



SeqOne platform

Istruzioni per l'uso

Valido da SeqOne Platform v2.1

1. Introduzione	7
2. Contattare il produttore e il rappresentante svizzero	8
3. Vigilanza sui dispositivi medici	8
4. Uso previsto della piattaforma	8
4.1. Uso previsto	8
4.2. Indicazioni d'uso	9
4.3. Descrizione	9
4.4. Popolazione di pazienti: popolazione di pazienti	10
4.5. Utenti previsti	10
4.6. Caratteristiche	10
4.6.1. Prestazioni analitiche	12
4.6.1.1. Moduli bioinformatici	13
4.6.1.2. Moduli di annotazione	22
4.6.2. Prestazioni cliniche	23
5. Precauzioni	24
5.1. Uso conforme	24
5.2. Avvertenze	24
5.3. Limiti	26
5.4. Malfunzionamenti	26
6. Requisiti minimi	27
6.1. Hardware	27
6.2. Rete	27
7. Autenticazione	28
7.1. Accesso	28
7.2. Disconnessione	29
7.3. Pagina iniziale	29
7.4. Gestione dell'account utente	30
7.4.1. Gestione di "My Account"	31
7.4.2. Gestione delle impostazioni utente - definizioni	31
Entity	31
Users	31
Worksets	31
Manifest	32
Panels - Transcript	32
Panels - Gene	32
Marker	32
Protocols	32
Reporting - Global settings	33
Reporting - Template list	33

Tags	33
API Keys	33
7.4.3. Gestione "Admin" - presentazione dell'interfaccia	34
7.4.4. Passaggio tra le modalità Main e Preview	37
8. Descrizione delle fasi operative	38
9. Creazione del progetto	39
9.1. Creare un progetto	39
9.2. Gestione dei progetti	40
9.3. Eliminazione di un progetto	40
9.4. Dati sul DNA (sequenziamento del DNA, sequenziamento del ctDNA)	40
9.5. Dati sull'RNA (sequenziamento dell'RNA)	42
9.6. Impostazioni del progetto	43
9.6.1. Configurare le metriche predefinite del progetto QC	43
9.6.2. Impostazione del profilo di filtro predefinito del progetto	43
10. Campione di caricamento	44
10.1. Caricamento dei campioni	44
Da un PC	44
Tramite SDS (SeqOne Data Sync)	44
Sample upload status	44
Capacità di caricamento	45
10.2. Formato del file di ingresso	45
10.2.1. File di analisi primaria	45
10.2.2. File di analisi secondaria	46
10.3. Metadati campione	48
10.4. Modifica di un campione	49
10.5. Eliminare un campione	49
11. Analisi	50
11.1. Avvio di una nuova analisi	50
11.1.1. Scelta del worksets d'analisi	51
Worksets di analisi disponibili da file FASTQ	51
GermlineVar	52
GermVar	52
GermlineVarLR	52
GermlineFamily	53
GermVar Family	53
CNVCapture	54
CNVExome	54
SomaVar	54
SomaCNVCapture	55
SomaRNA	55

SomaMSI	56
SomaHRD	56
SomaVar LF	57
SomaCGP	57
SomaHemato	58
SomaLBx	58
SomaMethyl	59
Worksets di analisi disponibili da file di analisi secondaria	59
GermVar Tertiary	59
GermVar Tertiary Family	60
SomaVar Tertiary	61
11.1.2. Scelta della modalità di elaborazione UMI	62
Presentazione delle modalità di elaborazione dell'UMI	62
Processo di trattamento UMI	63
Compatibilità con i worksets di analisi	63
11.2. Stato dell'analisi	64
11.3. Qualità del campione	64
12. Controllo della qualità dei dati	66
12.1. Metriche e report sul QC	66
12.1.1. Metriche e verifiche QC	66
12.1.2. Rapporto QC	66
12.1.3. Rapporto 'ID-Check'	70
12.1.4. Rapporto tecnico	71
Punteggio UPD	72
13. Interpretazione dei risultati dell'analisi	74
13.1. Risultati dell'analisi	74
13.1.1. Gestione del flusso dell'interpretazione dell'analisi	74
13.1.1.1. Assegnazione di analisi per l'interpretazione	74
13.1.1.2. Analisi del bloccaggio	75
13.1.1.3. Pannello laterale per la gestione delle analisi	76
Vista modale Workflow	77
Vista modale Patient	78
Vista modale HPOs	79
Vista modale Panel	81
13.1.1.4. Pannello laterale per la gestione delle valutazioni	82
13.1.2. Panoramica dei risultati dell'analisi nella scheda Dashboard	83
13.1.3. Copertura dettagliata per esone	84
Metodologia	84
Vista e interpretazione	85
13.1.4. Piccole varianti	88

Metodologia	88
Visualizzazione e interpretazione	92
Ordinamento delle varianti	106
Configurazione delle colonne della tabella delle varianti	108
Modifica dei metadati del campione dal pannello di analisi	108
Varianti di filtraggio	110
Filtri Genes & Variants	110
Sezione Samples & Quality	110
Sezione Frequencies	112
Sezione Genetics	113
Sezione Databases	114
Sezione Transmission & Phenotypes	115
Sezione Classification	118
Sezione Scores	119
Selezionare un'indicazione	121
Utilizzo di un profilo di filtro predefinito	122
Modifica e salvataggio dei profili di filtro	122
Profili dei filtri preferiti	122
Combinazione di profili di filtro	123
Profili dei filtri di SeqOne	123
Funzione "Mark as viewed"	124
Selezione delle varianti	124
Valutazione delle varianti	125
Pagina della variante	128
Generazione di un referto clinico	136
13.1.5. CNV	137
13.1.5.1. CNV in pannello di geni	137
Metodologia	137
Vista e interpretazione	138
13.1.5.2. CNV nell'analisi di SomaCGP e SomaLBx	143
Metodologia	143
Visualizzazione e interpretazione	143
13.1.5.3. CNV negli esomi e nei genomi	144
Metodologia per gli esomi	144
Metodologia per il genoma	145
Vista e interpretazione	145
Tabella dei risultati del CNV	145
Pagina dettagliata CNV	150
13.1.6. Visualizzatore di varianti grandi per CNVs e SVs (LVV)	154
Tabella dei risultati di CNVs & SVs	154

Ordinamento, Filtraggio, Visualizzazione e Interpretazione	156
Pannello laterale CNV/SV (Details Drawer)	157
13.1.7. Firme nell'analisi di SomaCGP	158
Metodologia di calcolo del punteggio MSI	158
Metodologia per il calcolo del punteggio TMB Score	158
Visualizzazione e interpretazione	159
13.1.8. Brevi ripetizioni in tandem (STRs)	161
Metodologia	161
Vista e interpretazione	161
13.1.9. Geni di fusione	163
13.1.9.1. Fusioni rilevate dal DNA	163
Metodologia	163
Vista e interpretazione	163
13.1.9.2. Fusioni rilevate dall'RNA	165
Metodologia	165
Vista e interpretazione	166
13.1.10. Clonalità	169
13.1.11. Varianti di splicing	170
Metodologia	170
Visualizzazione e interpretazione	170
13.1.12. Stato MSI/MSS	172
Metodologia	172
Vista e interpretazione	172
13.1.13. Stato HRD/HRP	174
Metodologia	174
Visualizzazione e interpretazione	176
13.1.14. Dati di metilazione per SomaMethyl	182
14. Rapporto	185
14.1.1. File di uscita	185
15. Risoluzione dei problemi e messaggi di errore	186
16. Garanzia	188
17. Avviso legale	188
17.1. Condizioni d'uso	188
17.2. Limitazione di responsabilità	188
17.3. Diritti d'autore e marchi di fabbrica	188

1. Introduzione

Si consiglia di leggere attentamente le istruzioni prima di utilizzare il software. Una versione cartacea di queste istruzioni può essere fornita su richiesta al team di supporto di SeqOne (support@seqone.com). Per ottenere i migliori risultati nell'uso della piattaforma, leggere

attentamente tutte le precauzioni evidenziate con il simbolo . I vari simboli normativi e altri simboli utilizzati in questa guida e nel software sono i seguenti:

Simbolo	Significato
	Produttore
	Indica il rappresentante autorizzato con sede in Svizzera.
	Data di produzione (data di rilascio del software)
	Dispositivo medico-diagnostico in vitro.
	Identificatore univoco del dispositivo
	Riferimento del catalogo del produttore
	Conforme ai requisiti del Regolamento europeo (UE) 2017/746. Marcatura CE con numero di identificazione dell'organismo notificato
	Consultare le istruzioni per l'uso
	Avvertenze Questo simbolo di sicurezza indica che è necessario prestare particolare attenzione durante l'utilizzo del dispositivo.

L'etichetta (comprese le informazioni sulla versione del software SeqOne e le informazioni di contatto) e questo documento che descrive le istruzioni per l'uso (in varie lingue) sono disponibili nella sezione "About".

2. Contattare il produttore e il rappresentante svizzero



SeqOne S.A.S

Web: <https://seqone.com/>

22 rue Durand
34000 Montpellier
FRANCIA

Supporto per l'utente :

support@seqone.com



Arazy Group Swiss GmbH

Bruderholzallee 53

4059 Basilea, Svizzera

swiss.ar@arazygroup.com

3. Vigilanza sui dispositivi medici

Se si verifica o è probabile che si verifichi un incidente grave durante l'utilizzo del dispositivo, gli utenti sono invitati a segnalarlo senza indugio al produttore e all'autorità competente del Paese in cui si è verificato o è probabile che si verifichi l'incidente.

Le normative applicabili e l'elenco delle autorità competenti nell'UE sono disponibili al seguente indirizzo:

https://health.ec.europa.eu/medical-devices-sector/new-regulations/contacts_en



4. Uso previsto della piattaforma

4.1. Uso previsto

SeqOne Platform è un dispositivo medico-diagnostico *in vitro* basato su software, progettato per supportare gli operatori sanitari nell'analisi e nell'interpretazione dei dati genomici e di sequenziamento grezzi di pazienti sottoposti ad analisi NGS o CGH array. Fornendo approfondimenti qualitativi e operativi, la piattaforma funziona come un sistema di supporto alle decisioni, aiutando gli operatori sanitari a prendere decisioni cliniche.

4.2. Indicazioni d'uso

SeqOne Platform supporta l'analisi secondaria e terziaria a partire dai dati di sequenziamento grezzi, nonché la sola analisi terziaria. Le sue capacità analitiche indipendenti dalla patologia trovano applicazione in diversi campi medici, tra cui l'oncologia e le malattie rare.

4.3. Descrizione

SeqOne Platform è un dispositivo medico-diagnostico *in vitro* basato su software che integra una suite di applicazioni che lavorano di concerto per fornire agli utenti esperti un potente kit di strumenti per l'interpretazione di dati genomici e di sequenziamento complessi.

SeqOne Platform integra informazioni consolidate provenienti da un'ampia gamma di **fonti esterne**, tra cui database pubblici, letteratura scientifica e conoscenze proprie. Queste fonti di dati esterne forniscono un contesto e un significato alle varianti genomiche identificate.

I file di input per varie applicazioni all'interno della piattaforma vengono elaborati utilizzando una combinazione personalizzata di moduli bioinformatici e di annotazione, garantendo che ogni analisi sia ottimizzata per i suoi requisiti specifici.

I moduli bioinformatici conducono analisi secondarie utilizzando metodi computazionali come l'allineamento e la chiamata di varianti per identificare varianti genetiche nei dati di sequenziamento di DNA e RNA confrontandoli con un genoma di riferimento (hg19 o hg38). I dati, che possono essere utilizzati per analisi sia germline che somatiche, possono essere generati dal sequenziamento dell'intero genoma, dal sequenziamento dell'esoma o di pannelli di geni.

I moduli di annotazione conducono un'analisi terziaria per annotare le varianti con dettagli quali il gene e la sua funzione nota, la posizione della variante all'interno del gene ed eventuali pubblicazioni o rapporti clinici che menzionano il gene o la variante. Questa annotazione determinerà quali regole di classificazione ACMG e AMP/CAP/ASCO si applicano a ciascuna variante, e da qui si può giungere a un verdetto di patogenicità. L'annotazione viene eseguita anche attraverso "DiagAI", una suite di strumenti alimentati dall'intelligenza artificiale che aiuta ad analizzare le varianti genetiche per identificare rapidamente le potenziali varianti causali.

Questa architettura modulare consente a SeqOne di essere altamente adattabile a diversi tipi di analisi e domande di ricerca. I file di input per le varie applicazioni all'interno della piattaforma vengono elaborati utilizzando una combinazione personalizzata di questi moduli bioinformatici e di annotazione, garantendo che ogni analisi sia ottimizzata per i suoi requisiti specifici. Grazie all'integrazione di queste diverse fonti di dati e strumenti di analisi, SeqOne consente agli utenti esperti di acquisire conoscenze approfondite sul complesso panorama dei dati genomici e sulle

loro implicazioni per la salute e la malattia umana.

La piattaforma opera come soluzione SaaS basata sul cloud, con tutte le funzioni accessibili tramite un browser web. Include SeqOne Flow, uno strumento opzionale che può essere installato nel cloud AWS o sui server del cliente per facilitare il trasferimento dei dati e l'integrazione dei flussi di lavoro.

Per quanto riguarda la sicurezza, tutti i dati sono criptati a riposo utilizzando la crittografia AES-256 e il traffico tra gli utenti e la piattaforma è protetto da certificati SSL.

4.4. Popolazione di pazienti: popolazione di pazienti

La soluzione è adatta a qualsiasi paziente che richieda un'analisi genomica. I pazienti target possono essere affetti da cancro, malattie rare, malattie ereditarie, predisposizione genetica al cancro o persone sane.

4.5. Utenti previsti

SeqOne Platform è progettata per gli operatori sanitari responsabili delle analisi genomiche cliniche di routine. Ciò include l'analisi e l'interpretazione dei file grezzi generati dal sequenziatore a partire da campioni biologici sequenziati. Gli utenti target della piattaforma sono genetisti, bioinformatici, biologi, ingegneri e tecnici di piattaforme NGS.

4.6. Caratteristiche

Ogni applicazione (workset) è costituita da più **moduli bioinformatici e di annotazione**, ciascuno responsabile di specifici compiti analitici. In questo capitolo, le prestazioni vengono valutate a livello di modulo per garantire una valutazione dettagliata dell'accuratezza e dell'affidabilità di ciascun componente.

Gli analiti rilevati dalla SeqOne Platform rientrano in due categorie di alterazioni genomiche:

- Varianti genetiche quali SNV, indel, MNV, CNV, UPD, fusioni geniche, varianti strutturali, varianti splicing, inserzioni Alu e FLT3-ITD (duplicazione interna in tandem).
- Firme o biomarcatori complessi quali lo stato IGHV, l'instabilità dei siti microsatellitari (MSI), il carico mutazionale tumorale (TMB), la metilazione o i difetti nella riparazione delle rotture a doppio filamento tramite ricombinazione omologa (HRD, Homologous Recombination Deficiency).

Modulo di bioinformatica	Workset	Tipo di analita
Rilevamento CNV Exome Germline	CNVExome, GermVar, GermVar Family	Varianti genetiche
Rilevamento CNV Pannello Germline	CNVcapture	Variazione genetica
SNV/Indel Germline	GermlineVar, GermlineFamily	Variante genetica
SV Germline	GermlineVar, GermlineFamily	Variante genetica
UPD	GermVar, GermVar Family, GermVar Tertiary, GermVar Tertiary Family	Variante genetica
Inserzioni Alu	GermlineVar, GermlineFamily, SomaVar, Somavar LF, SomaCGP, SomaHemato, GermVar, GermVar Family	Variante genetica
MNV	GermlineVar, GermlineFamily, GermVar, GermVar Family, SomaVar Tertiary, GermVar Family Tertiary, SomaVar, SomaVar LF, SomaCGP, SomaHemato, SomaLBx, SomaVar Tertiary	Variante genetica
SNV/Indel somatico	SomaVar, SomaHemato, SomaRNA, SomaCGP	Variante genetica
SNV/Indel somatico LF	SomaVar LF	Variazione genetica
CNV somatica Pannelli estesi	SomaCGP	Varianti genetiche
Indice di metilazione	SomaMethyl	Firma
Fusione del DNA somatico	SomaCGP, SomaVar, SomaVar LF, SomaHemato	Variante genetica
Fusione del DNA somatico LBx	SomaLBx	Variante genetica
TMB	SomaCGP	Firma
MSI	SomaCGP, SomaMSI	Firma
Probabilità GI (GiP)	SomaHRD	Firma
Sentieon SNV/Indel Germline	GermVar, GermVar Family	Varianti genetiche
Varianti mosaiche Germline	GermVar, GermVar Family	Variazione genetica
CNV WGS	GermVar, GermVar Family	Variante genetica
Sentieon SNV/Indel Somatico	SomaLBX	Variante genetica
RNA di fusione	SomaRNA	RNA di splicing
RNA di splicing	SomaRNA	Variazione genetica
CNV somatica Pannello ridotto	SomaCNVcapture	Variante genetica
Stato IGHV	SomaHemato	Firma
Sottogruppi IGHV (subsets)	SomaHemato	Firma
CFH	GermlineVarLR	Variante genetica
FLT3	SomaVar, SomaHemato	Variante genetica

4.6.1. Prestazioni analitiche

Le prestazioni dichiarate in queste istruzioni per l'uso (IFU) si riferiscono alle prestazioni delle ultime versioni IVD dei worksets, elencate nella tabella sottostante:

Nome del workset	Versione
CNVCapture	2.6
CNVExome	2.2
GermlineVar	2.6
GermlineFamily	2.6
GermlineVar LR	1.2
GermVar	3.12
GermVar Family	3.12
GermVar Tertiary	1.16
GermVar Tertiary Family	1.14
SomaCGP	1.6
SomaCNVCapture	2.4
SomaHemato	0.6
SomaHRD	1.5
SomaLBx	0.7
SomaMethyl	0.3
SomaMSI	1.1
SomaRNA	3.3
SomaVar	2.9
SomaVar LF	2.1
SomaVar Tertiary	2.4

4.6.1.1. Moduli bioinformatici

- **Rilevamento CNV Esoma Germline e Panel Germline**

Tipo di variante	Versione del genoma	Controlli positivi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità
Rilevamento CNV Exome Germline (GermlineVar)	GRCh37	8	8	0	100%	100%
Rilevamento CNV Exome Germline (GermVar)	GRCh37	8	8	0	100%	100%
Rilevamento CNV Panel Germline	GRCh37	15	15	0	100%	100%
Rilevamento CNV Panel Germline	GRCh38	11	11	0	100%	100%

- **SNV / Indel Germline - Dati del pannello**

Versione del genoma	Tipo di variante	Criteri di accettazione	Sensibilità prima del filtro	Sensibilità dopo il filtro
GRCh37	SNV	100%	98%	100%
GRCh38	SNV	100%	98%	100%

Versione del genoma	Tipo di variante	Criteri di accettazione	Precisione prima del filtro	Precisione dopo il filtro
GRCh37	SNV	100%	90%	100%
GRCh38	SNV	100%	90%	100%

- **SNV / Indel "Germline" - Dati dell'esoma**

Versione del genoma	Tipo di variante	Criteri di accettazione*	Sensibilità prima del filtro	Sensibilità dopo il filtro
GRCh37	SNV	> 98%	98,7%	98,5%
	Indel	> 80%	88,0%	87,5%
GRCh38	SNV	> 98%	98,5%	98,3%
	Indel	> 80%	84,9%	84,3%

* per il rilevamento di varianti situate nelle regioni ad alta affidabilità di NA12878 e NA24385.

Versione del genoma	Type of variant	Criteri di accettazione*	Precisione prima del filtro	Precisione dopo il filtro
GRCh37	SNV	> 90%	92,1%	94,4%
	Indel	> 80%	84,7%	89,0%
GRCh38	SNV	> 90%	94,2%	95,4%
	Indel	> 80%	82,2%	87,6%

* per il rilevamento di varianti situate nelle regioni ad alta affidabilità di NA12878 e NA24385.

- **SNV / Indel Germline - Ripetibilità e Riproducibilità**

- **Ripetibilità**

Tipo di variante	Sensibilità	Precisione
SNV	98,59% ± 0%	95,92% ± 0%
indel	87,75% ± 0%	81,63% ± 0%
tutto	98,38% ± 0%	95,63 ± 0%

- **Riproducibilità**

Tipo di variante	Sensibilità	Precisione
SNV	98,57% ± 0,03%	95,93% ± 0,08%
indel	86,58% ± 0,42%	78,97% ± 0,96%
(all)	98,34% ± 0,03%	95,58 ± 0,09%

- **Varianti mosaiche Germline**

- **Sensibilità SNV**

Versione del genoma	Intervalli VAF	# varianti vere	Criteri di accettazione	Sensibilità 100x
GRCh38	[0,05,0,1[	3704	40%	47,6%
GRCh38	[0,1, 0,2[	747	70%	73,1%
GRCh38	[0,2; 0,3]	685	90%	90,7%
GRCh38	[0,3,1,0[	76	90%	93,4%

- **Sensibilità Indels**

Versione del genoma	Intervalli VAF	Numero di varianti vere	Criteri di accettazione	Sensibilità 100x
GRCh38	[0,05,0,1[	87	40%	43,7%
GRCh38	[0,1, 0,2[	14	70%	78,6%
GRCh38	[0,2; 0,3]	15	90%	100%
GRCh38	[0,3,1,0[	0	'N/A'	"N/A"

- **SV Germline**

Tipo di variante	Versione del genoma	Controlli positivi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità
SV Germline	GRCh37	2	2	0	100%	100%

- **Inserzioni Alu**

Versione del genoma	Controlli positivi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità
GRCh37	1	1	0	100%	100%
GRCh38	1	1	0	100%	100%

- **MNV**

Versione del genoma	Tipo di variante	Criteri di accettazione	Precisione dell'Exome	Genoma di precisione
GRCh37	MNV	> 85%	87,0%	86,9%
GRCh38	MNV	> 85%	93,0%	91,5%

- **SNV / Indel somatico**

- **Sensibilità**

Versione del genoma	Tipo di variante	Controlli positivi	Veri positivi	Falsi negativi	Criteri di accettazione*	Sensibilità
SNV	SNV ($\geq 5\%$)	121	121	0	100%	100%
	Indel ($\geq 5\%$)	21	21	0	100%	100%
GRCh38	SNV ($\geq 5\%$)	121	121	0	100%	100%
	Indel ($\geq 5\%$)	21	21	0	100%	100%

**Il criterio di accettazione per il rilevamento delle varianti è definito come una sensibilità del 100% per SNV e indel con un VAF previsto superiore al 5%.*

○ Veridicità

			Veridicità			
			GRCh37		GRCh38	
Gene	Variant	Frequenza prevista	Frequenza media *	Distorsione	Frequenza frequenza *	Distorsione
KIT	p.D816V	10%	9,07% +/- 0,79%	0,97%	9,07% +/- 0,79%	0,97%
PIK3CA	p.E545K	8,80%	7,88% +/- 0,84%	1,04%	7,87% +/- 0,83%	1,05%
EGFR	p.E746_A750del	1,90%	1,53% +/- 0,5%	0,40%	1,53% +/- 0,5%	0,40%
KRAS	p.G12D	6,30%	6,27% +/- 0,76%	0,57%	6,27% +/- 0,76%	0,57%
KRAS	p.G13D	15%	14,59% +/- 0,86%	0,69%	14,59% +/- 0,86%	0,69%
EGFR	p.G719S	24,50%	24,5% +/- 1,39%	1,15%	24,5% +/- 1,39%	1,15%
PIK3CA	p.H1047R	17,50%	17,47% +/- 1,04%	0,77%	17,47% +/- 1,04%	0,77%
EGFR	p.L858R	2,80%	3,58% +/- 0,51%	0,83%	3,58% +/- 0,51%	0,83%
NRAS	p.Q61K	12,50%	11,27% +/- 1,19%	1,27%	11,27% +/- 1,19%	1,27%
EGFR	p.T790M	0,90%	1,33% +/- 0,15%	0,43%	1,33% +/- 0,15%	0,58%
BRAF	p.V600E	10,70%	11,68% +/- 0,62%	0,98%	11,68% +/- 0,62%	0,98%

* La media delle frequenze è ottenuta su 12 repliche del campione HD200.

● SNV/Indel Somatic LF

Genoma versione	Controlli controlli	Vero Falsi	Falsi negativi	Criteri di accettazione*	Sensibilità
GRCh37	42	42	0	100%	100%
GRCh38	42	42	0	100%	100%

*Il criterio di accettazione per il rilevamento delle varianti è definito come una sensibilità del 100% per SNV e indel con un VAF previsto superiore allo 0,1%.

● Somatic Large Panels for CNV

Versione del genoma	Tipo di variante	Controlli positivi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità
GRCh37	Gain	27	26	1	90%	96,3%
	Loss	20	19	1	90%	95,0%
	LOH	11	11	0	90%	100,0%
GRCh38	Gain	27	26	1	90%	96,3%
	Loss	20	19	1	90%	95,0%
	LOH	11	10	1	90%	90,9%

- **Indice di metilazione**

Versione del genoma	Criteri di accettazione	Concordanza ponderata
GRCh37	>99%	99,88%

- **Fusione del DNA somatico**

Genoma versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh37	9	9	0	100%	100%
GRCh38	9	9	0	100%	100%

- **Somatic DNA LBx fusion**

Versione del genoma	Frazione tumorale	Controlli positivi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità
GRCh38	1%	22	21	1	85%	95,5%
	0,5%	22	19	3	85%	86,4%
	0,1%	22	21	1	85%	95,5%

- **TMB**

Versione del genoma	Criteri di accettazione	Sensibilità	Precisione
GRCh37	>80%	83%	85%

- **MSI**

Versione del genoma	Controlli positivi	Controlli negativi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità	Precisione
GRCh37	22	24	22	0	100%	100%	100%
GRCh38	6	17	6	0	100%	100%	100%

- **GiP**

Genoma versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh37	171	165	6	90%	96,5%

Genoma versione	Controlli controlli	Vero Falsi	Falsi positivi	Accettazione Criteri	Precisione
GRCh37	171	165	7	90%	95,9%

- **Sentieon SNV / Indel Germline**

Versione del genoma	Tipo di variante	Criteri di accettazione*	Sensibilità Exome	Sensibilità Genoma
GRCh37	SNV	> 90%	98,9%	98,5%
	Indel	> 80%	96,5%	98,5%
GRCh38	SNV	> 90%	98,5%	99,3%
	Indel	> 80%	95,5%	99,1%

* per il rilevamento di varianti situate nelle regioni ad alta affidabilità di NA12878 e NA24385

Versione del genoma	Tipo di variante	Criteri di accettazione*	Precisione Exome	Genoma di precisione
GRCh37	SNV	> 90%	98,7%	98,5%
	Indel	> 80%	96,5%	98,6%
GRCh38	SNV	> 90%	99,6%	99,6%
	Indel	> 80%	97,2%	99,5%

* per il rilevamento di varianti situate nelle regioni ad alta affidabilità di NA12878 e NA24385.

- **CNV WGS**

Versione del genoma	Controlli positivi	Veri positivi	Falsi positivi	Criteri di accettazione	Sensibilità
GRCh38	Duplicazione	1 kb - 5 kb	8	10%	37,5%
	Duplicazione	≥ 5 kb	3	10%	100%
	Delezione	1 kb - 5 kb	321	80%	61,7%
	Delezione	≥ 5 kb	132	80%	92,4%

Sebbene la sensibilità alle piccole delezioni (1-5 kb) sia inferiore ai criteri di accettazione, questa riduzione rappresenta un compromesso accettabile al fine di massimizzare il rilevamento di eventi di grandi dimensioni e duplicazioni più rilevanti.

Versione del genoma	Controlli positivi	Vero positivi	Falsi positivi	Criteri di accettazione	Precisione
GRCh38	Duplicazione	1 kb - 5 kb	8	"NA"	30%
	Duplicazione	≥ 5 kb	3	"NA"	75%
	Delezione	1 kb - 5 kb	321	"NA"	94,3%
	Delezione	≥ 5 kb	132	"NA"	92,4%

- **Sentieon SNV/Indel Somatic**

Versione del genoma	Tipo di variante	Controlli positivi	Veri positivi	Falsi positivi	Criteri di accettazione	Sensibilità
GRCh38	SNV	39	37	2	85%	95%
	Indel	21	16	5	75%	76%

- **RNA di fusione**

Genoma Versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh37	20	20	0	100%	100%
GRCh38	20	19	1	95%	95%

- **RNA di splicing**

Genoma Versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh37	3	3	0	100%	100%
GRCh38	3	3	0	100%	100%

- **CNV Pannello ridotto Somatic**

Genoma Versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh37	4	4	0	100%	100%
GRCh38	4	4	0	100%	100%

- **Status IGHV IMGT**

Genoma versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh37	122	116	6	95%	95,9%

- **Stato IGHV igblast**

Versione del genoma	Controlli positivi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità
GRCh37	122	117	5	95%	95,1%

- **Sottogruppi IGHV**

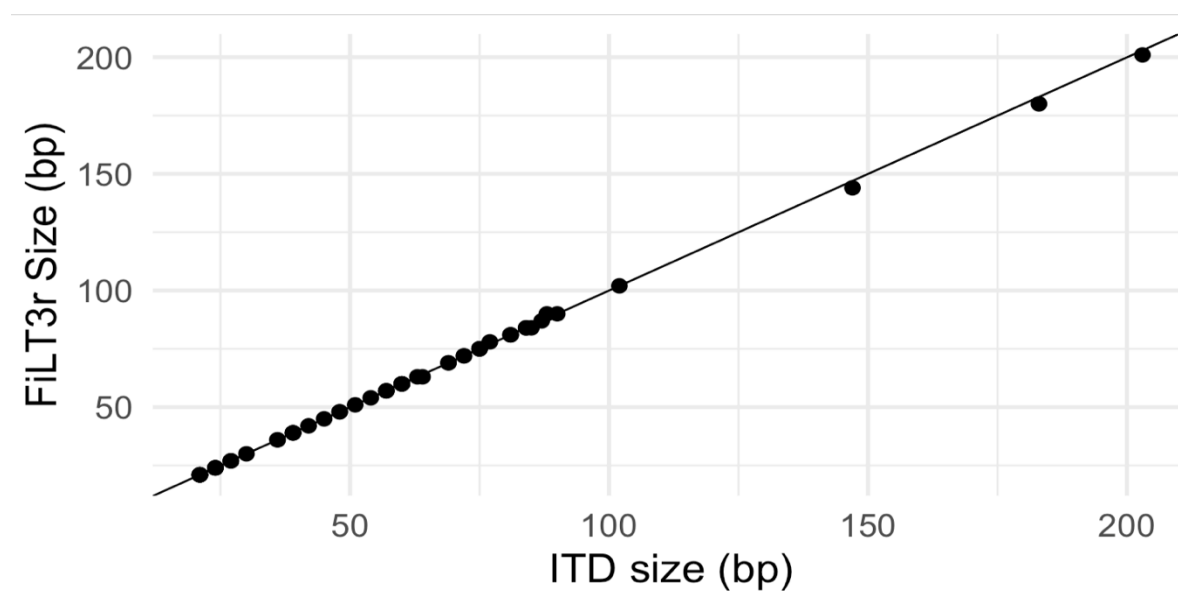
Genoma versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh37	61	10	0	95%	100%

Genoma versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Precisione
GRCh37	61	10	0	95%	100%

- CFH

Genoma versione	Controlli controlli	Vero Falsi	Falsi negativi	Accettazione Criteri	Sensibilità
GRCh38	4	4	0	100%	100%

- FLT3



Dimensione osservata vs teorica degli eventi FLT3-ITD. Coefficiente di Pearson $R^2 > 0,99$

I risultati mostrano una sensibilità del 100% nel rilevare gli eventi FLT3-ITD e una correlazione quasi perfetta tra le dimensioni degli eventi attese e quelle osservate dall'analisi.

- **UPD**

- **Sensibilità**

Tipo di analisi	Versione del genoma	Controlli positivi	Controlli negativi	Veri positivi	Falsi negativi	Criteri di accettazione	Sensibilità
Singleton	GRCh37/38	7	23	7	0	80%	100%
Trio	GRCh37/38	7	2	6	1	80%	86%

- **Precisione**

Tipo di analisi	Versione del genoma	Controlli positivi	Controlli negativi	Veri positivi	Falsi positivi	Criteri di accettazione	Precisione
Singleton	GRCh37/38	7	23	7	0	80%	100%
Trio	GRCh37/38	7	2	6	0	80%	100%

4.6.1.2. Moduli di annotazione

- **DiagAI Score**

Set di dati	Analisi del tipo	Metrico valutato	Accettazione Criteri	Metrica valore
Esomi positivi per malattie rare	"Germline"	Rang mediano	= 1	1
Esomi positivi per malattie rare	'Germline'	top_10	> 90%	95%
Esomi CNV per malattie rare	"Germline"	Rang mediano	= 1	1
CNV positivi per malattie rare	'Germline'	top_1	> 80%	83,3%

- **Punteggio somatico DiagAI**

Set di dati	Tipo di analisi	Metrica valutata	Criteri di accettazione	Valore della metrica
TSO500 positivo per il cancro	"Somatic"	top_10	> 90%	100%

- **Selezioni DiagAI**

Set di dati	Analisi del tipo	Metrica valutata	Accettazione Criteri	Valore della metrica
Esomi positivi per malattie rare	"Germline"	Sensibilità	> 90	95%
Esomi positivi per malattie rare	'Germline'	Lunghezza media	< 25	18,475
Esomi positivi per malattie rare	"Germline"	Lunghezza mediana	< 25	17

- **SmartPick di DiagAI**

Set di dati	Analisi del tipo	Metrica valutata	Accettazione Criteri	Valore della metrica
Esomi positivi per malattie rare	Sensibilità	Sensibilità	> 40%	45%
Esomi positivi per malattie rare	'Germline'	precisione	> 80%	78%

- **ACMG CNV**

Set di dati	Tipo di analisi	Metrica valutata	Criteri di accettazione	Valore della metrica
Set di dati CNV annotato da esperti ACMG	'Germline'	Sensibilità	> 70%	76%

- **ACMG SNV**

Set di dati	Analisi del tipo	Metrica valutata	Criteri di accettazione	Valore della metrica
Set di dati annotato da esperti ACMG	"Germline"	Accuratezza	> 50%	56%

- **DiagAI UP2**

Set di dati	Analisi del tipo	Metrica valutata	Criteri di accettazione	Metrica valore
Varianti patogene	Sensibilità	Sensibilità	90%	92%

- **DiagAI UP2 SV**

Set di dati	Analisi del tipo	Metrica valutata	Criteri di accettazione	Metrica valore
Varianti patogene	Sensibilità	Sensibilità	90%	98%

- **DiagAI Phenogenius**

Set di dati	Analisi del tipo	Metrica valutata	Criteri di accettazione	Metrica valore
Associazioni tra geni e sintomi	'Germline'	top_100	50%	49,8%
Associazioni tra geni e sintomi	'Germline'	top_200	60%	60,7%
Associazioni tra geni e sintomi	'Germline'	top_500	75%	77,0%
Associazioni tra geni e sintomi	'Germline'	Mediana nella top 100	<20	15
Associazioni tra geni e sintomi	'Germline'	Media nella top 100	<30	25

- **Compermed**

Set di dati	Type of analysis	Metrica valutati	Criteri di accettazione	Metrica valore
Cancro positivo TSO500	'Somatic'	sensibilità	90%	91%

4.6.2. Prestazioni cliniche

La SeqOne Platform non è destinata alla valutazione di una particolare condizione clinica o stato fisiologico. Sebbene fornisca informazioni sullo stato fisiologico di un paziente, come l'identificazione di determinate varianti, non è destinata alla valutazione di condizioni cliniche o stati fisiologici specifici. Pertanto, la SeqOne Platform di per sé non richiede una dimostrazione delle prestazioni cliniche.

5. Precauzioni

5.1. Uso conforme

(Vedi sezione **Utenti previsti**)

È vietato qualsiasi uso improprio della piattaforma.

Gli utenti sono tenuti a verificare che la piattaforma sia sempre operativa e a garantire le condizioni necessarie per il corretto funzionamento della stessa.

5.2. Avvertenze

 I risultati forniti non devono costituire l'unica base per modificare la cura del paziente. Di conseguenza, gli utenti devono prestare la dovuta attenzione nell'utilizzo del software e dei suoi risultati. Le diagnosi devono essere confermate attraverso vari esami clinici e medici.

L'indagine sulle malattie tramite NGS può rivelare risultati secondari o incidentali, come la predisposizione al cancro, da geni non correlati alla condizione primaria (Kuhlen e Borkhardt 2015). Nonostante le dichiarazioni politiche di agenzie come AMP e ACMG (Green et al. 2013; Allyse e Michie 2013; Consiglio di Amministrazione dell'ACMG 2015; Hegde et al. 2015), le direttive nazionali sulla segnalazione dei risultati incidentali variano. Pertanto, è fondamentale ottenere un documento di "consenso informato" (Anderson et al. 2016). Il medico curante ha la responsabilità di assicurarsi che il paziente o il suo rappresentante legale comprenda i potenziali esiti e rischi prima di prescrivere qualsiasi test NGS.

Questo prodotto non garantisce il rilevamento di tutte le varianti genetiche presenti in un campione. La probabilità di rilevare una variante di interesse dipende intrinsecamente dalla qualità e dalle caratteristiche del campione in esame (ad esempio, integrità del campione, preparazione della libreria, profondità e uniformità del sequenziamento, filtraggio bioinformatico e tipo o contesto della variante). Un risultato negativo non esclude definitivamente la presenza di una variante patogena se la sua frequenza (VAF) è inferiore alle soglie di rilevamento del modulo analitico utilizzato. Non è garantito il rilevamento affidabile di riarrangiamenti genomici su larga scala (varianti strutturali, incluse CNVs e delezioni di interi cromosomi). Ciò è dovuto alla limitazione intrinseca dei dati di sequenziamento di nuova generazione (NGS) a lettura breve, che rende difficile l'identificazione di questi eventi. Poiché le letture NGS sono brevi (circa 150-200 bp), questo vincolo è una caratteristica della tecnologia di sequenziamento stessa e non costituisce una limitazione della SeqOne Platform. (Vedi Tan R. et al., Hum Mutat. Luglio 2014; 35(7):899-907. doi: 10.1002/humu.22537 e riferimenti ivi contenuti). Gli utenti devono valutare l'idoneità del workset selezionato e la qualità del sequenziamento rispetto alle varianti previste per garantire che sia stato coperto il contesto clinico appropriato.

Le decisioni cliniche basate sui risultati dello screening neonatale o del Carrier screening devono essere prese con estrema cautela e non devono mai basarsi esclusivamente sui risultati di questo software. Per lo screening neonatale, un risultato positivo richiede un'immediata correlazione clinica con i sintomi del neonato e test di conferma urgenti prima di avviare trattamenti che alterano la vita o restrizioni dietetiche. Per lo screening dei portatori, un risultato negativo non elimina il rischio di essere portatori, ma lo riduce piuttosto a un livello di rischio residuo, poiché alcune varianti genomiche potrebbero rientrare al di sotto del limite di rilevanza del software. Poiché i risultati di questo dispositivo influenzano direttamente scelte riproduttive irreversibili e cure neonatali critiche, tutti i risultati che richiedono un intervento devono essere verificati utilizzando un metodo alternativo convalidato e interpretati da un professionista qualificato nel contesto della storia clinica completa del paziente e dei dati biochimici o fenotipici rilevanti.

Le varianti di recente scoperta che non sono ancora state documentate in modo approfondito nella letteratura scientifica potrebbero essere classificate in modo errato.

I risultati sono accurati al momento dell'analisi e riflettono le attuali conoscenze scientifiche. Tuttavia, l'esecuzione dell'analisi su una versione più recente della piattaforma potrebbe produrre risultati diversi.

I risultati della classificazione della SeqOne Platform mirano ad allinearsi al consenso comune. Si tenga presente che qualsiasi risultato generato dopo la modifica delle impostazioni di classificazione esula dall'ambito CE-IVD e non è quindi convalidato.

Questo software incorpora componenti di intelligenza artificiale (AI). Gli utenti devono essere consapevoli delle seguenti limitazioni inerenti a questi strumenti basati sull'AI e tenerne conto nell'interpretazione dei risultati:

- I dati di addestramento derivano prevalentemente da popolazioni di origine europea/caucasica a causa della composizione delle attuali banche dati cliniche di riferimento;
- Le prestazioni possono variare per le varianti rare con rappresentazione limitata nei dati di addestramento;
- I punteggi di confidenza riflettono l'incertezza del modello, non la certezza clinica; devono essere interpretati in combinazione con il contesto clinico;
- Le prestazioni potrebbero peggiorare a causa della deriva dei dati proveniente da nuove piattaforme di sequenziamento o protocolli di preparazione delle librerie non rappresentati nei dati di addestramento;
- Le prestazioni di prioritizzazione genica di DiagAI Phenogenius sono valutate sulle associazioni gene-sintomo e dipendono dalla qualità e dalla completezza dell'annotazione dei termini HPO da parte del medico che effettua l'ordine.

5.3. Limiti

Lo screening dei portatori per le condizioni genetiche è soggetto a specifiche leggi nazionali e normative etiche. È responsabilità del laboratorio e dell'operatore sanitario garantire che tutti i requisiti legali, tra cui la consulenza genetica obbligatoria, la richiesta da parte di un medico e il consenso informato, siano soddisfatti all'interno della loro specifica giurisdizione.

5.4. Malfunzionamenti

In caso di malfunzionamento:

- Interrompere immediatamente l'utilizzo del software
- Cercare di identificare o eliminare le cause seguendo la procedura di risoluzione dei problemi descritta in questa guida (vedere la sezione **Risoluzione dei problemi e messaggi di errore**)
- Se non si riesce a identificare o eliminare la causa utilizzando questa guida, inviare un'email al team di assistenza all'indirizzo support@seqone.com.

6. Requisiti minimi

6.1. Hardware

I moduli software della piattaforma SeqOne sono installati come parte di un sistema client-server. Il fornitore del servizio deve essere un server Linux per garantire la compatibilità con la maggior parte degli strumenti comunemente utilizzati in bioinformatica. Ogni elemento del sistema è incapsulato in un ambiente di runtime come il container Docker®, che può essere distribuito su qualsiasi piattaforma Docker® conforme, indipendentemente dall'architettura hardware sottostante. Gli utenti devono disporre di un computer con una connessione a Internet e di un browser web (vedi sotto).

I requisiti minimi di sistema per un computer client sono i seguenti:

- 4 GB RAM
- Processore dual-core
- Accesso a Internet
- Browser Chrome (dalla versione 36.0.1985 in poi) o Firefox (dalla versione 31 in poi) o Microsoft Edge (tutte le versioni basate su Chromium).

6.2. Rete

La velocità minima consigliata di upload/download per l'utilizzo della piattaforma su Internet è di 100 Mbps. Per la connessione alla piattaforma si utilizza il protocollo HTTPS. Per accedere correttamente alla piattaforma, si consiglia di chiedere al proprio reparto IT di autorizzare i seguenti indirizzi:

*.eu.seqone.com

*.common.seqone.com

7. Autenticazione

7.1. Accesso

Per motivi di sicurezza, gli utenti possono accedere alla piattaforma solo dopo aver effettuato l'accesso nella schermata iniziale.

La schermata iniziale è disponibile tramite il seguente URL:

- Per l'Europa e l'Africa: <https://auth.eu.seqone.com/>
- Per la Svizzera: <https://auth.ch.seqone.com/>
- Per l'Asia e l'Oceania: <https://auth.au.seqone.com/>
- Per le Americhe: <https://auth.us.seqone.com/>

Quando si accede alla schermata iniziale, è necessario inserire il proprio nome:

- Accesso o indirizzo e-mail nel campo "Username" o "email".
- Password nel campo "Password".

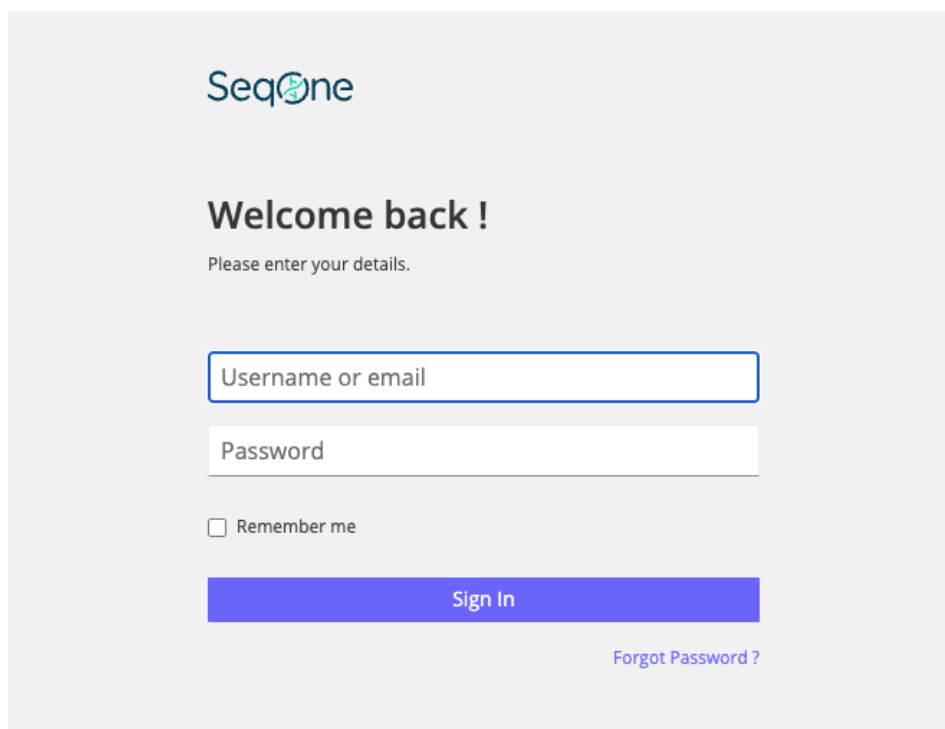


Figura 1: Pagina di accesso

Se l'autenticazione a due fattori è attivata per il vostro account, dovrete inserire un codice di verifica a 6 cifre inviato via e-mail all'indirizzo associato al vostro account SeqOne. Questo codice è valido per 30 minuti.

Se si ha accesso a più workspace o entità, verrà chiesto di selezionare l'entità a cui collegarsi tramite un menu a discesa in "Choose a workspace".

È possibile modificare la password facendo clic su "Forgot Password". Seguire le istruzioni inserendo l'indirizzo e-mail associato al proprio account SeqOne e si riceverà un'e-mail che consentirà di reimpostare la password.

Il criterio delle password segue le seguenti regole:

- Lunghezza minima: 12 caratteri.
- Complessità: le password devono includere almeno una lettera maiuscola, una lettera minuscola, un numero e un carattere speciale. Più ampio è il set di caratteri, più forte è la password.
- Cronologia delle password: nessun riutilizzo di password precedenti.
- Nessuna lunghezza massima: per consentire l'uso di frasi segrete.

Gli utenti non possono modificare il proprio ID utente.

7.2. Disconnessione

Gli utenti possono uscire cliccando sul logo dell'account utente nell'angolo in alto a destra della pagina e poi su "Logout".

7.3. Pagina iniziale

Dopo l'autenticazione, l'utente accede alla pagina del dashboard, suddivisa in due parti:

- un banner blu di ricerca e azione, accessibile in tutta la navigazione, da sinistra a destra:
 - l'icona SeqOne per tornare al dashboard in qualsiasi momento,
 - una barra di ricerca per accedere direttamente a una determinata analisi,
- l'icona VKB search  per la ricerca di varianti valutate nel database VKB (vedere la sezione **Valutazione delle varianti**),
- un'icona  che apre la "To do list" dell'utente, da cui può accedere alle analisi che gli sono state assegnate (vedere la sezione **Assegnazione di analisi per l'interpretazione**),
- una barra dei menu, disponibile tramite l'icona , che dà accesso ai parametri dell'utente, al pannello dell'amministratore, al manuale dell'utente, alle condizioni d'uso, all'indirizzo e-mail dell'assistenza e che consente all'utente di effettuare il logout (vedere la sezione **Gestione dell'account utente**).
- la schermata principale per la gestione dei progetti e delle analisi.

