



SeqOne platform

Instrucciones de uso

Válido a partir de SeqOne platform v2.1

1. Introducción	7
2. Contactar con el fabricante y el representante suizo	8
3. Vigilancia de los productos sanitarios	8
4. Uso previsto de la plataforma	8
4.1. Uso previsto	8
4.2. Indicación de uso	9
4.3. Descripción	9
4.4. Población de pacientes	10
4.5. Usuarios previstos	10
4.6. Características	10
4.6.1. Rendimiento analítico	12
4.6.1.1. Módulos bioinformáticos	13
4.6.1.2. Módulos de anotación	21
4.6.2. Rendimiento clínico	23
5. Precauciones	24
5.1. Uso conforme	24
5.2. Advertencias	24
5.3. Limitaciones	26
5.4. Averías	26
6. Requisitos mínimos	27
6.1. Hardware	27
6.2. Red	27
7. Autenticación	28
7.1. Conexión	28
7.2. Desconexión	29
7.3. Página principal	29
7.4. Gestión de cuentas de usuarios	30
7.4.1. Gestión de "My Account"	31
7.4.2. Gestión de la configuración del usuario - definiciones	31
Entity	31
Users	31
Worksets	31
Manifest	32
Panels - Transcript	32
Panels - Genes	32
Marker	32
Protocols	32
Reporting - Global settings	33
Reporting - Template list	33

Tags	33
API Keys	33
7.4.3. Gestión administrativa - presentación de la interfaz	34
7.4.4. Cambiar entre los modos Main y Preview	37
8. Descripción de las etapas operativas	38
9. Creación de proyectos	39
9.1. Crear un proyecto	39
9.2. Gestión de proyectos	40
9.3. Suprimir un proyecto	40
9.4. Datos de ADN (DNA Sequencing, ctDNA Sequencing)	40
9.5. Datos de ARN (RNA Sequencing)	42
9.6. Configuración del proyecto	43
9.6.1. Configurar las métricas QC del proyecto por defecto	43
9.6.2. Establecer el perfil de filtro predeterminado del proyecto	43
10. Carga de muestras	44
10.1. Cargar muestras	44
Desde un PC	44
A través de SDS (SeqOne Data Sync)	44
Estado de la carga de muestras	44
Capacidad de carga	45
10.2. Formatos de archivo de entrada	45
10.2.1. Archivos de análisis primarios	45
10.2.2. Archivos de análisis secundarios	46
10.3. Metadatos de muestra	49
10.4. Editar una muestra	50
10.5. Borrar una muestra	50
11. Análisis	51
11.1. Iniciar un nuevo análisis	51
11.1.1. Elección del workset de análisis	52
Workset de análisis disponibles a partir de archivos FASTQ	52
GermlineVar	53
GermVar	53
GermlineVarLR	53
GermlineFamily	54
GermVar Family	54
CNVCapture	55
CNVExome	55
SomaVar	56
SomaCNVCapture	56
SomaRNA	57

SomaMSI	57
SomaHRD	57
SomaVar LF	58
SomaCGP	58
SomaHemato	59
SomaLBx	59
SomaMethyl	59
Workset de análisis disponibles a partir de archivos secundarios	61
GermVar Tertiary	61
GermVar Tertiary Family	61
SomaVar Tertiary	62
11.1.2. Elegir el modo de tratamiento UMI	63
Presentación de los modos de tratamiento UMI	63
Proceso de tratamiento UMI	64
Compatibilidad con los workset de análisis	64
11.2. Estado del análisis	65
11.3. Calidad de las muestras	65
12. Comprobación de la calidad de los datos	67
12.1. Métricas e informes de control de calidad	67
12.1.1. Métricas QC y controles QC	67
12.1.2. Informe QC	67
12.1.3. Informe 'ID-Check'	71
12.1.4. Informe técnico	72
Puntuación UPD	73
Gráfico de frecuencia del alelo B (BAF) del informe UPD	74
13. Interpretación de los resultados de los análisis	75
13.1. Resultados de los análisis	75
13.1.1. Gestión del flujo de trabajo de interpretación de análisis	75
13.1.1.1. Asignación de análisis para interpretación	75
13.1.1.2. Bloqueo de análisis	76
13.1.1.3. Panel lateral para la gestión de análisis	77
Vista modal Workflow	78
Vista modal Patient	79
Vista modal HPOs	80
Vista modal Panel	81
13.1.1.4. Panel lateral para gestionar las evaluaciones	83
13.1.2. Resumen de los resultados del análisis en la pestaña Dashboard	84
13.1.3. Cobertura detallada por exón	85
Metodología	85
Visualización e interpretación	86

