un pannello di parametri normali specifici per i propri protocolli di laboratorio. Se si utilizzano questi file per il test iniziale, le librerie devono essere preparate per i campioni di casi seguendo lo stesso protocollo utilizzato per preparare il pannello dei parametri normali. Le deviazioni nel protocollo possono portare a differenze nei profili di copertura, riducendo l'efficacia del pannello predefinito dei parametri normali. I file e le informazioni aggiuntive sono disponibili sul sito di supporto Illumina.

\*ACMG, American College of Medical Genetics and Genomics.

Tabella 4: rilevamento delle CNV ordinato secondo le dimensioni approssimative dell'evento in base alle coordinate dell'evento previste

ID Coriell	Cromosoma	Gene (esoni interessati) o posizione cromosomica	Evento previsto	Dimensione approssimativa dell'evento (kb) <sup>a</sup>	Rilevamento a 50 milioni di letture
NA04315	X	<i>DMD</i> (44)	Perdita	0,15	✓
NA05115	X	<i>DMD</i> (45)	Perdita	0,18	✓
NA05117	X	<i>DMD</i> (45)	Perdita	0,18	✓
NA23599	X	<i>MECP2</i> (3-4)	Perdita	2,25	✓
NA18949	17	<i>BRCA1</i> (15-16)	Perdita	3,59	✓
NA21939	15	<i>FBN1</i> (42-43)	Perdita	3,70	✓
HG03857	16	<i>PALB2</i> (5-7)	Perdita	4,23	✓
HG00343	22	<i>CHEK2</i> (9-10)	Perdita	4,25	✓
NA23127	X	<i>DMD</i> (27-28)	Guadagno	7,47	✓
NA04520	16	<i>TSC2</i> (1-15)	Perdita	16,28	✓
NA04327	X	<i>DMD</i> (5-7)	Guadagno	22,70	✓
NA10283	X	<i>DMD</i> (72-79)	Perdita	54,38	✓
HG00500	2	<i>SPAST</i> (4-17)	Guadagno	58,84	✓
NA23675	X	<i>MECP2</i> (tutto)	Guadagno	76,14	✓
NA23087	X	<i>DMD</i> (2-30)	Guadagno	608,45	✓
NA06870	18	18p11.32-18p11.1	Guadagno	15.390,21	✓
NA20125	10	10q23.1-10q26.3	Guadagno	52.877,99	✓

a. In base alle coordinate degli eventi previsti per l'esone o la posizione cromosomica interessati.

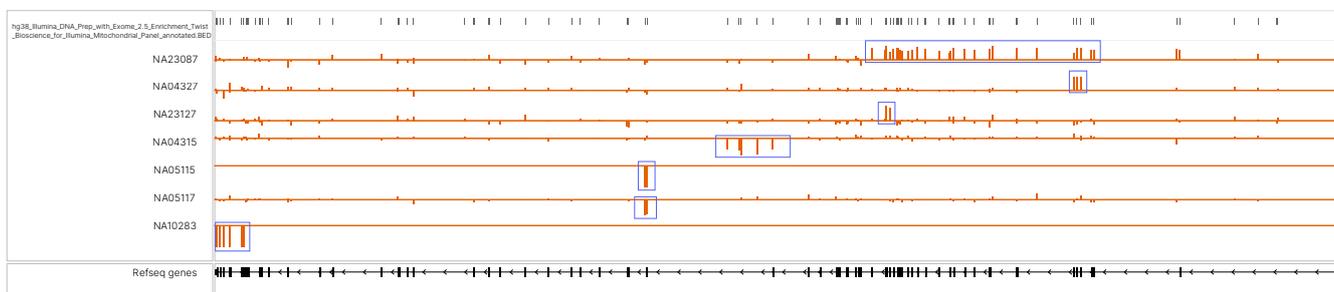


Figura 2: visualizzazione delle CNV rilevate su sette campioni contenenti gli eventi previsti nel gene *DMD*. I file che forniscono la rappresentazione BigWig del segnale normalizzato tangente (\*.tn.bw) vengono creati come parte dell'analisi dell'app DRAGEN Enrichment e mostrano guadagni e perdite in base alle regioni target nel pannello di arricchimento Twist Bioscience for Illumina Exome 2.5. Le tracce sono mostrate in scala automatica come grafico a barre.

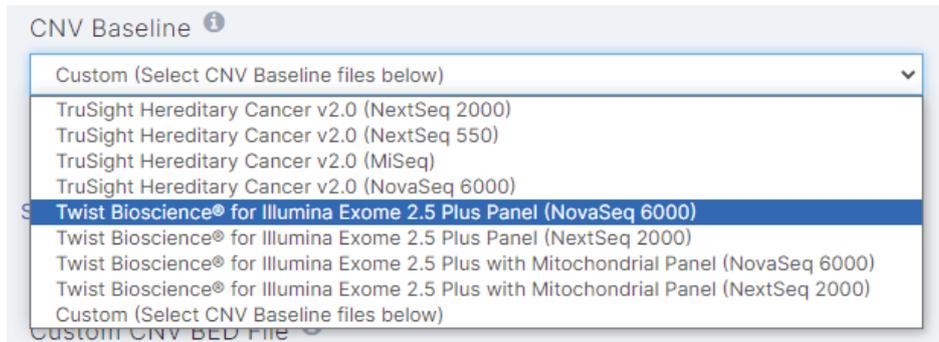


Figura 3: menu a discesa del pannello dei parametri normali nell'app DRAGEN Enrichment v4.3 in BaseSpace Sequence Hub.