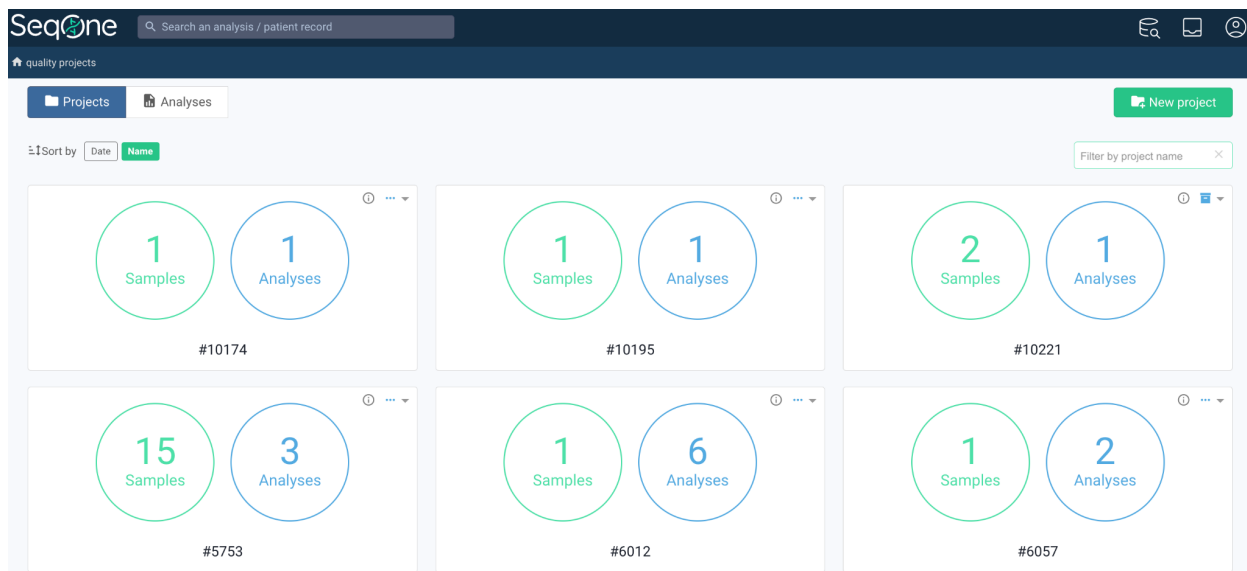


Figura 2: Dashboard" dell'utente

È possibile accedere al cruscotto utente in qualsiasi momento facendo clic sul logo SeqOne .

7.4. Gestione dell'account utente

È possibile gestire il proprio account utente cliccando sul logo nell'angolo in alto a destra della barra delle applicazioni.

In questa sezione, l'utente ha la possibilità di:

- accedere alle impostazioni dell'account cliccando su "My account"
- accedere ai parametri utente cliccando su "Entity Settings"
- cambiare l'entità a cui l'utente è connesso cliccando su "Entity" e selezionando l'entità desiderata sotto "Choose your workspace",
- accedere al pannello di amministrazione cliccando su "Admin", consentendo l'accesso all'elenco degli utenti con accesso all'entità e la creazione, modifica o cancellazione degli account utente.
- cliccare su "About" per accedere al manuale d'uso in varie lingue, alle informazioni sulla versione del software SeqOne e ai dati di contatto,
- accedere ai termini della licenza SeqOne cliccando su "Terms of Use",
- cliccare su "Support" per accedere all'indirizzo e-mail di supporto per qualsiasi domanda o problema riscontrato sulla piattaforma,
- accedere alla scorciatoia cliccando su "Keyboard Shortcuts"

- passare dalla versione standard alla versione preview dell'applicazione cliccando su "SeqOne Mode" e selezionando "SeqOne" (standard mode) o "SeqOne Preview" (preview mode). La modalità attualmente attiva è indicata da un segno di spunta.
- effettuare il logout cliccando su "Logout".

7.4.1. Gestione di "My Account"

Nella sezione "My account", l'utente può:

- caricare la propria firma per includerla nel referto clinico,
- cambiare la propria password cliccando su "Change your password" (apre una nuova scheda) nella sezione "Manage your credentials".

7.4.2. Gestione delle impostazioni utente - definizioni

Nelle impostazioni utente, l'utente può gestire le impostazioni, i file e le configurazioni descritti di seguito.

Entity

L'entità si riferisce alla struttura o al team che ha sottoscritto una licenza per la piattaforma SeqOne. La piattaforma è progettata per essere utilizzata da più membri di uno stesso gruppo. Dalle impostazioni dell'utente, la scheda *Account* è utilizzata per gestire le informazioni relative all'entità.

Users

Per questa scheda, fare riferimento alla **Gestione "Admin" - presentazione dell'interfaccia**.

Worksets

Questa scheda elenca i diversi worksets di analisi disponibili sull'entità a cui l'utente è collegato. Poiché sulla piattaforma SeqOne possono coesistere diverse versioni di determinate pipeline, l'utente può definire da questo menu quale versione deve essere preselezionata all'avvio dell'analisi.

Le modifiche apportate a questo elenco sono tracciate, in modo da visualizzare automaticamente l'utente che ha definito la versione da lanciare per impostazione predefinita di un workset, nonché la data della modifica.

Manifest

Il manifest visualizza le regioni genomiche di interesse (ROI) in un file BED, che può essere accompagnato dalle annotazioni associate. È possibile caricare un nuovo manifest tramite la scheda *Manifest* nelle impostazioni utente, facendo clic su "Add manifest" e selezionando il file manifest che si desidera caricare. Una volta caricato, questo manifest sarà disponibile per tutti gli utenti quando si crea un nuovo progetto e quando si applica un panel 'in silico' in un'analisi.

Panels - Transcript

Gli utenti possono caricare un nuovo elenco di transcripts di riferimento (RefSeq) tramite la scheda *Transcript*.

Panels - Gene

Questa scheda consente di importare o generare pannello di geni (Gene Panel) basati su NCBI GeneID facendo clic su "Add gene list".

Il formato del file previsto è un file tabulare (ad esempio un file .tsv) senza intestazioni di colonna, che è dettagliato sulla piattaforma. La prima colonna del file previsto è dedicata agli identificatori dei geni (NCBI Entrez gene ID), la seconda colonna (opzionale) contiene il transcript di interesse (RefSeq ID) e la terza colonna (opzionale) contiene il simbolo 'Gene & Transcript' del gene. Questo file può essere generato anche inserendo l'elenco desiderato di geni, separati da una virgola, nel campo "Gene symbols".

Marker

Questa scheda viene utilizzata per importare un elenco di marcatori di identità genetica facendo clic su "Add marker config". Il formato del file previsto è un file tabulare (file .tsv o .txt) con un'intestazione contenente 5 colonne che specificano: cromosoma (CHR), identificatore della variante o rsID (ID), posizione genomica (POS), allele di riferimento (REF) e allele alternativo (ALT). Il file non deve contenere più di 100 righe, compresa l'intestazione.

Protocols

Un protocollo di analisi è una combinazione predefinita di attività progettata per standardizzare l'interpretazione delle analisi. La relativa scheda consente agli utenti di visualizzare l'elenco dei protocolli disponibili sull'entità, le loro versioni e di crearli.

Per creare un protocollo, fare clic su "New protocol", quindi aggiungere e ordinare le attività che corrispondono alle SOP del laboratorio. Per modificare un protocollo esistente, fare clic sull'icona della matita accanto ad esso e salvare la nuova versione del protocollo.

Selezionare un protocollo al lancio dell'analisi o allegarlo successivamente dalla scheda Workflow nel riquadro di analisi. Contrassegnare le attività come completate direttamente dal riquadro di analisi. Ogni attività completata registra automaticamente l'utente e la data per una completa tracciabilità.

Reporting - Global settings

La sezione Global settings della scheda Reporting consente agli utenti di abilitare o disabilitare il blocco automatico dell'analisi quando lo stato passa a "Reporting Validating" e di personalizzare le sezioni di classificazione e reporting. Per farlo, sotto Germline / Somatic, fare clic su "New section", inserire un nome per la sezione e selezionare un tag da associare alla classificazione desiderata. Le varianti che corrispondono a uno dei tag o delle classificazioni selezionati vengono aggiunte automaticamente a questa sezione. È necessario prima creare i tag nella scheda corrispondente per utilizzare questa funzione.

Reporting - Template list

La scheda Reporting sotto "Template list" consente agli utenti di configurare i modelli utilizzati per l'esportazione dei risultati delle analisi. Da questa scheda è possibile personalizzare le informazioni incluse nei report generati.

Tags

La scheda "Tags" consente agli utenti di gestire etichette personalizzate per organizzare e filtrare le varianti all'interno della VKB e del report. Gli utenti possono creare tag cliccando su "Add tag", oppure rinominare ed eliminare i tag disponibili per tutti gli utenti all'interno dell'entità utilizzando le corrispondenti icone della penna e del cestino.

API Keys

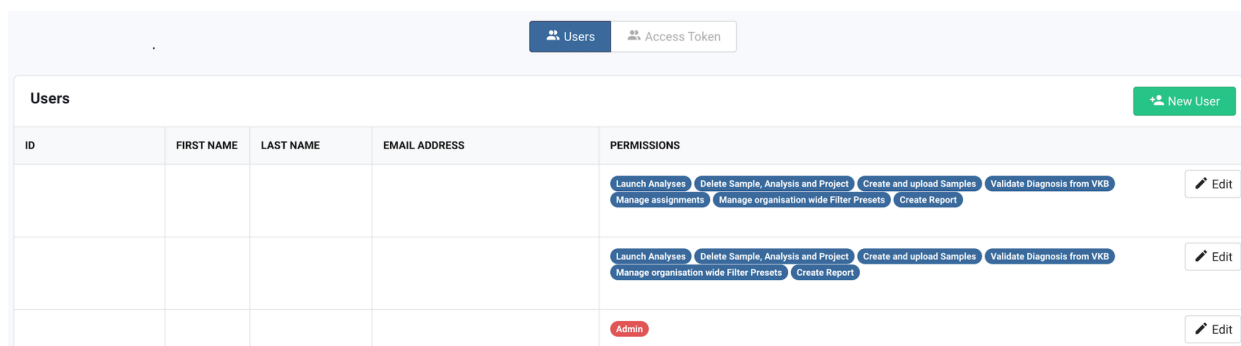
Una chiave API è un token di autenticazione univoco che consente ad applicazioni o script esterni di interagire con la piattaforma SeqOne in modo programmatico. Dalla scheda API Keys, è possibile visualizzare tutte le chiavi esistenti insieme alle informazioni associate.

Creare una nuova chiave cliccando su "+ New API Key". Le chiavi scadute possono essere visualizzate o nascoste utilizzando l'interruttore "Show expired key". Ogni chiave può essere eliminata singolarmente. Le chiavi API devono essere mantenute riservate, poiché garantiscono l'accesso alla piattaforma con i relativi permessi associati.

7.4.3. Gestione "Admin" - presentazione dell'interfaccia

Questa interfaccia è dedicata agli amministratori degli account e consente loro di :

- accedere all'elenco degli utenti dell'entità,
- creare nuovi utenti,
- modificare i diritti degli utenti: creare progetti e caricare campioni, lanciare analisi, cancellare campioni, analisi o progetti, valutare le varianti in VKB, creare report, filtrare la gestione dei profili, gestire gli incarichi.
- attivare la doppia autenticazione,
- disattivare un account utente.



The screenshot shows the 'Users' management interface. At the top, there are tabs for 'Users' and 'Access Token'. Below the tabs is a 'Users' section with a 'New User' button. The main part of the interface is a table with columns: ID, FIRST NAME, LAST NAME, EMAIL ADDRESS, and PERMISSIONS. The table contains three rows. The first two rows show users with various permissions like 'Launch Analyses', 'Delete Sample, Analysis and Project', 'Create and upload Samples', 'Validate Diagnosis from VKB', 'Manage assignments', 'Manage organisation wide Filter Presets', and 'Create Report'. The third row shows an 'Admin' user with full permissions. Each row has an 'Edit' button.

ID	FIRST NAME	LAST NAME	EMAIL ADDRESS	PERMISSIONS	
				Launch Analyses Delete Sample, Analysis and Project Create and upload Samples Validate Diagnosis from VKB Manage assignments Manage organisation wide Filter Presets Create Report	Edit
				Launch Analyses Delete Sample, Analysis and Project Create and upload Samples Validate Diagnosis from VKB Manage organisation wide Filter Presets Create Report	Edit
				Admin	Edit

Figura 3: Pannello dell'amministratore

L'ID corrisponde al nome utente dell'utente. Deve essere in minuscolo e non deve contenere accenti o spazi; sono accettati caratteri speciali come il punto e il trattino basso. È consigliabile seguire la nomenclatura "firstname.lastname" quando si crea un nuovo utente.

Per quanto riguarda l'indirizzo e-mail dell'utente, sconsigliamo vivamente di utilizzare un indirizzo e-mail generico per motivi di sicurezza.

Poiché le autorizzazioni sono disattivate per impostazione predefinita, spetta all'amministratore attivarle manualmente:

Permissions

Admin privileges

Grant full access to all system features, users management and settings



Create and upload Samples



Create Report



Delete Sample, Analysis and Project



Edit restricted panels



Launch Analyses



Manage assignments



Manage organisation wide Filter Presets



Manage VKB Knowledge Base



Protocols Management



Report template



Unlock Analyses



Update project settings (lock filters for now)



Validate Diagnosis from VKB



Advanced

Activate 2FA

Enable email two-factor authentication for enhanced security



Temporarily suspending the access. The data is retained, but the user cannot log in or use the system until re-enabled.

 Disable User

Figura 4: Autorizzazioni per gli utenti

È inoltre possibile aumentare la sicurezza attivando la doppia autenticazione. Questa opzione invia un codice unico all'utente ogni volta che si connette alla piattaforma.

Il ruolo "Manage organization wide Filter Presets" consente all'utente di modificare i profili di filtro pubblici. Se non si dispone di questo ruolo, è possibile modificare solo i profili di filtro privati.

I permessi sono gestiti da un amministratore dell'entità dal pannello Admin. Quando viene creato un utente, tutti i permessi sono disabilitati per impostazione predefinita e l'amministratore deve abilitare manualmente i diritti richiesti per quell'utente. I diritti assegnati a un utente si applicano a tutte le entità (spazi di lavoro) a cui l'utente ha accesso sulla piattaforma SeqOne (ad esempio, se l'utente può accedere a più entità, lo stesso set di permessi si applica a tutte).

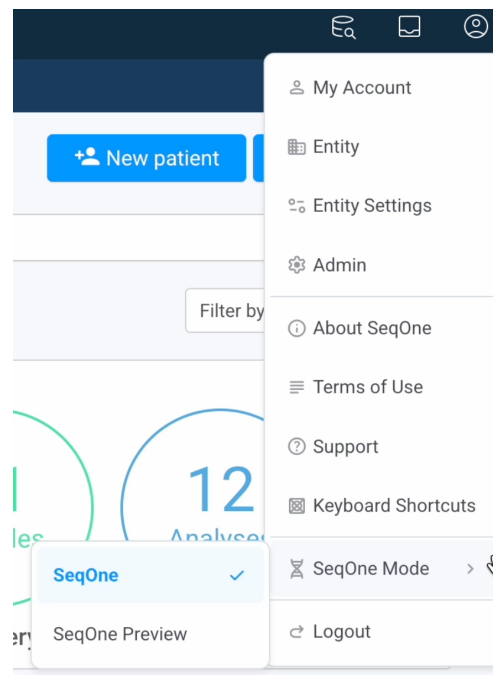
Se si riceve un messaggio di errore, contattare l'assistenza clienti all'indirizzo support@seqone.com.

7.4.4. Passaggio tra le modalità Main e Preview

La connessione a SeqOne tramite l'URL di accesso predefinito porta gli utenti alla modalità "Main" della loro entità. Questa modalità predefinita ospita workset con marchio CE e altre certificazioni locali.

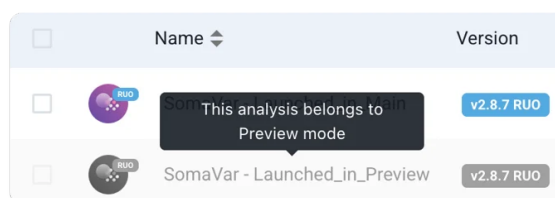
Per l'accesso anticipato a nuove funzionalità e versioni dei workset, passa alla modalità "Preview" (RUO: Research Use Only) andando in fondo al menu delle impostazioni e, sotto SeqOne One, selezionando "Preview".

Le modalità "Main" e "Preview" condividono gli stessi dati principali (progetti, campioni, analisi), permessi, risorse (file, template, manifest, liste di geni), conoscenza (VKB) e frequenze (dati di frequenza dei campioni, occorrenze).



Durante la navigazione in modalità "Preview", viene visualizzato il logo "SeqOne Preview RUO" per ricordare che questa modalità contiene funzionalità e workset RUO.

In modalità "Main", le analisi avviate in "Preview" appaiono in grigio. Se fai clic su una di queste analisi in grigio, la piattaforma ti chiede di passare alla modalità Preview per visualizzarla.



8. Descrizione delle fasi operative

Passo	Descrizione
1	Creazione del progetto Le analisi condotte sulla piattaforma SeqOne sono organizzate all'interno di progetti, che definiscono le caratteristiche dei campioni analizzati attraverso metadati inseriti dall'utente (tipo di sequenziamento, metodo di preparazione, geni di interesse).
2	Sample Upload I campioni vengono caricati come file di sequenziamento grezzi (FASTQ), che devono essere inseriti nel progetto designato. Una volta caricati i file, una fase di verifica assicura l'integrità dei file inviati prima di qualsiasi analisi.
3	L'avvio dell'analisi Una volta caricati i dati sui server di SeqOne, l'utente può scegliere tra una serie di applicazioni o "worksets", a seconda del tipo di campione analizzato e dei tipi di alterazioni genomiche ricercate (SNV, CNV, fusioni geniche, varianti di splicing, ecc.). Queste applicazioni eseguono analisi bioinformatiche.
4	Controllo della qualità dei dati Le applicazioni disponibili sulla piattaforma SeqOne includono diversi moduli dedicati all'analisi della qualità dei dati, utilizzando varie metriche di sequenziamento, allineamento e profondità di copertura.
5	Interpretazione dei risultati dell'analisi Le variazioni genomiche rilevate nel campione vengono visualizzate in varie schede o "Samples", dove l'utente può cercare la variante diagnostica responsabile della patologia del paziente.
6	Ottenere raccomandazioni terapeutiche e studi clinici associati alla patologia del paziente.
7	Rilascio di un rapporto I risultati dell'analisi possono essere esportati come report in formato PDF, che può essere scaricato.

9. Creazione del progetto

Le analisi condotte sulla piattaforma SeqOne sono organizzate all'interno di progetti, che definiscono le caratteristiche dei campioni analizzati attraverso metadati inseriti dall'utente (tipo di sequenziamento, metodo di preparazione, geni di interesse).

9.1. Creare un progetto

Dopo aver fatto clic su "New project", viene richiesto di inserire i valori di diversi tipi di informazioni, o metadati, utilizzati per descrivere il progetto e i campioni associati.

- **Descrizione del progetto**
- **Sequenziamento:**
 - Assay type: tipo di analisi (Germline o Somatic),
 - I progetti Germline (*Assay type Germline*) presentano una casella di controllo UMI Trimming, che consente di elaborare i campioni contenenti UMI. Selezionando questa opzione, gli UMI verranno tagliati dalle sequenze FASTQ, come nella modalità UMI disabled per i workset somatici (vedere la sezione **Scelta della modalità di elaborazione UMI**).
 - Sequencing platform: la piattaforma di sequenziamento
 - Library type: tipo di sequenziamento delle read eseguito (*Paired-end* o *Single-end*),
 - Data type: DNA, RNA, ctDNA o Methylation Sequencing
 - Se si seleziona DNA per il tipo di dati, è necessario specificare il metodo di sequenziamento tramite il menu a discesa "*DNA Sequencing method*", nonché il metodo di amplificazione (Capture o Amplicon) tramite "*Amplification method*".
 - File input type: *File Fastq, VCF files*".
 - Genome: la versione del genoma di riferimento (GRCh37 o GRCh38),
 - Low-Quality Mapping Region: la chiamata di varianti in regioni di allineamento di bassa qualità considerando le reads MAPQ0,
 - Sequencing method: Gene Panel, Whole exome sequencing, Comprehensive Genomic Profiling, Shallow WGS, Whole genome sequencing, CGH MicroArrays,
 - Secondary analysis provider: AmoyDX, Pillar, Thermofisher,
 - Manifest: il manifesto nel caso di un pannello di geni,
 - ID Marker (opzionale): il file del marcatore di identità.
- **Campioni:**
 - (*Facoltativo*) È possibile aggiungere nuovi metadati del campione facendo clic su "*You can also create your own metadata.*" Quando viene richiesto, inserire il nome e il valore del metadato.

Una volta inseriti i diversi metadati per il progetto, gli utenti sono invitati a creare il loro progetto facendo clic sul pulsante "Create project".

9.2. Gestione dei progetti

Attraverso la dashboard, gli utenti possono gestire i diversi progetti creati sulla piattaforma. In particolare, possono:

- Accedere ai progetti tramite la scheda "Projects", nonché ai campioni e alle analisi per ciascun progetto facendo clic rispettivamente sui loghi "Samples" e "Analysis".
- Accedere alle analisi direttamente tramite la scheda "Analysis".
- Creare un nuovo progetto facendo clic su "Create project" (vedere la sezione **Creare un progetto**).

9.3. Eliminazione di un progetto

Per eliminare un progetto, gli utenti devono:

- Selezionare il progetto
- Fare clic sulla scheda "Settings" e poi su "Delete project".
- Confermare selezionando "Yes" nel messaggio visualizzato. Questa azione è irreversibile.

9.4. Dati sul DNA (sequenziamento del DNA, sequenziamento del ctDNA)

Low-Quality Mapping Region. Quando si sequenziano pannelli di geni, esomi o genomi, l'utente può includere regioni di allineamento di bassa qualità quando chiama le varianti. Queste regioni corrispondono alle cosiddette "MAPQ0" reads, la cui qualità di allineamento è uguale a 0, come gli pseudogeni e le regioni omologhe del genoma.

L'opzione è applicabile a specifici geni o regioni target. Per utilizzare questa opzione, l'utente è invitato a caricare il manifesto che ha come target le regioni desiderate o a selezionarlo tramite un menu a discesa.

Questa opzione può ridurre la qualità e l'accuratezza dei risultati. Può introdurre rumore di fondo nella frequenza allelica e avere un impatto negativo sulla prioritizzazione delle varianti. Può anche aumentare il tempo di analisi a causa di un maggiore carico computazionale.

Exome. Nel caso del sequenziamento dell'esoma, gli utenti non devono specificare il manifesto da utilizzare; sono invitati a selezionare il produttore del kit utilizzato facendo clic su una delle icone nella categoria Exome kit provider. Un menu a discesa consente all'utente di selezionare il manifesto desiderato. Se il manifesto desiderato non è disponibile sulla piattaforma, l'utente è invitato a contattare il supporto (support@seqone.com).

Gene panel. Nel caso di un pannello di geni, agli utenti viene richiesto di selezionare il *metodo di amplificazione* e il manifesto (in formato BED) corrispondente ai geni di interesse in un menu a discesa. Se il manifest richiesto non è disponibile sulla piattaforma, può essere caricato selezionando "Upload new manifest" e seguendo le istruzioni sullo schermo.

Shallow WGS, CGH MicroArrays, Whole genome sequencing. Nel caso del sequenziamento del genoma, l'utente non deve specificare quale manifesto utilizzare.

Capture. Scegliendo il metodo di amplificazione *capture*, è disponibile l'icona UMI (*Unique Molecular Identifiers*). Gli identificatori molecolari unici sono indici aggiunti prima dell'amplificazione per aiutare a eliminare i duplicati della PCR e ridurre il rumore di fondo dopo l'analisi.

Amplicon. Quando si analizzano gli ampliconi, occorre prestare particolare attenzione ai primer di amplificazione. Quando si crea un progetto per i dati di ampliconi, si consiglia agli utenti di selezionare un BED in cui le regioni corrispondenti ai target non includano le sequenze dei primer. Se tale BED non è disponibile, è possibile selezionare un'opzione per *tagliare* una lunghezza fissa alla fine delle sequenze.

Identity monitoring. Per i metodi di sequenziamento corrispondenti a gene panel, exome, genomic profiling (CGP) o genome (shallow WGS) da file FASTQ, l'utente è invitato a selezionare il file dei marcatori di Identity Monitoring da un menu a discesa. Se il file desiderato non è disponibile sulla piattaforma, può essere importato selezionando 'Upload new markers' e seguendo le istruzioni sullo schermo (vedere anche il paragrafo **Marker** nella sezione **Gestione delle impostazioni utente**).

9.5. Dati sull'RNA (sequenziamento dell'RNA)

Se si seleziona RNA per il tipo di dati, gli utenti devono specificare se il protocollo mantiene le informazioni sul filamento originale tramite il menu a discesa *Stranded protocol*. Se il protocollo è specifico per il filamento, è necessario specificare l'ordine di sintesi dei filamenti (*Forward-reverse* o *reverse-forward*) utilizzando l'opzione *Read orientation*. Infine, è necessario specificare il metodo di sequenziamento (gene panel).

Read orientation. Sono disponibili due opzioni per specificare l'orientamento della sequenza:

- **Forward-Reverse:** la prima lettura (R1) per ogni coppia corrisponde alla sequenza dell'RNA messaggero trascritto (filamento senso).
- **Reverse-Forward:** la prima lettura (R1) per ogni coppia corrisponde alla sequenza complementare dell'RNA messaggero trascritto (filamento antisenso).

In generale, l'orientamento dei read è il seguente per i kit menzionati di seguito:

- Agilent: Reverse-Forward
- Illumina: Unstranded
- Roche Kapa Hypercap: Reverse-Forward
- Twist: Reverse-Forward

In caso di dubbi, contattare l'assistenza (support@segone.com).

Una volta specificato il tipo di dati e configurato correttamente il progetto, agli utenti viene richiesto di convalidare facendo clic sul pulsante "Next step".

9.6. Impostazioni del progetto

Dalla scheda "Settings" del progetto, è possibile :

- modificare il nome e la descrizione del progetto nei campi "*Project name*" e "*Project description*",
- modificare il profilo del controllo qualità (QC),
- modificare il profilo di filtro predefinito nella tabella delle varianti,
- cancellare definitivamente il progetto.

9.6.1. Configurare le metriche predefinite del progetto QC

L'utente può scegliere di utilizzare le metriche QC predefinite o di definire valori soglia personalizzati dalla scheda "Settings" del progetto. L'utente è invitato a contattare l'assistenza tecnica (support@seqone.com) per definire i valori desiderati.

Verranno applicate metriche QC personalizzate :

- *QC checks*' nella vista *Analyses* del progetto,
- *Controlli QC* nella sezione "*QC auto-validation*" della scheda QC dell'analisi,
- alla scheda "*GeneCov*" dell'analisi.

9.6.2. Impostazione del profilo di filtro predefinito del progetto

L'utente può scegliere di impostare il profilo di filtro predefinito applicato alla tabella delle varianti predefinita dalla scheda "Settings" del progetto (vedere anche la sezione **Utilizzo di un profilo di filtro predefinito**).

Per ogni workset è possibile selezionare un profilo di filtro tramite un menu a discesa, purché sia pubblico. Il profilo di filtro sarà quindi applicato per impostazione predefinita quando si visualizza la tabella delle varianti, per tutte le analisi di progetto, per tutti gli utenti dell'organizzazione.

Se questo parametro non viene modificato dall'utente, alla creazione del progetto il valore predefinito sarà il profilo definito da SeqOne.

I profili filtro predefiniti possono essere bloccati cliccando sull'interruttore "*Lock Variants Filters*".

10. Campione di caricamento

10.1. Caricamento dei campioni

Gli utenti possono importare i loro campioni tramite l'icona "Add samples" in ogni progetto. I campioni possono essere importati nella piattaforma SeqOne in due modi:

Da un PC

Gli utenti possono caricare i propri file FastQ direttamente da una cartella del proprio PC facendo clic sulla scheda "From your computer" dall'icona "Add files". Gli utenti possono anche trascinare e rilasciare i file dal proprio PC all'area dedicata dell'interfaccia.

Tramite SDS (SeqOne Data Sync)

SDS è un modulo premium opzionale per la creazione automatica di progetti, il trasferimento di dati, l'avvio di analisi e il recupero di dati strutturati (TSV) da integrare nel sistema di gestione delle informazioni di laboratorio (LIMS).

Questo modulo compatibile con Linux, MacOS e Windows è un eseguibile che consente agli utenti di autenticarsi e interagire con le API di SeqOne in un ambiente predefinito e altamente sicuro.

Requisiti minimi:

- Linux 1.19 minimo
- MacOS 10.15 con CPU Intel minimo
- MacOS con CPU Apple Silicon (chip M1) minimo
- Windows 10 64 bit minimo
- 4 GB di RAM (8 GB per Windows)
- Processore quad-core
- Accesso a Internet e alla piattaforma SeqOne (app.eu.seqone.com)
- Accesso al sistema di archiviazione dei dati (direttamente o tramite rete)

Per utilizzare questa soluzione o per ulteriori informazioni, contattate il nostro team di supporto (support@seqone.com).

Sample upload status

La piattaforma SeqOne utilizza diversi status per consentire agli utenti di seguire l'avanzamento del caricamento dei campioni:

- *Upload pending*: importazione in corso,
- *Verifying sample...*: conformità del campione in corso di verifica,
- *Mapping...*: fase di allineamento dei campioni,
- *Verified sample*: campione verificato,
- *Upload failed*: importazione fallita.

Capacità di caricamento

La capacità di caricamento dei file (in particolare dei file fastq.gz) dipende direttamente dalla connessione Internet dell'utente. Pertanto, la velocità della connessione internet deve essere presa in considerazione quando si caricano i campioni. Per un'idea più chiara dei tempi di caricamento stimati, consultare la seguente tabella:

Estimation du temps d'attente				
	ADSL		Fibre optique (Mbits/s)	
	Ethernet : 10 Mbits/s	Wifi : 7,5 Mbits/s	Ethernet : 100 Mbits/s	Wifi : 90 Mbits/s
Poids du fichier fastq				
100Mo	1min 20s	1min 45s	8s	9s
1Go	13min 20s	17min 45s	1min 20s	1 min 30s
10Go	2h 13min 20s	2h 57mn	13min 20s	15 min
100Go	22h 22min	30h	2h 13min 20s	2h 30min

Figura 5: Tempo di caricamento stimato in base alle dimensioni del file e alla velocità della connessione Internet.

10.2. Formato del file di ingresso

10.2.1. File di analisi primaria

Il formato del file da caricare per ciascun campione dipende dalla piattaforma di sequenziamento utilizzata. I nomi dei file devono essere conformi alla convenzione di denominazione dei file di ciascun sequenziatore.

I file devono essere compressi con l'estensione .gz (gzip) prima di poter essere importati nella piattaforma. Nel caso del sequenziamento "Paired-end", sono previsti due file, corrispondenti alle reads di entrambe le estremità di ciascuna molecola.

Quando un campione è stato correttamente rilevato dalla piattaforma al momento dell'importazione, viene visualizzato il nome completo del file. È quindi possibile rinominare il file facendo clic su di esso. I file "Paired-end" vengono automaticamente combinati in un unico campione.

La piattaforma supporta i seguenti sequenziatori:

- [Illumina](#): formato fastq.gz, solo "Paired-end",
- [Element Biosciences](#) : formato fastq.gz, solo paired-end,

- [MGI](#) : formato fastq.gz, solo paired-end. I nomi dei file previsti sono: sample_1.fq.gz e sample_2.fq.gz o flowcellID_L01_sample_1.fq.gz e flowcellID_L01_sample_2.fq.gz

10.2.2. File di analisi secondaria

Il formato del file da caricare per ciascun campione dipende dalla piattaforma di sequenziamento e dall'applicazione utilizzata per l'analisi secondaria. I file devono essere compressi con l'estensione .gz (gzip) prima dell'importazione dalla piattaforma, se non diversamente specificato.

Il formato di importazione dei file di analisi secondaria generati con applicazioni Illumina, Oxford Nanopore Technologies, Agilent, AmoyDx, Pillar, Thermofisher dipende dal workset utilizzato (vedere la sezione **Worksets di analisi disponibili dai file di analisi secondaria**).

Quando la piattaforma rileva correttamente un campione durante l'importazione, visualizza il nome completo del file. È possibile rinominare il file facendo clic su di esso. I file che condividono lo stesso nome di base vengono automaticamente combinati in un unico campione.

[File DRAGEN](#). I file e i formati previsti per ciascun workset sono i seguenti, quando i file VCF - Secondary Analysis data e i metadati Illumina sono selezionati al momento della creazione del progetto:

GermVar Tertiary, GermVar Tertiary Family sample.hard-filtered.vcf.gz
sample.cnv.vcf.gz
sample.sv.vcf.gz
sample.repeats.vcf.gz

[File Oxford Nanopore Technologies](#). I file e i formati previsti sono i seguenti quando si selezionano i file VCF - Secondary Analysis data, e i metadati di Oxford Nanopore Technologies durante la creazione del progetto. Gli utenti possono importare un file di varianti (SNP) da solo o combinarlo con file supplementari per analisi CNV, STR o SV:

GermVar Tertiary, GermVar Tertiary Family sample.wf_snp.vcf.gz
sample.wf_cnv.vcf.gz
sample.wf_str.vcf.gz
sample.wf_sv.vcf.gz

Thermofisher/Affymetrix files. Il formato di importazione per i file di analisi secondaria CGH è il formato .txt generato dall'applicazione Affymetrix di Thermofisher. Questo formato è supportato solo quando durante la creazione del progetto vengono selezionati sia i metadati Thermofisher/Affymetrix che quelli CGH MicroArrays. I file e i formati previsti per ciascun workset sono i seguenti:

GermVar Tertiary ARRAY_SampleID.txt

File CGH MicroArrays. Il formato di importazione dei file di analisi secondaria CGH è il formato .xls generato dall'applicazione Cytogenomics di Agilent. Questo formato è supportato solo se durante la creazione del progetto vengono selezionati sia i metadati Agilent che quelli CGH MicroArrays. I file e i formati previsti per ciascun workset sono i seguenti:

GermVar Tertiary IntervalBasedReport_sample.xls

File di Alissa Reporter. Il formato di importazione dei file di analisi secondaria di Alissa Reporter è il formato .xls generato dall'applicazione Alissa Reporter di Agilent. Questo formato è supportato solo quando durante la creazione del progetto vengono selezionati i metadati Agilent e CGH MicroArrays. I file e i formati previsti per ciascun workset sono i seguenti:

GermVar Tertiary IntervalBasedReport_sample.xls
ABERRATION_&_LOH_INTERVAL_sample.xls

PacBio-HiFi files. Quando si crea un progetto, se vengono scelti i file VCF (Secondary Analysis data) e i metadati PacBio-HiFi, i file e i formati previsti per ogni workset sono i seguenti:

GermVar Tertiary, GermVar Tertiary Family sample.hard-filtered.vcf.gz
sample.cnv_sv.vcf.gz
sample.repeats.vcf.gz

File VCF per i pannelli di geni somatici. Il formato di importazione per i file SNV, CNV, Fusioni di analisi secondaria è il formato .vcf.gz generato dalle applicazioni Illumina, AmoyDx, Pillar, Thermofisher. I nomi dei file originali devono essere convertiti in modo da corrispondere al formato di denominazione richiesto descritto di seguito.

I file e i formati previsti sono i seguenti quando si selezionano i file VCF - Dati dell'analisi secondaria e metadati del "Gene Panel" durante la creazione del progetto somatico. Gli utenti possono importare un file di varianti (SNV) da solo o combinarlo con file supplementari per l'analisi di CNV o fusioni:

SomaVar Tertiary sample.snv.vcf.gz

sample.cnv.vcf.gz
sample.fus.vcf.gz

I file e i formati previsti sono i seguenti quando si selezionano i "VCF files - Secondary Analysis data" e "Comprehensive Genomic Profiling" metadati durante la creazione di un progetto somatico. Gli utenti possono importare un file multi-varianti VCF come il file VCF "All Variants" da Oncomine™ Comprehensive Assay V3/Plus.

SomaVar Tertiary Oncomine_<LibPrepID>_<analysisID>.vcf

10.3. Metadati campione

Per ogni campione caricato, gli utenti possono assegnare i valori ai metadati selezionati al momento della creazione del progetto o definire nuovi valori facendo clic sull'icona nell'angolo superiore destro dell'elenco dei campioni . 

Alcuni metadati del campione possono essere caricati direttamente come file di valori separati da tabelle selezionando l'icona  .

Questi criteri svolgono un ruolo importante per garantire la tracciabilità dei dati e per effettuare riferimenti incrociati tra le diverse analisi eseguite sulla piattaforma. Alcuni metadati, come sesso, malattia e ID HPO, sono coinvolti nel processo di identificazione delle varianti.

Quando il campo Assay type è compilato con "Germline", i metadati associati al campione per impostazione predefinita sono "HPO", "Sex" e gruppo "GnomAD". Quando il campo "Assay type" è compilato con "Somatic", i metadati associati al campione per impostazione predefinita sono "Disease" e "Sex".

Per aggiungere o modificare gli HPO in un campione, l'utente può cliccare su "+ Add" nella colonna HPO dopo aver selezionato un campione. Si apre una finestra pop-up che consente di aggiungere HPOs uno per uno, cercando un fenotipo o un codice HPO nella barra di ricerca. È

anche possibile importare un file di testo contenente i codici HPO facendo clic sull'icona  . È quindi disponibile un riepilogo dei termini inseriti, con la possibilità di eliminare i termini se necessario.

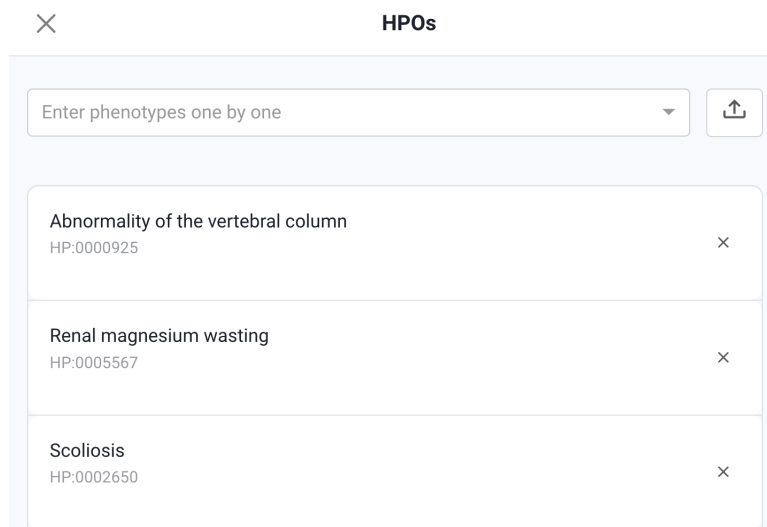


Figura 6: Finestra pop-up per l'importazione di HPO

10.4. Modifica di un campione

Il nome e i metadati associati a un campione possono essere modificati dall'utente in qualsiasi momento, selezionando il campione nell'elenco "Samples" e cliccando sull'icona . Le modifiche devono essere convalidate per essere efficaci.

10.5. Eliminare un campione

Gli utenti possono eliminare i file FastQ di un campione dalla piattaforma facendo clic sull'icona  dopo aver selezionato il campione desiderato. Questi file libereranno spazio di memoria e i risultati saranno ancora disponibili su SeqOne.

Per eliminare un campione, fare clic sull'icona  nella scheda "Samples" dopo aver selezionato il campione desiderato. Il campione può essere eliminato definitivamente solo dopo aver cancellato tutte le analisi ad esso associate.

11. Analisi

11.1. Avvio di una nuova analisi

Gli utenti possono lanciare una nuova analisi facendo clic su "New analysis" nella schermata Overview di ciascun progetto.

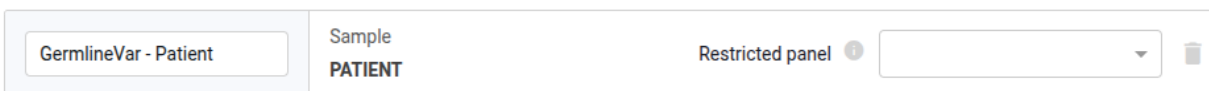
Il lancio dell'analisi avviene in tre fasi distinte, descritte di seguito.

Il passo "1. Workset" consente di scegliere il workset di analisi appropriato (vedere la sezione **Scelta del worksets d'analisi**).

La fase "2. Select Samples" è suddivisa in due parti:

- 2.1. *Impostare le opzioni*:
 - descrive il workset e il genoma di riferimento selezionato,
 - permette di aggiungere *padding* attraverso la funzionalità "Extra calling padding", se applicabile (vedere la sezione **Scelta del worksets d'analisi**),
 - consente di scegliere la modalità di elaborazione UMI (vedere la sezione **Scelta della modalità di elaborazione UMI**), se applicabile,
 - consente di scegliere la baseline di CNV, la baseline di MSI o di regolare la soglia minima VAF, se applicabile.
 - Consente di scegliere il protocollo.
- 2.2. *Select sample(s)*: consente di selezionare i campioni da analizzare. Gli utenti possono selezionare in modo specifico i nuovi campioni, contrassegnati come 'New', utilizzando la funzione 'Select All'.

Il passaggio "3. Confirm Analysis" riassume i campioni selezionati. Per ogni analisi da lanciare, è possibile applicare un pannello ristretto selezionando uno o più pannelli di geni tramite l'opzione "Restricted panel" (vedere anche la sezione **Vista modale Panel**).



L'opzione "Restricted panel" applica un filtro all'analisi sotto forma di uno o più pannelli di geni. Questa opzione non influisce sulla chiamata di varianti, che è determinata dal manifesto del progetto e dai *padding* aggiunti. La scelta di "Restricted panel" è reversibile; l'utente può modificare questi pannelli dalla scheda *Panel* del pannello laterale di gestione dell'analisi dopo l'analisi (vedere la sezione **Vista modale Panel**).

11.1.1. Scelta del worksets d'analisi

Sono disponibili diversi worksets di analisi per rispondere a diverse domande biologiche. I worksets proposti quando si crea una nuova analisi dipendono dai metadati definiti al momento della creazione del progetto e possono quindi variare a seconda dei dati.

Worksets di analisi disponibili da file FASTQ

I seguenti worksets sono disponibili quando i metadati *Fastq files - Primary Analysis data* sono selezionati durante la creazione del progetto.

Escludendo i progetti in cui il metodo di amplificazione è amplicone, alcuni worksets (cioè GermlineVar, GermlineFamily, SomaVar, SomaVar LF (*versioni* ≥ 2.0) e SomaCGP (*versione* ≥ 0.6), GermVar, GermVar Family, SomaLBx, SomaMethyl) consentono di estendere l'area di ricerca per la chiamata di varianti e di aggiungere tra 0 e 100 bp di padding al manifesto iniziale.

Extra calling padding (between 0 and 100bp)  Apply extra padding on the manifest for calling, this won't affect the computation of QC metrics.

L'opzione è disponibile nel passaggio 2 (Select Sample) del lancio dell'analisi tramite la funzione "Extra calling padding". Per impostazione predefinita, il valore è impostato su 0 e deve essere un numero intero naturale.

Alcuni workset (ad esempio SomaVar (*versioni* ≥ 2.6) e SomaHemato (*versioni* ≥ 0.2)) offrono un'opzione per regolare la soglia minima VAF per il rilevamento delle varianti. Questa soglia può essere impostata tra lo 0,01% e l'1%, con l'1% come valore predefinito.

Minimal VAF threshold (between 0.01% and 1%)  Apply a custom minimal VAF threshold for short variant detection. Sensitivity and precision cannot be guaranteed when deviating from the default pipeline value (1% for SomaVar)

Per ulteriori informazioni su questa opzione, contattare l'assistenza utenti all'indirizzo support@seqone.com.

GermlineVar



GermlineVar è utilizzato per rilevare mutazioni germline basandosi sul sequenziamento di un panel di geni o dell'esoma del paziente.

Quando si esegue un'analisi GermlineVar, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select",
- Selezionare i campioni da analizzare e fare clic su "Next step",
- Avviare l'analisi facendo clic su "Start analysis".

GermVar



GeneVar consente di rilevare mutazioni germline, sulla base del sequenziamento di un panel di geni, dell'esoma o del genoma di un paziente.

Quando si esegue un'analisi GermVar, gli utenti sono invitati a:

- Selezionare la versione del workset che si desidera eseguire, quindi fare clic su "Select",
- Selezionare il modello da utilizzare,
- Selezionare i campioni da analizzare e fare clic su "Next step",
- Avviare l'analisi facendo clic su "Start analysis".

GermlineVarLR



GermlineVarLR è utilizzato per rilevare mutazioni germline nei dati di sequenziamento long-read provenienti dal sequenziamento di un panel di geni.

Quando si esegue un'analisi GermlineVarLR, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select",
- Selezionare il modello da utilizzare e il/i campione/i da analizzare, quindi fare clic su "Next step".
- Avviare l'analisi facendo clic su "Start analysis".

GermlineFamily



GermlineFamily è utilizzato per rilevare mutazioni ereditarie sulla base del sequenziamento di un panel di geni, dell'esoma o del genoma di un paziente e dei membri della sua famiglia. Questo worksets può essere utilizzato anche per il "Carrier screening".

Quando si esegue un'analisi GermlineFamily, gli utenti devono:

- Selezionare la versione del workset da eseguire, quindi fare clic su "Select",
- Selezionare almeno 2 campioni.
- Nella sezione "Set conditions", indicare se il campione è del paziente o di un familiare definendo il legame familiare in "Family link". Il paziente viene automaticamente contrassegnato come "affected". In caso di "Carrier screening", indicare per ogni campione il legame familiare con "Mother" o "Father" senza selezionare la casella "Is affected?".
- Verificare che i parametri corrispondano, quindi fare clic su "Next step",
- Avviare l'analisi facendo clic su "Start analysis".

GermVar Family



GermVar Family consente di rilevare mutazioni germline, sulla base del sequenziamento di un pannello di geni, dell'esoma o del genoma di un paziente.

Quando si esegue un'analisi di GermVar Family, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire, quindi fare clic su "Select",
- Selezionare il modello da utilizzare,
- Selezionare almeno 2 campioni,
- Nella sezione "Set conditions", indicare se il campione è del paziente o di un familiare definendo il legame familiare in "Family link". Il paziente viene automaticamente contrassegnato come "affected". In caso di "Carrier screening", indicare per ogni campione il legame familiare con "Mother" o "Father" senza selezionare la casella "Is affected?".
- Verificare che i parametri corrispondano, quindi fare clic su "Next step",
- Avviare l'analisi facendo clic su "Start analysis".

CNVCapture



CNVCapture è utilizzato per rilevare variazioni del numero di copie (CNVs) a livello germinale dai dati di cattura del pannello di geni.

Quando si esegue un'analisi CNVCapture, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire, quindi fare clic su "Select",
- Selezionare un numero minimo di 8 campioni da analizzare, quindi fare clic su "Next step",
- Eseguire l'analisi facendo clic su "Start analysis".

CNVExome

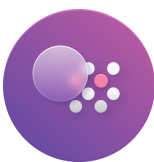


CNVExome è utilizzato per rilevare le variazioni del numero di copie (CNVs) dai dati di cattura dell'esoma.

Durante l'esecuzione di un'analisi CNVExome, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select".
- Selezionare il modello da utilizzare e il/i campione/i da analizzare, quindi fare clic su "Next step".
- Avviare l'analisi facendo clic su "Start analysis".

SomaVar



SomaVar rileva mutazioni somatiche da campioni di tumore. Questo workset può essere utilizzato con dati di cattura e ampliconi.