13.1.4. Pequeñas variantes	89
Metodología	89
Visualización e interpretación	93
Clasificar las variantes	106
Configuración de las columnas de la tabla de variantes	108
Editar los metadatos de la muestra desde el panel de análisis	109
Filtrar las variantes	111
Filtros Genes & Variants	111
Sección Samples & Quality	111
Sección Frequencies	112
Sección Genetics	113
Sección Databases	115
Sección Transmission & Phenotypes	115
Sección Classification	117
Sección Scores	118
Seleccionar una indicación	120
Utilizar un perfil de filtros predefinidos	121
Editar y guardar perfiles de filtro	121
Perfiles de filtro favoritos	121
Combinación de perfiles de filtro	122
Perfiles de filtro SeqOne	122
Funcionalidad Mark as viewed	123
Selección de variantes	124
Evaluar las variantes	124
Página variante	127
Generar un informe clínico	136
13.1.5. CNV	137
13.1.5.1. CNV en paneles de genes	137
Metodología	137
Visualización e interpretación	138
13.1.5.2. CNV en el análisis SomaCGP y SomaLBx	143
Metodología	143
Visualización e interpretación	143
13.1.5.3. CNV en exomas y genomas	144
Metodología para el exoma	144
Metodología para el genoma	145
Visualización e interpretación	145
Tabla de resultados de la CNV	145
Página detallada de la CNV	149
13.1.6. Visor de variantes grandes para CNVs y SVs (LVV)	154

Tabla de resultados de CNVs & SVs	154
Clasificación, Filtrado, Visualización e Interpretación	156
Panel lateral CNV/SV (Details Drawer)	157
13.1.7. Firmas en el análisis SomaCGP	158
Metodología para calcular la puntuación del MSI	158
Metodología para calcular la puntuación TMB	158
Visualización e interpretación	159
13.1.8. Repeticiones cortas en tándem (STR)	161
Metodología	161
Visualización e interpretación	161
13.1.9. Genes de fusión	163
13.1.9.1. Fusiones detectadas a partir del ADN	163
Metodología	163
Visualización e interpretación	163
13.1.9.2. Fusiones detectadas a partir de ARN	165
Metodología	165
Visualización e interpretación	166
13.1.10. Clonality	169
13.1.11. Variantes de splicing	169
Metodología	169
Visualización e interpretación	170
13.1.12. Estado MSI/MSS	171
Metodología	172
Visualización e interpretación	172
13.1.13. Estado HRD/HRP	174
Metodología	174
Visualización e interpretación	176
13.1.14. Datos de metilación para SomaMethyl	182
14. Informe	185
14.1.1. Archivos de salida	185
15. Solución de problemas y mensajes de error	185
16. Garantía	187
17. Menciones legales	187
17.1. Condiciones de uso	187
17.2. Limitación de responsabilidad	188
17.3. Derechos de autor y marcas registradas	188

1. Introducción

Le recomendamos que lea atentamente las instrucciones antes de utilizar el software. Puede solicitar una versión impresa de estas instrucciones al equipo de soporte de SeqOne (support@seqone.com). Para obtener los mejores resultados al utilizar la plataforma, lea

atentamente todas las precauciones resaltadas con el símbolo . Los distintos símbolos normativos y otros símbolos utilizados a lo largo de esta guía y del software son los siguientes:

Símbolo	Significado
	Fabricante
	Indica el representante autorizado establecido en Suiza.
	Fecha de fabricación (Fecha de lanzamiento del software)
	Dispositivo médico de diagnóstico in vitro.
	Identificador único de dispositivo
	Referencia del catálogo del fabricante
	Cumple los requisitos del Reglamento europeo (UE) 2017/746 Marcado CE con el número de identificación del organismo notificado
	Consulte las instrucciones de uso
	Advertencia Este símbolo de seguridad indica que se requiere un cuidado especial al utilizar el aparato.

La etiqueta (que incluye información sobre la versión del software SeqOne y datos de contacto) y este documento que describe las instrucciones de uso (en varios idiomas) están disponibles en la sección "About".