Quando si utilizza BaseSpace Sequence Hub, l'app DRAGEN Enrichment v4.3 fornisce un menu a discesa di opzioni per eseguire l'analisi CNV iniziale con il pannello di arricchimento Twist Bioscience for Illumina Exome 2.5 per più riferimenti (grafico e non grafico hg19, hg38 e hs37d5) su più piattaforme (NextSeq 2000 System e NovaSeq 6000 System) utilizzando molteplici librerie (pannello di arricchimento Twist Bioscience for Illumina Exome 2.5 con e senza aggiunta di Twist Bioscience for Illumina Mitochondrial Panel) (Figura 3). I campioni dei casi utilizzati con questo flusso di lavoro di analisi devono essere preparati come descritto nella sezione Metodi della presente nota sull'applicazione.

## Riepilogo

Il flusso di lavoro dell'esoma NGS per l'analisi delle CNV utilizza Illumina DNA Prep with Exome 2.5 Enrichment per la preparazione delle librerie, il sequenziamento con NovaSeq 6000 System o NextSeq 2000 System e le applicazioni DRAGEN per l'analisi dei dati. I risultati dimostrano l'applicazione di un pannello di parametri normali come riferimento per l'analisi CNV di campioni di casi utilizzando il flusso di lavoro WES descritto. L'elevato richiamo di CNV da parte del software di analisi secondaria DRAGEN è perfettamente correlato con altri metodi. Questo flusso di lavoro per l'analisi CNV, compreso l'uso del pannello dei parametri normali, può anche essere adattato ad altre piattaforme di sequenziamento Illumina. Il software Emedgene consente ai laboratori di eseguire ulteriori analisi, come l'interpretazione delle varianti e la generazione di report di ricerca.

## Maggiori informazioni

[Dati di riferimento del pannello dei parametri normali](#)

[Analisi delle varianti del numero di copie](#)

[Illumina DNA Prep with Exome 2.5 Enrichment](#)

[Piattaforme di sequenziamento Illumina](#)

[Analisi secondaria DRAGEN](#)

[Applicazione DRAGEN Germline](#)

[Software Emedgene](#)

## Bibliografia

1. Zhang F, Gu W, Hurles ME, Lupski JR. [Copy number variation in human health, disease, and evolution](#). *Annu Rev Genomics Hum Genet.* 2009;10:451-481. doi:10.1146/annurev.genom.9.081307.164217.
2. Mehawej C, Maalouf JE, Abdelkhalik M, Mahfouz P, Chouery E, Megarbane A. [CNV Analysis through Exome Sequencing Reveals a Large Duplication Involved in Sex Reversal, Neurodevelopmental Delay, Epilepsy and Optic Atrophy](#). *Genes (Basel).* 2024;15(7):901. doi:10.3390/genes15070901.
3. Fang X, Ma M, Rong W, et al. [Exome sequencing confirms the clinical diagnosis of both joubert syndrome and klinefelter syndrome with keratoconus in a han Chinese family](#). *Front Genet.* 2024;15:1417584. doi:10.3389/fgene.2024.1417584.
4. Zeng Y, Ding H, Wang X, et al. [High positive predictive value of CNVs detected by clinical exome sequencing in suspected genetic diseases](#). *J Transl Med.* 2024;22(1):644. doi:10.1186/s12967-024-05468-1.
5. Harrison PW, Amode MR, Austine-Orimoloye O, et al. [Ensembl 2024](#). *Nucleic Acids Res.* 2024;52(D1):D891-D899. doi:10.1093/nar/gkad1049.
6. Kozareva V, Stroff C, Silver M, Freidin JF, Delaney NF. [Clinical analysis of germline copy number variation in DMD using a non-conjugate hierarchical Bayesian model](#). *BMC Med Genomics.* 2018;11(1):91. doi:10.1186/s12920-018-0404-4.
7. Greenwood Genetic Center. GGCC reduces turn around time on genomic analysis by 75% with Emedgene's AI platform. [ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform](https://ggc.org/in-the-news-app/ggc-reduces-turn-around-time-on-genomic-analysis-by-75-with-emedgenes-ai-platform). Pubblicato il 12 settembre 2019. Consultato il 14 agosto 2024.
8. Dati in archivio. Illumina, Inc. 2024.



Numero verde 1.800.809.4566 (U.S.A.) | Tel. +1.858.202.4566  
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. Tutti i diritti riservati. Tutti i marchi di fabbrica sono di proprietà di Illumina, Inc. o dei rispettivi proprietari.  
Per informazioni specifiche sui marchi di fabbrica, visitare la pagina web [www.illumina.com/company/legal.html](http://www.illumina.com/company/legal.html).  
M-GL-01453 ITA v1.0