Quando si esegue un'analisi SomaVar, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select".
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**).
- Selezionare il/i campione/i da analizzare e cliccare su "Next step".
- Avviare l'analisi facendo clic su "Start analysis".

SomaCNVCapture



SomaCNVCapture è utilizzato per rilevare variazioni somatiche del numero di copie (CNVs) dai dati di cattura del pannello di geni.

Quando si esegue un'analisi "SomaCNVCapture", gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select".
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**).
- Selezionare un numero minimo di otto campioni da analizzare, quindi fare clic su "Next step".
- Eseguire l'analisi facendo clic su "Start analysis".

SomaRNA



SomaRNA può essere utilizzato per rilevare fusioni geniche dai dati di acquisizione RNA-seq.

Quando si esegue un'analisi di SomaRNA, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select".
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**).
- selezionare il/i campione/i da analizzare, quindi fare clic su "Next step",
- Eseguire l'analisi facendo clic su "Start analysis".

SomaMSI



SomaMSI rileva lo stato delle regioni microsatelliti (MSS o MSI).

Quando si esegue un'analisi SomaMSI, gli utenti sono invitati a:

- Selezionare un numero minimo di otto campioni da analizzare, quindi fare clic su "Next step".
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**).
- selezionare un numero minimo di 8 campioni da analizzare, quindi fare clic su "Next step",
- Eseguire l'analisi facendo clic su "Start analysis".

SomaHRD

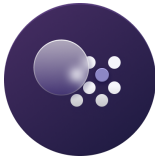


SomaHRD consente di rilevare lo status di HRD da dati WGS poco profondi da soli o accompagnati da un panel di geni, tra cui BRCA1 e BRCA2.

Quando si esegue un'analisi SomaHRD, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select".
- Inserire i metadati sulla cellularità (% tumor cells) per ciascun campione (vedere la sezione **Metadati del campione**).
- Selezionare i campioni da analizzare e cliccare su "Next step",
- Selezionare l'analisi SomaVar o SomaCGP (GRCh37 o GRCh38) appropriata per ciascun campione utilizzando l'opzione "Select panel analysis (Optional)",
- Eseguire l'analisi facendo clic su "Start analysis".

SomaVar LF

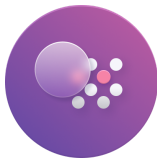


Rilevamento di varianti brevi e strutturali a bassa frequenza, nonché di geni di fusione, da campioni di tumore (biopsie solide o liquide, DNA tumorale circolante) arricchiti da capture su un panel di geni.

Quando si esegue un'analisi SomaVar LF, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select",
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**),
- Selezionare i campioni da analizzare e cliccare su "Next step",
- Eseguire l'analisi facendo clic su "Start analysis".

SomaCGP



SomaCGP consente di rilevare mutazioni somatiche, varianti strutturali, fusioni geniche, nonché lo status MSI e TMB da campioni tumorali. Questo workset è dedicato all'analisi di DNA FFPE preparato con il pannello SureSelect Cancer CGP Assay di Agilent ed è disponibile solo quando i metadati Comprehensive Genomic Profiling sono selezionati alla creazione del progetto. Questo workset può essere utilizzato con i dati di acquisizione.

Quando si esegue un'analisi di SomaCGP, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select",
- Se richiesto, selezionare una "CNV baseline" e/o una "MSI baseline" (se non viene selezionata alcuna linea di base, l'analisi verrà eseguita senza rilevamento CNV/MSI),
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**).
- Selezionare i campioni da analizzare e cliccare su "Next step",
- Avviare l'analisi facendo clic su "Start analysis".

SomaHemato



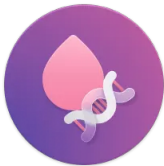
SomaHemato consente di rilevare mutazioni somatiche da campioni con specificità ematologiche. Questo workset può essere utilizzato con i dati di acquisizione.

Quando si esegue un'analisi di SomaHemato, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select".
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**).
- Selezionare i campioni da analizzare e cliccare su "Next step".

Avviare l'analisi facendo clic su "Start analysis".

SomaLBx



SomaLBx consente di rilevare mutazioni somatiche a bassa frequenza, varianti strutturali e fusioni geniche da ctDNA. Questo workset è dedicato all'analisi del ctDNA e può essere utilizzato con i dati di acquisizione.

Quando si esegue un'analisi SomaLBx, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select",
- Selezionare una "CNV baseline" (se non viene selezionata alcuna linea di base, l'analisi verrà eseguita senza rilevamento CNV),
- Nelle opzioni, selezionare la modalità di elaborazione UMI applicabile tra le modalità ammissibili per l'analisi (vedere la sezione **Scelta della modalità di elaborazione UMI**),
- Selezionare i campioni da analizzare e cliccare su "Next step".

Avviare l'analisi facendo clic su "Start analysis".

SomaMethyl



SomaMethyl consente di rilevare le isole CpG metilate da campioni di DNA tumorale circolante. Questo workset è dedicato all'analisi della metilazione nel ctDNA e può essere utilizzato con i dati di acquisizione del bisolfato.

Quando si esegue un'analisi Somamethyl, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire e fare clic su "Select",
- Selezionare una "Methyl baseline" corrispondente alla coorte di controllo (se non viene selezionata alcuna coorte, l'analisi verrà eseguita senza rilevamento della soglia e metilazione differenziale),
- Selezionare i campioni da analizzare e cliccare su "Next step",
- Avviare l'analisi facendo clic su "Start analysis".

Worksets di analisi disponibili da file di analisi secondaria

I seguenti worksets sono disponibili quando si seleziona *VCF files - Secondary Analysis data* durante la creazione del progetto o quando si selezionano entrambi i metadati Agilent e CGH MicroArrays (vedere la sezione **Secondary analysis files**). In questo caso, la piattaforma riceve i dati di output dall'applicazione del fornitore selezionato, procede all'annotazione e all'iniezione delle varianti nelle rispettive interfacce di visualizzazione.

GermVar Tertiary



Pipeline per l'analisi di campioni singoli di varianti rilevate dall'applicazione Illumina Germline o dall'applicazione EPI2ME di Oxford Nanopore Technologies, a partire da dati WGS e WES o di varianti rilevate dall'applicazione Cytogenomics di Agilent a partire da dati CGH MicroArray.

Quando si esegue un'analisi GermVar Tertiary, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire, quindi fare clic su "Select",
- Selezionare il/i campione/i da analizzare, quindi fare clic su "Next step".
- Avviare l'analisi facendo clic su "Start analysis".

GermVar Tertiary Family

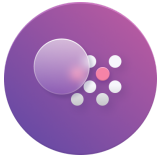


Pipeline di analisi multi-campione delle varianti rilevate dall'applicazione Illumina Germline o dall'applicazione EPI2ME di Oxford Nanopore Technologies, da dati WGS e WES.

Quando si esegue un'analisi di GermVar Tertiary Family, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire, quindi fare clic su "Select",
- Selezionare almeno 2 campioni,
- Nella sezione "Set conditions", indicare se il campione è del paziente o di un familiare definendo il legame familiare in "Family link". Il paziente viene automaticamente contrassegnato come "affected".
- Verificare che le impostazioni corrispondano.
- Eseguire l'analisi facendo clic su "*Start analysis*".

SomaVar Tertiary



SomaVar Tertiary è una pipeline per l'analisi terziaria basata su VCF che elabora sia i file VCF di Illumina WES che i file VCF di pannelli di geni provenienti da vari fornitori di kit (vedere la sezione **Formato dei file di input**).

Quando si esegue un'analisi SomaVar Tertiary, gli utenti sono invitati a:

- Selezionare la versione del workset da eseguire, quindi fare clic su "Select",
- Selezionare il/i campione/i da analizzare, quindi fare clic su "Next step".
- Avviare l'analisi facendo clic su "Start analysis".

11.1.2. Scelta della modalità di elaborazione UMI

Presentazione delle modalità di elaborazione dell'UMI

In pratica, la piattaforma propone uno o più metodi per elaborare le UMI durante l'esecuzione di un'analisi:

1. **Standard mode** (consigliato): i consensi sono generati dai duplicati PCR che presentano la stessa UMI quando il loro numero è maggiore o uguale a 2. Vengono mantenute anche le UMI rappresentate da una singola sequenza (singleton). La frequenza allelica minima delle varianti riportate e il numero minimo di reads a supporto della mutazione (AO, Alternate Observation) variano a seconda del workset utilizzato:
 - In SomaVar, SomaHemato, SomaRNA, SomaCNVCapture, SomaCGP e SomaCMSI, la frequenza allelica minima è dell'1% e l'AO è un minimo di 3 reads.
 - In SomaVar LF e SomaLBx, la frequenza allelica minima è dello 0,1% e l'AO minimo è di 2 reads. Questo worksets consente di rilevare varianti con una frequenza allelica inferiore all'1%. È consigliato quando la profondità di sequenziamento è superiore a 5000x e per applicazioni quali il sequenziamento di ctDNA.
2. **UMI disabled** (disponibile solo in SomaVar, SomaRNA, SomaHemato, SomaLBx, SomaCNVCapture, SomaCGP e SomaMSI): Le UMI non vengono utilizzate per la deduplicazione e vengono tagliate dalla fine delle sequenze prima dell'analisi. Pertanto, i dati generati da questa pipeline di analisi sono grezzi (non deduplicati). La frequenza allelica minima delle varianti riportate è pari all'1% e il numero minimo di reads che supportano la mutazione (AO) è pari a 3.

Le differenze di impostazione tra queste modalità sono riassunte nella tabella seguente:

	UMI disabled	UMI standard
Qualità minima mediana/media della base mutata (punteggio Phred)*.	30	30
Numero minimo di reads per consenso	N/A	2
Qualità minima di ogni base nel consenso (Score Scores.)	N/A	30
Reads senza consensi filtrati	N/A	no

Tabella 1: Impostazioni delle diverse configurazioni di elaborazione UMI

* Questi filtri si applicano solo alle varianti non selezionate dal modello SVM dopo la chiamata delle varianti (pipeline SomaRNA e SomaVar).

Processo di trattamento UMI

La struttura delle UMI dipende da ciascun kit del fornitore e i seguenti kit sono attualmente disponibili sulla piattaforma SeqOne:

- QIAGEN QIAseq
- Agilent XTBS/Low input
- Agilent XTBS V2
- Illumina TruSight Oncology 500
- Twist UMI Adapter System
- Roche KAPA HyperCap
- QIAseq targeted DNA Pro

 **Warning:** La configurazione UMI "QIAseq Targeted DNA Pro" deve essere utilizzata solo per SomaVar, con una lunghezza di lettura di almeno 150 bp. Per ulteriori informazioni, contattare il supporto.

L'elaborazione di UMI si basa sull'uso della suite di strumenti fgbio, in grado di gestire UMI e read taggati con UMI. La nostra pipeline di analisi UMI consiste nelle seguenti fasi:

1. Se necessario, tagliare gli UMI dalla fine delle sequenze e generare un BAM non allineato dalla coppia di file FASTQ,
2. Allineamento di sequenze senza UMI con BWA-MEM2 (DNA/ctDNA) o STAR (RNA)
3. Unire i file BAM contenenti le UMI con le sequenze allineate usando ZipperBam,
4. Raggruppa i read se hanno la stessa posizione e condividono lo stesso UMI,
5. Generare la sequenza di consenso per ogni gruppo di UMI,
6. Filtrare ogni base in base alla sua qualità: la fiducia nella sequenza di consenso aumenta con il numero di reads per gruppo,
7. Filtrare le sequenze di consenso: se viene filtrato più del 20% delle basi della sequenza di consenso, la sequenza di consenso non viene mantenuta,
8. Conversione di queste sequenze consensus filtrate in FASTQ e riallineamento finale sul genoma di riferimento con BWA-MEM2 (DNA/ctDNA) o STAR (RNA).

Compatibilità con i worksets di analisi

I worksets SomaVar, SomaHemato, SomaCGP, SomaLBx, SomaMethyl, SomaVar LF, SomaCNVCapture, SomaMSI e SomaRNA sono compatibili con i dati UMI. Le opzioni di elaborazione disponibili all'avvio dei diversi worksets sono descritte nella tabella seguente:

	UMI disabled	UMI standard
SomaVar	Sì	Sì
SomaCNVCapture	Sì	Sì
SomaMSI	Sì	Sì
SomaRNA	Sì	Sì
SomaVar LF	No	Sì

SomaCGP	Si	Si
SomaLBx	Si	Si
SomaHemato	Si	Si
SomaMethyl	No	Si

Tabella 2: Opzioni di elaborazione UMI disponibili quando si lanciano i worksets compatibili.

11.2. Stato dell'analisi

Dopo aver lanciato un'analisi, la piattaforma SeqOne utilizza diversi "Status" per consentire agli utenti di seguire l'avanzamento dell'analisi:

- **Bioinformatics Pending:** la piattaforma è in attesa del termine di un'altra analisi
- **Bioinformatics Running:** l'analisi è in corso; questo stato mostra anche quale fase dell'analisi è in corso.
- **Bioinformatics Finished:** l'analisi è stata completata.
- **Analysis failed:** si è verificato un errore durante l'analisi; in questo caso, gli utenti sono invitati a contattare il team di supporto all'utente
- **Interpretation - new:** l'analisi è terminata e pronta per essere interpretata, nessun utente l'ha ancora aperta
- **Analysis failed - error:** L'analisi ha riscontrato un errore. In questo caso, l'utente è invitato a contattare l'assistenza.
- **Analysis failed - canceled:** l'analisi non è stata avviata perché il campione collegato ha riscontrato un errore durante il caricamento o la verifica.

11.3. Qualità del campione

La scheda QC del progetto è suddivisa in 2 sottotabelle: 'General statistics' e 'ID check summary'.

General statistics. Questa scheda visualizza una tabella che rappresenta la qualità complessiva dell'allineamento e della copertura per ciascun campione del progetto. L'utente può esportare la tabella come file TSV facendo clic su .

Il calcolo si basa sullo strumento SeQC, un programma sviluppato da SeqOne per calcolare le statistiche di allineamento e copertura di tutte le reads allineate al genoma.

L'utente può scegliere di utilizzare le metriche QC predefinite o di definire valori soglia personalizzati dalla scheda "Settings" del progetto (vedere la sezione **Configurare le metriche predefinite del progetto QC**).

Le metriche proposte sono le seguenti:

- numero totale di reads
- lunghezza media delle reads
- dimensione mediana dei frammenti di DNA nella libreria (inserti)
- percentuale di reads di qualità >30
- percentuale di reads allineate con il genoma di riferimento
- percentuale di reads sul target
- percentuale di reads duplicate
- tasso di transizione/trasversione
- Copertura media
- Frazione di bersagli coperti a diverse soglie a seconda del contesto (Amplicon, Capture).

I target del pannello o dell'esoma sono determinati dal file BED selezionato al momento della creazione del progetto.

'ID check summary'. Questa scheda riporta in dettaglio i genotipi di tutti i marcatori inclusi nelle analisi per ciascun campione (vedere il paragrafo **Marker** nella sezione **'Managing user parameters'**). Una riga corrisponde a un campione e ogni colonna rappresenta un marcatore con il suo ID associato. L'utente può esportare questa tabella come file TSV facendo clic su .



12. Controllo della qualità dei dati

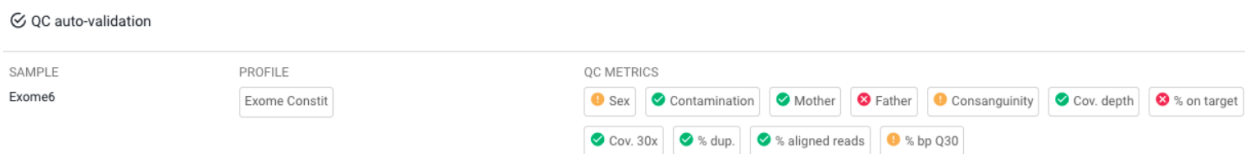
12.1. Metriche e report sul QC

12.1.1. Metriche e verifiche QC

Dalla scheda QC, l'utente accede ai controlli di qualità applicati all'analisi, sotto forma di sotto-tabelle descritte nelle sezioni seguenti: 'QC report', 'ID check' (se applicabile) e 'Technical report'.

Nel caso di un'analisi CGH MicroArrays, la scheda "QC" non è disponibile. Tuttavia, la relazione tecnica è disponibile nella scheda *Files* dell'analisi in formato "report.html".

Nella parte superiore della scheda QC, una sezione "QC auto-validation" permette all'utente di colorare le metriche QC come *Passed*, *Warning* o *Failed*. Facendo clic sull'icona associata, l'utente accede alle soglie applicate e ai dettagli del QC: sesso, contaminazione, parentela e consanguineità, profondità di copertura, % sul target, copertura 30X o 100X, % duplicati, % reads allineate con il genoma di riferimento o % bp Q30.



12.1.2. Rapporto QC

Il "QC report" globale del processo di sequenziamento contiene varie metriche che descrivono la qualità e l'allineamento delle sequenze, la copertura delle regioni nel manifesto e alcune statistiche relative alle varianti. Questo rapporto è composto dalle seguenti sezioni:

[Alignment details \(Dettagli dell'allineamento\)](#) (SomaMethyl). Letture totali elaborate, letture allineate in modo univoco e non univoco e tassi di mismatch.

[Deduplication \(Deduplicazione\)](#) (SomaMethyl). Metriche di duplicazione per le letture allineate, evidenziando potenziali problemi con la preparazione della libreria.

[Cytosine coverage & methylation \(Copertura della citosina e metilazione\)](#) (SomaMethyl). Tassi di conversione da citosina a uracile per ogni contesto (CpG, CHG, CHH).

[Statistiche dei gruppi UMI](#) (solo SomaVar, SomaVar LF, SomaMethyl e SomaRNA UMI). Vedere il paragrafo **Scelta della modalità di elaborazione UMI**.

Il "QC report" delle analisi di SomaVar, SomaVar LF, SomaMethyl e SomaRNA offre una visione più dettagliata della composizione del campione, in particolare della distribuzione degli UMI in base al numero di sequenze che li ospitano.

Le statistiche descrivono in particolare:

- La frazione di singleton (UMI trasportati da una sequenza unica)
- La dimensione media e mediana dei gruppi UMI, nonché la deviazione standard
- La dimensione del gruppo più grande e la frazione di sequenze totali che comprende.

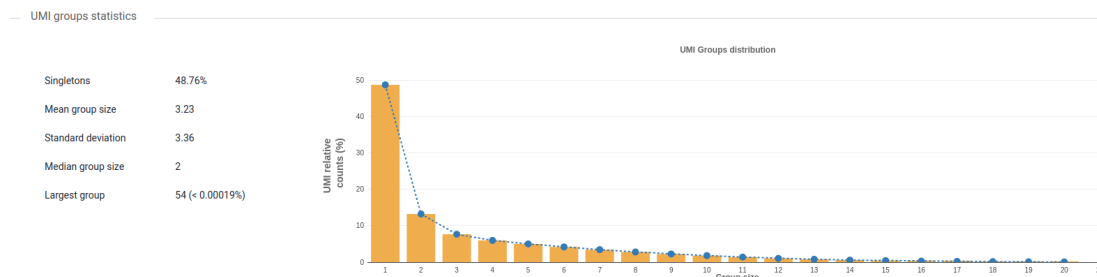


Figura 7: Statistiche relative alla distribuzione UMI in termini di numero di sequenze all'interno di un campione.

Anche le metriche globali di "QC" per il progetto sono più dettagliate e suddivise in tre sezioni:

- Metriche QC dopo l'elaborazione UMI, che indicano l'andamento di ciascuna metrica (in percentuale) in base ai dati grezzi.
- Metriche "Raw QC metrics" calcolate da tutte le sequenze (prima dell'elaborazione UMI).
- Distribuzione dell'UMI in termini di numero di sequenze all'interno di ciascun campione.

'*Cross-contamination estimate*' (solo SomaVar LF). La Cross-contamination' viene stimata individualmente utilizzando lo strumento GATK CalculateContamination'. Quest'ultimo utilizza il segnale degli alleli di riferimento rilevati nei loci omozigoti per calcolare la frazione di sequenze nell'individuo provenienti da cross-contamination.

Sample integrity and identity (solo analisi Germline). Questa sezione elenca le metriche relative all'integrità e all'identità del campione:

- **Family link'** : metadati definiti dall'utente
- **Original sex**: metadati relativi al sesso inseriti dall'utente (Male/Female/Unknown)
- **Detected sex**: sesso rilevato (Male/Female/Unknown)
- **Relationship coefficient**": coefficiente di relazione calcolato con lo strumento Somalier per ciascun campione nell'analisi della famiglia; questa metrica visualizza N/A' per il campione stesso.
- **Contamination**': percentuale stimata di contaminazione del campione da parte di DNA estraneo, utilizzando lo strumento verifyBamID2'.

Sono inoltre elencati il numero di varianti omozigoti ed eterozigoti, il rapporto tra questi due valori (Hom/Het ratio), il numero di varianti omozigoti ed eterozigoti sul cromosoma X, il rapporto equivalente sul cromosoma X (Het/Hom ratio on chr X) e il numero di varianti sul cromosoma Y.

Le metriche "*Family link*" e "*Relationship coefficient*", utilizzate per verificare la relazione tra gli individui di una famiglia, vengono visualizzate solo nelle analisi di famiglia a tre (GermVar Family o GermVar Tertiary Family).

Le metriche di genere sono utilizzate per verificare la corrispondenza tra il genere inserito e quello rilevato nei dati.

La metrica di contaminazione non viene calcolata per i workset GermVar Tertiary e GermVar Tertiary Family.

I controlli "QC" sulla corrispondenza di genere, sulla contaminazione e sulla parentela sono disponibili a livello di progetto o di analisi per identificare facilmente eventuali corrispondenze o discrepanze.

Alignment metrics. Questa sezione elenca le principali metriche calcolate alla fine dell'allineamento, come ad esempio:

- Il numero totale di sequenze nel campione,
- La dimensione mediana degli inserti, calcolata dalla distanza tra le due reads della stessa coppia, la cui lunghezza rispettiva è inclusa,
- La lunghezza media della sequenza,
- La frazione di sequenze allineate con successo al genoma di riferimento,
- Il tasso di duplicazione (solo Capture),
- La frazione di basi la cui qualità è superiore a 30,
- La copertura media delle regioni mirate, determinata dal manifesto dell'analisi.

Coverage. Grafico che mostra la frazione di regioni target coperte a diverse soglie, indicate sull'asse delle ascisse. I valori di copertura possono essere visualizzati in forma tabellare

View as table



facendo clic sul pulsante che consente all'utente di copiare/incollare tali valori.

QC graphs. Grafici dell'analisi della qualità FastQC, che illustrano in particolare il contenuto di GC delle sequenze e la qualità delle basi in base alla loro posizione nelle sequenze.

'Variant statistics.' Un elenco di statistiche relative alle varianti rilevate nell'analisi, che includono in particolare:

- Il numero di transizioni (Ts) e di transversioni (Tv) tra le sostituzioni rilevate,

- Il rapporto T_s / T_v ,
- Il numero totale di mutazioni rilevate,
- La frazione di SNV, MNV e indel in questo totale.

12.1.3. Rapporto 'ID-Check'

La scheda 'ID check' è dedicata ai risultati del monitoraggio dell'identità nella scheda QC di worksets che chiamano varianti.

I risultati vengono visualizzati sotto forma di tabella con le seguenti informazioni:

- Una descrizione di ciascun marcatore basata sulle coordinate genomiche, ad esempio: chr-start-ref-alt,
- un'etichetta (rsID, etichetta personalizzata) se fornita dall'utente,
- genotipo (GT) o frequenza allelica (VAF) osservati,
- la profondità di copertura nella posizione del marker (COV),
- la qualità del database (QUAL.).

Se nell'analisi sono inclusi più campioni, le colonne specifiche del campione (GT/VAF, QUAL, COV) vengono ripetute per ogni singolo campione.

Selezionando una variante nella tabella si apre un browser IGV in una nuova scheda, centrato sulle coordinate genomiche della variante.

12.1.4. Rapporto tecnico

La relazione tecnica elenca gli strumenti utilizzati durante l'analisi e la loro versione. È possibile accedervi dalla scheda QC di un'analisi, quindi da "Technical report".

In particolare, descrive :

- informazioni generali sulla configurazione e sui metadati del progetto e dell'analisi, come il tipo di sequenziamento, la versione del genoma e la versione del worksets utilizzata;

[QC report](#) [ID check](#) [Technical report](#)

Technical Report

This report describes the back-end process used for the secondary analysis.



General Infos

Project Name	VerificationTest - ID Marker Multiple samples
Experiment Type	dna-seq
Sequencing Type	paired-end
Sequencing Equipment	Illumina
Sequencing Method	panel
Amplification Method	capture
UMI Mode	standard
Genome Version	GRCh37
Analysis Name	SomaVar - toy_fusions48
Workset Name	SomaVar
Workset Description	Detection of somatic variants based only on the index sample
Workset Version	v2.5.11

- dettagli del manifesto utilizzato, come nome, dimensione, lunghezza (bp) e padding aggiunto all'analisi;
- dettagli sulle banche dati e sugli strumenti bioinformatici utilizzati.

12.1.5. Rapporto UPD

Il Rapporto UPD è disponibile nella sezione QC delle analisi germinali compatibili (ad es., GermVar e flussi di lavoro correlati). Riassume i risultati del rilevamento UPD e fornisce elementi visivi per supportare l'interpretazione.

Il Rapporto UPD in genere include:

- Panoramica dei risultati: indicazione se l'analisi ha rilevato un evento di tipo UPD e il tipo di evento (ad es., completo o segmentale; nelle analisi di trio, il rapporto può anche supportare l'origine genitoriale e i sottotipi di eventi come UPD | Iso / UPD | Het).
- Elenco degli eventi: un elenco degli eventi rilevati con la loro posizione cromosomica e dimensione, e i punteggi associati utilizzati dall'algoritmo.
- Grafici di visualizzazione: grafici basati sui genotipi SNP (ad es., frequenza dell'allele B (BAF) e segnale di omozigosi/ROH) che aiutano a confermare la presenza di lunghe regioni omozigoti e dei loro confini.
- Informazioni sul controllo di qualità: metriche relative all'affidabilità del rilevamento (ad es., densità SNP / numero di SNPs) e avvisi quando i dati non sono sufficienti per una chiamata sicura.

Gli eventi UPD rilevati sono visualizzati anche nel Large Variant Viewer (LVV) con etichette specifiche (ad es., UPD | Iso e UPD | Het).

Punteggio UPD

Il punteggio UPD è un indicatore numerico utilizzato per quantificare quanto fortemente il campione mostri un segnale di tipo UPD lungo il genoma. Il rilevamento UPD in SeqOne si basa sui modelli di genotipo SNP:

- In un'analisi singola (solo), l'algoritmo ricerca ampie regioni di omozigosi (ROH) e cambiamenti improvvisi nella distribuzione degli SNPs lungo i cromosomi.
- In un'analisi di trio, il modello SNP viene inoltre utilizzato per determinare l'origine genitoriale dell'evento e per distinguere l'isodisomia (UPD | Iso) dall'eterodisomia (UPD | Het).

In pratica, il punteggio UPD riflette la forza e la coerenza del segnale a supporto di un evento UPD (ad es., un chiaro modello ROH/BAF e confini ben delimitati). Un punteggio più alto indica che l'evento ha maggiori probabilità di essere un vero evento UPD, mentre valori più bassi indicano segnali più deboli o più rumorosi.

Il punteggio UPD deve essere sempre interpretato insieme agli elementi forniti nella scheda Rapporto UPD (sezione QC), tra cui:

- i grafici di visualizzazione UPD (ad es., frequenza dell'allele B / segnale di omozigosi),
- la dimensione e la posizione cromosomica dell'evento,
- le metriche di qualità dell'analisi (ad es., densità SNP / numero di SNPs).
- Grafico della frequenza dell'allele B (BAF) dal rapporto UPD

Il grafico della frequenza dell'allele B (BAF) è una visualizzazione dei genotipi SNP lungo il genoma e viene utilizzato per valutare lo squilibrio allelico e la perdita di eterozigosi (LOH).

Per ogni posizione SNP, la BAF corrisponde alla frazione di letture di sequenziamento che supportano l'allele "B" e varia da 0 a 1:

- Valori vicini allo 0 corrispondono a un genotipo omozigote (AA).
- Valori vicini allo 0,5 corrispondono a un genotipo eterozigote (AB).
- Valori vicini a 1 corrispondono a un genotipo omozigote (BB).

In una regione diploide normale, il grafico BAF mostra in genere tre bande intorno a 0, 0,5 e 1. Nelle lunghe regioni di omozigosi (ROH) o isodisomia UPD, la banda eterozigote (intorno a 0,5) è ridotta o assente, il che supporta la presenza di un segnale di tipo UPD.

Limitazioni importanti:

- Un segnale di tipo UPD può essere osservato anche nel contesto della consanguineità genitoriale (più regioni ROH attraverso il genoma).
-
- Il calling UPD si basa principalmente sulla distribuzione degli SNPs e potrebbe non essere sufficiente da solo per distinguere alcuni modelli UPD da eventi relativi al numero di copie; quando necessario, effettuare un controllo incrociato con altre prove disponibili (ad es., copertura / risultati CNV).

13. Interpretazione dei risultati dell'analisi

13.1. Risultati dell'analisi

Gli utenti possono accedere alle analisi di un progetto facendo clic sull'icona "Analyses" nel riquadro associato o facendo clic sull'icona omonima nell'angolo in alto a sinistra della dashboard utente.

Quando un'analisi è stata completata ed è disponibile per la visualizzazione, viene visualizzata con lo status *"Interpretation - New"*.

I risultati dell'analisi sono disponibili sotto forma di diversi report, che possono essere aperti direttamente facendo clic su di essi.

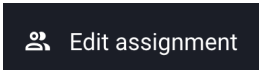
Esistono diverse varianti dello stato *"Interpretation"*, a seconda dei progressi raggiunti nell'interpretazione dei risultati:

- **'New'** - la tabella delle varianti non è ancora stata visualizzata
- **'In progress'** - la tabella delle varianti è stata visualizzata, ma non è ancora stata convalidata alcuna variante.
- **'Evaluating'** - almeno una variante è stata convalidata
- **'Reporting validated'** - è stato generato un report per l'analisi

13.1.1. Gestione del flusso dell'interpretazione dell'analisi

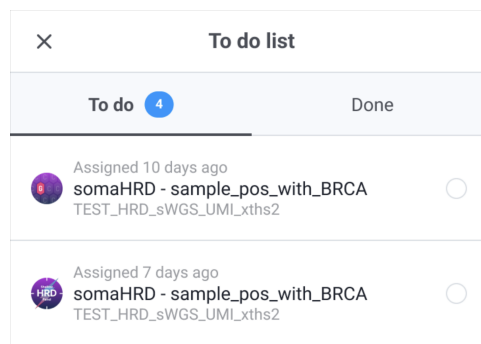
13.1.1.1. Assegnazione di analisi per l'interpretazione

Gli amministratori delle entità possono scegliere di assegnare una o più analisi da interpretare a diversi utenti dell'entità dalla scheda "Analyses" del progetto. A tal fine, l'amministratore può

selezionare una o più analisi, quindi fare clic sul pulsante  in fondo alla pagina.

La finestra "Assignment" consente di assegnare uno o più utenti alle analisi selezionate, cercando l'identificativo dell'utente (nome utente) tramite l'apposita barra di ricerca. È anche possibile modificare l'assegnazione aggiungendo o eliminando utenti. Gli utenti a cui è assegnata un'analisi sono elencati nella colonna "Assigned to". Gli utenti che non hanno mai effettuato l'accesso a SeqOne non possono essere assegnati finché non creano le proprie credenziali.

Gli utenti possono accedere alle analisi assegnate (scheda To do) e a quelle convalidate (scheda Done) nella loro "To do list", accessibile tramite l'icona  in alto a destra della pagina. Per ogni nuova assegnazione viene visualizzata una notifica.

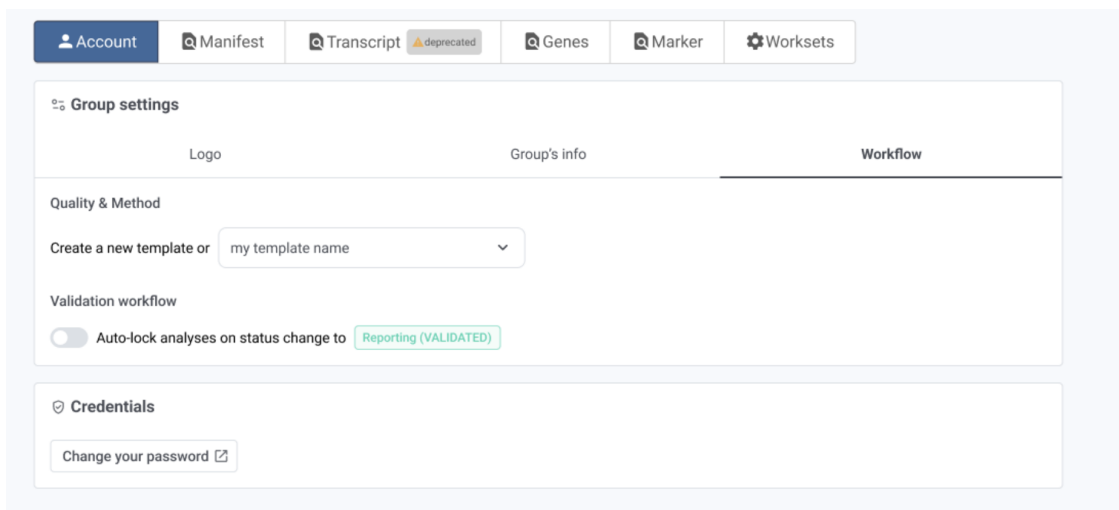


Gli utenti possono accedere direttamente alle analisi assegnate facendo clic su di esse. Possono convalidare il loro compito facendo clic sulla casella di controllo corrispondente nella scheda "To do" per spostarlo nella scheda "Done". Questa azione è reversibile.

Le analisi vengono visualizzate nel seguente ordine in base al loro status: completate, in corso o "Failed". Per ogni analisi, vengono dettagliati la data di assegnazione, il nome dell'analisi e il nome del progetto.

13.1.1.2. Analisi del bloccaggio

L'amministratore di un'entità può bloccare o sbloccare un'analisi. Nelle *Impostazioni* dell'entità, dalla scheda "Account" e poi "Workflow", può attivare il blocco automatico delle analisi convalidate utilizzando la funzione "Auto-lock analyses on status change to Reporting (VALIDATED)":



Al termine del lavoro di interpretazione, quando un utente modifica lo stato dell'analisi in "Reporting - validated", l'analisi viene automaticamente bloccata. Le analisi bloccate visualizzano lo stato "Reporting - Validated & Locked", visibile con l'icona  Reporting VALIDATED & LOCKED.

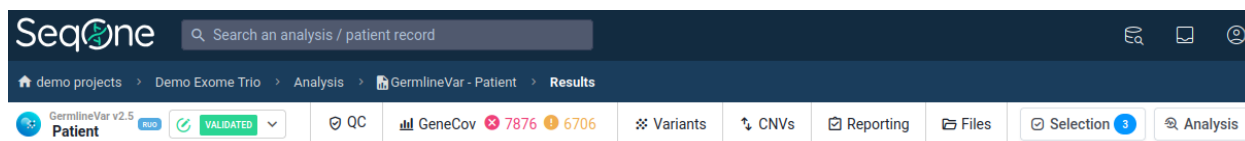
Quando un'analisi è bloccata, gli utenti non possono eseguire le seguenti azioni dall'interfaccia:

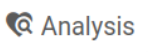
- aggiungere o modificare le 'Evaluations' delle varianti in "VKB",
- aggiungere o modificare i commenti associati alle varianti,
- generare un nuovo rapporto dalla finestra *Selection*,
- generare un nuovo rapporto HRD,
- modificare manualmente lo status dell'analisi,
- eseguire una delle seguenti azioni nel pannello di analisi:
 - modificare l'assegnazione del saggio (in "Assigned To.")
 - modificare il sesso del paziente ("modifier")
 - aggiungere o modificare gli HPOs:
 - modificare i "Panel" di analisi
 - eliminare il rapporto clinico

Se un utente tenta una di queste azioni, gli verrà chiesto di sbloccare l'analisi, se ha l'autorizzazione, o di contattare l'amministratore, se non ce l'ha. L'utente che desidera sbloccare un'analisi deve disporre dell'autorizzazione "Unlock analyses".

13.1.1.3. Pannello laterale per la gestione delle analisi

Il pannello laterale Analisi, noto anche come "Analysis Drawer", è accessibile da qualsiasi scheda di analisi.



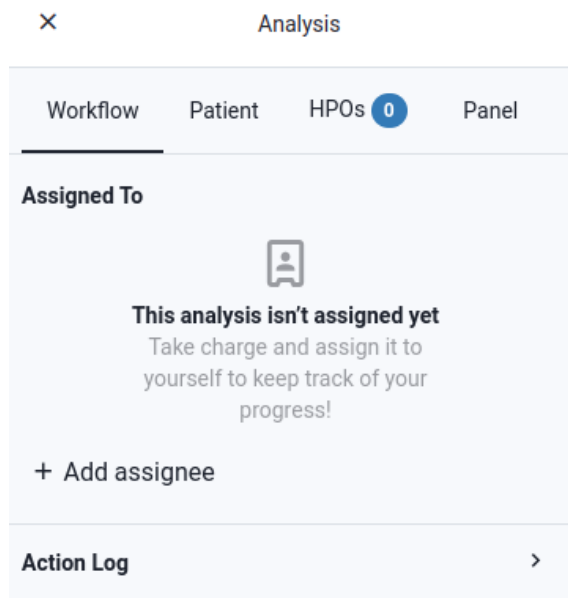
Facendo clic sul pulsante  in alto a destra nel banner delle schede di analisi, l'utente può visualizzare e gestire direttamente i dati di analisi e gli HPOs: senza lasciare la schermata corrente.

Il "Analysis Drawer" contiene 4 viste:

- **Workflow:** consente di organizzare e tracciare il flusso di lavoro per l'analisi.
- **Patient:** consente di visualizzare e modificare i metadati del paziente.
- **HPOs:** consente di aggiungere o modificare i termini HPO nell'analisi, disponibile solo per le analisi in contesto costituzionale.
- **Panel:** consente di applicare un elenco di geni all'analisi e di filtrare i risultati dell'analisi sulla base di questo elenco.

Vista modale Workflow

Questa vista consente di organizzare e tenere traccia dei flussi di lavoro degli utenti per l'analisi in corso di interpretazione.



"Assigned To." (Assegnato a). Questa vista visualizza i compiti assegnati per l'analisi corrente. Solo gli amministratori o gli utenti con l'autorizzazione "Manage assignments" possono assegnare compiti (vedere la sezione **Gestione degli amministratori - Presentazione dell'interfaccia**).

Quando i compiti vengono assegnati all'analisi corrente, viene visualizzato un elenco di tutti gli utenti assegnati, con la data e l'ora di completamento. Questo elenco viene aggiornato automaticamente non appena un utente completa un compito.

Gli utenti possono modificare i propri compiti solo selezionando o deselezionando la casella corrispondente. Quando si passa il mouse su un'attività di sola lettura, un tooltip indica "You can only complete yours".

Se non sono stati assegnati compiti, la vista visualizza "This analysis isn't assigned yet".

Se un utente non autorizzato tenta di assegnare un'attività facendo clic sul pulsante "+ Add assignee", il messaggio di errore indica "Per assegnare un'analisi è necessario disporre dei privilegi di amministratore. Se necessario, contattare l'amministratore della piattaforma per richiedere l'accesso appropriato".

Action Log. Il banner "Analysis Action Log (Audit Trail)" visualizza la cronologia di tutte le azioni eseguite sullo stato HRD nell'analisi SomaHRD, sui filtri, sui profili dei filtri e sui "Restricted panel" nella tabella delle varianti, specificando il tipo di azione, l'utente, il timestamp e una descrizione dell'azione.

Le azioni più recenti appaiono in cima all'elenco. La paginazione consente agli utenti di navigare in modo efficiente nella cronologia delle azioni. L'intera cronologia può essere scaricata in formato .csv tramite l'icona "Download Log".

Vista modale Patient

Questa vista consente di visualizzare e modificare i metadati del paziente, come il sesso (Female, Male o Unknown), il gruppo di ascendenza genetica GnomAD (solo Germline) o l'indicazione (solo Somatic).

La visualizzazione e la modifica dell'indicazione sono disponibili solo per le analisi somatiche:

- Se l'indicazione è stata inserita prima dell'avvio dell'analisi (vedere **Metadati campione**), viene definita per impostazione predefinita per il campione, è contrassegnata da una stella e appare in prima posizione nell'elenco.
- Se l'indicazione viene modificata durante l'interpretazione, un messaggio di avviso indica che l'indicazione selezionata è diversa da quella originale e la colorazione della pillola nella sottocolonna "A" della colonna *Clinical Evidences* viene aggiornata (vedere "**Selection**").
- Se non è disponibile alcuna indicazione per l'analisi, verrà visualizzato il messaggio "No available indication".

La visualizzazione e la modifica del gruppo di ascendenza genetica GnomAD è disponibile solo per le analisi germinali. Il valore predefinito è vuoto e viene considerato come la popolazione GnomAD combinata.

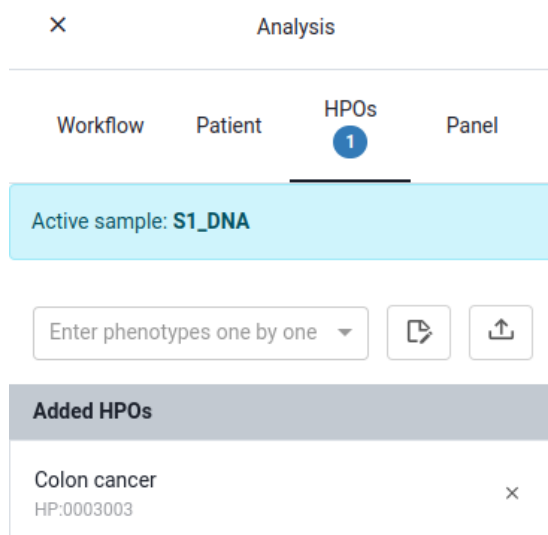
Le modifiche apportate in questa vista devono essere salvate o annullate facendo clic sulle corrispondenti icone "Save modifications/Cancel" nella parte inferiore del riquadro, per poter cambiare scheda.

Vista modale HPOs

Questa vista, disponibile solo nei progetti Germline, consente di visualizzare, modificare e assegnare i termini HPO relativi al campione, manualmente o a partire da una descrizione clinica del paziente.

L'utente può visualizzare l'elenco dei termini HPOs già associati al campione prima dell'analisi. In assenza di codici HPOs, viene visualizzato il messaggio "No HPOs yet".

Per aggiungere manualmente i termini, l'utente può inserirli dalla barra di ricerca "Enter phenotypes one by one", cercando un fenotipo o un codice HPO. Può anche importare un file di testo contenente i codici HPO cliccando sull'icona  (si veda anche la sezione **Metadati campione**).



Panel 8: vista degli HPOs: pannello di analisi

L'utente può generare termini HPO in base alla descrizione clinica del paziente utilizzando DiagAI. Facendo clic sull'icona , appare una finestra di dialogo che richiede all'utente di descrivere il profilo clinico del paziente (la descrizione è limitata a 4000 caratteri) e quindi di fare clic su "Extract HPOs with DiagAI" per generare un elenco di HPOs trovati. I termini HPO sono evidenziati a colori e l'utente può associarli al paziente passandovi sopra il mouse e facendo clic su "Add", oppure aggiungere tutti i termini HPO trovati facendo clic su "Add all" (vedere anche la sezione **Modifica dei metadati del campione con il pannello di analisi**).

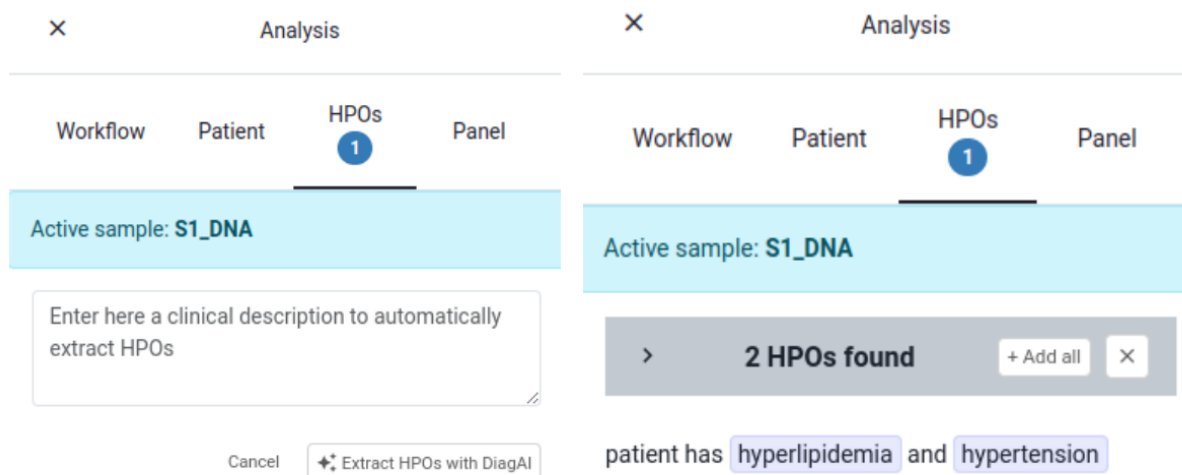
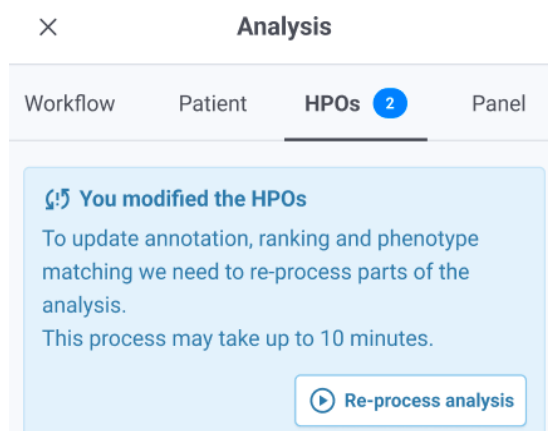


Figura 9: "Extract HPOs terms with DiagAI" dalla vista HPOs nell'Analysis Drawer.

Quando si aggiungono o modificano HPOs da questa vista, appare un banner blu che suggerisce di rielaborare l'analisi con i nuovi HPOs.

Selezionando "Re-process analysis", l'utente può aggiornare il DiagAI Score e la classificazione delle varianti (Short list o SmartPick) in base ai nuovi HPOs inseriti.

Questa fase di rielaborazione congela la tabella delle varianti per diversi minuti per aggiornare la classificazione delle varianti e la colonna "DiagAI".

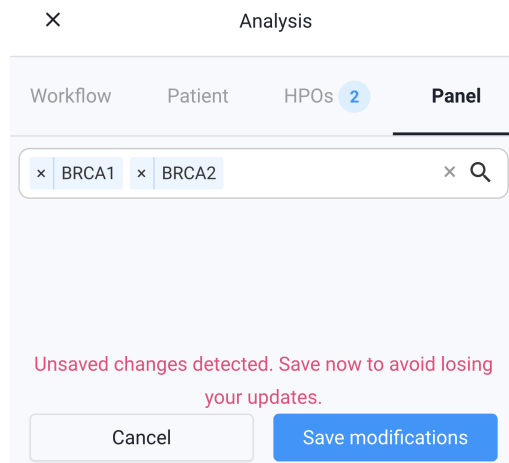


Vista modale Panel

Questa vista è disponibile per tutti i worksets, ad eccezione di quelli per l'analisi delle CNV (CNVExome, CNVCapture e SomaCNVCapture).

Consente di accedere alla funzione *"Restricted panels"*, che permette di limitare e bloccare l'analisi a geni selezionati, come un *"in silico panel"*.

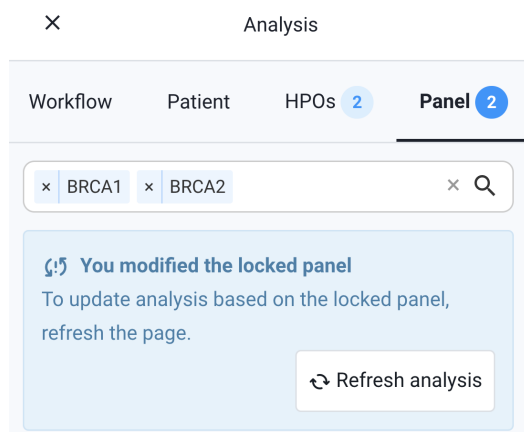
A tal fine, l'utente può selezionare uno o più elenchi di geni da applicare, tramite un menu a tendina, tra quelli precedentemente importati (vedere il paragrafo **Genes**, nella sezione **Gestione delle impostazioni utente**). Salvando i pannelli selezionati tramite *"Save modifications"*, viene applicato un filtro all'intera analisi in base al o ai pannelli di geni selezionati.



I *"Restricted panels"* possono essere selezionati all'avvio dell'analisi (vedere la sezione **Avvio di una nuova analisi**) o dalla vista Panel dopo l'analisi.

Quando vengono aggiunti o modificati pannelli genici, un banner blu avvisa l'utente della modifica (*"You modified the locked panel"*) e suggerisce di aggiornare l'analisi con i nuovi pannelli tramite la funzione *"Refresh analysis"*.

Facendo clic su *"Refresh analysis"*, l'utente sceglie di aggiornare la tabella delle varianti, la scheda CNVs (se applicabile) e la scheda GeneCov' in base al nuovo panel.



13.1.1.4. Pannello laterale per la gestione delle valutazioni

Il pannello  **Selection** è accessibile da qualsiasi scheda di un'analisi. Consente all'utente di visualizzare le piccole varianti e le CNVs annotate in VKB per l'analisi corrente, di generare un file JSON ("Generate file") e di modificare un report clinico ("Generate report").

Le CNVs validate in "VKB" sono disponibili anche in questa sezione, solo nel caso di CNVs provenienti da analisi dell'esoma, del genoma e CGH MicroArray.

Le piccole varianti convalidate in VKB sono differenziate dal logo  , mentre le CNVs sono visualizzate con il logo  e le parole "gain" o "loss".

×

Selection 3

↓ Generate file

📄 Generate report

Variants

⌘ SLC25A22

5

NM_001191061.2 | chr11:g.792146_792147del

📄

The frameshift variant c.813_814del (p.Ala272GlnfsTer144) in the gene SLC25A22 (NM_001191061.2) is not found in control...

⌘ SLC25A22

5

NM_001191061.2 | chr11:g.792142C>T

📄

The missense variant c.818G>A (p.(Arg273Lys)) in the gene SLC25A22 (NM_001191061.2) is extremely rare, with onl...

CNVs

↕ 6q14.3 Loss

4

103.25kb | SNX14, SYNCRIP

13.1.2. Panoramica dei risultati dell'analisi nella scheda Dashboard

Per alcuni worksets, una scheda "Dashboard" che riassume tutti i risultati e le metriche QC è accessibile agli utenti e appare per impostazione predefinita all'apertura di un'analisi. La scheda "Dashboard" offre una panoramica dei risultati dell'analisi in più sezioni. Le sezioni disponibili variano a seconda del tipo di workset. Questa scheda è disponibile per SomaCGP, SomaVar Tertiary e SomaLBx.

Patient. Questa sezione riassume i metadati del campione: nome del campione (*Sample ID*), sesso e indicazione della patologia (Indication).

Quality Controls. Questa sezione fornisce informazioni su diverse metriche chiave relative alla qualità dei campioni e del sequenziamento:

- Purezza del tumore,
- Ploidy" del campione,
- la percentuale di basi la cui qualità è maggiore o uguale a 30 (% basi Q30),
- copertura media delle regioni di interesse (copertura media delle ROI) prima e dopo l'elaborazione UMI,
- percentuale di regioni di interesse con copertura superiore a 200X (ROI > 200X) prima e dopo l'elaborazione UMI.

Signatures. La sezione dedicata alle firme mostra le cifre chiave per le analisi TMB, MSI e HRD:

- Calcolo del TMB (Tumor Mutational Burden): numero di mutazioni/Mb.
- Microsatellite instability: percentuale di siti instabili.
- HRD status (Homologous Recombination Deficiency).

Gli status dei punteggi TMB e MSI possono essere selezionati dall'utente tramite i menu a tendina presenti sotto ciascun punteggio.

Selection. In questa sezione vengono visualizzate le varianti valutate nel database VKB per l'analisi corrente.

Suggested by SeqOne. Le alterazioni genomiche rilevate nell'analisi sono presentate come un elenco ridotto di varianti clinicamente rilevanti. Le varianti vengono selezionate se la loro classe ComPerMed. è 5, o se hanno un'asserzione di livello I corrispondente all'indicazione del paziente, o un'asserzione di livello II per terapie sperimentali correlate a questa indicazione, o un'asserzione di livello I per un'altra indicazione. L'utente può aggiungere varianti di questa categoria alla sezione Selection cliccando su "Add".

Il file JSON può essere generato facendo clic su "Generate JSON file" e contiene tutte le informazioni della pagina "Dashboard".

13.1.3. Copertura dettagliata per esone

Nella scheda "GeneCov", gli utenti possono visualizzare il tasso di copertura dei geni inclusi nel manifesto di analisi per ciascun esone, sia esso un pannello di geni o un esoma.

Metodologia

Questo modulo interseca il manifesto dell'utente con le annotazioni RefSeq per tutti i geni inclusi nel panel o nell'esoma, mantenendo 10bp su ciascun lato per coprire le giunzioni di splice, quindi procede a calcolare la profondità di copertura per ciascuna base inclusa in questa intersezione utilizzando lo strumento mosdepth.

Vengono impostate diverse soglie per assegnare a ciascun esone uno status (covered, warning, failed) in base alla percentuale di basi al di sotto di tali soglie di copertura.

Lo status è definito in base alla copertura delle basi nell'esone, comprese le 10bp adiacenti. Se si utilizzano le UMI, la copertura viene calcolata dopo la deduplicazione.

Le soglie applicabili ai diversi tipi di analisi sono riportate nella tabella seguente*:

	GermlineVar / Germline Family/ GermVar / GermVar Family				SomaVar		SomaRNA		SomaVar LF		SomaCGP	
	Panel		Exome		Panel		Panel		Panel		Panel	
	Warning	Failed	Warning	Failed	Warning	Failed	Warning	Failed	Warning	Failed	Warning	Failed
Capture	Cov. 100x < 100%	Cov. 30x < 100%	Cov. 30x < 100%	Cov. 10x < 100%	Cov. 500x <100%	Cov. 200x <100%	Cov. 200x <100%	Cov. 50x <100%	Cov. 5000x <100%	Cov. 2500x <100%	Cov. 100x < 100%	Cov. 50x < 100%
Amplicon	Cov. 100x < 100%	Cov. 30x < 100%			Cov. 1000x < 100%	Cov. 500x < 100%	Cov. 200x <100%	Cov. 50x <100%				

Tabella 3: Soglie di copertura per il modulo Gene Coverage in base al tipo di analisi.

*Se l'utente ha predefinito metriche QC personalizzate dai parametri del progetto, queste verranno applicate nella scheda GeneCov (vedere la sezione **Configurare le metriche predefinite del progetto QC**).

Vista e interpretazione

Ogni gene può essere visualizzato su una rappresentazione grafica e i risultati di "Gene Base" possono essere filtrati in base ai geni, ai "Gene panels" e agli elenchi di trascritti. Gli esoni bersagliati sono rappresentati in verde (*Covered*), arancione (*Warning*) o rosso (*Failed*) in base alle soglie descritte sopra.

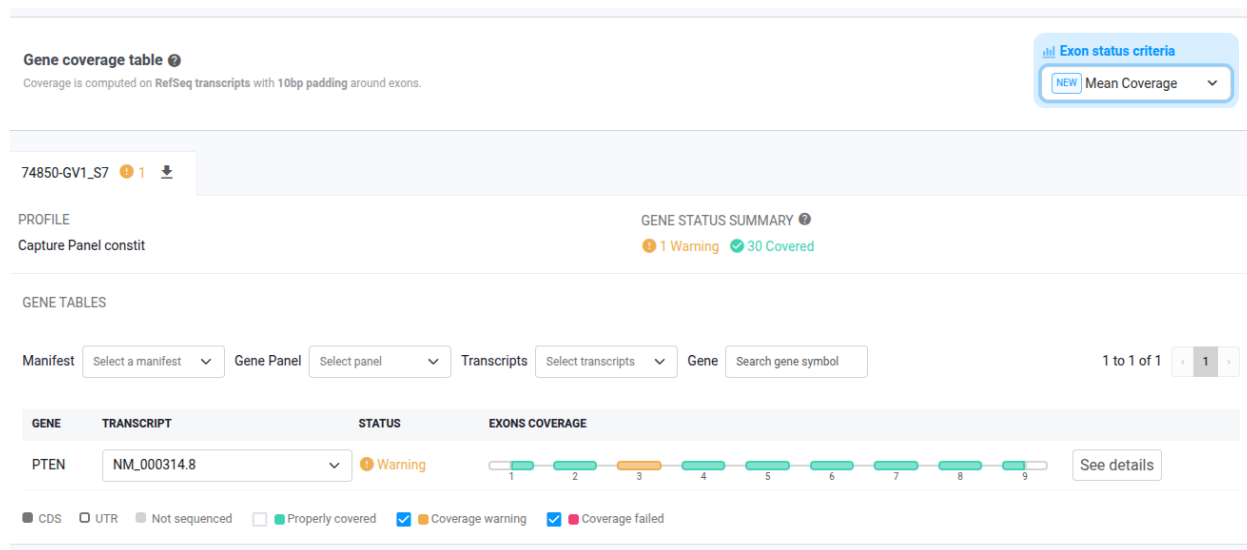


Figura 10: dettagli della scheda "GeneCov"

Opzioni del filtro

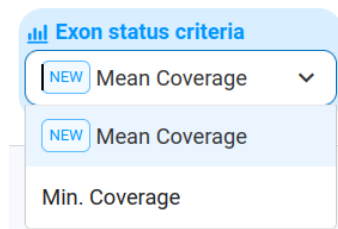
Un menu di filtri consente di scegliere tra

- limitare la visualizzazione a un sottoinsieme di regioni utilizzando un file .bed precedentemente importato nell'interfaccia, tramite il filtro "Manifest",
- applicare un elenco di geni tramite il filtro "Gene Panel",
- applicare un elenco di trascrizioni tramite il filtro "Transcripts",
- cercare un gene specifico tramite il filtro "Gene".

Metodi di calcolo della copertura.

L'utente può selezionare il metodo di calcolo della copertura da applicare utilizzando il menu a discesa "Exon status criteria" in alto a destra della pagina:

1. **Mean Coverage** (predefinito): Utilizza la profondità media di copertura dell'intero esone per determinare lo stato di copertura.
2. **Minimum Coverage**: utilizza la frazione minima di basi che soddisfano una specifica soglia di copertura. In questo modo si garantisce che una certa percentuale di basi abbia almeno la profondità di copertura minima richiesta.



Il menu a discesa "**Gene status criteria**" consente agli utenti di applicare questi metodi di calcolo a livello di gene:

- **Mean coverage (default):** Utilizza la profondità di copertura media di tutti gli esoni del gene per determinare lo stato della copertura.
- **Worst exon status:** Determina lo stato di copertura del gene in base all'esone con le prestazioni peggiori. Se un esone non soddisfa la soglia di copertura, il gene è considerato non coperto. Questo era l'unico metodo supportato nelle versioni precedenti.

Dati visualizzati

Gli esoni non coperti o parzialmente coperti dal disegno (< 80% di sovrapposizione tra l'intervallo manifesto e i confini RefSeq dell'esone) sono considerati non mirati. Questi esoni sono rappresentati in grigio (*Not sequenced*) nel grafico.

Una vista dettagliata di ciascun gene può essere utilizzata per visualizzare le metriche di copertura degli esoni facendo clic su "*See details*": copertura media, mediana, percentuale di ciascuna regione coperta a 10x, 30x, 50x, 100x, 200x, 300x e 500x.

Per ogni esone, una finestra IGV permette di visualizzare la copertura della sequenza e l'allineamento in modo più dettagliato.

Esportazione dei dati "GeneCov"

I dati completi dell'analisi di copertura sono disponibili come file `sample_genecov.v2.tsv.gz`, accessibile dalla scheda Files dell'analisi.

Dalla scheda "GeneCov", l'utente può esportare l'analisi della copertura in base al metodo di calcolo della copertura (Mean Coverage o Min

Coverage), tramite l'icona  accanto al nome del campione nell'angolo in alto a sinistra della scheda del campione.

L'utente può scegliere di esportare i dati in base allo stato di copertura del gene (Failed, Warning, Covered) e di esportare solo il trascritto canonico (Canonical) o tutti i trascritti disponibili (All) prima di cliccare su Download per ottenere il file desiderato.

I dati vengono esportati come file .tsv contenente le seguenti colonne:

Download genes from S1_DNA as .tsv file

Genes status

☒ Failed ☒ Warnings ☐ Covered

Transcripts

☒ Canonical ☐ All

Close

Download

- Gene (simbolo del gene),
- Numero del gene secondo HGNC (gene_id),
- Trascrizione canonica (is_canonical_transcript),
- Stato del transcript (transcript_status)
- Exon (numero_esone)
- Tipo di esone CDS o UTR (exon_type)
- Stato dell'esone coperto o non coperto (exon_status)
- Regione coperta o meno dal manifesto (target)
- Posizione genomica (intervallo_id)
- Numero di cromosomi (chr)
- Posizione di partenza (start)
- Posizione finale (fine)
- Strand (filamento) forward o reverse (filamento)
- % di copertura a 1, 10, 20, 30, 50, 100, 200, 300, 500 e 1000X (cov_)
- Copertura media (cov_avg)
- Copertura mediana (cov_med)

13.1.4. Piccole varianti

Metodologia

Il rilevamento di piccole varianti è gestito da una combinazione di diversi strumenti, che possono variare da un workset all'altro.

Che si tratti di analisi di esomi o del pannello di geni, la chiamata di varianti è determinata dal manifest (in formato BED) associato al progetto, compresi gli intervalli delle regioni di interesse e il padding applicato a tali regioni (vedere la sezione **Creare un progetto**).

Inoltre, le varianti sono chiamate tenendo conto della qualità di allineamento delle reads sul genoma umano (versione GRCh37 o GRCh38). Le cosiddette reads "MAPQ0" con una qualità di allineamento (mappatura) pari a 0, poiché si allineano in diversi siti del genoma, non vengono prese in considerazione per la chiamata di varianti per i worksets di analisi somatica. Nelle analisi "Germline", le reads MAPQ0" vengono considerate per la chiamata di varianti se, al momento della creazione del progetto, viene inserito un manifesto che specifica le regioni di allineamento di bassa qualità nei metadati Low-Quality Mapping Region. Le varianti rilevate dalle letture MAPQ0' sono identificate nella tabella delle varianti da un tooltip "_mapq0" visibile nella colonna CAL (Caller).

SNVs / indels. Per il rilevamento di SNV e indel :

- *Freebayes*¹ è utilizzato come unico variant caller dai workset GermlineVar (esoma), GermlineFamily (esoma), SomaRNA e SomaVar LF. Le soglie di rilevamento inferiori sono le seguenti:
 - VAF $\geq$ 10% in GermlineVar - Exome,
 - VAF $\geq$ 5% in GermlineVar per i pannelli di geni,
 - VAF' $\geq$ 1% in SomaRNA,
 - VAF $\geq$ 0,1% in SomaVar LF,
- È accompagnato da *GATK HaplotypeCaller*² nel workset GermlineVar (Panel).
- La sua controparte somatica, *Mutect2*³, è utilizzata insieme a Freebayes' nei workset SomaVar, SomaHemato e SomaCGP. La soglia minima di rilevamento è VAF $\geq$ 1% per questi due chiamanti in SomaVar.
- *Sentieon*⁴ è utilizzato come principale variant caller dai workset GermVar, GermVar Family e SomaLBx.
- *Clair3*⁵ è utilizzato come unico variant caller nel workset GermlineVarLR.

¹ <https://github.com/freebayes/freebayes>

² <https://gatk.broadinstitute.org/hc/en-us/articles/360037225632-HaplotypeCaller>

³ <https://gatk.broadinstitute.org/hc/en-us/articles/360037593851-Mutect2>

⁴ <https://github.com/Sentieon>

⁵ <https://github.com/HKU-BAL/Clair3>

Le analisi GermVar e GermVar Family si basano sui componenti bioinformatici Sentieon per i passaggi chiave dell'analisi secondaria. In particolare:

- L'allineamento viene eseguito utilizzando l'implementazione ottimizzata di BWA-MEM2 di Sentieon e le letture duplicate vengono contrassegnate.
- Il calling germinale (SNVs/Indels) viene eseguito utilizzando Sentieon DNAScope⁶.
- Sentieon TNScope⁷ può essere utilizzato per rilevare varianti a mosaico con VAF basso (a seconda della configurazione GermVar).

Nota sulle regioni a bassa mappabilità (MAPQ0): in regioni altamente omologhe (ad es., SMN1/SMN2), le varianti potrebbero essere filtrate dal filtraggio basato sull'apprendimento automatico del caller. Quando necessario, SeqOne fornisce un modulo MAPQ0 opzionale (modulo per regioni a bassa qualità) basato su Freebayes con soglie adattate per aumentare la sensibilità nelle regioni definite dall'utente.

Le analisi SomaLBx si basano sui componenti bioinformatici Sentieon per i passaggi chiave dell'analisi secondaria della biopsia liquida. In particolare:

- L'allineamento delle letture viene eseguito con Sentieon (BWA-MEM) e la pipeline include l'elaborazione/consenso UMI basato su Sentieon per supportare i flussi di lavoro ctDNA ad alta sensibilità.
- Il calling somatico di varianti corte (SNVs/Indels) viene eseguito utilizzando Sentieon TNScope^{7ID-Check} in modalità solo tumore (tumor-only), con supporto per frazione allelica molto bassa (VAF $\geq 0,1\%$ a seconda della configurazione).

Nota: i genotipi delle varianti sono riportati come eterozigoti per impostazione predefinita nei risultati SNV/indel SomaLBx perché Sentieon TNScope non esegue la previsione del genotipo.

SVs - Structural Variants. Le inserzioni e le delezioni la cui lunghezza è prossima o superiore a quella di una lettura di sequenziamento vengono rilevate utilizzando il caller SV GRIDSS⁸ nei workset GermlineVar (panel/exome), SomaVar, SomaVar LF SomaHemato e SomaCGP. Questi eventi vengono rilevati utilizzando il chiamante SV Delly⁹ nei worksets GermVar e GermVar Family. Negli worksets somatici, vengono rilevate solo inserzioni e delezioni di dimensioni inferiori a 300 bp, mentre nel workset GermlineVar (per versioni ≥ 2.0) GermVar e GermVar Family worksets (versioni ≥ 3.5) viene applicato un limite di 10 kb. Le varianti rilevate da GRIDSS vengono rilevate indipendentemente dal VAF.

⁶ <https://github.com/Sentieon/sentieon-dnascope-ml>

⁷ <https://github.com/Genomic-Medicine-Linkoping/TNScope>

⁸ <https://github.com/PapenfussLab/gridss>

⁹ <https://github.com/dellytools/delly>

FLT3-ITD. Un modulo di rilevamento specifico per le duplicazioni interne in tandem del gene FLT3 (FLT3-ITD) è disponibile solo nei worksets SomaVar e SomaHemato, a partire dalla versione v2.2. Questo modulo si basa sullo strumento *FiLT3r*¹⁰.

Alu Insertions. Infine, SeqOne ha sviluppato un modulo per il rilevamento delle inserzioni di elementi trasponibili (*AluMEI*), come gli elementi Alu. È utilizzato dai workset GermlineVar (panel/exome), GermlineFamily (panel/exome), GermVar (panel/exome/genome), GermVar Family (panel/exome/genome), SomaHemato e SomaVar. Non viene applicata alcuna soglia VAF inferiore per le varianti rilevate da AluMEI.

Le varianti combinate di *Freebayes*, *HaplotypeCaller* (o *Mutect2*) e *FiLT3r* sono disponibili nel file *final.vcf.gz* dell'analisi. Le varianti specifiche di *GRIDSS*, *FiLT3r* o *AluMEI* sono disponibili nei rispettivi file VCF (***gridss.vcf.gz***, ***filt3r.vcf.gz*** e ***variants.alumei.vcf.gz***), anch'essi scaricabili dalla scheda *Files* dell'analisi.

Nel caso specifico di *FiLT3r*:

- Le varianti *FiLT3r* comuni a *Freebayes* e *Mutect2* sono disponibili nel VCF finale dell'analisi (***final.vcf.gz***),
- Le varianti esclusive di *FiLT3r* e le varianti comuni a *FiLT3r* e *GRIDSS* sono elencate nel VCF di *FiLT3r* (***filtr.vcf.gz***).

Due moduli aggiuntivi sviluppati da SeqOne consentono poi:

- Ritenzione di varianti multinucleotidiche, o MNV, che sono codificanti e limitate a un singolo codone durante il processo di decomposizione delle varianti delineate.
- Ricalibrazione della frequenza allelica delle varianti indels e delins rilevate da *Freebayes* e *GATK* (strumento *Kraf*).

Questi due moduli riguardano solo i workset GermlineVar (panel/exome), SomaVar, SomaRNA e SomaVar LF.

Kraf utilizza una strategia a k-mers singoli, senza allineamento. La frequenza allelica ricalibrata di indels e delins viene mantenuta solo se è superiore alla frequenza allelica iniziale della variante.

Le varianti vengono quindi annotate principalmente con lo strumento *'Variant Effect Predictor'*¹¹ (VEP) e vengono aggiunte ulteriori informazioni dai seguenti database:

- 'ClinVar'
- OMIM (versione commerciale) + PanelApp Australia + PanelApp England".
- CKB
- dbSNP

¹⁰ <https://gitlab.univ-lille.fr/filt3r/filt3r>

¹¹ <https://www.ensembl.org/info/docs/tools/vep/index.html>

- COSMIC (versione commerciale)
- GnomAD (dati esoma + genoma).
- Dfam
- DGV
- LOVD

Inoltre, sono disponibili i seguenti punteggi *"in silico"*:

- Punteggi di splicing (*MaxEntScan, GeneSplicer, dbSCSNV*)
- Punteggio di conservazione (*GERP++, PhastCons100way, PhastCons20way, PhyloP100way Vertebrati, PhyloP20way Mammiferi, Scores*).
- Functional Prediction (*REVEL, DANN, FATHMM, FATHMM-MKL, LRT, MetaLR, MetaSVM, MutationAssessor, MutationTaster 2, PROVEAN, SIFT, PLI*)

Visualizzazione e interpretazione

La scheda Variants  consente di visualizzare, filtrare e annotare le varianti ottenute durante un'analisi. Queste vengono visualizzate sotto forma di tabella, con ogni riga corrispondente a una diversa variante. Il numero di varianti selezionate dopo il filtraggio (selected variants) e il numero totale di varianti rilevate (total number of variants) sono visualizzati nell'angolo superiore destro dell'intestazione della tabella delle varianti mediante la seguente icona:

 . Una paginazione in fondo alla pagina consente di navigare da una pagina di risultati alla successiva. Se l'analisi utilizza un "Restricted panel", l'intestazione della tabella includerà anche il "total number of variants" nel "Restricted panel" (*numero totale di varianti nel "Restricted panel"*).

La navigazione è facilitata da scorciatoie da tastiera che utilizzano i tasti ← e → della tastiera.

Per impostazione predefinita, nella tabella vengono visualizzate diverse colonne descrittive. Gli utenti possono esportare la tabella come file TSV facendo clic su .

Per ogni variante possono essere disponibili più trascritti, il cui numero è indicato a sinistra di ogni riga. Le varianti sono annotate sui transcripts RefSeq, ad eccezione delle varianti mitocondriali che sono annotate su *Ensembl*.

I transcripts di MANE Select e MANE Clinical Plus sono identificati da

un'icona, rispettivamente  o , accompagnata da un tooltip.

TRANSCRIPT	HGVS
MANE Select transcript	
NM_005962.5	c.536T>G p.V179G
NM_015252.5	c.-296+2T>G
MANE Plus Clinical transcript	
NM_000132.4	c.5375T>G p.V1792G

Per la stessa variante, ad eccezione delle varianti mitocondriali, i trascritti sono ordinati secondo l'ordine seguente:

- l'impatto ['high' > 'moderate' > 'low' > 'modifier']
- tipo di transcripts ['NM', 'XM', 'NR', 'XR'] (per i transcripts RefSeq)
- Tipo di transcript: "Mane Select > MANE Clinical Plus > non MANE" [Mane Select > MANE Clinical Plus > non MANE]
- ordine alfanumerico

Per le varianti mitocondriali, i trascritti sono ordinati nel seguente ordine:

- impact ['high' > 'moderate' > 'low' > 'modifier']
- se l'impatto è "modifier":
 - l'effetto "non_coding_transcript_exon_variant".
 - il biotipo [protein coding > Mt_rRNA > Mt_tRNA] [protein coding > Mt_rRNA > Mt_tRNA].

c) ordine alfanumerico

DiagAI. Questa colonna mostra il punteggio da 0 a 100 in base alla suite DiagAI che utilizza l'apprendimento automatico. Il punteggio DiagAI Score. si basa su quattro componenti:

- L'algoritmo UP², che predice la patogenicità della variante
- La matrice PhenoGenius, che valuta la corrispondenza tra il gene e gli 'HPOs' del paziente (se questi sono stati inseriti durante l'analisi)
- La concordanza tra il genotipo osservato e la modalità di trasmissione della malattia descritta nei database GenCC, OMIM e PanelApp (Australia e England).
- La qualità della Variant.

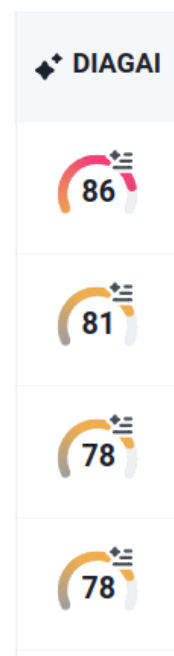
Più il punteggio è vicino a 100, più si prevede che la variante sia patogena e di interesse diagnostico. Più il punteggio è vicino a 0, più si prevede che la variante sia benigna.

Due icone richiamano inoltre l'attenzione su alcune varianti di interesse:

- L'icona  indica che la variante ha una maggiore probabilità di essere la causa della malattia (AI Smart Pick).
- L'icona  indica che la variante è presente nella short list delle varianti con un'alta probabilità di essere la causa della malattia del paziente.

I termini HPO possono essere inseriti prima dell'avvio dell'analisi modificando i metadati del campione (vedere la sezione **Metadati del campione**). Possono anche essere modificati durante l'interpretazione utilizzando la modal view HPOs nel pannello laterale dell'analisi (vedere la sezione **HPOs modal view**). In questo caso, l'utente può rielaborare il DiagAI Score e la classificazione delle varianti (Short list o SmartPick) in base ai nuovi HPOs inseriti, utilizzando la funzione "Re-process with new HPOs".

Se l'utente fa clic sull'icona del DiagAI Score, la pagina della variante si apre direttamente nella scheda DiagAI con i dettagli del punteggio (cfr. DiagAI nella sezione Variant.).



ACMG. La 'Pathogenicity classification' di ogni variante è calcolata nell'ambito di un processo semi-automatico secondo le raccomandazioni dell'*American College of Medical Genetics and Genomics*. Gli utenti possono rivedere la classificazione in una scheda dedicata nella pagina della variante, che viene poi utilizzata per generare la valutazione finale di una variante.

ComPerMed. La 'Pathogenicity classification' di ogni variante è calcolata nell'ambito di un processo semi-automatico, secondo le raccomandazioni del comitato di esperti belga "ComPerMed" (Personalised Medicine Commission). Gli utenti possono accedere a informazioni dettagliate sulla classificazione in una scheda dedicata nella pagina della variante.

AMP. (versione di analisi SomaVar v1.8 o successiva, solo SomaVar LF o SomaCGP). Questa colonna viene visualizzata per impostazione predefinita nelle analisi che contengono annotazioni CKB e mostra il livello AMP di ciascuna variante, se disponibile (livello I o II). Il livello AMP visualizzato si basa sulle migliori evidenze cliniche per tutti i tipi, indipendentemente dall'indicazione selezionata.

VKB. Le valutazioni effettuate dall'utente durante le analisi precedenti vengono memorizzate nella sua base di conoscenze. Questa valutazione appare nella colonna VKB della tabella per ogni nuova occorrenza della variante. Un tooltip visualizza rapidamente la classe e il commento associati alla variante valutata o precedentemente valutata.

☐	Sort by: DiagAI RUO							
☐								
☐	5	81	5	4	BRCA1			NM_007294.4

Gene. Denominazione del gene secondo [HGNC](#) e flag OMIM associato:

- Bandiera rossa: "OMIM gene" **con** fenotipi e caratteristiche cliniche",
- Contorno rosso: "OMIM gene" **con** fenotipi ma **senza** caratteristiche cliniche,
- Punto rosso: "OMIM gene" **senza** fenotipi o caratteristiche cliniche".

	GENE	TRANSC
☐	RPE65	NM_000
☐	OMIM	
☐	Decipher	1_206
☐	Orphanet	
☐	MedGen	1_005
☐	GenCC	

I geni noti come geni di imprinting secondo "PanelApp Australia" sono evidenziati dall'icona . Il tooltip associato indica se il gene è espresso maternamente, paternamente, sconosciuto o se si fa riferimento a più imprinting.

Facendo clic sulla bandierina accanto al nome del gene si accede direttamente alla pagina OMIM. Facendo clic sul nome del gene, appare un menu a discesa che fornisce collegamenti ai database OMIM¹², Decipher¹³, Orphanet¹⁴, MedGen¹⁵ e GenCC¹⁶. Ogni link porta direttamente alla pagina corrispondente al gene selezionato.

Le informazioni relative a OMIM e ai database sono disponibili solo per le analisi germline.

HGVSC. Questa colonna fornisce l'annotazione HGVS della variante sul DNA complementare (cDNA) e sulla proteina.

HGVSC
c.10580G>A p.R3527Q

Effect. Questa colonna descrive l'impatto della variante sulla proteina e l'effetto sul transcript.

L'impatto è calcolato in base a 4 livelli:

- **High** (high impact, in rosso): la variante ha un elevato impatto distruttivo sulla proteina, causando probabilmente il troncamento della proteina, la perdita di funzione o innescando il suo decadimento nonsense mediato.
- **Moderate** (moderate impact, in giallo): variante non distruttiva in grado di modificare l'efficacia della proteina.
- **Low** (low impact, in grigio): si presume che la variante sia innocua o che non modifichi il comportamento della proteina.
- **Modifier** ("modifying impact", in blu): in genere varianti non codificanti o che interessano regioni o geni non codificanti, per le quali le previsioni sono difficili o non vi è alcuna evidenza di impatto.

EFFECT
Moderate ● missense
Low ● splice polypyrimidine tract + 1
Moderate ● missense
Moderate ● missense
Moderate ● missense
Moderate ● missense

Per ogni variante, nella colonna Effect possono essere disponibili diversi effetti sul transcript. In questo caso, viene visualizzato solo l'effetto con la conseguenza più elevata secondo il VEP¹⁷, seguito da un numero (ad esempio +2) per gli altri effetti disponibili. I dettagli di questi ultimi sono mostrati nel tooltip corrispondente.

¹² <https://www.omim.org/>

¹³ <https://www.deciphergenomics.org/genes>

¹⁴ <https://www.orpha.net/fr/disease/gene>

¹⁵ <https://www.ncbi.nlm.nih.gov/medgen/>

¹⁶ <https://search.thegencc.org/>

¹⁷ http://www.ensembl.org/info/genome/variation/prediction/predicted_data.html

Transmission. Questa colonna indica la modalità di trasmissione delle malattie associate ai geni riportati nei database GenCC e OMIM. I badge colorati sono visibili per identificare rapidamente i modelli di trasmissione basati sulle previsioni DiagAI:

- "De Novo"
- "Compound Het - Verified"
- "Recessive Hom"
- "Compound Het - Suspected"

I badge "De Novo", "Compound Het - Verified" e "Recessive Hom" vengono visualizzati solo nelle analisi familiari.

È possibile utilizzare un filtro per selezionare le varianti con questi contrassegni (vedere la **sezione "Selection variants" - Transmission & Phenotypes**).

TRANSMISSION	
AD AR	CompHet Verified Nail disorder, nonsyndromic ... +1
AD	CompHet Suspected Coffin-Siris syndrome 2
AD	De Novo Advanced sleep phase syndrome..
AR	Recessive Hom Advanced sleep phase syndrome..
AD	Tarsal-carnal coalition syndrome +3

Transmission
AR, AD

Leber congenital amaurosis 9
OMIM:7638289

AD, AR OMIM
AD 3 2 GenCC
AR 3 GenCC

Leber congenital amaurosis 2
OMIM:0019200

AD, AR OMIM
AR 3 GenCC

Leber congenital amaurosis 5
MONDO:739002

AR 3 GenCC

Retinis pigmentosa
OMIM:10939229

AD, AR OMIM

Definitive to Moderate
Supportive
Limited to Unknown

Facendo clic sui link alle modalità di trasmissione in questa colonna si aprirà una finestra popup in cui gli utenti potranno trovare le malattie elencate con i loro nomi e i link ai database (OMIM, MONDO - se OMIM non è disponibile - o GenCC).

Le malattie sono organizzate in una gerarchia specifica: quelle trovate sia in OMIM che in GenCC appaiono per prime, seguite dalle voci solo OMIM e poi da quelle solo GenCC.

Ogni voce comprende etichette con codice colore: Le etichette OMIM hanno bordi verdi, mentre quelle GenCC utilizzano un sistema a semaforo (verde/giallo/rosso) per indicare la forza delle prove:

- Verde: Definitivo + Strong ("forte") + Moderato ("moderato")
- Giallo: "Supporting"
- Red: Limitato, Animale, Unknown)

Clinical Evidences. (Analisi SomaVar v1.8 o successive, solo SomaVar LF o SomaCGP). Questa colonna viene visualizzata per impostazione predefinita nelle analisi contenenti annotazioni CKB. È suddivisa in tre sottocolonne, ognuna delle quali contiene il numero totale di un determinato tipo di evidenza: A (Actionability), D (Diagnosis) e P (Prognosis). La "Columns" mostra anche il numero di asserzioni terapeutiche sensibili (a sinistra) e resistenti (a destra) come badge.

CLINICAL EVIDENCES		
A	D	P
81 81 8	0	0
54 53 2	0	0
54 53 2	0	0

I valori di queste sottocolonne sono codificati a colori:

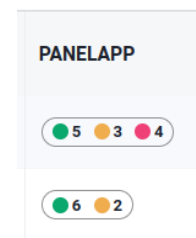
- La resistenza, se trovata (> 0 asserzioni), viene sempre visualizzata in rosso, indipendentemente dall'indicazione, e in grigio altrimenti.
- La sensibilità è evidenziata in verde solo quando è selezionata l'indicazione corrispondente e in grigio altrimenti.

Se almeno un'evidenza di questo tipo corrisponde all'indicazione selezionata nella tabella delle varianti (pannello Indication), nel tooltip compare un segno di spunta davanti all'etichetta dell'indicazione. Se non viene trovata alcuna corrispondenza, l'etichetta non appare.

SEO-027 IFU IT V6 Pubblicato il 2026-04-29 98/188

PanelApp. Questa colonna integra le annotazioni del database dei pannelli *PanelApp Australia*¹⁹ e *PanelApp England*²⁰. Indica il numero di panel associati al gene per il quale è stata rilevata una variante, secondo il livello di classificazione descritto di seguito.

Se un gene non viene trovato in un panel, la colonna visualizzerà "N/A".



La finestra pop-up associata visualizza i dettagli dell'associazione gene/fenotipo, nonché la modalità di trasmissione associata al fenotipo, la modalità di trasmissione associata al fenotipo e il database da cui è derivato il panel, PanelApp Australia (**AUS**) o PanelApp England (**ENG**) (**Figura 12**).

A ciascun gene verranno assegnati uno o più panel secondo un sistema di classificazione a tre livelli basato sull'evidenza dell'associazione tra gene e fenotipo: 'High' in verde, 'Moderate' in giallo e 'Low' in rosso.

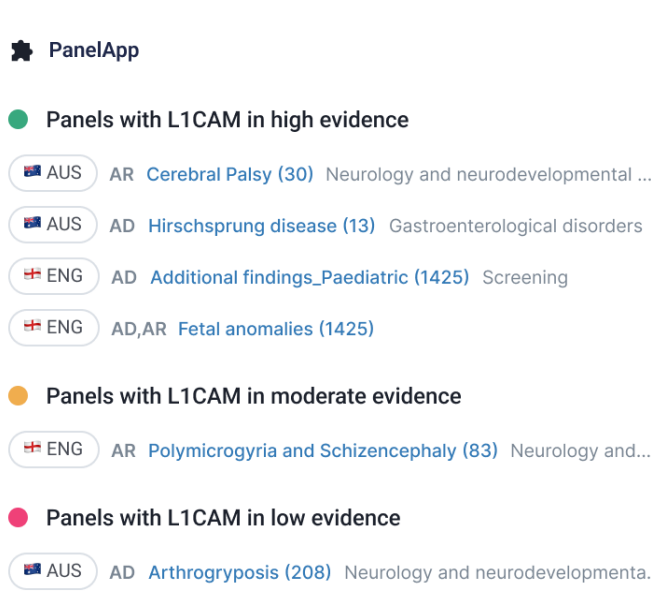
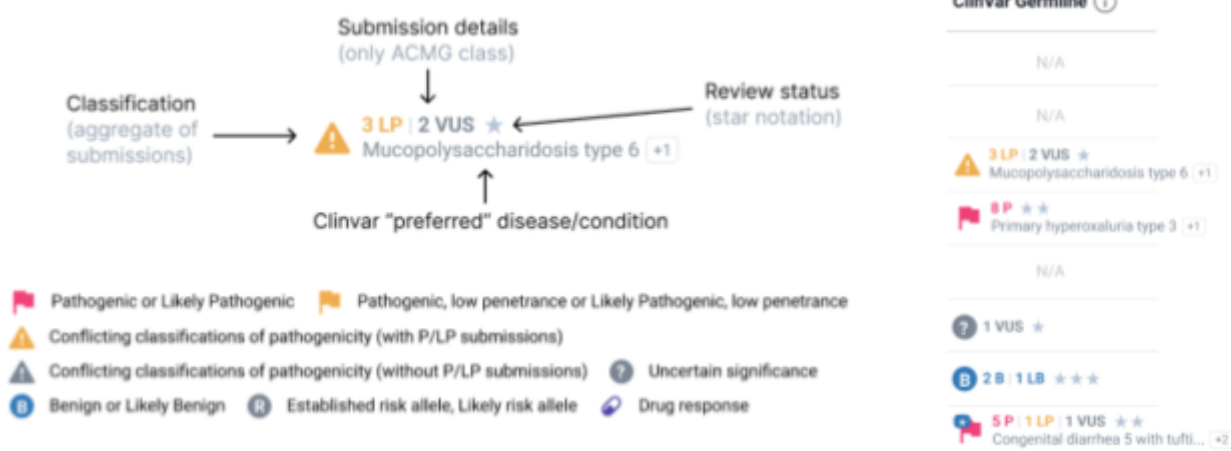


Figura 12: Visualizzazione delle informazioni del PanelApp.

¹⁹ <https://panelapp-aus.org>

²⁰ <https://panelapp.genomicsengland.co.uk/>

ClinVar Germline. Questa colonna visualizza le informazioni estratte da "ClinVar" durante l'analisi. Fornisce una serie di informazioni, tra cui la classe di patogenicità, i dettagli e il numero di invii per ogni classe ACMG, lo stato di revisione e la condizione/malattia preferita da ClinVar. Facendo clic sulla cella della tabella delle varianti, l'utente viene reindirizzato alla pagina ClinVar' corrispondente.



Per ulteriori informazioni, vedere "[ClinVar](#)".

GnomAD. Questa colonna visualizza 6 elementi del database GnomAD per una specifica variante nella popolazione: la frequenza allelica (AF), il numero di alleli (AC), il numero di omozigoti (HOM), il numero di emizigoti (HEM), la frequenza dell'allele maggiore (MAF) e, per i progetti GRCh37, la versione di GnomAD per cui si trova questa variante (VER).

Facendo clic sul valore visualizzato nella colonna AF, l'utente viene reindirizzato al database di "GnomAD" nella sua versione v4 o, in alternativa, nella sua versione v2. Se una variante non è presente nel database, il valore visualizzato in questa colonna sarà "N/A".

GNOMAD				PROJ
	AC	HOM	VER	
e-1	47775	9503	v2	4 / 4

Nella colonna VER viene visualizzato un avviso se la variante è presente in "GnomAD", ma non supera tutti i filtri di qualità per almeno uno dei due set di dati (esoma o genoma). In questo caso, i valori visualizzati si adattano ai valori disponibili per questa variante e vengono visualizzati i filtri che hanno innescato l'avviso, insieme a un breve messaggio che menziona i set di dati incriminati.

Score UP². Lo Score UP² (Universal Pathogenicity Prediction), che va da -1 a 1, classifica la patogenicità delle varianti utilizzando un approccio di Machine Learning secondo tre categorie: **P** (Pathogenic: punteggio > 0,1), **"VUS"** (Variant of Unknown Significance: punteggio ≥ -0,1 e ≤ 0,1) o **B** (Benign: punteggio < -0,1). Il punteggio UP² è visualizzato di default nella tabella delle varianti, ma l'utente può disattivare la colonna UP² dal pannello di gestione delle colonne

facendo clic sul pulsante  .

'CI-SpliceAI'. Questa colonna incorpora i punteggi di predizione calcolati utilizzando un approccio di Machine Learning per identificare le varianti con un impatto sullo splicing alternativo alla maniera di "CI-SpliceAI".²¹ .

SpliceAI				
AG	DG	AL	DL	Link
0.01	0.02	0.62	0.84	↗
0.01	0.22	0.01	0.09	↗
0.01	0.02	0.12	0.01	↗
0.01	0.62	0.02	0.01	↗
N/A	N/A	N/A	N/A	↗
No splicing impact				
SpliceAI score was not significant (Delta score <0.2)				

Gli score sono calcolati per un sottoinsieme di varianti che soddisfano i seguenti criteri: varianti di splicing secondo la VEP (splice_*) con una frequenza 'GnomAD' inferiore a 10⁻³e che soddisfano i seguenti criteri di qualità (QA > 20, AO > 10, DP > 20).

Se calcolata, la colonna fornisce 4 punteggi delta per la variante in questione, per guadagno dell'accettore (AG), perdita dell'accettore (AL), guadagno del donatore (DG) e perdita del donatore (DL).

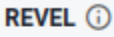
Il punteggio delta di una variante, che va da 0 a 1, corrisponde alla probabilità che la variante modifichi lo splicing. Alle soglie del delta score viene applicato un codice colore: **verde** (>0,2; alto richiamo), **giallo** (>0,5; raccomandato) e **rosso** (>0,8; alta precisione). Se il punteggio delta è inferiore a 0,2, consideriamo il punteggio insignificante. La colonna specifica quindi "No splicing impact".

Per ogni variante, viene fornito un link all'interfaccia corrispondente di 'SpliceAI' lookup²² .

²¹ https://github.com/YStrauch/CI-SpliceAI__Annotation

²² <https://spliceailookup.broadinstitute.org/>

REVEL. Il punteggio di predizione REVEL si basa su un metodo complessivo per la previsione della patogenicità delle varianti missense che utilizza una combinazione di punteggi di 13 strumenti individuali: MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP e phastCons.

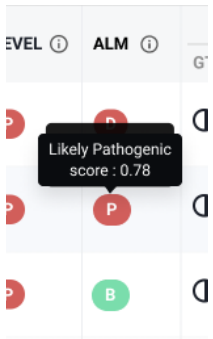


B

P

Il punteggio può variare da 0 a 1 per una variante missenso, con punteggi più alti che riflettono una maggiore probabilità che la variante causi la malattia. Le soglie applicate sono le seguenti: Punteggio REVEL > 0,5 per **P** (Pathogenic) e punteggio REVEL < 0,5 per: **B** (Benign). Il punteggio può essere visualizzato dall'utente nella tabella delle varianti selezionando la casella REVEL nell'inserto "Columns", nella sezione "Functional Prediction".

ALM. Il punteggio di predizione di AlphaMissense²³ è disponibile nella colonna ALM. AlphaMissense è un modello di previsione che valuta la patogenicità delle varianti missense utilizzando il deep learning. Il modello si basa sulle previsioni strutturali di AlphaFold per migliorare la sua precisione.



AlphaMissense integra la conservazione evolutiva e il contesto strutturale per valutare il potenziale impatto delle varianti missense sulla funzione delle proteine.

Il punteggio varia da 0 a 1, con valori più alti che indicano una maggiore probabilità che una variante sia deleteria. Le soglie applicate sono: Punteggio ALM > 0,564 per **P** (Likely Pathogenic) e punteggio ALM < 0,34 per **B** (Likely Benign). Negli altri casi (0,34 < punteggio ALM < 0,564), sarà etichettata come **VUS** (Ambiguous).

Sample Freq. Questa colonna mostra per ogni variante la sua frequenza nel progetto (sottocolonna PROJECT) e in tutti i progetti dell'entità con metadati simili (sottocolonna "ALL").

La Sample frequency del progetto (PROJECT) corrisponde al numero di campioni in cui viene rilevata la variante rispetto al numero totale di campioni analizzati nel progetto.

SAMPLE FREQ.	
PROJECT	ALL
2 / 13	7.41%
2 / 13	7.41%

²³ <https://alphamissense.hegelab.org/>

La Sample Freq. ALL corrisponde alla percentuale di campioni analizzati in cui la variante è presente nell'entità, all'interno di progetti di metadati simili. Facendo clic sul valore visualizzato nella colonna "ALL", l'utente può accedere ai nomi dei progetti e al numero di campioni in cui è presente la variante, nonché ai dettagli di tutti i progetti di metadati simili (opzione "Show all projects").

Per impostazione predefinita, la sottocolonna ALL è nascosta nella tabella delle varianti, ma può essere riattivata dal pannello di gestione delle colonne facendo clic sull'icona  e attivando il pulsante "Sample Frequency Entity (Entity)".

Exon. Questa colonna indica il numero totale di esoni del gene in base al trascritto in questione, nonché il numero di esoni interessati dalla mutazione, sotto forma di frazione. Un tooltip indica la distanza tra la variante e il sito di splicing più vicino. Se la variante è a monte del sito di splicing, il valore sarà positivo, mentre se è a valle sarà negativo.

GT (Genotype). Questa colonna mostra il genotipo previsto dai programmi di rilevamento delle varianti. I vari pittogrammi possibili sono descritti nella sezione **Filtraggio delle varianti**.

In un'analisi di trio (GermlineFamily), se è possibile determinare l'origine di una variante eterozigote, questa viene indicata con il seguente codice colore:

- **pink** → Origine materna eterozigote
- **blue** → origine paterna eterozigote
- **red** → *De novo* variante eterozigote
- **arancione** → violazione mendeliana

Non viene applicato alcun colore se la trasmissione è indeterminata (ad esempio, se entrambi i genitori sono eterozigoti) o per le varianti omozigoti.

Patient				Father				Mother			
GT	VAF	QUAL	COV	GT	VAF	QUAL	COV	GT	VAF	QUAL	COV
	50.8%	34	61		N/A	N/A	82		44.6%	33	83
	53%	34	83		N/A	N/A	49		N/A	N/A	100
	46.9%	34	64		48.8%	33	84		N/A	N/A	86

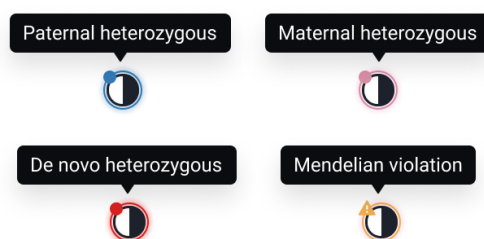


Figura 13: Codice colore applicato ai genotipi eterozigoti in base alla trasmissione GermlineFamily.

La fasatura delle varianti è disponibile per le analisi GermVar Tertiary e GermVar Tertiary Family che utilizzano i file VCF dei sequenziatori Oxford Nanopore Technologies. Quando le varianti sono fasate, le icone del genotipo appaiono come segue : . Il segno di spunta indica che la variante è fasata. Le varianti fasate con la parte nera sullo stesso lato e appartenenti allo stesso "Phase ID" sono in fase, cioè sullo stesso filamento.

È possibile fare clic sul genotipo per ottenere un elenco di varianti cis o trans, se presenti. Vengono presentate solo le 5 varianti classificate più in alto da DiagAI.

L'elenco completo delle varianti è disponibile cliccando su "Browse all variants".

Variants in trans
Phase set ID 283928 X

	GT
Compound heterozygous candidates for	
SLC25A22 chr11:g.792142C>T	
SLC25A22 chr11:g.792146_792147del	

VOF (Variant Allele Frequency). Questa colonna indica la "Variant Allele Frequency", ossia la proporzione di reads che supportano l'allele alternativo (AO) rispetto al numero totale di reads in quella posizione (COV), espressa in percentuale.

Le analisi che utilizzano i worksets SomaVar versione ≥ 2.5 , SomaCGP versione ≥ 0.7 e SomaVar LF versione ≥ 2.1 mostrano le seguenti informazioni aggiuntive:

- un tooltip mostra i valori di 'VOF', 'AO' e 'COV' calcolati dai due 'variant callers' Freebayes' e 'Mutect2' che rilevano la variante.
- un'icona di avvertimento  avvisa l'utente di eventuali disaccordi tra i due "variant callers" per le varianti rilevate con un 'low' VOF. Pertanto, per le varianti che mostrano un VOF inferiore al 5% nella colonna VOF, l'avviso viene visualizzato in caso di discrepanza VOF > 10% tra i VOF calcolati da Freebayes e Mutect2.

AO (Alternate Observation). Questa colonna indica il numero di reads portatori della variante.

CAL (Caller). Questa colonna indica il nome del (o dei) *variant callers* che hanno rilevato la variante.

Cov (Coverage). Questa colonna indica il numero di read totali nella posizione della variante.

SOR. Lo Strand Odds Ratio (SOR) è una metrica del bias di filamento derivata dal test dell'Odds Ratio Simmetrico che misura lo squilibrio delle letture a supporto della variante tra i filamenti forward e reverse. Un SOR più elevato suggerisce una maggiore probabilità di un artefatto, mentre un SOR più basso suggerisce un supporto del filamento più bilanciato ed è più coerente con una variante reale.

Ordinamento delle varianti

Nella scheda Variant sono disponibili diverse opzioni per l'ordinamento delle varianti. Queste sono disponibili in un menu a discesa nell'angolo superiore destro della tabella. Per ogni opzione di ordinamento, è possibile ordinare in ordine crescente o decrescente facendo clic sull'icona di ordinamento . Per impostazione predefinita, tutte le opzioni sono ordinate in ordine decrescente, ad eccezione dell'opzione "Locus", che è ordinata in ordine crescente.

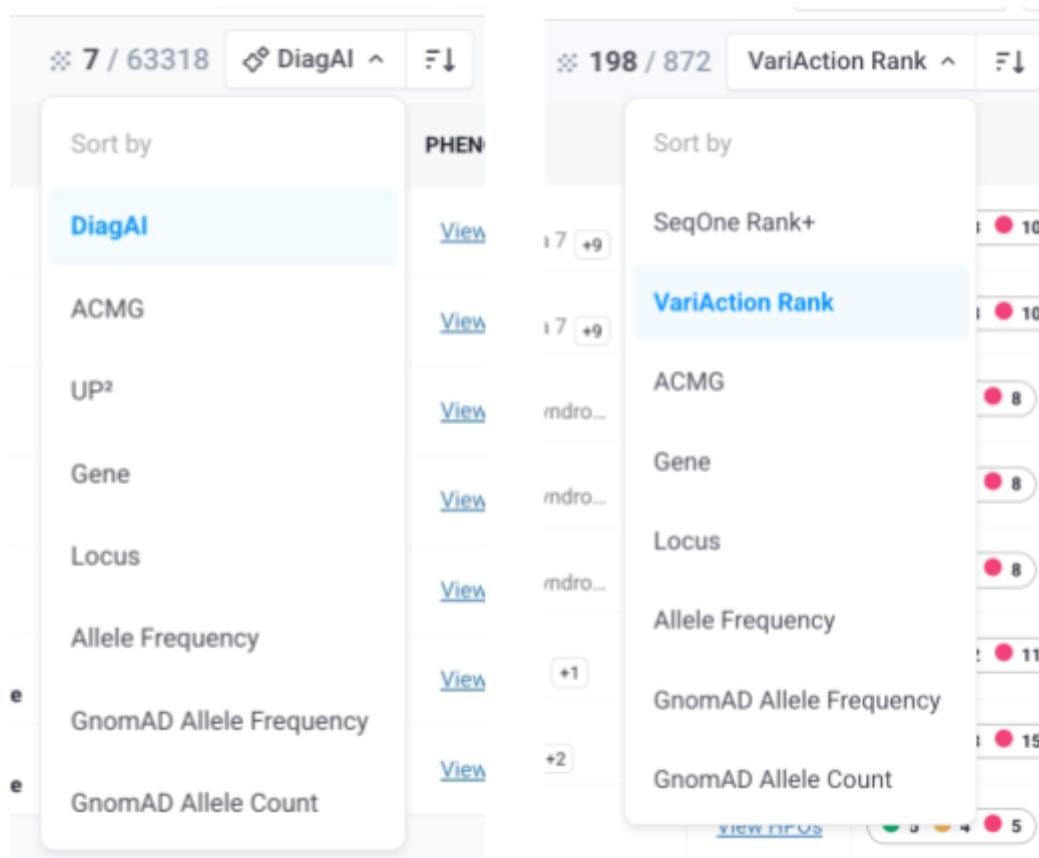


Figura 14: Finestra di selezione dell'opzione di ordinamento della variante

DiagAI Score. Questa opzione di ordinamento viene applicata per impostazione predefinita a tutte le analisi germinali e consente di ordinare le varianti in base al loro "DiagAI Score" (sostituisce "SeqOne Rank+" nelle analisi effettuate dopo il 2 ottobre 2024).

SeqOne Rank+. L'algoritmo di prioritizzazione "SeqOne Rank+", basato su un modello di intelligenza artificiale più vecchio di "DiagAI Score", rimane disponibile e viene applicato di default a tutte le analisi effettuate prima del 2 ottobre 2024.

Si basa su 4 componenti principali:

- Una nuova implementazione della classificazione ACMG, basata sull'apprendimento automatico addestrato su "ClinVar" (punteggio UP²).
- Il "phenotypes" del paziente (se al campione sono associati termini di HPO)
- La concordanza tra il genotipo osservato e la modalità di trasmissione della patologia descritta nei database GenCC, OMIM e PanelApp (Australia).
- Qualità Variant.

VariAction Rank. L'algoritmo "VariAction Rank" per la prioritizzazione delle varianti somatiche è proposto nelle analisi di SomaVar dalla versione v1.8. Questo sistema utilizza l'annotazione JaxCKB per calcolare i punteggi di priorità delle varianti in base alla loro azionabilità clinica. Questi punteggi di priorità tengono conto dell'impatto della variante, della sua patogenicità utilizzando il punteggio "ComPerMed", dell'AMP Tier, della specificità della variante con il profilo molecolare, nonché della corrispondenza con l'indicazione, se la malattia è stata inserita prima dell'avvio dell'analisi.

ACMG. Le varianti possono essere ordinate in base alla loro classificazione ACMG selezionando questa opzione. Per impostazione predefinita, le varianti sono ordinate in ordine decrescente dalla classe Pathogenic 5 alla classe Benign 1.

Locus. Le varianti possono essere ordinate in base alla loro posizione genomica selezionando questa opzione.

Allele frequency. Le varianti possono essere ordinate in base alla frequenza allelica selezionando questa opzione. Per impostazione predefinita, le varianti vengono ordinate in ordine decrescente. È possibile ordinare in ordine crescente facendo clic sull'icona di

ordinamento . 

GnomAD Allele Frequency, GnomAD Allele Count. Le varianti possono essere ordinate in base alla frequenza allelica (Allele Frequency : AF) o al conteggio allelico (Allele Count : AC) di GnomAD in ordine crescente o decrescente (se i dati GnomAD sono disponibili).

UP². Le varianti possono essere ordinate in base al punteggio di patogenicità UP², un componente di DiagAI Score.

Gene. Le varianti possono essere ordinate in ordine alfabetico crescente o decrescente in base al simbolo del gene.

Per le varianti che interessano geni sovrapposti, il simbolo del gene utilizzato per l'ordinamento è quello associato alla variante con l'impatto più elevato, in base alle regole di prioritizzazione dei trascritti descritte sopra. Quando si espande l'elenco dei trascritti, l'utente potrebbe vedere il secondo gene sovrapposto, che potrebbe apparire non ordinato. L'ordinamento potrebbe anche sembrare non applicato quando alla tabella delle varianti vengono applicati più filtri.

Configurazione delle colonne della tabella delle varianti

Le colonne visualizzate nella tabella delle varianti possono essere configurate dal menu

"Organise your columns", accessibile tramite il pulsante  in alto a destra della tabella delle varianti. La visualizzazione delle colonne può essere modificata dall'utente facendo clic sulle icone  per attivare le colonne desiderate nella sezione di sinistra. La sezione di destra visualizza le colonne attive nell'analisi e consente agli utenti di creare e salvare profili di colonne da un menu a discesa facendo clic sull'icona "Select profile".

Le colonne attive possono essere riordinate facendo clic e trascinandole nella posizione desiderata. Per impostazione predefinita, le colonne sono organizzate in base ai profili "Germline" e "Somatic".

Quando gli utenti passano il mouse sul nome di un profilo di colonna nel menu a discesa delle preimpostazioni, un tooltip mostra quale workset ha creato quel profilo. Questo avverte gli utenti se il loro profilo potrebbe avere colonne incompatibili con il workset corrente.

Per creare un nuovo profilo di colonna, è sufficiente inserire un nome nel campo "Create a new preset" e fare clic su "Save".

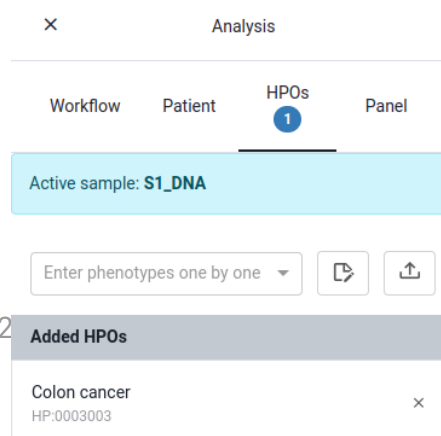
Modifica dei metadati del campione dal pannello di analisi

L'utente può modificare i metadati del campione direttamente dall'analisi e inserire il sesso del paziente o gli HPOs (solo per le analisi germinali) dal cassetto Analisi, accessibile facendo clic sul pulsante  Analysis in alto a destra della pagina (vedere la sezione **Pannello laterale per la gestione delle analisi**).

Dalla vista Patient, l'utente accede alla modifica del genere (vedere la sezione **Vista modale Patient**).

Nelle analisi dei progetti Germline, dalla vista HPOs, l'utente ha accesso a diverse funzioni di modifica dei termini HPOs:

- Visualizzare i termini già inseriti prima di lanciare l'analisi
- Aggiunta manuale di codici HPO o fenotipi tramite la barra di ricerca "Enter phenotypes one by one".



Analisi

Workflow Patient HPOs 1 Panel

Active sample: S1_DNA

Enter phenotypes one by one

Added HPOs

Added HPOs
Colon cancer HP:0003003

- Importazione di un file di testo contenente i codici HPO
- Generazione di termini HPO in base alla descrizione clinica del paziente, richiamando "DiagAI" tramite l'icona . Questa funzione richiede all'utente di descrivere il profilo clinico del paziente in inglese (descrizione limitata a 4000 caratteri). Quindi, dopo aver cliccato su "Extract HPOs with DiagAI", viene generato un elenco di HPOs evidenziati a colori. Infine, l'utente può associare tutti i termini HPO proposti al paziente cliccando su "Add all", oppure selezionarli singolarmente passando il mouse sui termini desiderati e cliccando poi su "Add".

L'aggiunta e la modifica di HPOs da questa vista offre all'utente la possibilità di rielaborare il DiagAI Score e la classificazione delle varianti (Short list o SmartPick) in base ai nuovi HPOs inseriti tramite la funzionalità "Re-process analysis".

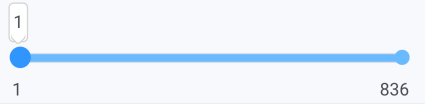
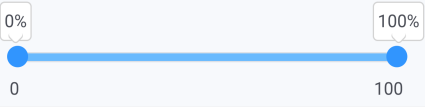
Varianti di filtraggio

È possibile affinare le ricerche nella scheda *Variant* utilizzando le categorie di filtri descritte di seguito. È possibile accedere al pannello dei filtri tramite l'icona  situata in alto a sinistra nella tabella delle varianti o premendo il tasto F sulla tastiera. I filtri sono disponibili nella scheda "Basic". Nella parte superiore del pannello dei filtri, la sezione "Active Filters" mostra il numero di filtri attualmente applicati.

Filtri Genes & Variants

Nome e panoramica del filtro	Descrizione
v Genes ⓘ Search gene symbols	Limita l'elenco delle varianti a uno o più geni.
v Variants ⓘ Search by HGVS	Limita l'elenco delle varianti a una o più varianti secondo la nomenclatura HGVS o HGVS.

Sezione Samples & Quality

Nome e panoramica del filtro	Descrizione
v Coverage ⓘ 	Definisce una soglia minima per la profondità di copertura (COV). Il valore massimo visualizzato da questo filtro corrisponde al valore di copertura più alto identificato tra le varianti.
v Allele Frequency ⓘ 	Filtra le varianti in base alla frequenza dell'allele mutante nel campione analizzato.
v Quality ⓘ 	Qualità media (punteggio phred) di ciascuna base portatrice dell'allele mutato, che indica la fiducia nel rilevamento della variante: - <i>Medium</i> ($\geq \text{QUAL}=20$): 1 possibilità su 100 di chiamata errata, - <i>High</i> ($\geq \text{QUAL}=30$): 1 possibilità su 1000 di chiamata errata. Questo filtro si applica a tutti i campioni.
v Quality ⓘ 	Per le analisi GermVar Tertiary e GermVar Tertiary Family sui dati di Oxford Nanopore Technologies, la qualità basata sul punteggio Phred indica la fiducia nel rilevamento delle varianti. $\geq \text{QUAL}=8$ (Low): 1 possibilità su 8 di chiamata errata, $\geq \text{QUAL}=20$ (Medium): 1 possibilità su 100 di chiamata errata,

	≥ QUAL=30 (High):: 1 possibilità su 1000 di chiamata errata. Questo filtro si applica a tutti i campioni.
Genotype ⓘ Homozygous 387 Heterozygous 1163 None 1942 Biallelic Variant 222 Biallelic Variant 84 Unknown 76 Unknown 0	Varianti Filters" in base al genotipo osservato.
Alternate Observations ⓘ 0 0 458	Definisce un numero di sequenze portatrici della mutazione (Alternate Observation o AO). Il valore massimo visualizzato da questo filtro corrisponde al valore AO più alto identificato tra le varianti.
Strand Odds Ratio ⓘ 0 0 1 2 3 4 ≥5 min 0 — max 5	Filtra le varianti in base al punteggio strand odds ratio (SOR) calcolato nel sito della variante.

Sezione Frequencies

Nome e panoramica del filtro	Descrizione
▼ Sample frequency ⓘ all all rare	Elimina le varianti in base alla loro ricorrenza nel progetto in esame, dalla più frequente (all) alla meno frequente (rare).
▼ Relative pop frequency ⓘ all 10-2 10-3 10-4 10-5 10-6 absent	Seleziona le varianti in base alla frequenza del loro allele mutante nella popolazione combinata dei progetti ExAc e gnomAD.
▼ Absolute pop frequency ⓘ 0 1609917	Seleziona le varianti in base al numero di alleli nella popolazione combinata di ExAc e GnomAD.
▼ Absolute homozygous pop frequency ⓘ 0 803576	Seleziona le varianti in base alle loro frequenze di omozigosi nella popolazione combinata ExAc e GnomAD.

Sezione Genetics

Nome e panoramica del filtro	Descrizione
▼ Impact ⓘ Low Moderate High ☒ Keep <i>Modifier</i>	L'impatto della mutazione sulla proteina. <i>Modifier</i> ("modificatore"): influisce sui geni non codificanti.
▼ Effects ⓘ > Loss of Function 10 • Transcript ablation 10 • Splice acceptor 10 • Splice donor 10 • Stop gained 10 • Frameshift 10 • Stop lost 10 • Start lost 10 • Transcript amplification 10 ☑ Loss of Function ☑ Transcript ablation ☑ Splice acceptor ☑ Splice donor ☑ Stop gained ☑ Frameshift ☑ Stop lost ☑ Start lost ☑ Transcript amplification	termini SO (Sequence Ontology) che descrivono le conseguenze più dannose a livello genico. Questo filtro consente di selezionare l'impatto della mutazione sul gene e le opzioni sono organizzate in categorie logiche per migliorare la navigazione: <i>Loss of Function</i>, <i>Coding</i>, <i>Splicing region</i>, <i>Silent variant</i>, <i>Intronic</i>, <i>UTR</i>, <i>ncRNA</i>, <i>Upstream/Downstream</i>, <i>Regulatory</i>, <i>Others</i>.
▼ Gene panel ⓘ Select panel ▼	Selezionare e applicare uno o più "Gene panel". Questo filtro visualizza tutti i "Gene panel" disponibili per l'entità. Questo filtro non è disponibile se all'analisi è stato applicato un pannello ristretto dalla scheda <i>Panel</i> del riquadro di gestione dell'analisi (vedere la sezione Panel modal view). Un messaggio informa l'utente che il filtro "Gene panel" non può essere utilizzato con un pannello ristretto.
▼ Panel in-silico ⓘ Dyslipidemia_panel_1_Twist★ ▼	Seleziona e applica un pannello "in silico". Questo filtro visualizza tutti i file BED disponibili per il gruppo. Il nome del file utilizzato durante l'analisi è evidenziato con una stella. Solo le varianti la cui posizione iniziale (start) si trova all'interno di uno degli intervalli del file letto selezionato vengono mantenute da questo filtro.

PanelApp ⓘ Epi ENG Blepharophimosis ptosis and epicanthus ENG Early onset or syndromic epilepsy ENG Epidermodysplasia verruciformis	Consente di filtrare le varianti in base a un elenco di pannelli in silico presenti nei cataloghi PanelApp Australia e PanelApp England. Un badge identifica il database da cui proviene il pannello "in silico": AUS (Australia) o ENG (Inghilterra). In entrambi i cataloghi, i superpanel non sono disponibili per la selezione, così come i panel che non includono almeno un gene. Ad esempio, un pannello contenente solo STRs non sarà selezionabile.
Gene occurrences ⓘ Usable only if they are fewer than 2000 selected variants.	Definisce una soglia per il numero di occorrenze di un dato gene nella tabella delle varianti.
Chromosomes & MT ⓘ X 25 Y 0 M 0	Mantiene le varianti localizzate sui cromosomi sessuali e/o sul cromosoma mitocondriale (M)
RefSeq transcripts ⓘ BARD1_2transcripts.txt	Seleziona un elenco di trascritti RefSeq per filtrare le varianti.

Sezione Databases

Nome e panoramica del filtro	Descrizione
Variant Databases ⓘ dbsnp 31 cosmic 31 civic 30 lovd 10 clinvar 3	Seleziona le varianti annotate dai seguenti database: dbsnp, cosmic, lovd, clinvar e civic.

Clinvar ⓘ P or LP 3 P/LP low penetrance 0 ⚠ Conflicting (with P/LP subs) 1 ⚠ Conflicting (without P/LP subs) 0 ? Uncertain significance 4 B B or LB 18 R Risk allele 0 💊 Drug response 2 ☐ Keep others	Consente di filtrare le varianti in base al loro stato di patogenicità su ClinVar. La casella 'Keep others' serve a mantenere le varianti senza annotazione ClinVar.
OMIM ⓘ 🚩 omim morbid clinic 111 ● omim without phenotypes 23 🚩 omim morbid 6	Permette di filtrare su OMIM morboso clinico, OMIM "without phenotypes" e OMIM morboso varianti. Questo filtro ha un pulsante "in/out" che può essere utilizzato per selezionare o eliminare uno stato.
SeqOne Artifacts ⓘ ☒ Remove artifacts	Pulsante per rimuovere le varianti ricorrenti presenti nel 5% di una coorte di 500 esomi considerati come artefatti secondo l'algoritmo SeqOne.

Sezione Transmission & Phenotypes

Nome e panoramica del filtro	Descrizione
▼ Description keywords ⓘ	Utilizza parole chiave per caratterizzare i fenotipi associati alle varianti interrogando i database OMIM, Medgen e Litvar. Questa funzione non tiene conto degli HPOs: ad essi è dedicato un filtro specifico (vedi sotto). Sono disponibili due modalità di ricerca: unione e intersezione.

▼ HPO Phenotypes ⓘ

Filter mode [See gene list \(10\)](#)

Default Extended

Filtering on HPOs [Select all](#)

Prostate cancer
HP:0012125 Sample ☒

Quadriceps muscle weakness
HP:0003731 Sample ☐

Zonular cataract
HP:0010920 Sample ☐

Zygotic cleavage failure
HP:0033336 Sample ☐

Consente di filtrare le varianti sulla base dei termini "Sample" precedentemente inseriti, sia prima dell'analisi del campione (see **Metadati campione**), sia dopo l'analisi dalla scheda HPO del riquadro di analisi (see **Vista modal HPOs**).

Le corrispondenze vengono identificate utilizzando il nostro approccio PhenoGenius, che si basa su una matrice HPO-gene per selezionare i geni.

In modalità "Filter" sono disponibili 2 opzioni:

- Filtraggio predefinito che bilancia l'accuratezza e il numero di geni restituiti.
- Filtraggio esteso che massimizza l'accuratezza con un maggior numero di geni.

L'opzione "See gene list" fornisce un elenco di geni associati agli HPOs selezionati in questo filtro. Questa funzione è disponibile solo se i termini HPO sono selezionati.

Si apre una finestra pop-up che mostra l'ID, il simbolo, il rango e il punteggio di Phenogenius di ciascun gene. È possibile scaricare l'elenco dei geni utilizzando l'icona Download o copiarlo utilizzando il pulsante Copia/Incolla.

▼ Transmission Mode ⓘ

▼ Dominant 10 ☒

AD Autosomal dominant inheritance 10 ☒

XLD X-linked dominant inheritance 10 ☒

DD XXX 10 ☒

PD XXX 10 ☒

> Recessive 10 ☐

> Other 10 ☐

Sources

OMIM 10 ☐

GenCC 10 ☒

Definitive 10 ☒

Supportive 10 ☒

Limited 10 ☒

Outside known sources 10 ☐

Consente agli utenti di filtrare le varianti utilizzando collegamenti di ereditarietà da malattie correlate alla genetica medica umana.

Le varianti possono essere filtrate in base alle modalità di trasmissione (Dominant, Recessive, Other) e alle fonti di trasmissione (OMIM, GenCC o altri database).

Transmission pattern ⓘ Compound Het - Suspected 10 ☒ Compound Het - Verified 10 ☒ Recessive Hom 10 ☒ De Novo 10 ☒	Consente di selezionare le varianti in base alle modalità di trasmissione eterozigote composita (sospetta o verificata), omozigote recessiva o de novo in base al DiagAI. Si noti che i filtri "Compound Het - Verified", "Recessive Hom" e "De Novo" sono disponibili solo nelle analisi familiari.
---	--

Sezione Classification

Nome e panoramica del filtro	Descrizione
▼ VKB ⓘ New Unknown variant 5 Pathogenic 4 Likely Pathogenic 3 Uncertain Significance 2 Likely Benign 1 Benign  Possible artifact ☐ Analysis only ⓘ	Permette di selezionare le varianti sconosciute non valutate (New - Unknown variant) o validate in "VKB" su tutte le analisi dell'entità. Se la casella 'Analysis only' è selezionata, le varianti saranno ricercate solo nell'analisi attualmente in corso di annotazione.
▼ ACMG ⓘ 5 Pathogenic 5 4 Likely Pathogenic 1 3 Unknown significance (VUS) 56 2 Likely Benign 49 1 Benign 6	'Pathogenicity classification' per le varianti germline, calcolata su una scala da 1 (benigne) a 5 (patogene).
▼ ComPerMed ⓘ 5 Pathogenic 7 4 Likely Pathogenic 6 3 VUS 102 2 Likely Benign 1 1 Benign 7	Seleziona le varianti somatiche in base alla loro classe di patogenicità "ComPerMed" su una scala da 1 (benigno) a 5 (patogeno).
▼ Evidence Type (CKB) ⓘ Actionable 1346 Diagnostic 0 Prognostic 0	Seleziona le varianti in base a 3 tipi di evidenza (CKB): A (Actionability), diagnostica e prognostica.

▼ CADD ⓘ Benign Pathogenic 1 99 Min 1 ^ ▼ Max 99 ^ ▼ ☒ Keep unknown Seleziona le varianti in base al punteggio CADD Phred.
▼ REVEL ⓘ Benign Pathogenic 0 0.5 1 Min 0 ^ ▼ Max 1 ^ ▼ ☒ Keep unknown Seleziona le varianti in base al punteggio REVEL.

Selezionare un'indicazione

In SomaVar, SomaVar LF e SomaCGP, è possibile selezionare un'indicazione patologica predefinita dal database JaxCKB da un menu a discesa disponibile nella scheda *Paziente* della barra laterale di gestione delle analisi (vedere la sezione **Vista modale del paziente**).

Il sistema ontologico è basato sulla *Disease Ontology*:²⁴ e comprende termini specifici per la JaxCKB. Le informazioni sulla sensibilità, nella sottocolonna "A" della colonna Clinical Evidences (vedi sezione **Pagina delle varianti**), sono evidenziate in verde solo quando l'indicazione per quella variante corrisponde all'indicazione selezionata. Se per una variante si trovano affermazioni di resistenza, la cella corrispondente nella colonna "A" sarà evidenziata in rosso, indipendentemente dall'indicazione. In caso contrario, sarà grigia.

Se la malattia è stata selezionata durante il caricamento del campione e prima dell'analisi (vedere **Metadati campione**), diventa l'indicazione predefinita definita per il campione e nella scheda Patient del pannello laterale di gestione dell'analisi. È contrassegnata dall'icona di una stella e appare per prima nell'elenco. La modifica dell'indicazione nel pannello laterale di gestione delle analisi non aggiorna la malattia preselezionata per il campione (metadati Disease).

The screenshot shows a modal window titled 'Analysis' with a close button (X) in the top left. It has four tabs: 'Workflow', 'Patient' (which is active), 'HPOs', and 'Panel'. Under the 'Patient' tab, there is a light blue box containing the text 'Active sample: CTDNA_toy'. Below this, there is a 'Sex' section with three buttons: 'Female', 'Male', and 'Unknown'. At the bottom, there is an 'Indication' section with a dropdown menu. The dropdown is open, showing a list of items, with the top item being '★ Colorectal cancer (Colorectal cancer)' which has a star icon and a close button (X) to its right.

²⁴ <https://disease-ontology.org/>

Utilizzo di un profilo di filtro predefinito

Gli utenti possono selezionare un profilo di filtro predefinito da un menu a discesa facendo clic sull'icona "Select profile". La piattaforma SeqOne offre diversi profili di filtro predefiniti per impostazione predefinita: "GermVar", "My Selection", "Research", "DiagAI".

Modifica e salvataggio dei profili di filtro

Gli utenti possono salvare la posizione attuale dei cursori per ciascun filtro come profilo di filtro, assegnando un nome al profilo e facendo clic su "Save as new". Il profilo di filtro sarà quindi disponibile tramite lo stesso menu a discesa nella categoria "Private".

Quando si salva un profilo di filtro, si apre un modulo di creazione e l'utente può selezionare il pulsante con il nome dell'entità per renderla accessibile agli altri utenti dell'entità.. Se l'utente seleziona il pulsante "Private", il profilo sarà accessibile solo all'utente che lo ha creato.

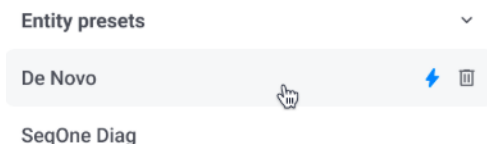
Solo l'utente che ha creato inizialmente il profilo del filtro potrà modificarlo cliccando su "Update"; se il profilo viene modificato da un altro utente dell'entità, a quest'ultimo verrà chiesto di salvarlo con un nome diverso. Se l'utente non ha il permesso di modificare un profilo di filtro pubblico, gli verrà chiesto di contattare l'amministratore della sua entità. I profili di filtro SeqOne non possono essere modificati e salvati sotto un altro profilo di filtro.

È possibile ripristinare uno o più filtri. È sufficiente fare clic sull'icona  a destra del filtro o nell'angolo in alto a destra della casella "Filters".

Profili dei filtri preferiti

Gli utenti possono impostare i profili dei filtri come preferiti, cliccando sull'icona  a destra del profilo, in modo che appaiano come scorciatoia direttamente nella barra di accesso rapido sopra la tabella delle varianti.

Per rimuovere un profilo di filtro dai preferiti, è sufficiente fare clic sulla stessa icona . Per impostazione predefinita, il profilo "My Selection" è definito come preferito.

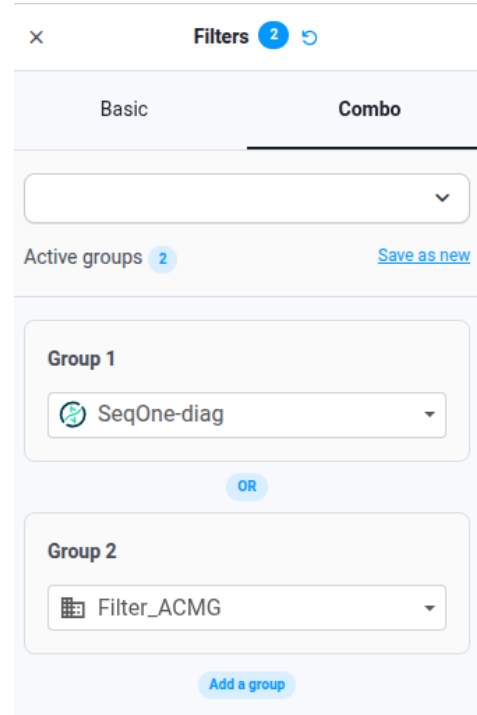


Combinazione di profili di filtro

I profili di filtro possono essere combinati sotto la condizione **OR** facendo clic sulla scheda Combo nel pannello dei filtri, quindi su "Create a combo". Per impostazione predefinita, il modulo di creazione mostra 2 gruppi di profili di filtro, ma gli utenti possono aggiungerne altri facendo clic su "Add a group".

Una volta selezionati i profili, gli utenti sono invitati a salvare il loro profilo combo cliccando su "Save as new". Come per i profili filtro, l'utente può rendere accessibile la sua combo a tutti gli utenti dell'entità nel modulo di creazione. Solo l'utente che ha creato inizialmente il profilo combo potrà modificarlo cliccando su "Update".

Se il profilo viene modificato da un altro utente dell'entità, a quest'ultimo verrà chiesto di salvarlo con un nome diverso. Se l'utente non ha il permesso di modificare il profilo di una combo pubblica, gli verrà chiesto di contattare l'amministratore della sua entità.



The screenshot shows a 'Filters' window with a 'Combo' tab selected. At the top, there's a 'Basic' tab and a 'Combo' tab. Below the tabs, there's a dropdown menu. Underneath, it says 'Active groups 2' and 'Save as new'. The main area shows two groups of filters. 'Group 1' contains a dropdown menu with 'SeqOne-diag'. Below this is an 'OR' operator. 'Group 2' contains a dropdown menu with 'Filter_ACMG'. At the bottom, there's a button 'Add a group'.

Profili dei filtri di SeqOne

My Selection. Questo profilo di filtro consente di visualizzare direttamente in VKB le varianti valutate come 3 (VUS), 4 (*Likely Pathogenic*) e 5 (*Pathogenic*) nell'analisi in corso di interpretazione.

DiagAI. Questo profilo di filtro, disponibile nelle analisi germline (GermlineVar, GermlineFamily, GermVar Tertiary, GermVar Tertiary Family), visualizza un elenco ridotto di varianti con un'alta probabilità di essere la causa della malattia del paziente. Questo profilo sfrutta l'algoritmo di prioritizzazione DiagAI e l'"Occurrence" della variante nel progetto (filtro Sample Frequency).

Può essere applicato quando i termini HPO sono o non sono associati al campione, ma la sua sensibilità aumenta in presenza di termini HPO. Nel caso di un'analisi familiare, il profilo del filtro si adatta automaticamente alle dimensioni della famiglia analizzata, poiché tiene conto del fatto che l'occorrenza di varianti nel progetto può essere maggiore. Gli utenti possono anche modificare il filtro "Sample frequency" nel progetto in qualsiasi momento.

Inserire gli HPOs: per visualizzare le varianti SmartPick con la più alta probabilità di causare malattie.

Family profiles. Nelle analisi GermlineFamily e GermVar Tertiary Family, i profili di filtro *De Novo (New)*, *Recessive Hom (New)* et *Compound Het (New)* saranno disponibili solo per le analisi di trio che includono un caso indice affetto, il padre non affetto e la madre non affetta. Altrimenti, questi profili saranno disabilitati.

Non è consigliabile modificare i profili dei filtri *De Novo (New)*, *Recessive Hom (New)* et *Compound Het (New)* quando si costruisce e si salva un profilo personale. Si consiglia di partire da zero quando si crea una variante di questi profili.

Carrier screening. Questo profilo di filtro è disponibile in GermlineFamily solo per le analisi duo (genitori senza campione indice) e identifica i geni recessivi con una variante eterozigote potenzialmente patogena in entrambi i genitori "non affetti" per i test preconcezionali.

Somatic FFPE. Questo profilo di filtro visualizza un elenco ridotto di varianti somatiche nelle analisi somatiche (SomaVar, SomaCGP) eseguite su DNA FFPE.

Somatic LF. Questo profilo di filtro visualizza un elenco ridotto di varianti somatiche nelle analisi SomaVar LF.

Funzione "Mark as viewed"

"Mark as viewed" (segna come visto) è una funzione che consente all'utente di nascondere selettivamente alcune varianti. A tal fine, è sufficiente selezionare la o le varianti desiderate e

fare clic sull'icona  Mark as viewed visualizzata nella parte inferiore della tabella delle varianti. La selezione delle varianti appare in grigio. Per rimuovere la selezione, l'utente può fare clic sull'icona  Reset all viewed nella parte superiore della tabella delle varianti per ripristinare la visualizzazione, oppure selezionare la variante o le varianti e fare clic su .

 Mark as not viewed

L'utente può nascondere il popup facendo clic sulla croce. In questo modo tutte le varianti verranno deselezionate, ma rimarranno in grigio.

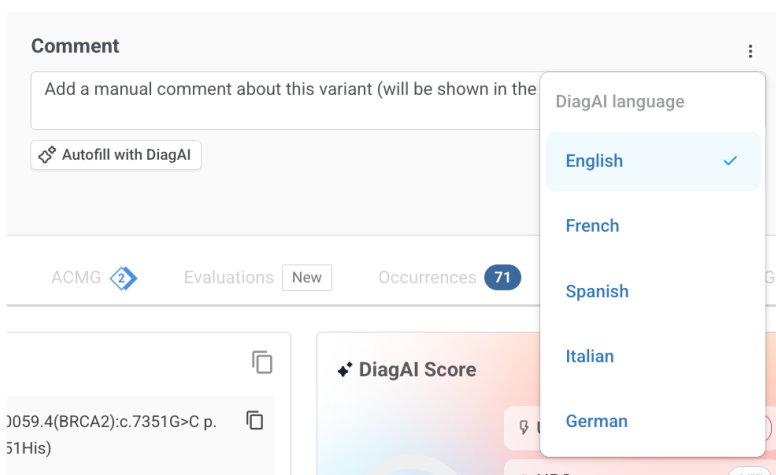
Selezione delle varianti

Le varianti possono essere aggiunte al pannello "Selection" insieme alle varianti valutate (vedere la sezione seguente). A tal fine, selezionare le varianti spuntando le caselle sul lato sinistro della tabella delle varianti, quindi fare clic sull'icona "Add to selection" nella parte inferiore della tabella.

Valutazione delle varianti

Le varianti possono essere valutate rapidamente dalla tabella delle varianti facendo clic nella colonna "VKB" e specificando la classe di patogenicità e un commento opzionale. Queste valutazioni e commenti sono accessibili a tutti gli utenti della piattaforma all'interno della stessa entità tramite l'icona "Selection" nell'analisi (vedi sezione **Pannello laterale per la gestione delle valutazioni**). Una variante può essere valutata anche dalla sua pagina dedicata.

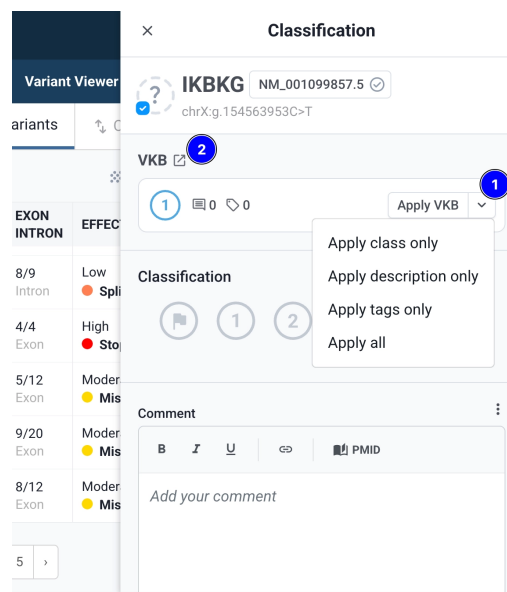
I commenti possono essere inseriti manualmente o generati automaticamente. L'interfaccia include un campo vuoto "Classification" per i commenti e un pulsante "Autofill with DiagAI" per la generazione automatica. Durante la generazione dei commenti viene visualizzato un indicatore. Per i commenti esistenti, l'opzione "Improve with DiagAI" consente di migliorarli. Ogni commento generato può essere approvato, modificato o azzerato. Una volta convalidati, i commenti vengono salvati con la loro classe di patogenicità, la data e il contenuto. Possono essere modificati o cancellati dopo la convalida, mantenendo la loro storia.



Per selezionare la lingua di DiagAI, clicca sui tre puntini nell'angolo in alto a destra della casella dei commenti e scegli la lingua che preferisci dal menu a tendina.

Le varianti possono anche essere classificate cliccando nella colonna di classificazione, la cui intestazione ha una bandierina. Per classificare una variante, cliccare sull'icona +. Si apre un pannello laterale di classificazione che consente all'utente di aggiungere la classe di patogenicità, un commento e un tag.

Per aggiungere questa classificazione al database VKB, attivare l'interruttore "Add to VKB". Se esiste già una classificazione nel database VKB, cliccare su "Apply VKB" per applicarla alla voce corrente. Opzioni aggiuntive sono disponibili cliccando sulla freccia accanto a "Apply VKB" (vedere il passaggio 1 nella figura sotto).



Convalidare la classificazione cliccando su "Save" in alto a destra del pannello laterale. Per modificare una classificazione direttamente nel database VKB, cliccare sul link di reindirizzamento VKB (vedere il passaggio 2 nella figura sotto) per aprire la finestra "VKB Entry".

Nella finestra "VKB Entry", cliccare su "Edit" in alto a destra per modificare la valutazione, quindi cliccare su "Save" per applicare le modifiche. La cronologia VKB e le occorrenze della variante sono accessibili navigando attraverso le schede dedicate.

The screenshot shows the "VKB Entry" window with a close button (X) and an "Edit" button. It features three tabs: "VKB Classification" (active, with a count of 1), "VKB History", and "Occurrences" (with a count of 1). The "VKB Classification" tab contains a "Classification" section with five circular buttons labeled 1 through 5. Button 1 is selected and highlighted in blue, and a "Benign" label is positioned below it. The "Variant Description" section below shows the text "No description found for this variant". The "Classification tags" section is empty.

VKB Search. Un pannello laterale è disponibile facendo clic sul pulsante "VKB" nella barra di ricerca dell'intestazione di SeqOne. Questo permette di cercare le varianti annotate nel database VKB in base al nome del gene, al transcript, all'annotazione HGVS o all'identificatore SeqOne (Genome_variant_chromosome_location_REF_ALT). La ricerca può essere filtrata per genoma di riferimento (GRCh37 o GRCh38) e/o classe di patogenicità.

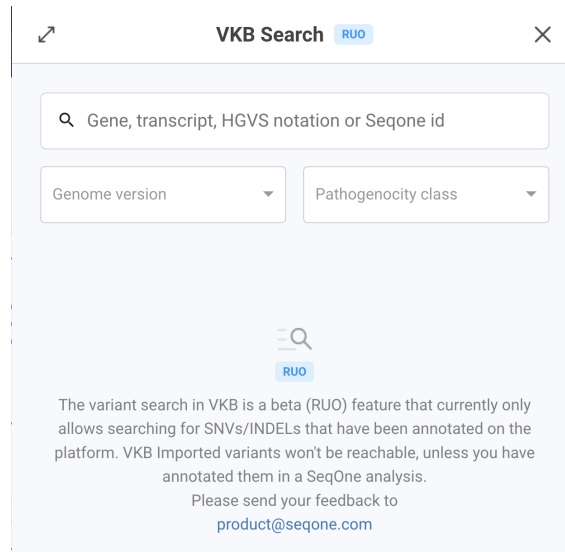


Figura 15: Finestra "VKB Search".

Quando si fa clic su una variante valutata, si apre una pagina più dettagliata con informazioni sulla variante, come la posizione genomica, le annotazioni HGVS sul DNA complementare (cDNA) e sulla proteina, le classi di patogenicità, le analisi e i relativi commenti in cui è stata valutata e un link alle analisi.

Figura 16 : Finestra dei dettagli della variante in "VKB".

Pagina della variante

Facendo clic su una variante nella tabella delle varianti si apre un pannello laterale contenente ulteriori informazioni, con diverse schede. Gli utenti possono aprire la pagina delle varianti in un'altra scheda facendo clic su "Open in new window" nell'angolo in alto a destra. Il pannello laterale può essere chiuso con il tasto "Esc".

In questa finestra gli utenti possono valutare e commentare una variante. Come le valutazioni e i commenti fatti dalla tabella principale delle varianti, queste valutazioni e commenti possono essere visti da tutti i membri del gruppo che utilizzano la piattaforma.

Da questa pagina, gli utenti possono anche accedere alle informazioni sulle varianti, tra cui il genotipo (GT), la frequenza allelica (VAF), la profondità di copertura (COV) e il punteggio di phred (QA).

Le schede della pagina Variant sono descritte di seguito.

Variant. La scheda è suddivisa in diversi blocchi, descritti di seguito.

- **Nomenclature:** vengono elencate le diverse nomenclature (g., c., p., voce VCF), con la possibilità di copiare negli appunti le informazioni dell'intero blocco o di ogni riga utilizzando l'icona .
- **DiagAI Score:** utilizza il blocco di riepilogo del calcolo del DiagAI Score disponibile nella scheda DiagAI Score della pagina della variante (vedere il paragrafo **DiagAI Score**).
- **ClinVar:** informazioni tratte dalla colonna "ClinVar Germline" della tabella delle varianti, che forniscono la classe di patogenicità assegnata alla variante, il suo stato clinico e il numero di voci in ClinVar (vedere **Visualizzazione e interpretazione** nella sezione **Piccole varianti**). L'icona  apre la pagina ClinVar' in una nuova scheda.
- **Annotations:** 'Transcript impact', esone interessato, distanza dal sito di splicing più vicino e tipo di mutazione.
- **Transmission:** collegamenti ai database OMIM e Genes. se una o più malattie sono associate al gene.
- **Compound Heterozygous Candidates**, questo blocco è diviso in 2 parti:
 - La sezione "*Short Variants*" raggruppa tutte le SNV candidate situate sullo stesso gene (in base all'ID del gene) e le classifica in base al loro DiagAI Score. La variante selezionata dalla pagina Variants viene visualizzata per prima ed è in grigio nell'elenco.
 - La sezione CNVs indica se sono state trovate CNVs, ordinate per classe ACMG. Questa sezione è disponibile solo se è stata eseguita un'analisi CNV, altrimenti viene visualizzato un messaggio di avviso "*CNVs not available*".

Facendo clic su uno dei risultati si aprirà una nuova scheda con la pagina della variante (se si tratta di una piccola variante) o della CNV (se si tratta di una variante di copia). Se non vengono trovate CNVs, viene visualizzato un messaggio di avviso con la dicitura "no CNVs found".

- **Tools:** collegamenti ad altre banche dati (LitVar, LOVD, VarSome, dbSNP, Cosmic, "GnomAD", CKB, UCSC, Decipher, Ensembl).
- **GnomAD:** se la variante è nota a questo database, viene visualizzata una tabella che riassume la frequenza della variante in diverse popolazioni e viene fornito l'accesso diretto a "GnomAD" tramite l'icona  .

Nomenclature

gDNA chr4(GRCh38)g.154611883G>C

cDNA NM_021870.3(FGG):c.323C>G

Protein NP_068656.2 p.Ala108Gly

VCF 4-154611883-G-C

Clinvar

Likely Pathogenic

Criteria Provided, Multiple Submitters, No Conflicts ★★☆☆

Review status **Reclassified by SeqOne**

Submissions **2 P | 9 LP | 2 VUS**

Disease Congenital Fibrinogen Deficiency, Thrombus

Annotations

Impact/Effect **Moderate** ● missense

Exon/Intron 4/9

Biotype protein coding

Transmission

AR

DISEASE	OMIM	MEDGEN
Hypofibrinogenemia, congenital	AR	N/A

Compound Heterozygous candidates

Short variants

79 4 New c.323C>G ● missen...

7 1 New c.*16C>T ● 3 prime ...

1 to 2 of 2 1

CNVs

No CNVs found

Tools

Variant database

[Ensembl](#)
[LOVD](#)
[ClinVar](#)
[dbSNP](#)
[UCSC](#)

Gene database

[OMIM](#)
[Ensembl](#)
[Cytoscape](#)

Genome database

[Genome Browser](#)
[DECIPHER](#)
[e!](#)

GnomAD

MAF 3.84e-3 ALL 3.03e-3

QC Exome | PASS QC Genome | PASS

Population	AF	AC	HOM
African/American population	6.13e-4	46	0
American population	6.67e-4	40	0
Amish	0.00e+0	0	0
Ashkenazi Jewish	1.52e-3	45	0
East Asian population	0.00e+0	0	0
Finnish population	1.06e-3	67	0
Middle Eastern population	6.61e-4	4	0
Non-Finnish European population	3.84e-3	4526	11
Other populations	1.63e-3	102	1
South Asian population	6.05e-4	55	0
Non-UKB combined population	2.38e-3	1854	5
Combined population	3.03e-3	4885	12

Figura 17: Scheda Variant nella pagina della variante

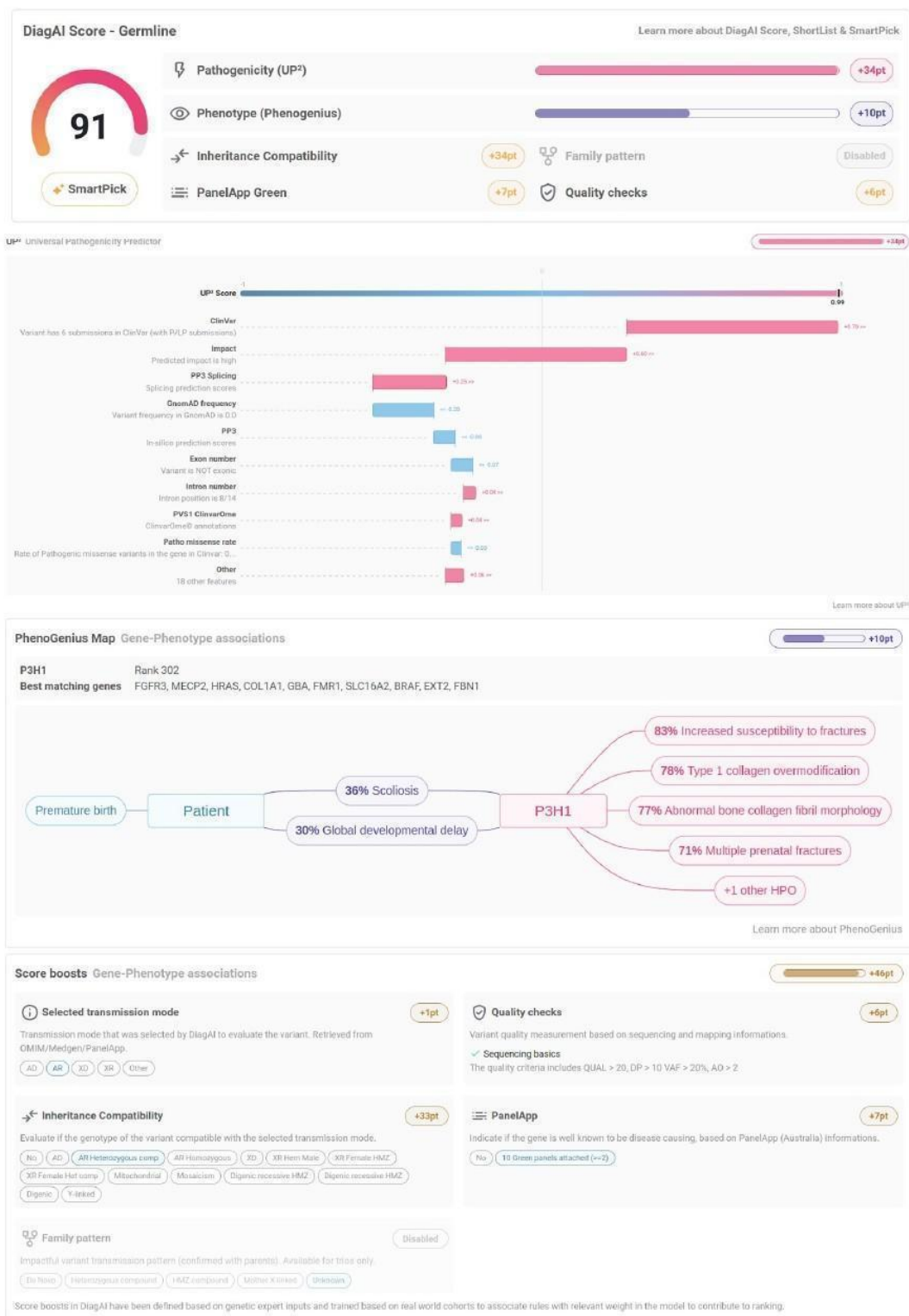


Figura 18: Visualizzazione della scheda "DiagAI" nella pagina delle varianti.

DiagAI. La scheda contiene spiegazioni sul DiagAI Score. assegnato alla variante. Questa scheda è disponibile solo per le analisi GermlineVar, GermlineFamily , GermVar Tertiary e GermVar Tertiary Family (dalla v1.8).

La scheda "DiagAI" è suddivisa in 4 sezioni principali:

- Un blocco di riepilogo '**DiagAI Score - Germline**' che mostra il punteggio, lo stato SmartPick e l'elenco delle principali caratteristiche che contribuiscono al punteggio complessivo e al numero di punti associato.
- Un grafico a cascata che spiega il punteggio **UP² (Universal Pathogenicity Predictor)**. Un messaggio di avviso  viene visualizzato quando l'utente consulta la pagina di una variante il cui trascritto differisce da quello utilizzato per calcolare l'UP².
- Un blocco di spiegabilità della **PhenoGenius Map**, colorato se i termini HPOs sono associati al campione. Il blocco riporta il rango del gene, un elenco dei 10 geni meglio associati e la "PhenoGenius Map", codificata a colori in base alla corrispondenza tra gli HPOs del paziente e il gene: blu per i soli HPOs del paziente, viola per gli HPOs del paziente e corrispondenti al gene, rosso per i soli HPOs. La sezione è in grigio se non sono stati inseriti termini HPO per il paziente.
- Una sezione "**Score boosts**" integra 5 sottocaratteristiche che possono contribuire o meno (disabilitate) al punteggio: Transmission', Compatibilità con l'ereditarietà, Modello di famiglia, PanelApp. e Controllo qualità.

ACMG. La scheda consente agli utenti di visualizzare e rivedere i criteri utilizzati durante la classificazione automatica ACMG. Possono modificare la selezione controllando i diversi criteri richiesti e quindi utilizzare la classificazione per valutare la variante cliccando su "Use for Evaluation". È inoltre possibile modificare il livello di evidenza (*Supporting, Moderate, Strong* e *Very Strong*) per assegnare un numero diverso di punti a ciascun criterio, che saranno presi in considerazione nel calcolo del punteggio finale ACMG. Ogni livello di evidenza assegnerà: 1 punto per *Supporting*, 2 punti per *Moderate*, 4 punti per *Strong* e 8 punti per *Very Strong*. I punti per ciascun criterio saranno poi sommati per assegnare uno status ACMG alla variante:

- '*Pathogenic*': 10 punti
- '*Likely Pathogenic*': da 9 a 6 punti
- '*Uncertain Significance*': Da 5 a 0 punti
- '*Likely Benign*': da -6 a -1 punti
- '*Benign*': da -7 punti

Per il criterio BA1, l'utente non può selezionare il livello di prova, poiché si tratta di una classe a sé stante.

TP53 protein coding

NM_000546.6 c.1083del p.S362AfsTer8 ● Frameshift

sample
GT / VAF ☐ 1.25%
COV / QA 319 34

1 2 3 4 5 Add a manual comment about this variant (will be shown in the output report) ... Add comment Cancel

Variant **ACMG** TP53 Compermed Evidences Evaluations New Occurrences 6 Scores IGV Gene

☒ Use ClinGen point-based system

Predicted:
Likely Pathogenic Use for evaluation

Automatic tags
PVS1 | Strong PM2 | Moderate
PP5 | Moderate

First level tags
No selected tags.

Second level tags
No selected tags.

Tags that do not apply
No selected tags.

Automatic tags (3 / 3)
These tags are automatically selected.

☒ **PVS1** Strong ⓘ
Predicted null variant in a gene where LOF is a known mechanism of disease
Non-sense or frameshift variant, NOT predicted to undergo NMD, but truncated region is critical to protein function (11 non-truncating pathogenic variants found in Clinvar).

Figura 19: Visualizzazione della scheda "ACMG" sulla pagina della variante

TP53. La scheda è disponibile solo se la variante è annotata sul gene TP53. Questa scheda fornisce informazioni sulle mutazioni di TP53 e sul loro significato clinico secondo il database IARC²⁵. La scheda "TP53" è suddivisa in diversi blocchi, descritti di seguito.

Un blocco di annotazioni (*Annotations*) visualizza le seguenti informazioni associate alla variante:

- Mutation Rate: tasso di sostituzione associato alla posizione della variante
- Hotspot (yes/no): presenza della variante in un sito di mutazione ricorrente (hotspot)
- Structural Motif : motivo strutturale affetto dalla mutazione
- Domain: dominio funzionale della proteina interessata dalla mutazione

Il blocco Dati funzionali (*Functional Data*) visualizza le classificazioni e le informazioni funzionali associate alla variante:

- DNE Class: effetto dominante-negativo (DNE) sulla transattivazione della proteina p53 wild-type
- DNElof Class: classificazione funzionale per la perdita di attività di soppressione della crescita e di attività dominanti negative
- Transactivation Class: classificazione funzionale basata sull'attività trascrizionale globale (TA)
- Structure Function Class: previsioni funzionali.

²⁵ <https://tp53.cancer.gov/>.

I blocchi di dettagli "Graphics" e "Sample" forniscono i dettagli dei campioni a cui fa riferimento il database IARC per la variante. L'utente può scegliere di visualizzare i dati associati a campioni somatici o germinali, selezionando Somatic o Germline.

Il blocco "Graphics" visualizza due grafici, con menu a discesa, che mostrano la distribuzione dei campioni in base alle loro caratteristiche.

Il blocco "Sample details" visualizza informazioni dettagliate sul campione in forma tabellare.

ComPerMed. (solo analisi somatica). La classe di patogenicità di ciascuna variante è calcolata in base alle raccomandazioni del comitato di esperti belga ComPerMed (Commission for Personalized Medicine). L'utente ha accesso a informazioni dettagliate su questa classificazione da una scheda dedicata nella pagina della variante.

Evidences. La scheda fornisce un elenco di terapie associate alla variante selezionata dal database JaxCKB. L'utente può scorrere ogni riga per trovare i dettagli delle terapie o delle indicazioni.

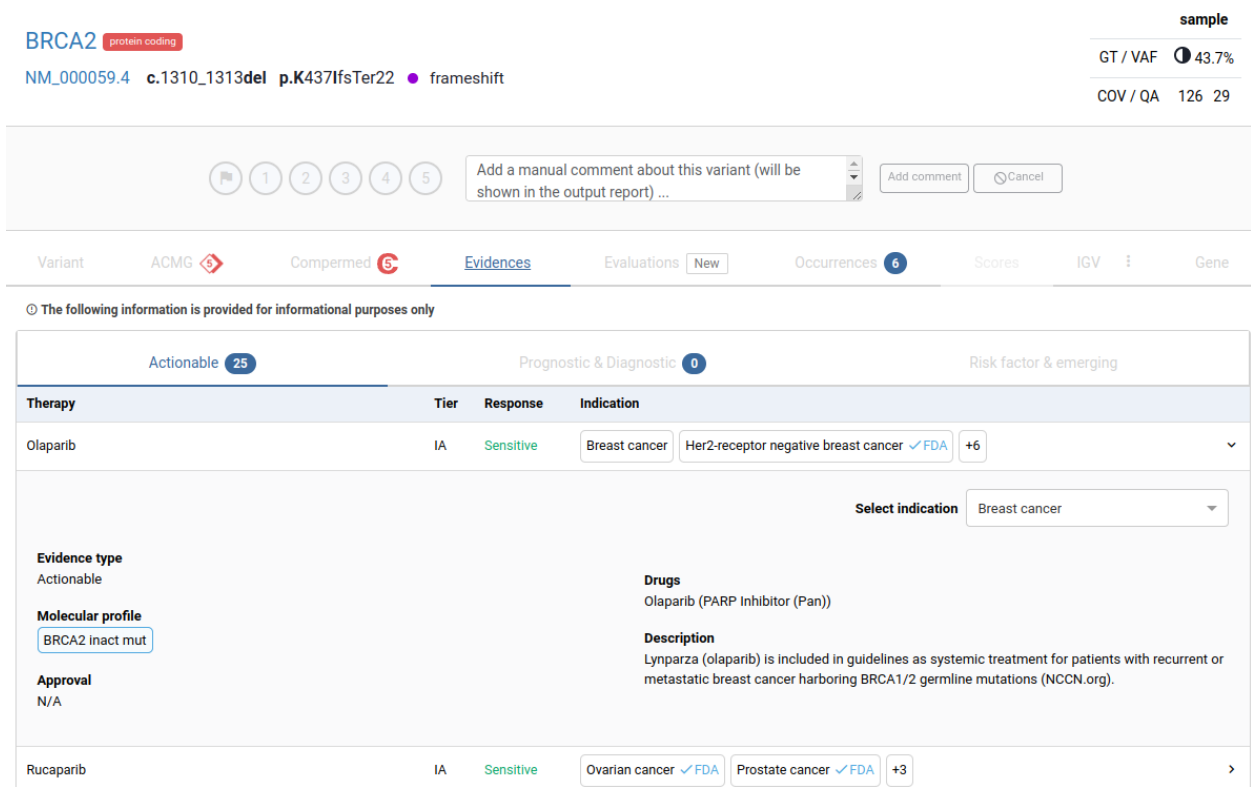


Figura 20: Visualizzazione della scheda "Evidences" nella pagina della variante

Evaluations. Questa scheda consente di accedere alla cronologia delle valutazioni e delle interpretazioni della variante nel tempo:

- Nell'analisi corrente (sezione sinistra della scheda)

- Nelle altre analisi svolte all'interno del gruppo (sezione destra della scheda)

Occurrences. La scheda identifica gli altri campioni contenenti la variante, indipendentemente da qualsiasi valutazione. Per ogni occorrenza, vengono inseriti il nome del campione, la frequenza allelica, il link all'analisi pertinente e il nome del progetto.

Publications (Pubblicazioni). Questa scheda fornisce l'accesso alla letteratura scientifica associata alla variante e al gene. Fare clic sulla sottoscheda "Variant" (Variante) o "Gene" per visualizzare le pubblicazioni PubMed che fanno riferimento alla variante o al gene. Un collegamento PMID reindirizza al record PubMed e il collegamento al testo completo apre l'articolo completo quando è disponibile.

Scores. La scheda è disponibile per le varianti con effetti missense, start lost e stop gained. Ad eccezione di questi casi, il punteggio sarà grigio. Questa scheda consente di visualizzare i risultati dei punteggi di predizione di diversi strumenti come *REVEL*, *DANN*, *FATHMM*, *FATHMM-MLK*, *LRT*, *MetaLR*, *MetaSVM*, *MutationTaster 2*, *PROVEAN*, *SIFT* e *PLI*. Fornisce inoltre accesso a punteggi di conservazione come *GERP++*, *PhastCons100way Vertebrate*, *PhastCons20way Mammalian*, *PhyloP100way Vertebrate*, *PhyloP20way Mammalian* e *SiPHhy*.

IGV. Questa scheda consente all'utente di visualizzare le reads allineate alla posizione della variante attraverso la versione web del Genome Browser di IGV. Nelle aree ad alta copertura, le letture sono sottocampionate per impostazione predefinita, ossia non vengono visualizzate tutte le letture. Vengono applicati i parametri di visualizzazione predefiniti, ossia un massimo di 100 reads per finestra di 50 nucleotidi. Queste regioni possono essere identificate dalla linea nera sotto il profilo di copertura (downsampling). Facendo clic su di essa viene visualizzato il

numero di reads rimosse dalla visualizzazione (# downsampled). Dall'icona  è disponibile un menu che offre tre opzioni:

- *Open in new tab*: apre la vista della variante IGV in una nuova scheda.
- *Track setting*: consente di aggiungere alla vista altri campioni dello stesso progetto o di progetti diversi.
- *Open local IGV*: reindirizza l'applicazione IGV Desktop alle coordinate della variante in corso di interpretazione. Per visualizzare SNV o altre varianti, il BAM del campione deve essere prima caricato nell'applicazione IGV Desktop. Questa opzione è compatibile solo con le versioni di IGV Desktop ≥ 17 .

'IGV' supporta le viste divise, ad esempio copiando due serie di coordinate nella barra di ricerca dei locus, separate da uno spazio. Sono disponibili varie impostazioni facendo clic sulle ruote dentate a destra del Genome Browser. Ad esempio, per visualizzare direttamente le reads non allineate (soft clips), selezionare l'opzione "Show soft clips".

Gene. La scheda contiene informazioni aggiuntive sul gene e la bibliografia associata.

Generazione di un referto clinico

L'utente può modificare un rapporto clinico esportabile in formato PDF e .docx facendo clic sul pulsante "Generate report" (Genera rapporto) dalla scheda "Selection" (Selezione) durante la navigazione nell'analisi.

Nella parte superiore della pagina di modifica, l'utente può selezionare un modello, impostare lo stato del rapporto su nuovo o bozza, selezionare un campione (è possibile selezionare più campioni per le analisi familiari) e selezionare la lingua di modifica. L'utente può anche passare dalla vista "Edition" (Modifica) alla vista "Preview" (Anteprima) prima di salvare una bozza o procedere al passaggio successivo utilizzando i pulsanti "Save Draft" e "Next Step".

I modelli di rapporto possono essere creati nella sezione "Reporting", sotto "Template list" nelle impostazioni dell'entità (Entity settings).

Nella vista di modifica, l'utente può modificare le diverse sezioni del rapporto e selezionare più opzioni. Nella vista di anteprima, l'utente può rivedere il rapporto prima di continuare la modifica, salvare la bozza o convalidare il rapporto.

L'utente può salvare una bozza e accedervi nuovamente dalla scheda "Selection", quindi su "Generate report" come quando si crea un nuovo rapporto. Quando viene generato un rapporto PDF, viene visualizzato automaticamente nella scheda "Reporting" (Rapporti) dell'analisi. Solo dopo la convalida il rapporto può essere esportato come .PDF o .DOCX. L'utente può eliminare il rapporto o creare rapporti aggiuntivi per la stessa analisi.

The screenshot shows the 'Edition' view of a clinical report form. At the top, there is a navigation bar with tabs for 'Selection' (active) and 'Analysis'. Below this, a toolbar contains buttons for 'Template' (set to 'None'), 'Report' (set to 'new'), 'Sample', 'Language' (set to 'English'), 'Save draft', and 'Next step'. The main form area is titled 'Header' and contains several input fields: 'Report title', 'Patient indication', and 'Patient HPO(s)'. Below these is a link 'Refresh HPO from sample' and a dropdown menu 'Select metadata to include from the list below (Optional)'. At the bottom, there is a checkbox 'Include bioinformatic information in the report' with a note: 'When enabled, reports will include sample ID, analysis name, date, user, and test information'.

Figura 21: Visualizzazione della scheda della relazione clinica.

13.1.5. CNV

13.1.5.1. CNV in pannello di geni

Metodologia

L'individuazione di CNVs nei dati del pannello di geni, in GermlineVar, GermVar/GermVar Family, SomaVar e SomaHemato, si basa sulla selezione di campioni di controllo dalla coanalisi, in base alla loro correlazione.

La correlazione è calcolata da una tabella di copertura, che a sua volta è calcolata da tutte le regioni del manifesto, escludendo però i cromosomi sessuali.

I criteri per l'ammissibilità di un campione all'analisi CNV sono i seguenti:

- Affinché una regione venga conservata nell'analisi CNV, la copertura minima è di 50 reads.
- Il valore di correlazione soglia per la selezione dei campioni di controllo è 0,90 per il contesto germline e 0,75 per il contesto somatico.
- Il numero minimo di controlli richiesti per ogni campione è di sei.

Il sesso dei campioni (metadati del campione) deve essere definito prima di lanciare CNVCapture o SomaCNVCapture per eseguire la chiamata sui geni localizzati sui cromosomi sessuali. Infine, per eseguire un'analisi CNVCapture o SomaCNVCapture su SeqOne sono necessari almeno otto campioni.

I criteri applicati all'utente manifesto affinché un gene del panel sia idoneo per l'analisi delle CNV sono i seguenti:

- Le regioni manifeste sono annotate a partire dal "Canonical" trascritto determinato in base alla risorsa NCBI MANE Selection. Se non è disponibile alcun trascritto MANE Select, l'annotazione si basa sulla risorsa NCBI RefSeq.
- vengono analizzati solo gli esoni per i quali esiste una sovrapposizione $\geq 80\%$ tra l'intervallo manifesto e l'esone. Se nessun esone del gene presenta una sovrapposizione $\geq 80\%$ con l'intervallo manifesto, il gene non viene incluso nell'analisi CNV.
- vengono analizzati solo gli introni per i quali la sovrapposizione tra l'intervallo manifesto e l'introne è $\geq 60\%$.

L'analisi CNVCapture (versione ≥ 2.4) include le regioni situate 20kb a monte e a valle dei geni coperti dal manifesto, se tali regioni sono incluse nel disegno del manifesto. In particolare, ciascuno dei 20 segmenti di 1kb che costituiscono la regione di 20kb viene analizzato se almeno il 20% del segmento si sovrappone al manifesto iniziale.

Vista e interpretazione

La scheda "CNVs" visualizza un grafico che rappresenta l'elenco dei geni in cui sono state identificate variazioni di copia. Questa scheda è disponibile nei worksets descritti nella sezione Metodologia.

La terminologia utilizzata per descrivere lo stato dei geni chiamati è la seguente:

- Gene: il gene è affetto da amplificazione parziale o totale
- Loss: il gene è portatore di una delezione parziale o totale
- Normal: il gene non è portatore di CNV.
- Partially not called: il gene è parzialmente chiamato.
- Not called: il gene è escluso dall'analisi CNV.

Lo stato "Partially not called" viene assegnato a un gene quando il suo stato di copia è normale ma almeno un esone non è stato chiamato (con lo stato "Failed").

Nonostante l'ammissibilità all'analisi CNV, un gene viene escluso (Not called) dall'analisi se :

- il gene si trova su un cromosoma sessuale e i metadati sul genere non sono stati compilati,
- tutti gli esoni falliscono in base ai criteri specificati di seguito: number of controls, Percentage of variation:". Lo status "Failed" di un gene è definito solo sul CDS. Le regioni introniche e quelle a monte e a valle sono escluse se vengono chiamati gli esoni CDS.

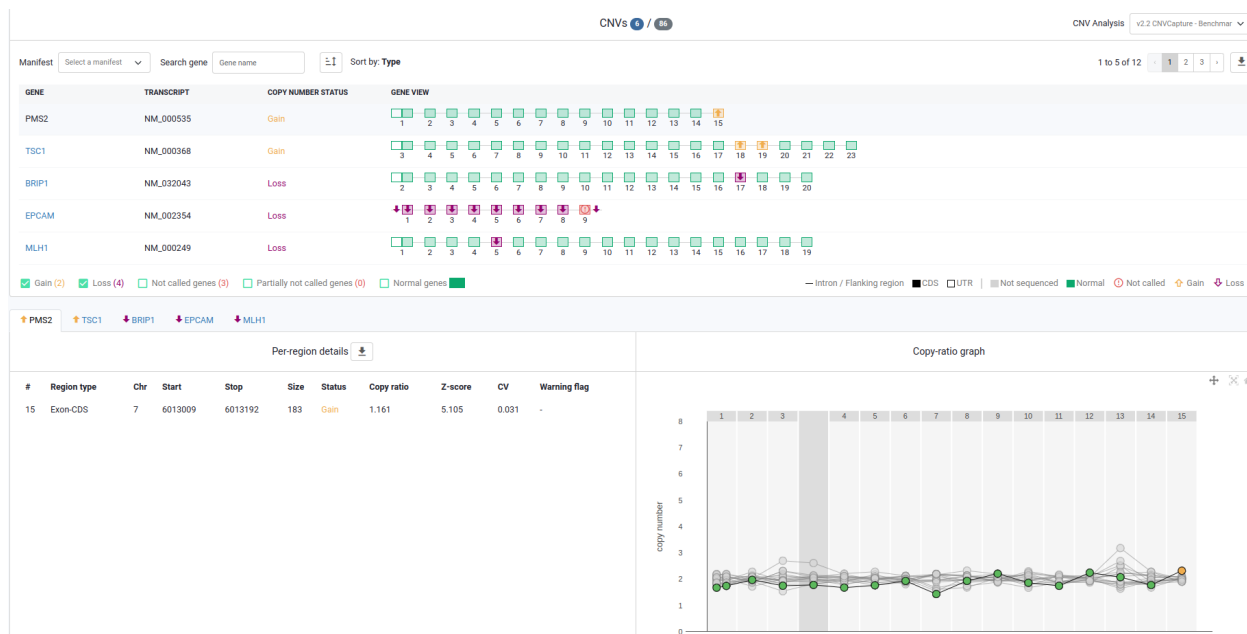


Figura 22: Visualizzazione della scheda "CNVs" in un'analisi GermlineVar.

Filtri e opzioni di ordinamento

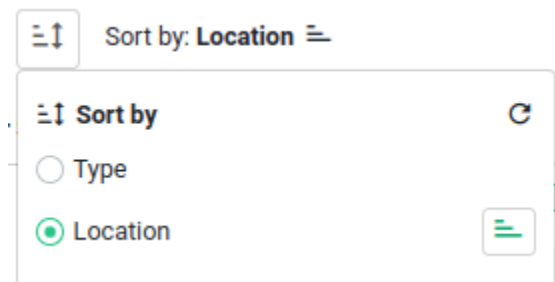
Il menu dei filtri può essere utilizzato per:

- Limitare la visualizzazione a un sottoinsieme di regioni/generi utilizzando un file BED precedentemente caricato nell'interfaccia con la funzione "*Manifest*".
- Cercare un gene specifico con la funzione "*Search gene*".

Sono disponibili due opzioni di selezione delle CNV:

- **Type** (default): logica di smistamento della corrente (Gain e poi Loss)
- **Location**: consente di ordinare le CNVs' in base ai cromosomi e alla loro posizione iniziale, in ordine crescente o decrescente, facendo clic

sull'icona di ordinamento.



L'utente può anche selezionare i geni in base al loro status: Gain, Loss, Normal genes, Partially not called genes o Not called genes.

☒ Gain (0) ☒ Loss (0) ☐ Not called genes (69) ☐ Partially not called genes (0) ☐ Normal genes

Dettagli per regione

La tabella dei dettagli per regione mostra i dettagli di ogni regione per ogni gene:

- **#**: numero di regione (esoni e introni)
- **Region type**: tipo di regione (upstream, utr-exonic, coding-exonic, intronic, downstream)
- **Chr**: cromosoma
- **Start/Stop**: posizione iniziale e finale di CNV
- **Size**: dimensione della CNV (bp)
- **Status**: stato della regione (gain = duplicazione, lost = delezione, normal = nessuna CNV, not called = la regione non è stata inclusa nell'analisi CNV)
- **Copy-ratio**: rapporto del numero di copie rispetto al normale (n=2).
- **Z-score**: numero di deviazioni standard che separano un risultato dalla media. Valori positivi o negativi elevati indicano una forte fiducia nella chiamata CNV.
- **CV**: coefficiente di variazione che misura la dispersione relativa all'interno dei campioni della coorte.
- **Warning flag**: Occasionalmente può essere visualizzato un messaggio di errore in base ai seguenti criteri:
 - **Low zscore**: Valore assoluto dello "Z-score" <4
 - **High variation**: coefficiente di variazione (CV) > 0,1
 - **Low coverage**: copertura della regione per questo campione <100

- **Number of controls:** Il numero di controlli è insufficiente per poter eseguire la chiamata su questa regione.
- **Percentage of variation:** La variazione tra i campioni è troppo grande perché questa regione possa eseguire la chiamata.
- **Read depth:** copertura della regione insufficiente (<50) e copy-ratio >0,1
- **Missing Gender:** la regione si trova su un cromosoma sessuale e non sono stati inseriti i metadati relativi al genere.

Warning: una regione ricca di soft clips, come una grande indel o un inserto Alu, può essere interpretata come un calo di copertura dai workset "SomaCNVCapture" e rilevata come una perdita di copia, se la dimensione dell'inserzione è dello stesso ordine di quella della regione chiamata.

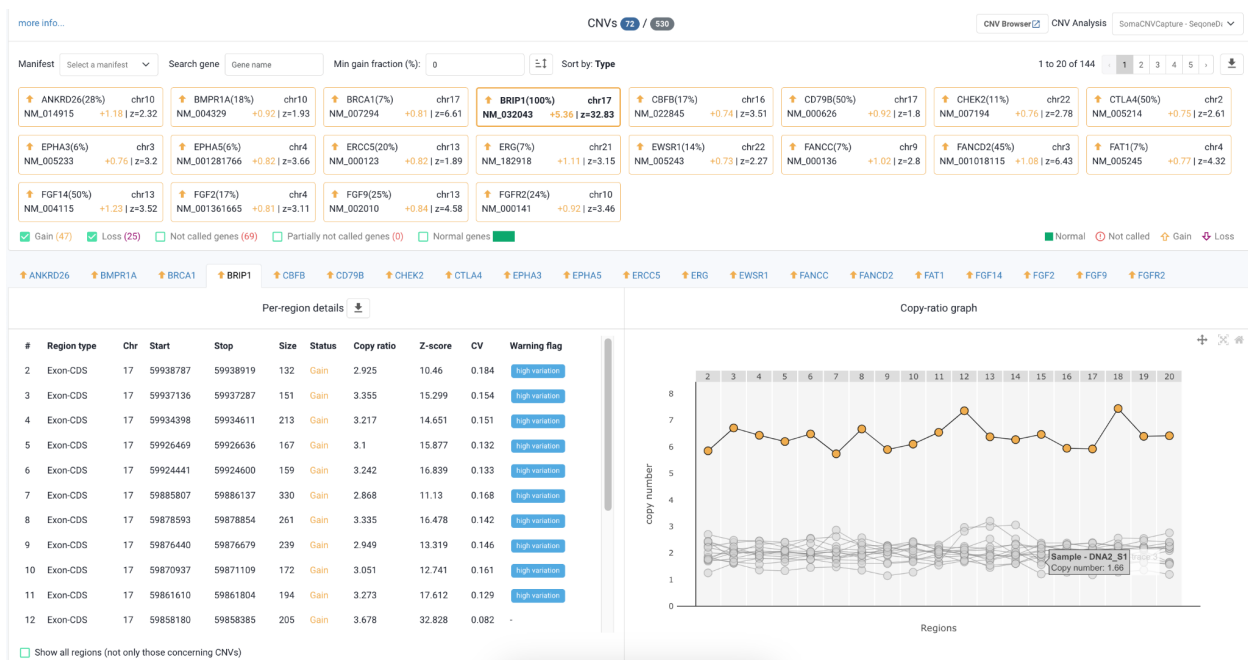
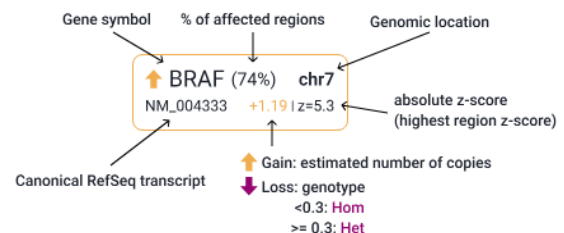


Figura 23: Visualizzazione della scheda "CNVs" in un'analisi SomaVar

Nel caso delle analisi di **SomaCNVCapture**, sebbene il metodo di rilevamento sia simile a quello di CNVCapture, la scala utilizzata per la rappresentazione grafica delle varianti è il gene, anziché l'esone. La rappresentazione è rappresentata da una serie di caselle. In questo modo vengono forniti il simbolo del gene, la posizione genomica, il "Transcripts" canonico, lo "Z-score" assoluto e, in caso di guadagno, il numero di copie o, in caso di perdita, il genotipo.



Una finestra pop-up che spiega queste caselle è disponibile sopra i filtri nella sezione "[more info...](#)".

La percentuale visualizzata in ogni casella riflette la frazione del gene (*min gain fraction*) interessata dalla CNV rilevata. La percentuale di guadagno può essere utilizzata per filtrare i risultati. Per impostazione predefinita, il filtro è impostato sullo 0% e si applica solo ai guadagni di copie.

Le delezioni rilevate da questa pipeline non sono soggette a questo filtro e devono essere interpretate con cautela. Infatti, la capacità di questo strumento di rilevare variazioni di una copia genica dipende dalla natura e dalla qualità del campione, in particolare dalla frazione tumorale.

Warning: Se sullo stesso gene vengono rilevate sia un'amplificazione che una delezione, il gene sarà considerato amplificato per impostazione predefinita e quindi soggetto al filtro "min gain fraction" descritto sopra.

CNV Browser.

Nella scheda di visualizzazione CNV delle analisi GermlineVar, GermlineFamily, SomaVar o

SomaHemato è disponibile un'icona  per aprire un CNV Browser in una nuova scheda. Il grafico CNV Browser visualizza la "B allele frequency" (BAF) degli SNP (Single Nucleotide Polymorphism) presenti nel manifesto del campione analizzato e presenti in un set di dati di array SNP di Illumina e SNP di altri backbone come OneSeq di Agilent, per facilitare la visualizzazione di alcuni tipi di eventi come 'CNV Browser'.

Il BAF è la frequenza dell'allele non di riferimento (allele B) osservata per un SNP eterozigote germinale.

Per impostazione predefinita, il grafico rappresenta il BAF delle varianti (asse y) sull'intero genoma (asse x), suddiviso per cromosoma (1-22, X e Y). Un filtro consente all'utente di visualizzare le varianti su un determinato cromosoma nell'angolo in alto a sinistra.

Nel contesto delle analisi CNV, queste informazioni BAF possono essere utilizzate per mostrare un'alterazione del numero di copie allele-specifiche, una perdita di eterozigosi o uno squilibrio allelico.

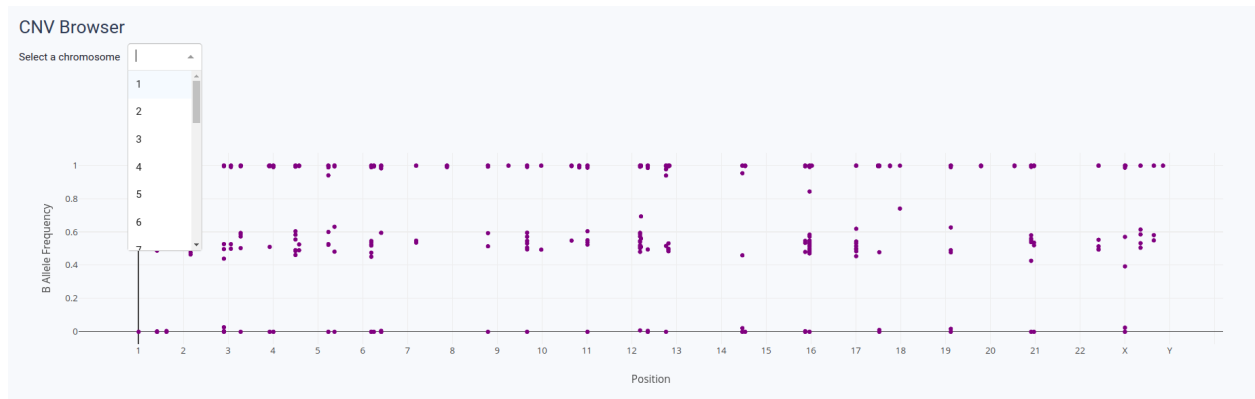


Figura 24: Visualizzazione della scheda "CNV Browser" in un'analisi GermlineVar.

Esportazione dei dati CNV

L'icona  in alto a destra può essere utilizzata per esportare i dati dell'analisi CNV sotto forma di un file .tsv contenente le seguenti colonne:

- gene (nome_gene),
- RefSeq transcript (refseq_id),
- stato del gene (status),
- cromosoma (Chr),
- coordinate genomiche della CNV: posizione iniziale e finale,
- frazione del gene interessato dalla CNV (gene_fraction_cnv),
- rapporto del numero di copie (Copy-ratio) rispetto al normale (n=2),
- numero di deviazioni standard che separano un risultato dalla media (zscore)
- numero di esoni che si sovrappongono alla CNV (esoni), nel caso di un'analisi CNVCapture.

Per ogni gene, l'icona  accanto a "Per-region details" può essere utilizzata per esportare specificamente i dati dell'analisi CNV per il gene selezionato sotto forma di file .tsv contenente i seguenti campi:

- gene (nome_gene),
- cromosoma (Chr),
- coordinate genomiche: posizioni iniziali e finali,
- numero di esoni o introni (exon_intron_number),
- Status",
- tipo (esone o introne, tipo),
- messaggi di errore associati (warning_flags, fail_flags),
- valore di probabilità critico aggiustato (padj),
- valore di probabilità critica (pvalue),
- rapporto tra il numero di copie e la norma (ratio),
- misura della dispersione relativa all'interno dei campioni (variabilità),
- deviazione standard del risultato rispetto alla media (zscore)

13.1.5.2. CNV nell'analisi di SomaCGP e SomaLBx

Metodologia

SomaCGP e SomaLBx utilizzano PureCN²⁶ per rilevare CNVs e LOHs (loss of heterozygosity). PureCN identifica le varianti somatiche dai dati di rilevamento delle varianti per valutare la ploidia e la purezza tumorale del campione. Queste informazioni vengono utilizzate per determinare il numero di copie di ciascun evento.

Visualizzazione e interpretazione

La scheda di visualizzazione delle CNV in SomaVar Tertiary analysis funziona in modo identico a quella di SomaCGP e SomaLBx, come descritto di seguito.

Le CNVs sono presentate in una tabella filtrabile. Ogni CNV visualizza le seguenti informazioni:

- **Gene:** nome del gene
- **Location:** cytoband
- **Coordinate dell'evento:** coordinate e dimensioni dell'evento.
- **Transcript:** transcript di riferimento
- **Exons:** Exons affected?
- **Type:** tipo di evento (LOHs, Loss o Gain)
- **Copy number:** numero assoluto di copie
- **Function:** funzione del gene, se applicabile, Oncogene o Tumor Suppressor Gene (TSG)

L'utente può cercare le CNVs che interessano un determinato gene tramite un campo di ricerca e scegliere se visualizzare o meno i geni presenti nelle regioni LOH selezionando la casella "Show genes in LOH regions". La tabella dei risultati può essere esportata utilizzando l'icona di download in alto a destra.

CNV Browser 

L'icona apre un *CNV Browser* in una nuova scheda. In questa scheda viene visualizzato un grafico che mostra la "B allele frequency" (BAF) degli SNP nel campione analizzato e degli SNP LOH provenienti da pubblicazioni selezionate e dal backbone Agilent di "Selection". Un secondo grafico rappresenta i dati logRatio generati da PureCN. Maggiori dettagli sono disponibili nella sezione **CNV in pannello di geni** in **CNV Browser**.

²⁶ <https://github.com/lima1/PureCN>

13.1.5.3. CNV negli esomi e nei genomi

La scheda di visualizzazione delle CNV per i dati esomici è disponibile nelle analisi GermlineVar, GermlineFamily, GermVar e GermVarFamily quando si utilizzano file FASTQ come input primario e si esegue un'analisi CNVExome. Questa scheda visualizza anche i risultati CNV per i dati del genoma nelle analisi GermVar e GermVar Family utilizzando file FASTQ, nonché nelle analisi GermVar Tertiary e GermVar Tertiary Family utilizzando file VCF o dati CGH MicroArrays.

Metodologia per gli esomi

Questa metodologia si applica specificamente allo strumento CNVExome. Proprio come le pipeline di analisi CNV su dati di pannello di geni, il processo di identificazione delle varianti in questo caso dipende da un'analisi della profondità di copertura. Tuttavia, l'analisi CNVExome con lo strumento [GATK GermlineCNVCaller](#) è individuale, poiché ogni campione viene confrontato con un modello HHMM (*Hierarchical Hidden Markov Model*) calcolato da SeqOne.

Per tenere conto delle variazioni tecniche legate alla preparazione dei campioni, questo modello si basa su un'ampia coorte che va da una a diverse centinaia di campioni sequenziati dal laboratorio che utilizza:

- Lo stesso protocollo di preparazione delle librerie
- Lo stesso kit di cattura.
- La stessa piattaforma di sequenziamento e la stessa lunghezza di read

Questi requisiti e il numero variabile di campioni necessari per impostare il modello possono essere spiegati dalla necessità di una coorte che offra un livello omogeneo di copertura. Si devono prendere in considerazione diversi fattori, come la profondità di sequenziamento, il grado di variabilità all'interno della coorte e la riproducibilità dei protocolli di preparazione dei campioni.

I metadati relativi al sesso (vedere **Metadati campione**) devono essere compilati prima di avviare l'analisi CNVExome per consentire la verifica della ploidia dei cromosomi sessuali. Se non viene compilato, sull'interfaccia di visualizzazione del CNV viene visualizzato un messaggio di avviso "Ploidy".

Limitazione: La documentazione di 'GATK' GermlineCNVCaller raccomanda di escludere le regioni PAR (PseudoAutosomal Regions), cioè le regioni PAR1 e PAR2, se il loro contributo alla copertura totale dell'X è sproporzionato. In effetti, è difficile calcolare la ploidia per le estremità del cromosoma X corrispondenti alle regioni PAR per molti modelli calcolati da SeqOne. Di conseguenza, la ploidia non viene calcolata correttamente per queste regioni, che vengono rilevate in modo artefatto a 1 copia, mentre l'intero cromosoma X viene rilevato a 2 copie.

Gli utenti sono invitati a contattare il team di supporto agli utenti di support@seqone.com per saperne di più sulla procedura di impostazione di CNVExome nel loro gruppo.

Metodologia per il genoma

Per le versioni fino alla v3.6, le analisi GermVar e GermVar Family sui dati del genoma integrano lo strumento Sentieon CNVScope²⁷ per il rilevamento delle CNV. A partire dalla versione v3.7, i workset GermVar e GermVar Family utilizzano lo strumento Cue²⁸ per l'identificazione delle CNV sui dati genomici.

I dati di duplicazione segmentale vengono rilevati utilizzando Sentieon's SegDup Caller, che è specificamente progettato per identificare eventi di numero di copie in regioni genomiche altamente omologhe su dati WGS solo utilizzando GermVar. Questo caller ha come target una serie curata di geni clinicamente rilevanti noti per essere impegnativi per gli approcci standard di rilevamento delle varianti. Gli eventi risultanti vengono automaticamente integrati nella pipeline di analisi e visualizzati nel visualizzatore di CNVs & SVs come CNV (sia guadagno che perdita), consentendo l'identificazione di varianti in regioni genomiche che erano precedentemente inaccessibili all'analisi standard.

Vista e interpretazione

Nella scheda **CNV** è disponibile un'interfaccia di interpretazione per le analisi CNVExome, GermVar e GermVarFamily, GermVar Tertiary o GermVar Tertiary Family. L'interfaccia è divisa in due sezioni: una tabella dei risultati CNV a sinistra e una pagina CNV dettagliata con più schede a destra.

Tabella dei risultati del CNV

Una tabella dei risultati mostra le CNVs identificate, insieme a un menu di filtri basati sulla loro presenza in "OMIM gene" morbosi [<https://omim.org/>], sulle dimensioni, sulla frequenza di comparsa delle CNVs nella coorte o sul LogRatio nel caso di analisi CGH MicroArray.

Il numero di CNV filtrate (CNV selezionate) e il numero totale di CNV rilevate (elementi in totale) sono visualizzati nella parte superiore della tabella CNV.

392 elements in total
CNVs 66 / 392

0 in locked panel
392 elements in total
CNVs 0 / 0

²⁷ <https://support.sentieon.com/appnotes/cnvscope/>

²⁸ <https://github.com/PopicLab/cue>

L'icona di un lucchetto indica che l'analisi utilizza un 'Restricted panel' e i dettagli delle CNVs appaiono nei tooltip associati.

Un messaggio di avviso  viene visualizzato quando viene rilevato un valore di ploidy' anomalo e/o quando i metadati relativi al genere (vedere **Metadati del campione**) non sono stati compilati prima dell'avvio dell'analisi.

Facendo clic sull'icona "Ploidy", l'utente può visualizzare la ploidia di ciascun cromosoma e il sesso del paziente. I valori anomali sono evidenziati in giallo.

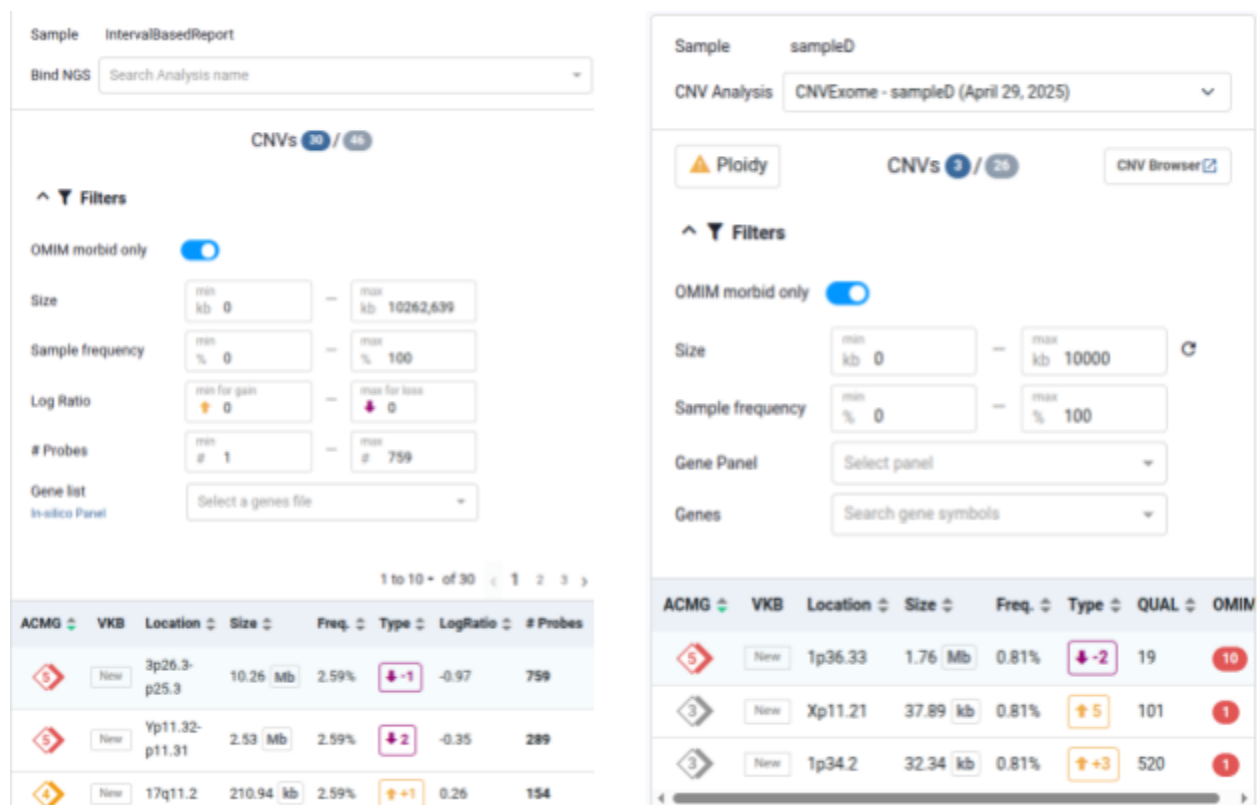


Figura 25: Panoramica delle 'CNVs' rilevate e menu dei filtri di un'analisi CGH (sinistra) e di un'analisi standard (destra).

Nel caso di un'analisi CGH MicroArrays combinata con un array SNP (CGH+SNP), gli eventi di loss of heterozygosity (LOH) rilevati sono riportati nella tabella dei risultati CNV. Questi eventi corrispondono alla perdita di eterozigosi neutrale (CN-LOH). In altre parole, la regione conserva due copie, ma identiche, caratterizzate dalla presenza di SNP omozigoti.

Nel caso di un'analisi CGH MicroArray, è possibile selezionare un'analisi SNV non-CGH (GermlineVar, GermlineFamily, GermVar Tertiary e GermVar Tertiary Family) dal menu a discesa "Bind NGS" per visualizzare le varianti che si sovrappongono alla CNV rilevata.

Filtri

OMIM morbid only. Questo filtro è attivato per impostazione predefinita per identificare facilmente le CNVs che interessano i geni clinici, annotati OMIM Morbid sulla base del loro identificatore NCBI (NCBI gene_id) o, in mancanza, sulla base del nome del gene e dei sinonimi (gene_symbol e gene_synonyms). È sufficiente sollevare questo filtro per visualizzare tutte le 'CNVs' rilevate nel campione.

Size. L'utente può filtrare le CNVs in base alla loro dimensione prevista (in kb).

Sample frequency. L'utente può filtrare le CNVs in base alla loro dimensione e frequenza (Sample frequency). Questa frequenza corrisponde all'occorrenza della CNV tra tutti i campioni analizzati fino alla data dell'analisi e confrontati con lo stesso modello, compresi i campioni utilizzati per costruire il modello. La Sample frequency si basa su una sovrapposizione del 95% e considera lo stesso tipo di CNV (gain/loss/deletion). Passando il mouse sul valore visualizzato in percentuale, l'utente vedrà apparire la frazione corrispondente. L'aggiornamento della frequenza può richiedere alcuni minuti dopo l'apertura dell'analisi.

Gene Panel. Come un pannello di geni in silico, questo filtro consente all'utente di selezionare e applicare un elenco di geni per filtrare le CNV rilevate nell'analisi utilizzando il menu a discesa (vedere **Gestione delle impostazioni utente**). Questo filtro non è disponibile se all'analisi è stato applicato un 'Restricted panel' dalla scheda Panel del pannello laterale di gestione dell'analisi (vedere la sezione **Vista Modale Panel**).

Genes. Permette di filtrare le CNVs rilevate contenenti i geni selezionati tramite questo filtro.

Log Ratio. Per le analisi basate sui dati CGH MicroArrays files, il Log Ratio (tra il segnale grezzo e il segnale di controllo) presente nei file di input Agilent viene visualizzato in una colonna aggiuntiva. Il filtro associato consente all'utente di selezionare i valori minimi per i guadagni di copie (min per il guadagno) e i valori massimi per le cancellazioni di copie (max per la perdita). Gli eventi LOH (loss of heterozygosity) vengono mantenuti anche dopo l'applicazione di questo filtro.

#Probes. Per le analisi basate su dati CGH MicroArrays, la colonna "Analyses" indica il numero di sonde utilizzate in ciascuna regione (esonica, intergenica o intronica). Questi valori vengono presi dal file di input IntervalBasedReport.xls e il filtro associato consente all'utente di selezionare il numero minimo e massimo di sonde utilizzate. In un'analisi combinata CGH+SNP, il numero di sonde indica il numero di SNP omozigoti nella regione LOH.

Smistamento CNV

L'elenco delle CNVs può anche essere ordinato per classe di patogenicità ACMG (ACMG), posizione genomica (*Location*), dimensione (*Size*), frequenza del campione (*Freq.*), numero di copie (*Type*), punteggio di qualità (QUAL) in GermVar Tertiary e per LogRatio in CGH MicroArray (*LogRatio*).

ACMG. La classe di patogenicità di ciascuna variante è calcolata in modo semiautomatico secondo le raccomandazioni dell'*American College of Medical Genetics and Genomics*. La classe di patogenicità è la modalità di ordinamento predefinita nella tabella CNV.

Gli eventi LoH (loss of heterozygosity) in CGH MicroArrays sono sempre assegnati alla classe 3 (VUS).

Type (Copy number). La variazione del numero di copie viene visualizzata (ad esempio: +2, -1, +3) per le CNV rilevate sui cromosomi autosomici (da 1 a 22), poiché il numero di copie è impostato su 2 per impostazione predefinita.

Tuttavia, per le CNVs rilevate sui gonosomi (cromosomi sessuali X e Y) viene visualizzato il numero di copie assoluto (es. 0, 1, 2) e non il fold change.

A differenza dell'analisi del pannello di geni, i metadati relativi al sesso del campione non sono necessari per includere i cromosomi sessuali nella ricerca di CNV nell'esoma.

Per le analisi combinate CGH+SNP, gli eventi CN-LOHs, che conservano due copie, sono rappresentati nella colonna Type dal distintivo **LOH**.

QUAL. Il punteggio di qualità Phred viene calcolato in CNVExome dal caller di GATK. Quando una CNV è composta da diversi intervalli genomici, il punteggio QUAL più alto tra tutti gli intervalli viene assegnato alla CNV. Per le analisi terziarie, il punteggio QUAL visualizzato è quello ottenuto dal file VCF di input (DRAGEN o ONT).

CNV Browser

L'icona apre un "CNV Browser" in una nuova scheda che visualizza la "B allele frequency" (BAF) degli SNP (Single Nucleotide Polymorphism) presenti nel campione analizzato. Il grafico è descritto nella sezione precedente (CNV in pannello di geni, sotto CNV Browser).

Pagina dettagliata CNV

Riepilogo CNV. Questa scheda iniziale presenta una serie di informazioni e annotazioni sulla CNV selezionata, tra cui:

- Dettagli della classificazione automatica ACMG.
- Le regioni cliniche e i geni inclusi nell'intervallo interessato dalla variante. L'annotazione delle regioni cliniche si basa sui database [ClinGen](#) e [OMIM](#). Per impostazione predefinita, vengono visualizzati solo i geni clinici annotati da OMIM Morbid. Per visualizzare tutti i geni clinici, l'utente può attivare la funzione "Show all genes" sulla destra della pagina.
- Sovrapposizioni tra la CNV rilevata e il database [DGV](#).
- Le piccole varianti (SNV e indel) rilevate nei geni affetti da CNV. Per impostazione predefinita, il filtro SeqOne-diag è attivato. Per disattivare questo filtro, fare clic su SeqOne-diag e deselezionare SeqOne-diag dai profili dei filtri. Per visualizzare tutte le varianti, gli utenti possono attivare la funzione "Show all variants" sulla destra della pagina.
- Una sezione dedicata alla "Variant Base" (VKB) per l'identificazione di varianti identiche oltre il 95% che sono già state validate nel gruppo e le rispettive valutazioni.

ACMG Classification (Automatic) ⓘ

Likely Pathogenic (+0.9)

Genomic content

Genes or Genomic regions

Gene number

Literature and databases

Inheritance patterns

1A 0

2C,1 +0.9

3A 0

No selected tags, only 40 is assessed.

Not assessed

Clinical regions

Region name	Overlap	Link	Description
CHROMOSOME 16p13.3 DELETION SYNDROME, PROXIMAL	0.29%	OMIM	
CHROMOSOME 16p13.3 DUPLICATION SYNDROME	0.29%	OMIM	
ALPHA-THALASSEMIA/IMPAIRED INTELLECTUAL DEVELOPMENT SYNDROME, DELETION TYPE	0.29%	OMIM	

Clinical genes

Show all genes

Gene symbol	Databases	Phenotype(s)	ClinGen haploinsufficiency score	pLI (GnomAD)	Exons	Fraction overlap	Variants
PKD1	omim	[AD] Polycystic kidney disease, ad...	3 (Sufficient evidence for dosage pathogenicity)	0.12051	12/46	48.39%	

DGV overlaps

ID	Fraction overlap	CNV type	CNV size	Localisation	Frequency	# Samples	# Studies
gssvG13891	100%	↑ Gain	51.76kb	16:2,138,711-2,190,473	20.24%	17	5
gssvG13898	88.78%	↑ Gain	25.42kb	16:2,165,630-2,191,053	0.13%	2	2
gssvL42059	44.04%	↓ Loss	25.71kb	16:2,147,425-2,173,136	100%	2	2
gssvL42061	32.63%	↓ Loss	10.27kb	16:2,178,475-2,188,740	7.46%	5	2

Variants

SeqOne-diag

0 / 51

Show all variants

No variant found

Figura 26: Panoramica della scheda di riepilogo CNV

- L'IGV del Genome browser in cui è possibile visualizzare le CNVs (*non disponibile nelle analisi GermVar Tertiary e GermVar Tertiary Family*). Per verificare la stabilità del segnale, è possibile visualizzare fino a cinque altri pazienti oltre all'individuo analizzato dal menu di configurazione  del grafico che rappresenta il numero di copie (DCN).

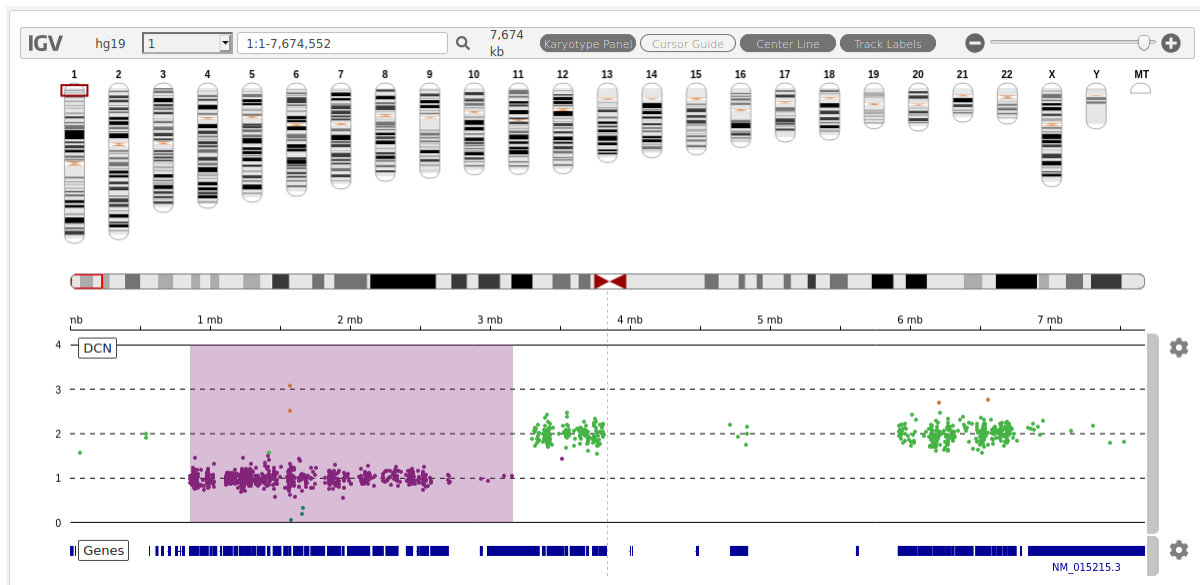


Figura 27: Vista delle CNVs da un browser integrato "IGV".

Il codice colore è il seguente per il paziente selezionato: blu (da 0 a 0,5 copie), viola (da 0,5 a 1,5 copie), verde (da 1,5 a 2,5 copie), arancione (da 2,5 a 3,5 copie) e rosso (>3,5 copie). Gli altri campioni sono visualizzati in grigio.

CNV Occurrence. Questa scheda viene utilizzata per identificare altri campioni contenenti la CNV, indipendentemente da qualsiasi valutazione. Per ogni occorrenza, vengono inserite le seguenti informazioni:

- il nome del campione (Sample),
- il nome dell'analisi e un link ad essa (Analysis),
- il nome del progetto (Project),
- la posizione genomica della CNV (location),
- la dimensione della CNV (Size),
- il tipo di CNV,
- la sovrapposizione delle frazioni.

I risultati sono ordinati in ordine decrescente in base alla percentuale di sovrapposizione. È possibile filtrare le CNVs visualizzate in base alla frazione minima di sovrapposizione

compilando il campo "min fraction overlap (%)" in cima all'elenco. Per impostazione predefinita, questo filtro è impostato sull'80%.

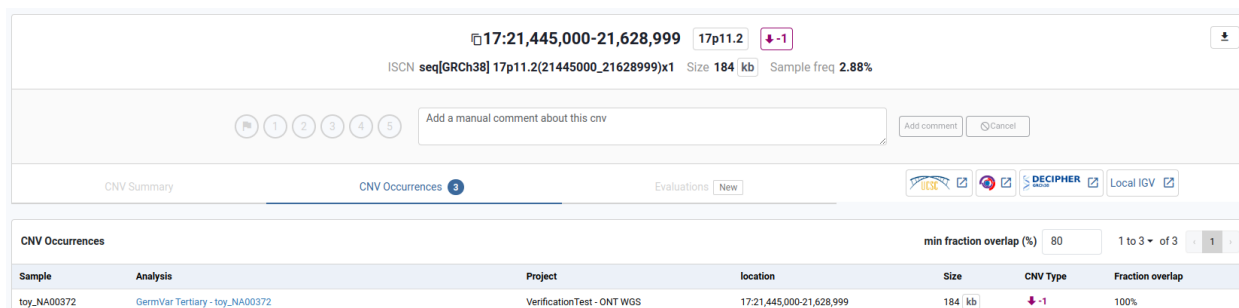


Figura 28: Panoramica della scheda "Occurrence" di CNVExome.

Evaluations. La 'Pathogenicity classification' per ogni CNV può essere definita dall'utente, accompagnata da interpretazioni sotto forma di commenti. Ogni nuova valutazione contribuisce inoltre a popolare la base di conoscenze personali del laboratorio.

UCSC. Un link esterno al "CNV Browser" dell'[UCSC](#) consente agli utenti di visualizzare la regione genomica della CNV. Questa pagina si apre in una scheda aggiuntiva del browser e include le impostazioni personali dell'utente per quel browser, offrendo così una visualizzazione personalizzata.

CNV-Hub. Un link esterno al classificatore [CNV-Hub](#) consente di visualizzare la classificazione delle CNV tenendo conto di più strumenti di classificazione. Il link apre un'ulteriore scheda del browser e integra le impostazioni personali del browser dell'utente, offrendo una visualizzazione personalizzata.

DECIPHER Genome Browser. Un link esterno al [DECIPHER](#) Genome Browser consente di visualizzare le "CNVs" sul genoma. Il collegamento apre un'ulteriore scheda del browser e integra le impostazioni personali del browser dell'utente, offrendo una visualizzazione personalizzata.

Local IGV. Questo link reindirizza l'applicazione alle coordinate della CNV in corso di interpretazione. Per visualizzare la CNV, è necessario aver precedentemente caricato il BAM del campione sull'applicazione IGV Desktop. Questa opzione è compatibile solo con le versioni di IGV Desktop ≥ 17 . *IGV Desktop*²⁹

Oltre ai file di allineamento, sono disponibili vari file da scaricare dalla scheda *Files* dell'analisi CNVExome:

- I risultati dell'analisi CNV in un file tabellare (cnv.tsv).

²⁹ <https://igv.org/doc/desktop/>

- La ploidia per ogni cromosoma (contig_ploidy.tsv)
- Il numero previsto di copie per ciascun esone nel manifesto (denoise.tsv)

Esportazione dei dati CNV

L'icona  in alto a destra consente di esportare un file TSV contenente i seguenti dati per il CNV in questione:

- il rating e il commento di "VKB",
- Valutazione ISCN,
- Classificazione ACMG,
- cromosoma, posizioni iniziali e finali e citobanda,
- Dimensione e tipo di CNV,
- Copy number: numero di copie
- Copy number: numero di copie
- Sample Freq.
- punteggio di qualità (scala Phred)
- Rapporto Log2
- numero di sonde coinvolte (Probes).
- numero di geni interessati dalla CNV.

Le informazioni sui geni associati sono descritte nelle colonne seguenti:

- nome del gene (Gene_Symbol),
- Fenotipo OMIM (OMIM_Phenotypes),
- probabilità di intolleranza genica al LoF (Pli_Score)
- Punteggi di aploinsufficienza e triplosensibilità ClinGen (HI_Score, TS_Score)
- il numero di esoni sovrapposti dalla CNV (Nb_overlap_exon),
- il numero totale di esoni del gene (Nb_total_exon).

Se alcuni campi descrittivi riguardano solo alcune CNVs (nel caso di CGH MicroArrays, ad esempio), questi campi hanno il valore "NA".

13.1.6. Visualizzatore di varianti grandi per CNVs e SVs (LVV)

Il visualizzatore di varianti grandi (Large Variant Viewer, LVV) consente all'utente di visualizzare e analizzare CNVs e varianti strutturali (SVs) in un'unica scheda. L'interfaccia è composta da una tabella centrale delle varianti, un pannello dei filtri a sinistra e un pannello laterale che si apre a destra quando viene selezionata una variante, simile alla scheda delle varianti (sezione Varianti piccole).

Nelle analisi compatibili, il visualizzatore si trova in una scheda etichettata "CNVs & SVs" (CNVs e SVs) o "SVs", a seconda del tipo di dati:

- La scheda CNVs & SVs mostra CNVs (Perdita/Guadagno), SVs (Delezione/Duplicazione) ed eventi LOH/UPD.
- Scheda "CNV" precedente: I CNVs vengono visualizzati anche nel visualizzatore precedente per mantenere la retrocompatibilità con i flussi di lavoro esistenti.
- I SVs sono accessibili solo nella scheda "CNVs & SVs" o "SVs".

La disponibilità dei dati nel visualizzatore di varianti grandi dipende dal tipo di analisi:

- WGS: Completamente compatibile (CNVs, SVs, UPDs).
- WES: Completamente compatibile (CNVs, SVs, UPDs). I CNVs vengono visualizzati solo se il modulo di chiamata CNV viene eseguito come parte dell'analisi.
- Pannelli: Parzialmente compatibile (solo SVs).
- Analisi terziaria (VCF/aCGH):
- aCGH: Vengono visualizzati tutti gli eventi relativi al numero di copie.
- VCF: Vengono visualizzati CNVs e SVs (alleli ALT simbolici). I SVs con sequenze ALT specifiche vengono visualizzati nel visualizzatore di varianti piccole.

Tabella dei risultati di CNVs & SVs

Simile alla tabella delle varianti piccole, la tabella principale mostra una riga per evento, con le seguenti colonne che forniscono metriche specifiche della variante:

DiagAI. Punteggio di priorità da 1 a 100, adattato per le varianti grandi, che classifica le varianti in base alla probabilità che siano causa di malattia e rilevanti a livello diagnostico nel contesto del caso (in particolare quando vengono forniti i termini HPO). Il punteggio DiagAI combina più componenti, tra cui: previsione della patogenicità tramite il punteggio UP², corrispondenza del fenotipo tramite PhenoGenius (guidata da HPO) e regole degli esperti/di ereditarietà e di qualità.

ACMG. L'indicatore di classificazione ACMG basato su varianti grandi è mostrato come un punteggio di classificazione da 1 a 5 nella tabella LVV. Nel cassetto dei dettagli, una scheda ACMG mostra i criteri sottostanti e un riepilogo della classificazione in tempo reale da Patogeno a Benigno (principalmente per CNV Guadagno/Perdita).

Classificazione (icona della bandiera) che consente all'utente di classificare la variante nella VKB.

Tipo di CNV/SV. Tipo di CNV/SV (guadagno o perdita), dimensione, posizione e citobanda, Regioni (esonica, intronica, UTR) e geni che si sovrappongono al CNV (nomi dei geni HGNC), presenza in OMIM, frazione di geni codificanti e dettagli sugli esoni sovrapposti. Le varianti sono annotate con geni sovrapposti e geni situati entro 5 kb dai punti di rottura (margine di 5 kb).

Geni associati a una variante sono ordinati per rilevanza clinica (ad es., stato morboso in OMIM, percentuale di sovrapposizione, punteggi di restrizione). I tre geni più rilevanti sono visualizzati nella tabella principale. L'elenco completo è disponibile nella scheda "Genes & Regions" (Geni e Regioni) del pannello laterale ("Details Drawer" (Cassetto dei dettagli)).

RMSK %. Percentuale dell'evento che si sovrappone alle annotazioni RepeatMasker (ad es., regioni ripetitive).

ClinVar & Decipher (Database clinici). Entrambe le colonne mostrano il numero di corrispondenze dal database clinico corrispondente. La corrispondenza migliore è definita come un evento in cui la sovrapposizione del database è > 50%. Le classificazioni ClinVar e Decipher sono mappate in Benigna (B), VUS o Patogena (P).

DGV Gold & GnomAD (Database di popolazione). La tabella mostra la frequenza di popolazione massima trovata per la variante. Gli eventi sono considerati corrispondenti se la sovrapposizione della variante è > 80% e il tipo di evento è equivalente. La scheda Frequenze nel pannello laterale elenca tutti gli eventi sovrapposti con una sovrapposizione > 1%.

Phenotypes (Fenotipi). Cliccando sui numeri visualizzati nelle colonne corrispondenti ai risultati dell'algoritmo Phenogenius, l'utente può accedere a:

- la percentuale di corrispondenza, che è il punteggio di corrispondenza più alto tra un singolo gene interessato e i termini HPO del paziente.
- la frazione di corrispondenza, che è il numero di termini HPO del paziente che corrispondono ad almeno un gene interessato.
- Per l'analisi familiare, la corrispondenza viene eseguita utilizzando i termini HPO del probando.

SampleFreq (Frequenza del campione interna). La frequenza della variante all'interno del progetto e dell'entità correnti è visualizzata nelle colonne corrispondenti.

- Le corrispondenze si basano su > 80% di sovrapposizione di variante e tipo di evento identico.

- Frequenza di aggiornamento: i dati si basano su un'istantanea delle analisi eseguite prima dell'avvio dell'analisi corrente. Un pulsante manuale "Refresh" (Aggiorna) è disponibile nel Details Drawer per aggiornare questi dati.

Colonne dei punteggi. L'utente può visualizzare più punteggi:

- UP²: punteggio di previsione della patogenicità che è un componente di DiagAI.
- pLI, LOEUF: intolleranza alla perdita di funzione.
- pTriplo, pHAPLO: Triplosensibilità e Aploinsufficienza.
- HI ClinGen, TS ClinGen.

Colonne del campione. Queste colonne mostrano le metriche degli eventi specifiche per l'origine dati:

- Per le analisi familiari (escluso aCGH), le metriche del campione vengono visualizzate per tutti i membri della famiglia.
- Genotype: per i CNVs in cui non viene fornito un genotipo dal chiamante (caller), il sistema deduce il genotipo in base al numero di copie assoluto relativo alla ploidia:
 - Omozigote (hmz) con punto completamente scuro
 - Eterozigote (htz) con punto scuro a metà
- VAF / #Reads: frequenza allelica della variante e letture di supporto (disponibile per - SVs).
- #Copy: numero di copie assoluto (disponibile per CNVs).
- LogRatio / Probes: disponibile per i dati aCGH.
- QUAL: punteggio di qualità del chiamante della variante o del file di input.

Ordinamento, Filtraggio, Visualizzazione e Interpretazione

Filtraggio. Il pannello dei filtri consente all'utente di filtrare le varianti in base ai dati della colonna, in modo simile al sistema di filtraggio della scheda Varianti piccole (vedere la sezione corrispondente).

- Pannello dei geni: filtra gli eventi che si sovrappongono a geni in un elenco di geni specifico.
- Qualità: filtra in base alla colonna QUAL. Per impostazione predefinita, gli eventi senza punteggio di qualità (N/A) sono nascosti a meno che non venga selezionato "Keep unknown" (Mantieni sconosciuti).
- Preimpostazioni: le configurazioni dei filtri possono essere salvate come preimpostazioni per un uso futuro.
- Per le analisi familiari, il filtro Qualità utilizza la logica "OR" (O). La variante viene visualizzata se almeno un membro della famiglia soddisfa la soglia di qualità.

Ordinamento. Per impostazione predefinita, la tabella è ordinata per DiagAI Score (decrescente). L'utente può ordinare per UP2, Size, Location o Sample Freq. Entity, in ordine crescente o decrescente.

Come nella precedente scheda CNV, l'utente può aprire le finestre di ploidia e il browser CNV nella parte superiore della tabella (vedere la sezione CNV: Visualizzazione e interpretazione). Un'icona per la gestione delle colonne è disponibile anche all'estrema destra dell'intestazione.

Pannello laterale CNV/SV (Details Drawer)

Simile alla pagina delle varianti (sezione Varianti piccole), la selezione di una variante nel Large Variant Viewer apre una pagina dettagliata sul lato destro dello schermo. Questa pagina contiene più schede:

Varianti. Questa scheda di riepilogo include la nomenclatura della variante, il punteggio DiagAI, l'elenco dei sospetti eterozigoti composti (Suspected Het Compound), le informazioni su gene e regione e i collegamenti ai database.

ACMG. Classificazione dettagliata.

Geni e regioni. l'elenco completo dei geni e delle regioni associati alla variante, con informazioni cliniche dettagliate come lo stato morboso OMIM, la percentuale di sovrapposizione, i punteggi di restrizione e la corrispondenza del fenotipo.

Classificazione. Classifica la variante corrente nel tuo VKB nella parte superiore del pannello laterale e salva. L'attuale classificazione e le occorrenze di questa variante in altre analisi possono essere trovate in questa pagina.

Frequenze. L'utente può accedere ai dettagli sulle frequenze delle varianti nei database DGV Gold e GnomAD, nonché per l'entità e il progetto corrente.

Database clinici. ClinVar & Decipher.

Viewer. L'utente può scegliere di visualizzare la variante nel SeqOne viewer o in IGV. Nella sottoscheda IGV, l'utente può visualizzare lo "Start" (Inizio) o l'"End" (Fine) dell'evento, o "Start to End" (Dall'inizio alla fine). Nella sottoscheda SeqOne, l'utente può accedere al Genome Browser e visualizzare l'evento in uno dei molteplici livelli disponibili: CNV, SV, UPD, Gene, CN (copy number), BAF, DGV.

13.1.7. Firme nell'analisi di SomaCGP

Metodologia di calcolo del punteggio MSI

L'analisi MSI è un'analisi individuale eseguita rispetto a un riferimento standardizzato. Questo riferimento consiste in un insieme di campioni negativi precedentemente selezionati e raggruppati. Il campione di tumore viene confrontato con questo riferimento utilizzando lo strumento *MSI/sensor'*. Viene quindi calcolato un punteggio, rappresentato dalla percentuale di regioni microsatelliti instabili "covered" dal pannello.

Il pannello di normali attualmente supportato da SomaCGP corrisponde al pannello Agilent SureSelect Cancer CGP. L'uso di un altro kit o di un design personalizzato derivato dal CGP richiederà la creazione di un pannello di normali dedicato.

Metodologia per il calcolo del punteggio TMB Score

Il TMB viene calcolato in tre fasi:

1. Calcolo delle dimensioni delle regioni di codifica incluse nel panel
2. Identificazione di varianti somatiche
3. Calcolo del numero di mutazioni somatiche per megabase (mut/Mb)

L'identificazione delle varianti somatiche si basa su una serie di filtri descritti di seguito:

- Tipo di variante e filtri effetto
 - Solo SNV (gli MNV sono esclusi)
 - Effetti conservati: guadagno di codoni di stop, spostamento della cornice di lettura, variante di splicing (accettore e donatore), ablazione del trascritto, delezione/inserzione in fase, variante missense, variante sinonima.

Le varianti sinonime sono escluse dal calcolo del punteggio TMB non sinonimo.

- Filtri di qualità sulla varianti
 - Varianti rilevate da Mutect2 che superano i filtri FilterMutectCalls
 - Profondità di copertura (DP) $\geq 50X$
 - Punteggio di qualità > 30
 - Rapporto di supporto allelico (ASR) $> 0,1$
- Filtri sulle varianti somatiche
 - Variant Allele Frequency (VAF) tra il 5% e il 40%.
 - Frequenza della popolazione generale ('GnomAD') $\leq 0,0001$

Rischi dell'uso di campioni di controllo TMB sintetici

- Distribuzione delle mutazioni innaturali:
 - Le mutazioni ingegnerizzate potrebbero non riflettere il contesto genomico dei tumori reali (ad esempio, raggruppamento delle mutazioni, bias dei filamenti, contesto della sequenza).
- Mancanza di eterogeneità tumorale:
 - I campioni sintetici non imitano la diversità clonale o la purezza del tumore, il che influisce sulla distribuzione della Variant Allele Frequency (VAF).
 - Nessun rumore di fondo o artefatti FFPE: i campioni reali presentano artefatti da sequenziamento, danni FFPE e variazioni del numero di copie, assenti nei materiali sintetici.
- Differenti dimensioni dei frammenti e comportamento chimico:
 - I costrutti sintetici si comportano spesso in modo diverso durante la preparazione delle librerie, la cattura degli ibridi e il sequenziamento.

Visualizzazione e interpretazione

La scheda "Signatures" visualizza i risultati delle analisi MSI e TMB in 2 sezioni:

Microsatellite Instability. Questa vista mostra informazioni come :

- il numero di regioni microsatelliti con una copertura maggiore o uguale a 50X (Sites above 50X coverage),
- il numero di siti somatici,
- Punteggio MSI in percentuale e intervallo di confidenza (CI) al 95%,
- Status" di MSI

Tumor Mutational Burden. Questa vista mostra le seguenti informazioni:

- la dimensione della regione codificante espressa in Mb (Coding region size),
- il numero di varianti somatiche e non sinonime che passano i filtri (Varianti somatiche che passano i filtri e Varianti somatiche non sinonime che passano i filtri),
- il punteggio TMB di tutte le varianti somatiche che passano i filtri in mut/Mb (TMB Score),
- il punteggio TMB delle varianti non sinonime che passano i filtri mut/Mb.

Ogni risultato può essere scaricato cliccando sull'icona  in alto a destra di ogni sezione come file tsv.

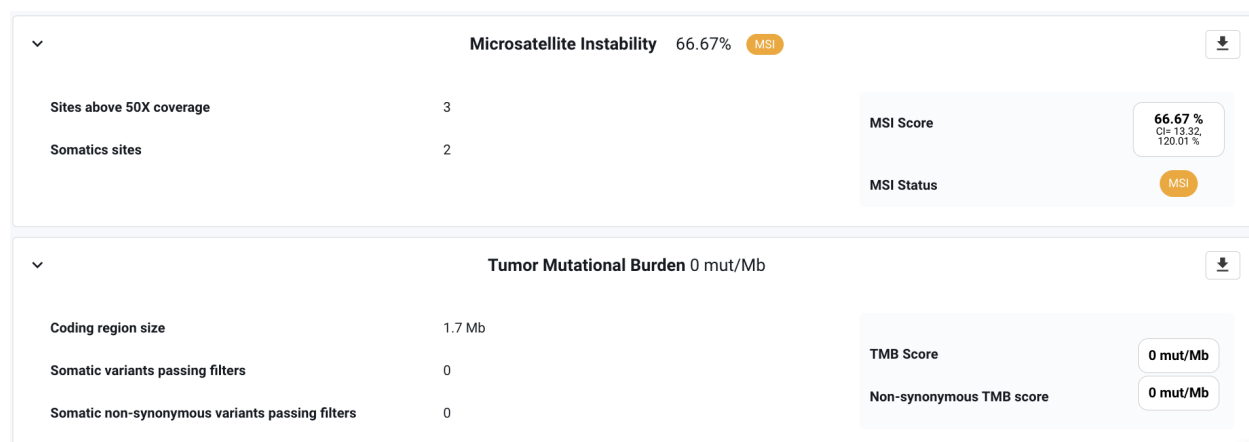


Figura 29: Visualizzazione della scheda "Signatures" di un'analisi SomaCGP.

13.1.8. Brevi ripetizioni in tandem (STRs)

La visualizzazione delle Short Tandem Repeats (STRs) è disponibile per i workset di analisi GermVar Tertiary e GermVar Tertiary Family dedicati all'analisi dei dati di Whole Genome Sequencing (WGS), se generati dal sequenziamento di Oxford Nanopore Technologies.

Il rilevamento delle brevi ripetizioni in tandem (STR) è disponibile anche nei workset di analisi GermVar e GermVar Family per WGS da dati FASTQ generati con Illumina, Element Biosciences, MGI e Oxford Nanopore Technologies.

Metodologia

Brevi ripetizioni in tandem da FASTQ

I dati delle brevi ripetizioni in tandem (STR) vengono rilevati utilizzando ExpansionHunter, che elabora il file BAM insieme a un genoma di riferimento e un catalogo di varianti che definisce i loci target. Il VCF risultante viene quindi annotato da Stranger, che classifica ciascun locus come Normale, Pre-mutazione o Patogeno in base al conteggio delle ripetizioni osservato e alle soglie di malattia note.

Entrambi gli strumenti sono integrati direttamente nelle pipeline GermVar e GermVar Family.

ExpansionHunter: <https://github.com/Illumina/ExpansionHunter>

Stranger: <https://github.com/moonso/stranger>

Brevi ripetizioni in tandem da VCF

I dati delle ripetizioni tandem sono estratti dal file VCF (sample.wf_str.vcf.gz) generato dal flusso di lavoro wf-human-variation v2.3.0 dell'applicazione EPI2ME (Oxford Nanopore Technologies) e fornito dall'utente. Poiché questi dati sono stati annotati per ciascun locus 'Locus', a partire da un catalogo di variazioni mantenuto da EPI2ME, la nostra pipeline non modifica le annotazioni. Lo status premutato o patogeno di ciascun allele è determinato in base al numero di ripetizioni sequenziate e alle soglie note di premutazione e patogenicità.

Vista e interpretazione

La scheda STRs consente di visualizzare le Short Tandem Repeats (ripetizioni in tandem) sotto forma di tabella che mostra i seguenti dettagli per ciascun locus genotipizzato:

Informazioni sul locus e annotazioni

- **Gene**: nome del gene (HGNC) e link corrispondente.
- **Site**: le coordinate genomiche della ripetizione.
- **Repeat Unit**: il motivo nucleotidico ripetuto.
- **Disease**: la malattia associata secondo il database OMIM e il link corrispondente.

- **Transmission**: la modalità di trasmissione nota.
- **Premutation range**: l'intervallo di premutazione corrispondente al numero minimo e massimo di ripetizioni che determinano lo stato di premutazione.
- **Pathogenic range**: la soglia inferiore di ripetizioni che determina lo status di patogenicità.

Dettagli specifici per ogni campione

- **GT**: il genotipo omozigote, biallelico o eterozigote del locus.
- **Allele repeats**: il numero di ripetizioni per allele. I loci omozigoti riportano un unico valore. Per i loci biallelici ed eterozigoti, viene indicato il numero di ripetizioni per entrambi gli alleli. Lo status di ciascun allele è determinato dal numero di ripetizioni: wild type in verde, premutato in arancione e patogeno in rosso.
- **Supp. reads**: il numero di reads di supporto per ciascun allele. I loci omozigoti riportano un singolo valore, mentre per i loci biallelici ed eterozigoti sono indicati i valori per 2 alleli.

STRs 34 ?										
Gene	Site	Repeat unit	Disease	Transmission	Premutation range	Pathogenic range	GT	Allele repeats	Supp. reads	
PABPN1	chr14:23321472-23321490	GCG	Oculopharyngeal Muscular Dystrophy	AD	10 - 11	> 11	●	16 - 6	0 - 0	
ATXN3	chr14:92071009-92071042	CTG	Machado-Joseph disease	AD	44 - 60	> 60	●	15	0	
ATXN10	chr22:45795354-45795424	ATTCT	Spinocerebellar Ataxia types 10	AD	32 - 800	> 800	●	14 - 12	0 - 0	
NIPA1	chr15:22786677-22786701	CGCGGG	Amyotrophic lateral sclerosis-parkinsonism/dementia complex, susceptibility to	AD	8 - 10000	> 10000	●	6 - 4	0 - 0	
NOP56	chr20:2652733-2652757	GCCTGG	Spinocerebellar Ataxia types 36	AD	14 - 650	> 650	●	8 - 5	0 - 0	
NOP56	chr20:2652757-2652775	GCCTGG	Spinocerebellar Ataxia types 36	AD	14 - 650	> 650	●	7 - 4	0 - 0	
ATN1	chr12:6936716-6936773	CAG	Dentatorubropallidoluysian Atrophy	AD	35 - 48	> 48	●	25 - 19	0 - 0	

Figura 30: Visualizzazione della scheda STRs di un'analisi GermVar Tertiary.

La tabella può essere scaricata come file repeats.tsv tramite l'icona  in alto a destra della tabella. Questo file è disponibile anche nella scheda Files dell'analisi.

13.1.9. Geni di fusione

13.1.9.1. Fusioni rilevate dal DNA

Metodologia

Il modulo [GRIDSS](#) (Genome Rearrangement IDentification Software Suite), specializzato nel rilevamento di riarrangiamenti genomici e implementato nel workset SomaVar e SomaVar LF, viene utilizzato per rilevare fusioni geniche dal sequenziamento del DNA, arricchite da capture mirate a un panel di geni.

Vista e interpretazione

La scheda Fusioni di un'analisi somatica (SomaVar, SomaVar LF, SomaCGP, SomaVar Tertiary) visualizza l'elenco delle fusioni che superano i vari filtri GRIDSS sotto forma di tabella con informazioni dettagliate sulla fusione e su ciascun partner coinvolto:

Informazioni sulla fusione

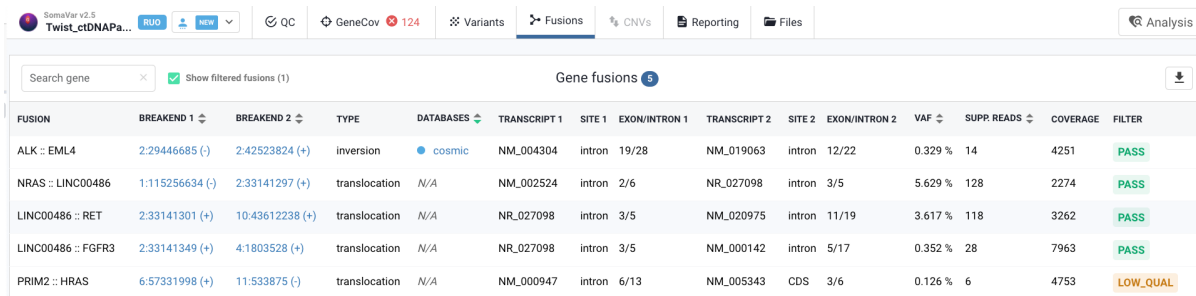
- fusione con i simboli dei geni di ciascun partner,
- il tipo di evento (traslocazione, duplicazione, inversione o delezione),
- l'occorrenza della fusione nel database COSMIC,
- la frequenza allelica (VAF),
- il numero di reads che supportano la fusione,
- copertura totale,
- filtri assegnati da GRIDSS.

Informazioni relative ai partner

- le coordinate genomiche del punto di fusione (breakend),
- il transcripts' RefSeq coinvolto,
- il sito corrispondente al punto di fusione (breakpoint): 5'UTR, 3'UTR, introne, esone (punto di rottura non codificante), CDS (punto di rottura codificante) o sito di splicing,
- la numerazione dell'esone o dell'introne coinvolto nella fusione, rispetto al numero totale di esoni o introni.

Un menu di filtri consente di

- cercare una fusione specifica o un gene coinvolto in una fusione tramite una barra di ricerca,
- visualizzare le fusioni che non hanno superato i filtri GRIDSS (LOW_QUAL) selezionando la casella 'Show filtered fusions'. Questa casella è deselezionata per impostazione predefinita.



FUSION	BREAKEND 1	BREAKEND 2	TYPE	DATABASES	TRANSCRIPT 1	SITE 1	EXON/INTRON 1	TRANSCRIPT 2	SITE 2	EXON/INTRON 2	VAF	SUPP. READS	COVERAGE	FILTER
ALK::EML4	2:29446685 (-)	2:42523824 (+)	inversion	cosmic	NM_004304	intron	19/28	NM_019063	intron	12/22	0.329 %	14	4251	PASS
NRAS::LINC00486	1:115256634 (-)	2:33141297 (+)	translocation	N/A	NM_002524	intron	2/6	NR_027098	intron	3/5	5.629 %	128	2274	PASS
LINC00486::RET	2:33141301 (+)	10:43612238 (+)	translocation	N/A	NR_027098	intron	3/5	NM_020975	intron	11/19	3.617 %	118	3262	PASS
LINC00486::FGFR3	2:33141349 (+)	4:1803528 (+)	translocation	N/A	NR_027098	intron	3/5	NM_000142	intron	5/17	0.352 %	28	7963	PASS
PRIM2::HRAS	6:57331998 (+)	11:533875 (-)	translocation	N/A	NM_000947	intron	6/13	NM_005343	CDS	3/6	0.126 %	6	4753	LOW_QUAL

Figura 31: Visualizzazione della scheda Fusioni per un'analisi SomaVar o SomaVar LF.

I dati della tabella possono essere ordinati in base alle coordinate genomiche (breakend), alla presenza della fusione in COSMIC (Databases), alla frequenza allelica (VAF) e al numero di reads che supportano la fusione (Supporting reads).

Il link associato alle coordinate genomiche di ciascun punto di rottura apre una vista divisa del punto di fusione su IGV. I tooltip mostrano in dettaglio il numero di split read e di coppie non corrispondenti nella colonna Supporting reads e la copertura in ciascun punto di rottura (Cov. 1 e Cov. 2) nella colonna Coverage.

La colonna "Filters" mostra i 3 filtri di qualità utilizzati da GRIDSS e fornisce un'indicazione della fiducia nella chiamata di fusione:

- PASS: la fusione ha superato i filtri di qualità. Il punteggio di qualità è maggiore o uguale a 500 ed è considerato una chiamata di fiducia media o alta (se > 1000).
- LOW_QUAL: la fusione ha un punteggio di qualità inferiore a 500 per almeno uno dei due breakend coinvolti nella fusione ed è considerata una chiamata a bassa confidenza.
- ASSEMBLY_ONLY: la fusione viene chiamata sulla base di un assemblaggio e non sulla base della read che lo supporta. L'assemblaggio si riferisce alla ricostruzione di sequenze intorno a potenziali punti di fusione, che vengono poi riallineate per identificare la fusione.

Se sono disponibili più filtri per la stessa fusione, il segno "Filters" visualizza tutti i filtri nella colonna Filtro.

Oltre ai risultati presentati nell'interfaccia, è possibile scaricare un file fusions.tsv con i dettagli delle fusioni rilevate dalla scheda Files o dall'icona  in alto a destra della pagina. Contiene le seguenti informazioni:

- nome della fusione con i simboli dei geni per ciascun partner (fusione)
- coordinate genomiche di ciascun punto di fusione (breakend_1 e breakend_2)

- trascritto canonico per ogni partner (Transcript_canonical_1 e 2)
- sito corrispondente al punto di fusione per ogni partner (sito_1 e sito_2)
- numerazione degli esoni o degli introni in base al transcript RefSeq associato (exon_1 e 2)
- 60 bp di sequenze fiancheggianti che circondano il punto di fusione (flanking_region_1 e 2)
- filtri di qualità assegnati da GRIDSS (filtro)
- tipo di evento (type)
- numero di split read (split_reads)
- numero di coppie di read discordanti (discordant_pairs)
- copertura per ogni partner (copertura_1 e copertura_2)
- frazione di reads che supportano la fusione (read_fraction)
- numero di reads che supportano la fusione (supporting_reads)
- copertura totale (copertura)
- Identificatore di fusione COSMIC (cosmic_id)

Questo file contiene:

- Fusioni visualizzate nell'interfaccia, che hanno superato i filtri di qualità (filtro = PASS),
- Fusioni rilevate ma con un punteggio di qualità inferiore a 500 per almeno uno dei due breakend coinvolti nella fusione (filtro = LOW_QUAL).

Il punteggio di qualità per ogni breakend si trova nel file gridss.raw.vcf.gz (colonna QUAL), scaricabile dalla scheda Files.

13.1.9.2. Fusioni rilevate dall'RNA

Strumento SomaRNA per l'analisi di dati RNA-seq provenienti da approcci mirati come la cattura e l'amplicon, che consente agli utenti di identificare e visualizzare eventi di fusione genica a livello di trascritto.

Metodologia

Per rilevare e visualizzare le fusioni di SomaRNA viene utilizzata una combinazione di due strumenti:

- [STAR](#): un allineatore di dati RNA-seq progettato per allineare sequenze non contigue direttamente al genoma di riferimento.
- [Gene](#): rileva fusioni di geni da allineamenti chimerici.

Sulla base degli allineamenti chimerici di STAR, Arriba applica una serie di filtri per eliminare gli artefatti noti e i trascritti osservati in contesti non patologici. Assegna un punteggio di confidenza a ogni potenziale evento di fusione che dipende da diversi criteri:

- Il numero di sequenze a supporto della fusione
- L'equilibrio tra split read e coppie discordanti
- La distanza tra i punti di rottura, la loro vicinanza (intragenica o meno)
- Tipo di evento

Le fusioni rilevate con una confidenza media o alta da Arriba sono visualizzate su SeqOne. Quelle con bassa confidenza sono elencate nei file "arriba_discarded" (vedi sotto). La confidenza di Arriba riflette la probabilità di un trascritto aberrante non osservato nei tessuti sani, la possibilità di un riarrangiamento genomico sottostante e l'esclusione di artefatti.

Maggiori dettagli sul livello di confidenza nella documentazione di Arriba: <https://github.com/suhrig/arriba/blob/master/documentation/08-Interpretation-of-results.md>

Per capire perché una fusione è "Low confidence", consultare la colonna 'Filters' (AA) nel file "discarded". Essa indica il filtro che ha scartato l'evento. Dettagli di ogni filtro: <https://github.com/suhrig/arriba/blob/master/documentation/11-Internal-algorithm.md>

Vista e interpretazione

Dalla scheda Fusioni di un'analisi SomaRNA, l'utente ottiene un elenco di previsioni di fusione, sotto forma di diverse schede.

Nella parte superiore della pagina sono riportati i dettagli della fusione e di ciascuno dei partner coinvolti:

Informazioni generali sulla fusione

- il numero di reads della giunzione di spanning, pari alla somma dei rispettivi split read di ciascuno dei geni 5' e 3' coinvolti nella fusione,
- il numero di compagni discordanti,
- il punteggio di fiducia assegnato da Arriba (High, Medium o Low),
- il tipo di evento che ha portato alla fusione (traslocazione, duplicazione, inversione o delezione),
- l'impatto della fusione sulla struttura di lettura,
- se del caso, l'occorrenza della fusione nel database COSMIC.

Informazioni su ciascun partner di fusione

- il simbolo di ciascun gene,
- il Transcript: o, in mancanza, ENSEMBL per ciascun gene,
- la numerazione degli esoni coinvolti nella fusione,
- le coordinate genomiche di ciascun breakpoint,
- la copertura totale nel punto di fusione (Coverage),

- la frazione read che supporta la fusione.

Gli split read sono letture allineate prevalentemente nel gene 1 o nel gene 2 e che si sovrappongono alla fusione. Nel caso di discordant pairs of reads, una delle reads si trova sul gene 1 e l'altra sul gene 2, senza necessariamente sovrapporsi alla fusione.

La frazione di lettura è la somma delle split read e delle discordant read pairs associate a un gene, divisa per la copertura totale. Il valore della frazione di lettura può essere superiore al 100% quando le split read associate a un gene fanno anche parte di discordant pairs. Queste letture vengono quindi contate due volte invece di una.

DHRX - TCEANC
E3 - E4

POLR2J2 - EFCAB13
E3 - E18

POLR2J2 - EFCAB13
E4 - E18

MEGF8 - CNFN
E3 - E4

General informations

of spanning junction reads

568

of discordant mates

246

Confidence

high

Fusion type

translocation

Reading frame

in-frame

Database

cosmic

Partners informations

5prim

3prim

Gene

NPM1

ALK

Transcript

ENST00000517671

NM_004304.4

Exons

5 / 12

20 / 29

Breakpoint

5 : 170818803 (+)

2 : 29446394 (-)

Coverage

14004

8455

Read fraction

3.9%

6.08%

Figura 32: Visualizzazione delle informazioni di fusione da un'analisi SomaRNA.

La parte inferiore della pagina mostra rappresentazioni grafiche della fusione, sia a livello di geni ed esoni, sia a livello di proteine. Quest'ultima sezione è suddivisa in due schede:

- La prima scheda presenta la struttura della fusione e i partner coinvolti, dove la copertura degli esoni coperti dal manifesto di analisi è rappresentata come un diagramma su sfondo giallo:

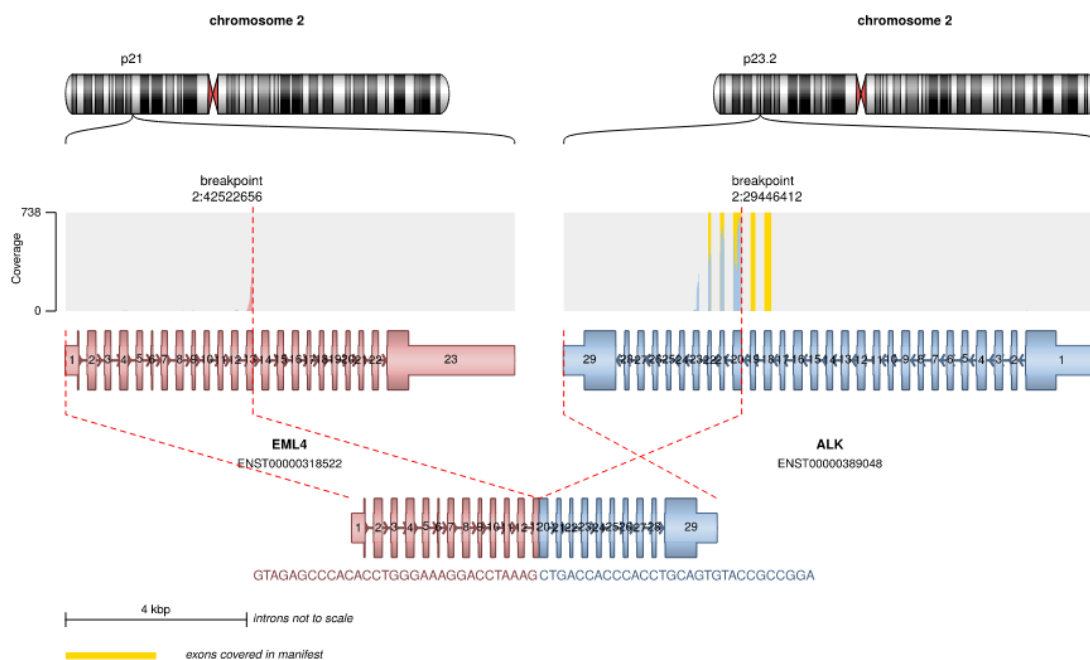


Figura 33: Vista di un evento di fusione tra due geni

- I domini proteici conservati dopo l'evento di fusione sono evidenziati in una seconda vista:

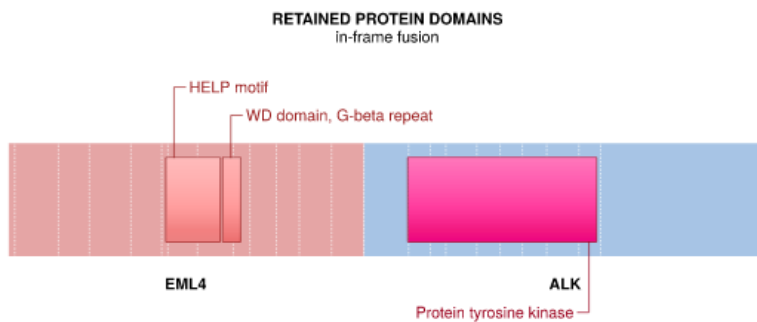


Figura 34: Vista del polipeptide risultante dalla fusione di due geni.

Oltre ai risultati presentati nell'interfaccia, sono disponibili vari file da scaricare:

- Le fusioni conservate da Arriba in formato TSV (file arriba.tsv),
- Le fusioni eliminate dai filtri Arriba in formato TSV (file arriba_discarded_standard.tsv),
- le fusioni eliminate dai filtri Arriba ed elencate in COSMIC in formato TSV (file arriba_discarded_cosmic.tsv),
- le varie figure visualizzate nell'interfaccia in formato SVG

13.1.10. Clonalità

La scheda 'IGHV mutational status report' (Rapporto IGHV mutational status) nella scheda "Clonality" (Clonalità) dell'analisi SomaHemato mostra i risultati dello stato mutazionale.

Viene visualizzata la sezione "IGHV Results":

- IGHV Stato mutazionale e percentuale di omologia:
 - Mutated: se l'omologia è <98% di identità con il germline'.
 - Unmutated: se l'omologia è ≥98% di identità con la germline': se l'omologia è ≥98% di identità con la germline'.
 - Borderline: per cloni nell'intervallo di identità del 97-98%.
 - Polyclonal: nessun clone principale al di sopra del 20%'.
 - Non-Conclusive: se uno qualsiasi dei QC essenziali fallisce
- Il sottoinsieme determinato (con impatto clinico) e confidenza: *2 (High risk), *8, or None
- Una notifica di avviso se nel rapporto compaiono avvisi.

La sezione "Clonality" elenca i cloni identificati con le loro frequenze e lo stato di produttività (Productive, Un-productive). Ogni clone ha una sezione dedicata che mostra il suo stato mutazionale, la frequenza, il link alla pagina IMGT_quest e ulteriori dettagli (omologia, completezza,) insieme ai controlli di qualità.

13.1.11. Varianti di splicing

Metodologia

Il workset SomaRNA utilizza la suite di strumenti *Regtools* per rilevare i salti esonici e identificare varianti di splice alternative dal sequenziamento dell'RNA, arricchito dalla cattura di un panel di geni. RegTools estrae le giunzioni esone-esone dal file BAM generato dall'allineatore di dati RNA-seq STAR e le annota in base a una struttura nota del trascritto, basata su un file GTF e sulle annotazioni di Ensembl.

Visualizzazione e interpretazione

La scheda Splice variants di un'analisi SomaRNA è incentrata sulla visualizzazione delle varianti di splicing dei seguenti geni clinicamente rilevanti e della loro principale variante azionabile:

- MET (Proto-Oncogene MET, Recettore Tirosina Chinasi) e la variante attivabile MET Δ 14
- EGFR (Epidermal Growth Factor Receptor) e la variante di splicing azionabile EGFRvIII
- AR (recettore degli androgeni) e la variante di splicing AR-V7, che può essere utilizzata.

Un elenco completo delle varianti di splicing rilevate nell'analisi SomaRNA, oltre a quelle sopra elencate, può essere scaricato direttamente come file TSV tramite l'icona Download nell'interfaccia o dalla scheda Files nell'analisi.

Per ciascun gene menzionato in precedenza, viene visualizzato lo stato di rilevamento (rilevato o non rilevato) della principale variante perseguibile, insieme al numero di altri eventi alternativi rilevati per quel gene.

Gli utenti possono visualizzare i dettagli delle principali varianti di splicing rilevate aprendo un inserto dedicato, suddiviso in tre parti:

- A sinistra, vengono mostrate le informazioni sulla principale variante di splice perseguibile: coordinate genomiche, nome del gene, trascritto RefSeq, tipo di variante rilevata con dettaglio degli esoni mancanti, fase di lettura, frazione di reads e numero di reads che supportano la variante. Infine, vengono specificate le letture medie di "Canonical Supporting". Questa corrisponde alla media delle reads 5' e 3' che supportano l'inclusione dell'esone.
- a destra, un grafico sashimi (Junction Plot) illustra le giunzioni rilevate tra gli esoni e il numero di reads che supportano ciascuna giunzione.
- in fondo, sono elencate le informazioni sugli eventi alternativi rilevati, ossia le isoforme diverse dalla variante principale di interesse descritta in precedenza: tipo di variante

rilevata (tra parentesi i numeri degli esoni mancanti), fase di lettura, frazione e numero di letture che supportano la variante, numero medio di letture che supportano il "Transcripts" canonico (Avg. canonical support), status PASS o FAILED secondo i filtri descritti di seguito.

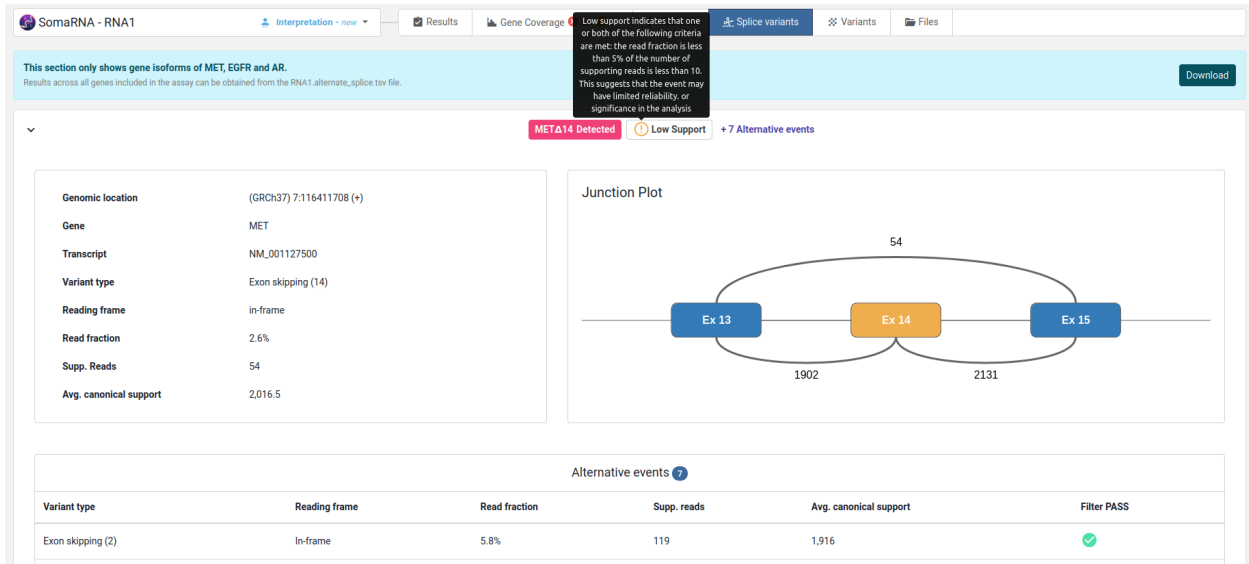


Figura 35: Visualizzazione della scheda Splice variants di un'analisi SomaRNA.

Il rilevamento delle varianti di splicing nei geni visualizzati nella scheda Varianti di splicing non è soggetto all'applicazione di filtri di rilevamento.

Tuttavia, l'icona "Low support" associata al rilevamento di METΔ14 e EGFRvIII indica all'utente che le metriche relative alle reads di supporto sono inferiori alle seguenti soglie: frazione di reads <5% o numero di reads di supporto <10.

Anche lo stato PASS o Failed di altri eventi alternativi utilizza queste soglie:

- **PASS** : frazione di letture $\geq 5\%$ e numero di letture ≥ 10 .
- **FAILED** : frazione di read $< 5\%$ o numero di reads < 10 .

Il file TSV scaricabile riporta le seguenti informazioni per tutti gli eventi di splicing rilevati: RefSeq transcript, gene, direzione del filamento (+ o -), frazione di letture e numero di letture a supporto della variante, numero di esoni mancanti e numeri di esoni corrispondenti, fase di lettura, coordinate genomiche delle giunzioni. Il file specifica anche in FILTER_PASS (yes/no) se la variante supera le soglie applicate alle letture di supporto (frazione di letture $\geq 5\%$ e numero di letture ≥ 10) e in Actionable (yes/no), l'azionabilità della variante.

13.1.12. Stato MSI/MSS

Il workset di analisi SomaMSI è disponibile quando si esegue una nuova analisi sui dati del pannello di geni e serve a determinare lo stato MSI o MSS di un campione di tumore.

Metodologia

Questa analisi di coorte non individuale è simile a un'analisi CNVCapture o SomaCNVCapture. Per eseguire l'analisi sono necessari almeno otto campioni.

Lo strumento genera un campione di riferimento dai campioni della coorte che si stima siano negativi. Poiché questo stato dei campioni non è noto in anticipo, la stima si basa sul calcolo del tasso di delezioni osservate nelle regioni dei microsatelliti: la metà dei campioni della coorte con il tasso più basso viene quindi selezionata per generare un campione di riferimento; ogni campione contribuisce quindi con una frazione delle sue sequenze.

Ogni campione di tumore viene successivamente confrontato con il campione di riferimento utilizzando lo strumento MSIsensor'. Viene calcolato un punteggio che rappresenta la percentuale di regioni microsatellite studiate che presentano instabilità. Le regioni studiate sono quelle microsatelliti coperte dal panel e rilevate da MSIsensor'.

In base a questo punteggio, lo stato MSI o MSS dei diversi campioni è dettato da una soglia che viene calcolata come segue: $\max(10, \text{mediana} + \max(3 \cdot \text{media-abs-dev}, 0, 3 \cdot \text{media}))$

Il principio alla base di questo punteggio è il seguente:

- In un'analisi che presenta campioni MSI e MSS, la soglia sarà generalmente pari a $\text{mediana} + 3 \cdot \text{mediana-abs-dev}$.
- Tuttavia, se tutti i campioni presentano lo stato MSS, se i punteggi sono molto bassi e se la loro distribuzione è estremamente vicina, la soglia sarà impostata su 10 per evitare di generare falsi positivi.

Vista e interpretazione

Una volta completata l'analisi, i risultati di tutti i campioni possono essere visualizzati dall'analisi SomaMSI. La scheda Risultati è suddivisa in due sezioni:

MSI graph: una rappresentazione grafica dei risultati in cui ogni campione è mostrato in base al suo punteggio. La soglia determinata dalla coorte di campioni crea due zone nel grafico che rappresentano i campioni considerati rispettivamente MSI e MSS.

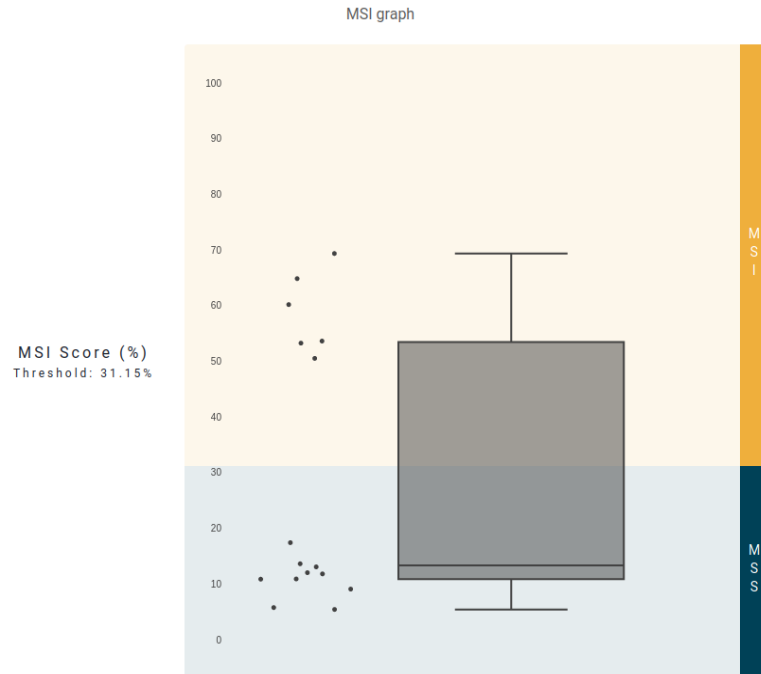


Figura 36: Box plot che rappresenta la distribuzione dei punteggi MSI per la coorte campione.

Per sample details: la ripartizione dei punteggi e lo stato attribuito a ciascun campione sono disponibili nel pannello di destra.



Figura 37: Ripartizione dei punteggi e dello stato di tutti i campioni inclusi nell'analisi SomaMSI.

13.1.13. Stato HRD/HRP

Metodologia

Il rilevamento dello status HRD proposto da SeqOne si basa sull'analisi di dati di sequenziamento del genoma a bassa copertura (sWGS), a cui possono essere aggiunti i risultati dell'analisi di un pannello di geni, se quest'ultimo include i geni BRCA1 e BRCA2.

Il principio è il seguente:

- Quantificare l'instabilità genomica dai dati sWGS e identificare grandi riarrangiamenti (LGA) e perdite di copie parentali (LPC), che rifletterebero una perdita di eterozygosity.
- Identificare le variazioni del numero di copie (CNVs) in geni chiave descritti come potenzialmente in grado di influenzare lo status di HRD.
- Rilevare facoltativamente le varianti patogene nei geni BRCA1, BRCA2 e TP53, se i risultati del panel sono disponibili sull'interfaccia SeqOne per il paziente corrispondente. In questo caso, i tipi di varianti rilevate dal workset SomaVar che vengono presi in considerazione sono SNV, indel e SV (rilevate da 'Freebayes', 'Mutect2' e 'GRIDSS').

Status HRD. Lo status positivo HRD è definito da almeno uno dei seguenti elementi (**solo per le versioni SomaHRD v1.2 e v1.3**): (1) almeno una mutazione BRCA1/2 patogena e/o (2) una HRD probability. maggiore o uguale a 0,5.

Dalla versione v1.4 del workset SomaHRD, lo stato dell'analisi è definito esclusivamente dal punteggio di instabilità genomica (GI score). Gli utenti possono selezionare uno status HRD (Positivo, Negativo o Non-Conclusive) accanto allo status GI per includerlo nel report al posto dello status GI.

HRD probability. (solo versioni SomaHRD v1.2 e v1.3) La Probabilità HRD di SeqOne corrisponde alla probabilità che il tumore sia positivo all'HRD sulla base delle caratteristiche di instabilità genomica LGA e LPC determinate dai dati sWGS. Le grandi alterazioni genomiche (Large Genomic Alterations, LGA) sono definite come il numero di copie dei breakpoint, dove un breakpoint corrisponde a un cambiamento nel numero di copie tra due segmenti genomici lunghi almeno 10 Mb e distanti non più di 3 Mb. Le perdite di copie parentali (LPC) sono definite come il numero di segmenti aploidi lunghi almeno 10 Mb. La probabilità di HRD è una combinazione lineare delle due caratteristiche di instabilità genomica, convertita utilizzando la regressione logistica per collocare i valori su una scala di probabilità compresa tra 0 e 1.

GI probability (GiP) (da SomaHRD v1.4) La probabilità di instabilità genomica (GI) rappresenta la probabilità che il tumore presenti instabilità genomica, sulla base dello Shallow Whole Genome Sequencing (sWGS). Il calcolo di questa probabilità dipende dai due tipi di caratteristiche di instabilità genomica LGA e LPC. LGA è definita come il numero di breakpoint del numero di

copie. Un punto di rottura corrisponde a una variazione del numero di copie tra due segmenti genomici sufficientemente grandi e vicini. Il punteggio LGA non è una singola misura, ma piuttosto un insieme di diverse variazioni. Queste variazioni sono calcolate utilizzando diversi parametri per la dimensione minima del frammento e la distanza tra i frammenti. Il punteggio LPC, invece, si riferisce al numero di segmenti aploidi lunghi almeno 10 Mb. La probabilità di instabilità genomica di SeqOne non è una funzione diretta dei punteggi medi LGA e LPC. Viene calcolata utilizzando un modello di boosting che integra diverse caratteristiche di LGA e LPC, migliorando così l'accuratezza della previsione. La probabilità finale è la media di 100 probabilità (comprese tra 0 e 1) calcolate applicando diversi parametri del modello.

Status GI. (da SomaHRD v1.4) Uno stato di instabilità genomica positivo (GI status) è definito da una probabilità di instabilità genomica superiore al 50%.

BRCA Status. Per l'individuazione di varianti patogene, che determinano lo status BRCA indipendentemente dalla HRD probability o dalla GI probability, vengono applicati i seguenti criteri alle varianti BRCA1 e BRCA2 dell'analisi SomaVar del paziente:

- Frequenza allele $\geq 15\%$ (**versioni SomaHRD v1.2 e v1.3**),
- Frequenza allelica $\geq 10\%$ (**da SomaHRD v1.4**)
- Copertura ≥ 35
- Qualità media di ogni base portatrice della variante (punteggio Phred) ≥ 20
- Osservazione alternativa (numero di sequenze portatrici della mutazione) ≥ 5
- Classificazione 4 (*Likely Pathogenic*) o 5 (*Pathogenic*) secondo il nostro strumento di classificazione della patogenicità delle varianti basato sui criteri ACMG.

TP53 variants. I seguenti criteri sono applicati alle varianti TP53 patogene dall'analisi SomaVar condotta sul paziente:

- Copertura ≥ 100
- Qualità media di ogni base portatrice della variante (punteggio Phred) ≥ 20
- Osservazione alternativa (numero di sequenze portatrici della mutazione) ≥ 5
- Classificazione 4 (*Likely Pathogenic*) o 5 (*Pathogenic*) secondo il nostro strumento di classificazione della patogenicità delle varianti basato sui criteri ACMG.

Queste varianti vengono visualizzate solo senza una soglia minima di frequenza allelica nel report dell'analisi SomaHRD, ma non sono determinanti per lo status HRD. Solo la variante 'TP53 Pathogenic' con la frequenza allelica più alta sarà riportata nel report HRD.

Cellularity. I metadati sulla cellularità (% di cellule tumorali nel campione) sono necessari per eseguire l'analisi SomaHRD. Può essere inserito al momento del caricamento del campione o prima dell'avvio dell'analisi, modificando i metadati del campione corrispondente (vedere **Metadati del campione**). Se il valore è inferiore al 20%, il risultato sarà 'NON CONCLUSIVE'. Tuttavia, lo status HRD può essere positivo se l'analisi è accompagnata da un pannello comprendente BRCA1 e BRCA2, che fornisce uno status BRCA positivo.

Procedura di avviso della frazione tumorale: L'analisi SomaHRD include un meccanismo di avviso per indicare quando un campione potrebbe avere un basso contenuto tumorale. Questo meccanismo si basa su una misura statistica chiamata "Copy-ratio", che valuta la somiglianza tra segmenti genomici consecutivi in termini di rapporto di numero di copie. I campioni con una sovrapposizione significativa hanno spesso frazioni tumorali inferiori, rendendo più difficile distinguere l'instabilità genomica.

Per stabilire una soglia affidabile per questo avviso, abbiamo correlato la statistica di sovrapposizione con la Variant Allele Frequency (VAF) delle mutazioni TP53 utilizzando dati clinici. Se la statistica di sovrapposizione supera una soglia predefinita, viene emesso un avviso. Questo avviso evidenzia potenziali discrepanze con la percentuale di contenuto tumorale fornita manualmente, che può indicare che la frazione tumorale è inferiore al 20%. I campioni con frazioni tumorali inferiori a questo livello presentano un rischio elevato di produrre falsi negativi. Tuttavia, tali campioni non vengono scartati e l'avviso serve come guida per un'attenta interpretazione dei risultati HRD.

Minimum coverage. La copertura minima richiesta per eseguire il test è 0,1X, ma si consiglia una copertura minima di 0,5X, idealmente 1X. Quando la copertura si riduce a 0,1X, il tasso di errore di classificazione aumenta del 5%.

Anche se i dati sWGS possono essere stati generati da un kit che presenta UMI, questi non vengono presi in considerazione durante l'analisi SomaHRD. Nel workset è inclusa una fase di trimming sia per gli adattatori che per gli UMI.

Visualizzazione e interpretazione

Una volta completata l'analisi, i risultati sono disponibili nell'analisi SomaHRD. Oltre alle metriche QC e al report tecnico, un report HRD è direttamente accessibile dalla scheda HRD Report (**versioni SomaHRD v1.2 e v1.3**).

Dalla **versione v1.4 del workset SomaHRD**, i risultati dell'analisi sono visualizzati in una scheda HRD dedicata. Ogni sezione della scheda riflette una sezione del report HRD descritto di seguito (HRD Summary, Sample quality, Genomic instability features, BRCA1/BRCA2 & TP53 mutations, Copy number variations).

Da questa scheda, l'utente può esportare i risultati come file CSV o come rapporto HRD in formato PDF. Nel caso di un'esportazione in formato PDF, l'utente può selezionare la lingua del rapporto HRD da un elenco di lingue disponibili: Francese, Spagnolo, Inglese, Italiano, Tedesco. La selezione predefinita è l'inglese.

Il rapporto HRD è suddiviso nelle seguenti sezioni.

Un'intestazione elenca i metadati relativi al paziente e all'analisi, il tipo di test eseguito e lo status HRD (**SomaHRD v1.2 e v1.3**) o lo status GI (**SomaHRD $\geq$ v1.4**).

Sample quality. Questa sezione riassume le principali metriche di qualità dell'analisi e le soglie minime richieste.

I controlli QC sono visualizzati sotto forma di tabella. Per ogni parametro sono disponibili le seguenti informazioni:

- **Name** : il nome del parametro
- **Value** : il valore del parametro, compreso :
 - Un'icona che indica se il filtro è *Passed*, *Warning* o *Failed*.
 - Il valore del parametro per il campione
- **Description** : Breve descrizione del parametro.

SAMPLE QUALITY

NAME	VALUE	DESCRIPTION
% of tumoral cells	✓ 90%	Low value of tumoral cells is likely to generate false negative results
Coverage	⚠ 0.43X	Average sequencing coverage of the genome
% correct mapping	✓ 96.5%	Proportion of reads correctly mapped
Confidence in genomic instability	✓ 1.00	Large values indicates high confidence in genomic instability computation
Low tumor fraction	✓ NORMAL	Low tumor fraction detected from the shallow WGS

SAMPLE QUALITY

NAME	VALUE	DESCRIPTION
% of tumoral cells	⚠ 30%	A percentage of tumoral cells below 20% is likely to generate false negative results.
Coverage	✓ 0.7X	The average sequencing coverage of the genome should be above 0.1x.
GI Dispersion	✓ 0.93	A large dispersion reflects high variability in the genomic instability computation, indicating low precision in the estimated score.
Low tumor profile	⚠ WARNING	Low cellularity detected from the genomic profile.

Figura 38 : Sintesi delle metriche QC relative al campione sWGS in SomaHRD v1.2 (in alto) e SomaHRD v1.5 (in basso).

I controlli di qualità vengono applicati ai seguenti parametri, in base alle soglie presentate alla fine del rapporto HRD (sezione '**SAMPLE QUALITY CHECKS**):

- la percentuale di cellule tumorali, basandosi esclusivamente sul valore fornito dall'utente.

- copertura media del sequenziamento del genoma,
- proporzione di reads allineate correttamente, calcolata dopo la fase di trimming dell'adattatore (**solo versioni SomaHRD v1.2 e v1.3**),
- la metrica della fiducia nell'instabilità genomica (LGA e LPC) basata sui dati di copertura (**solo per la versione SomaHRD v1.2**). Questo controllo di qualità viene utilizzato per misurare la fiducia dello status HRD restituito. Alcuni risultati positivi e negativi sono contrassegnati da un'etichetta "low confidence" per identificare i campioni con la più bassa fiducia nello stato.
- un intervallo di confidenza della probabilità HRD (HRD probability confidence interval) (**solo per la versione SomaHRD v1.3**). Questa metrica avrà uno stato di Warning quando l'errore standard (deviazione standard) è maggiore di 0,07; avrà uno stato di Failed se l'intervallo di confidenza della probabilità HRD include il valore 0,5.
- un intervallo di confidenza della probabilità GI (GI probability confidence interval) (**solo per la versione SomaHRD v1.4**). Questa metrica avrà uno stato di Warning quando l'errore standard (deviazione standard) è maggiore di 0,07; avrà uno stato di *Failed* se l'intervallo di confidenza della 'GI probability confidence' include il valore 0,5.
- una dispersione GI (**solo per la versione SomaHRD v1.5**) che mostra la dispersione dei 100 calcoli di probabilità GI. Questa metrica visualizzerà uno stato "Warning" con un tag "Moderate" accanto al valore quando la dispersione è > 2,6 e uno stato "Failed" con un tag "High" se > 2,7. In caso contrario, accanto al punteggio apparirà un tag "Low".
- in caso di sospetto di basso contenuto tumorale (*Low tumoral fraction*) in base alla sWGS, viene visualizzato il messaggio *WARNING*. In caso contrario, il controllo visualizza il messaggio *'NORMAL'*.

Se una qualsiasi delle seguenti metriche QC è inferiore alle soglie di guasto previste, il risultato dell'analisi sarà inconcludente (NON CONCLUSIVE):

- % cellule tumorali
- copertura
- % di reads allineate correttamente (**solo per le versioni SomaHRD v1.2 e v1.3**)
- Dispersione GI > 2,7 **e se** $0,37 < GiP < 0,59$ (**solo per SomaHRD v1.5**)

In questo caso, le metriche relative alla fiducia nell'instabilità genomica e alla "Low tumoral fraction" non vengono calcolate e vengono visualizzate come 'N/A' nelle rispettive sezioni (**per SomaHRD version < v1.5**).

Per SomaHRD v1.3, se la metrica della fiducia nell'instabilità genomica è "Failed", cioè inferiore alla soglia di fallimento prevista, il risultato dell'analisi sarà:

- POSITIVE LOW CONFIDENCE se la probabilità di HRD è maggiore o uguale al 50%.
- NEGATIVE LOW CONFIDENCE se la probabilità di HRD è inferiore al 50%.

In questo caso, il risultato dell'analisi è accompagnato da un'avvertenza.

Tuttavia, se l'analisi è accompagnata da un pannello che include BRCA1 e BRCA2 e il pannello restituisce uno status BRCA positivo, lo status HRD sarà positivo, anche se una metrica QC relativa al campione sWGS è inferiore alla soglia di fallimento (**solo per le versioni SomaHRD v1.2 e v1.3**).

Variazioni del numero di copie. Questa sezione è dedicata alle variazioni del numero di copie dei geni CCNE1 e RAD51 in termini di numero di copie rilevate. Queste non vengono prese in considerazione nel calcolo della probabilità HRD e della probabilità GI.

Caratteristiche di instabilità genomica. Questa sezione fornisce dettagli sulle caratteristiche calcolate per determinare l'instabilità genomica del tumore. Queste caratteristiche sono utilizzate per determinare la probabilità HRD di SeqOne (**SomaHRD v1.2 e v1.3**) e la probabilità GI (**SomaHRD versione $\geq$ v1.4**). Questa sezione ha lo scopo di far capire come viene calcolata la probabilità. Per ciascuna caratteristica, vengono fornite le seguenti informazioni:

- **Name** : Nome della caratteristica (es. LGA Status o LGA Score)
- **Value**: il valore calcolato per il paziente.
- **Description** : Breve descrizione delle caratteristiche

In caso di risultati "NON CONCLUSIVE", relativi a metriche QC inferiori alle soglie previste, le sezioni 'AMPLIFICATIONS' e 'GENOMIC INSTABILITY FEATURES' del report appaiono in grigio.

GENOMIC INSTABILITY FEATURES

The following variables are used to determine the HRD Probability.

NAME	VALUE	DESCRIPTION
LGA Status	23	Large Genomic Alteration Status
LPC Status	15	Loss of Parental Copy Status

Figura 39: Caratteristiche della probabilità di HRD legate all'instabilità genomica.

BRCA1/BRCA2 & TP53 mutations. Nel caso di un test HRD completo che include l'analisi SomaVar, forniamo una sezione aggiuntiva dedicata alle varianti patogene rilevate nei geni BRCA1, BRCA2 e TP53. Sono elencate tutte le varianti patogene di BRCA1 e BRCA2 corrispondenti ai criteri descritti nella sezione **Methodologia**. Tuttavia, verrà riportata solo la variante TP53 patogena o probabilmente patogena con la più alta frequenza allelica. Le altre varianti patogene di TP53 possono essere visualizzate nell'analisi SomaVar associata. Questa sezione è nascosta solo per i rapporti sWGS.

Le mutazioni rilevate sono riportate in una tabella con le seguenti colonne:

- **Gene & Transcripts**: Gene & Transcript: nome del gene HGNC e ID del trascritto RefSeq.
- **Variant**: nomenclature HGVSC e HGVSP. della variante

- **Evaluation:** Valutazione della variante secondo i criteri ACMG (*Pathogenic* o *Likely Pathogenic*).
- **VAF:** frazione allelica della variante.

Se nell'analisi del pannello collegato non sono state identificate mutazioni BRCA patogene, in questa sezione del referto viene visualizzato il seguente messaggio: "No pathogenic mutations have been identified on BRCA1 and BRCA2 genes."

Se nell'analisi del linked panel non sono state identificate mutazioni TP53 patogene, in questa sezione del referto viene visualizzato il seguente messaggio: "No pathogenic mutations have been identified on TP53 gene."

BRCA1/BRCA2 & TP53 MUTATIONS

GENE & TRANSCRIPT	VARIANT	EVALUATION	VAF
BRCA1 NM_007294.4	c.3116_3117del p.Ala1039GlufsTer6	Pathogenic	60.68%
BRCA1 NM_007294.4	c.3113_3117delinsG p.Glu1038GlyfsTer9	Pathogenic	58.18%
BRCA1 NM_007294.4	c.3113_3114del p.Glu1038GlyfsTer7	Pathogenic	60.08%
TP53 NM_000546.6	c.524G>A p.Arg175His	Pathogenic	39.29%

Figura 40: Varianti patogene BRCA1/BRCA2 e TP53.

HRD Summary. L'ultima sezione del rapporto è dedicata ai risultati. Questa sezione mira a fornire una visione dettagliata dei risultati, mostrando le seguenti informazioni:

- **HRD Probability (SomaHRD v1.2 e v1.3):** Il valore della probabilità HRD, compreso tra 0 e 1. Lo stato è positivo quando il valore è maggiore o uguale a 0,5.
- **GI probability (versione SomaHRD ≥ v1.4):** Valore della probabilità GI, compreso tra 0 e 1. Lo stato è positivo quando questo valore è maggiore o uguale a 0,5.
- **BRCA Status:** stato determinato dal rilevamento di almeno una mutazione BRCA patogena: MUTATA, WILD-TYPE o "N/A" in assenza di analisi del Panel.
- **HRD Status o GI Status:** "POSITIVE", "NEGATIVE", "NEGATIVE LOW CONFIDENCE", "POSITIVE LOW CONFIDENCE" o "NON CONCLUSIVE", come indicato nell'intestazione del rapporto.
- **Method :** Il metodo (o modalità) utilizzato per produrre il punteggio HRD:
 - HRD completo ⇒ "WGS SHALLOW + BRCA1/2 GENE PANEL"
 - Solo per la profondità ⇒ "SHALLOW WGS ONLY"

In caso di risultati "NEGATIVE LOW CONFIDENCE", "POSITIVE LOW CONFIDENCE" o "NON CONCLUSIVE", questi vengono visualizzati in grigio chiaro.

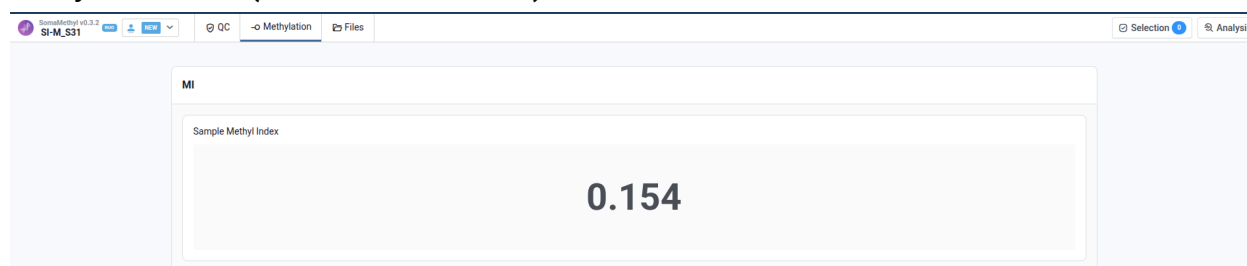
I dettagli sui diversi valori visualizzati nel report e sulla metodologia complessiva del test sono presenti nelle sezioni "**Footnotes**" e "**Methodology**", situate dopo il risultato.

Oltre al report HRD, i grafici dell'instabilità genomica associati all'analisi HRD sono disponibili per il download dalla scheda Files dell'analisi SomaHRD. Questi grafici rappresentano la copertura normalizzata su tutti i cromosomi dei dati sWGS.

In caso di risultati "**NON CONCLUSIVE**", relativi a metriche QC inferiori alle soglie previste, i grafici dell'instabilità genomica non vengono generati.

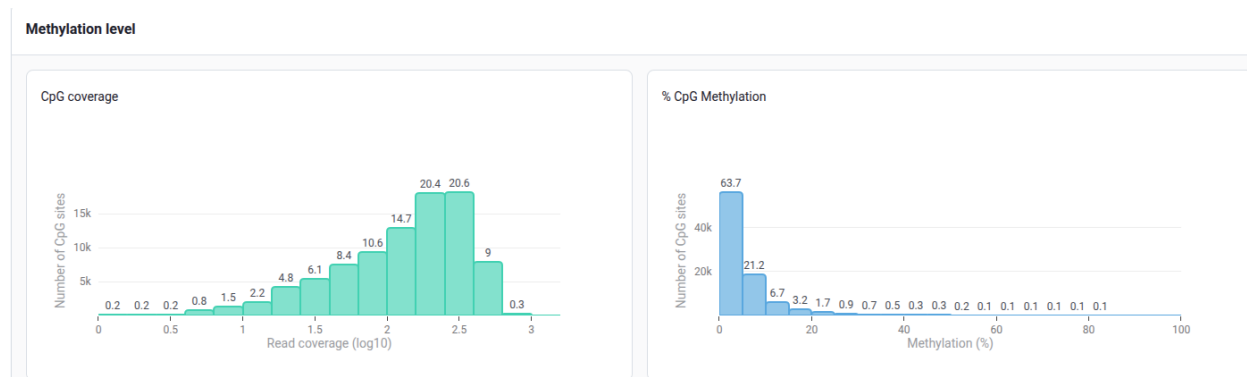
13.1.14. Dati di metilazione per SomaMethyl

Methylation Index (Indice di metilazione)



L'indice di metilazione (MI) è un singolo punteggio numerico che quantifica lo stato di metilazione complessivo di un campione biologico e riassume il segnale di metilazione associato al tumore. Può rilevare in modo sensibile anche bassi livelli di DNA derivato dal tumore in un background di DNA normale. In SomaMethyl, l'MI viene calcolato in modalità Standard utilizzando l'algoritmo proprietario di Agilent (MethylAlgo). Quando è disponibile una coorte di riferimento, l'MI del campione può essere confrontato con i valori MI della coorte e interpretato utilizzando una soglia definita dall'utente. Per valutare l'MI, Agilent ha fornito campioni di titolazione con livelli di metilazione attesi. I valori MI calcolati dal modulo SomaMethyl in modalità Standard vengono confrontati con i valori MI attesi forniti da Agilent e le prestazioni sono riassunte utilizzando una metrica di concordanza ponderata definita come 1 - wMAPE (Errore percentuale assoluto medio ponderato).

Methylation level (Livello di metilazione)



CpG coverage (histogram) (Copertura CpG (istogramma))

Una distribuzione della profondità di lettura attraverso i siti CpG all'interno delle regioni bersaglio (manifest).

Asse X (copertura di lettura log10): copertura per sito CpG (numero di letture che supportano la chiamata di metilazione in quel CpG, ovvero conteggi metilati + non metilati).

Asse Y (# siti CpG): numero di siti CpG nel manifest che hanno quella copertura.

È possibile verificare se i CpG bersaglio sono sufficientemente coperti e se la copertura è abbastanza uniforme da interpretare le percentuali di metilazione con sicurezza.

% CpG methylation (histogram) (% di metilazione CpG (istogramma))

Una distribuzione dei livelli di metilazione attraverso i siti CpG all'interno delle regioni bersaglio (manifest).

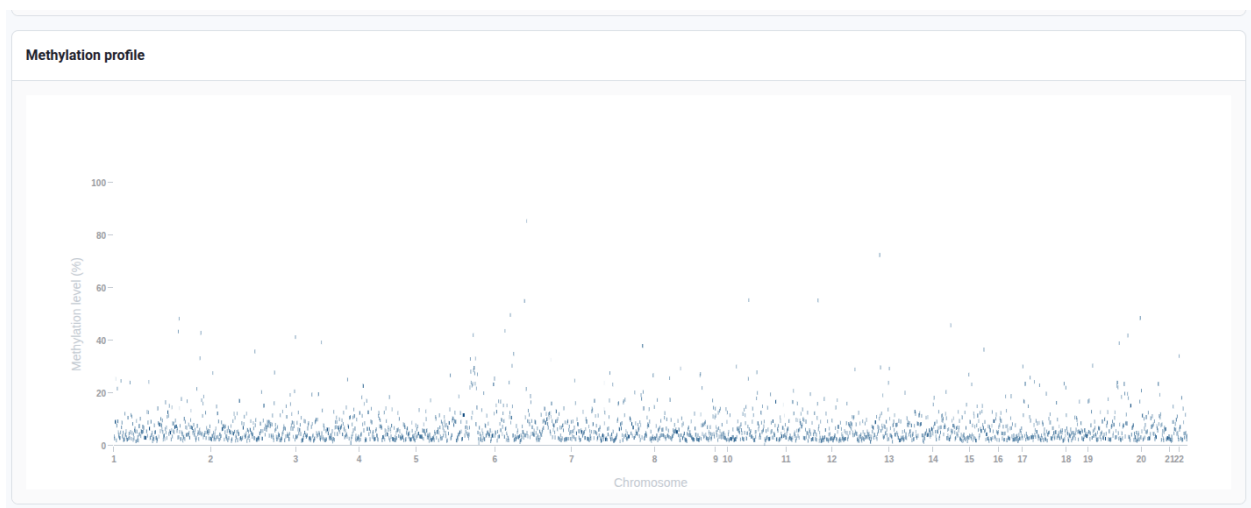
Asse X (% di metilazione): per ogni sito CpG, la frazione di letture metilate tra tutte le letture che coprono quel CpG.

Asse Y (# siti CpG): numero di siti CpG nel manifest che mostrano quel livello di metilazione.

Fornisce una vista a livello di campione degli stati di metilazione CpG attraverso le regioni bersaglio (spesso mostrando picchi vicino a metilazione bassa e alta, a seconda delle regioni catturate e della biologia del campione).

I siti CpG con una copertura molto bassa producono percentuali di metilazione meno affidabili perché un piccolo numero di letture può far oscillare la % verso l'alto o verso il basso. Per questo motivo, l'istogramma della % di metilazione CpG dovrebbe sempre essere interpretato insieme all'istogramma della copertura CpG (e alle altre metriche di controllo qualità).

Methylation profile (Profilo di metilazione)



Il grafico del profilo di metilazione mostra il segnale di metilazione CpG del campione in tutto il genoma, visualizzato per cromosoma. Ogni punto corrisponde alla percentuale media di metilazione CpG calcolata su finestre genomiche successive (come nell'output .mcpG.track.bedGraph). Questa vista aiuta a visualizzare ampi modelli di metilazione e a identificare le regioni con metilazione maggiore o minore rispetto al resto del genoma.

Methylation Table (Tabella di metilazione)

Table					
			Search genes	Locus	FT
REGION	METH. LEVEL	CpG NBR.	CpG REGIONS	GENOMIC FEATURE	REFSEQ CDS
C1orf94 chr1:34176744-34177348	11	53	Island +1	silencer	
chr1:35577191-35577675	3	42			
chr1:37033668-37034149	5	51			
chr1:37034542-37035023	3	61			
EPHA10 chr1:37735070-37735507	24	26		enhancer	NM_001099439.2
chr1:38559361-38559798	5	22			
chr1:39080823-39081222	12	28	Island +1		
chr1:39672120-39672554	3	28			
HPCAL4 chr1:39683853-39684273	11	47	Island		
chr1:40884160-40884512	5	43			
31 to 40 of 2217					
1 2 3 4 5 6					

La tabella di metilazione elenca le regioni genomiche (cromosoma:inizio-fine) su cui SomaMethyl riassume la metilazione CpG (finestre fisse o segmenti uniti, a seconda del passaggio). Per ogni regione, METH. LEVEL riporta il livello di metilazione aggregato calcolato dalle chiamate di metilazione CpG in quell'intervallo e CpG NBR. indica quanti siti CpG hanno contribuito al calcolo. Ogni riga viene quindi annotata con CpG REGIONS (isola, riva, piattaforma quando applicabile) e annotazioni funzionali, inclusi gli elementi regolatori GENOMIC FEATURE sovrapposti (promotore, enhancer, silenziatore) e le informazioni REFSEQ CDS (sequenza codificante).

14. Rapporto

14.1.1. File di uscita

La scheda *Files* elenca i file generati nelle diverse fasi del workset. Ogni file può essere scaricato facendo clic sull'icona  .

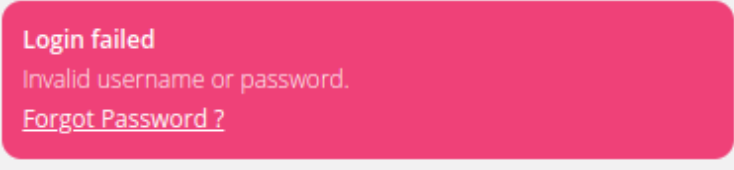
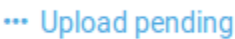
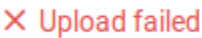
Alcuni file non includono il nome del campione come prefisso. Per scaricarli con questo prefisso, è necessario attivare l'opzione "Prefix with sample name" in alto a destra della pagina.

I file possono essere scaricati dalle diverse sezioni dell'interfaccia:

- I file FastQ sono disponibili nella scheda "*Samples*" del progetto, selezionando il campione pertinente e quindi l'icona  .
- I "Manifest" (file BED) sono disponibili nelle *Impostazioni* dell'utente tramite l'icona nell'angolo in alto a destra della dashboard, quindi nella scheda *Manifest*.
- I rapporti di analisi in formato PDF sono disponibili nella scheda *Reporting* dopo aver selezionato un'analisi.
- I file generati durante un'analisi (BAM, VCF, ecc.) e i diversi report HTML (qualità, tecnico, ecc.) sono disponibili nella scheda *Files* per ogni analisi.

15. Risoluzione dei problemi e messaggi di errore

Vengono utilizzati diversi messaggi per notificare agli utenti un errore

Messaggio di errore	Descrizione
 A pink rectangular box with rounded corners. It contains the text "Login failed" in bold, followed by "Invalid username or password." and a link "Forgot Password ?" in blue.	La password e/o il login non sono corretti. Gli utenti sono invitati a inserire nuovamente il login e la password, oppure a cliccare su "Forgot password?" per reimpostare la password.
 A blue text message that says "... Upload pending".	L'upload del campione è in sospeso. Si consiglia agli utenti di attendere, di non chiudere la finestra del browser e di non interrompere la connessione a Internet.
 A red text message that says "X Upload failed".	Il caricamento del campione non è riuscito. <i>Questo errore può verificarsi se la finestra del browser è stata chiusa o se la connessione Internet ha subito un'interruzione durante il caricamento.</i> Gli utenti sono invitati a riprovare a caricare i file.
 A dark blue horizontal bar. On the left, there is a green progress indicator showing 62%. To the right of the bar are three white icons: a magnifying glass, a document, and a person.	Indica l'avanzamento del caricamento corrente. Si consiglia agli utenti di attendere, di non chiudere la finestra del browser e di non interrompere la connessione a Internet.

	Caricamento della pagina. Gli utenti sono invitati ad attendere.
... Bioinformatics - <i>pending</i>	L'analisi è in corso. Gli utenti sono invitati ad attendere.
× Failed - <i>error</i>	L'analisi è fallita. Agli utenti viene richiesto di controllare le impostazioni e i dati utilizzati per eseguire l'analisi e di riavviare l'analisi.
⚠ Something went wrong Something went wrong, you should refresh this page before continue. Feel free to contact the support if the problem persist.	Si è verificato un errore. All'utente viene richiesto di fare clic sul tasto "Esc". Se il problema persiste, contattare l'assistenza all'indirizzo support@seqone.com.
× Missing fastq files	A questo campione mancano i file FastQ associati, il che significa che non è possibile avviare nuove analisi per quel campione.
unexpected EOF × Verification failed	Il messaggio "Unexpected EOF" (End Of File) corrisponde a un file troncato. L'utente è invitato a controllare i dati e a ricaricare i file interessati.

Gli utenti possono contattare il team di supporto facendo clic sul logo  in alto a destra della dashboard dell'utente, quindi sull'icona  **Support**

16. Garanzia

Consultare i termini e le condizioni della garanzia nel contratto di manutenzione.

17. Avviso legale

17.1. Condizioni d'uso

Le informazioni sulla licenza d'uso sono accessibili facendo clic sul logo dell'account utente seguito da *"Terms of use"*. Gli utenti possono visualizzare e scaricare le informazioni in formato PDF dalla finestra pop-up facendo clic su *"OPEN PDF"*.

17.2. Limitazione di responsabilità

SeqOne declina ogni responsabilità per qualsiasi uso improprio o non professionale della piattaforma SeqOne.

SeqOne è uno strumento di supporto alle decisioni cliniche e non è responsabile delle diagnosi derivate dai risultati ottenuti.

Di conseguenza, gli utenti devono usare la dovuta cautela nell'utilizzo del software e dei suoi risultati. Le diagnosi devono essere confermate da vari esami clinici e medici.

17.3. Diritti d'autore e marchi di fabbrica

Tutti i nomi di aziende e prodotti citati in questa guida possono essere marchi di fabbrica dei rispettivi proprietari.