2. Contactar con el fabricante y el representante suizo



SeqOne S.A.S

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

22 rue Durand
34000 Montpellier
FRANCIA

Soporte al usuario :

support@seqone.com



Arazy Group Swiss GmbH

Bruderholzallee 53

4059 Basilea, Suiza

swiss.ar@arazygroup.com

3. Vigilancia de los productos sanitarios

Si se produce o es probable que se produzca un incidente grave al utilizar el dispositivo, se pide a los usuarios que notifiquen el incidente sin demora injustificada al fabricante y a la autoridad competente del país donde se haya producido o sea probable que se produzca el incidente.

Las normas aplicables y la lista de autoridades competentes en la UE están disponibles en la siguiente dirección:

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



4. Uso previsto de la plataforma

4.1. Uso previsto

SeqOne platform es un dispositivo médico de diagnóstico *in vitro* basado en software, diseñado para apoyar a los profesionales de la salud en el análisis e interpretación de datos genómicos y de secuenciación brutos de pacientes sometidos a análisis NGS o CGH arrays. Al proporcionar información cualitativa y accionable, la plataforma funciona como un sistema de apoyo a la decisión, ayudando a los profesionales de la salud a tomar decisiones clínicas.

4.2. Indicación de uso

SeqOne Platform admite tanto el análisis secundario y terciario a partir de datos de secuenciación brutos, así como el análisis terciario por sí solo. Sus capacidades analíticas, independientes de la patología, tienen aplicaciones en diversos campos médicos, incluyendo la oncología y las enfermedades raras.

4.3. Descripción

SeqOne Platform es un dispositivo médico de diagnóstico *in vitro* basado en software que integra varias aplicaciones que trabajan en conjunto para proporcionar a los usuarios expertos una poderosa herramienta para interpretar datos genómicos y de secuenciación complejos.

SeqOne Platform integra información consolidada de una amplia gama de **fuentes externas**, incluyendo bases de datos públicas, literatura científica y bases de conocimiento propietarias. Estas fuentes externas de datos proporcionan contexto y significado a las variantes genómicas identificadas.

Los archivos de entrada para diversas aplicaciones dentro de la plataforma se procesan utilizando una combinación personalizada de módulos bioinformáticos y de anotación, lo que garantiza que cada análisis se optimice para sus requisitos específicos.

Los **Módulos de Bioinformática** realizan análisis secundarios mediante métodos computacionales, como el alineamiento y la identificación de variantes, para identificar variantes genéticas en datos de secuenciación de ADN y ARN, comparándolos con un genoma de referencia (hg19 o hg38). Los datos, útiles tanto para análisis germinales como somáticos, pueden generarse a partir de secuenciación del genoma completo, secuenciación del exoma o paneles de genes dirigidos.

Los **Módulos de Anotación** realizan el análisis terciario para anotar variantes con detalles como el gen y su función conocida, la localización de la variante dentro del gen y cualquier publicación o reporte clínico que mencione el gen o la variante. Esta anotación determina qué reglas de clasificación ACMG y AMP/CAP/ASCO aplican a cada variante y, a partir de ahí, se puede llegar a un veredicto de patogenicidad. La anotación también se realiza a través de DiagAI, un conjunto de herramientas impulsadas por IA que ayuda a analizar rápidamente variantes genéticas e identificar posibles variantes causales.

Esta arquitectura modular permite a SeqOne ser altamente adaptable a diferentes tipos de análisis y preguntas de investigación. Los archivos de entrada para las diversas aplicaciones dentro de la plataforma se procesan utilizando una combinación personalizada de estos módulos de Bioinformática y Anotación, asegurando que cada análisis esté optimizado para sus requisitos específicos. Al integrar estas diversas fuentes de datos y herramientas de

análisis, SeqOne permite a los usuarios expertos obtener una visión profunda del complejo panorama de los datos genómicos y sus implicaciones para la salud y la enfermedad humanas.

La plataforma funciona como una solución SaaS basada en la nube, con todas las funcionalidades accesibles a través de un navegador web. Incluye SeqOne Flow, una herramienta opcional que puede desplegarse en la nube de AWS o en los servidores del cliente para facilitar la transferencia de datos y la integración de flujos de trabajo.

Por seguridad, todos los datos se cifran en reposo usando cifrado AES-256 y el tráfico entre los usuarios y la plataforma está protegido mediante certificados SSL.

4.4. Población de pacientes

La solución es adecuada para cualquier paciente que requiera un análisis genómico. La población diana pueden ser personas con cáncer, enfermedades raras, enfermedades hereditarias, predisposición genética al cáncer o personas sanas.

4.5. Usuarios previstos

SeqOne Platform está diseñada para profesionales de la salud responsables del análisis clínico rutinario de datos genómicos. Esto incluye el análisis e interpretación de archivos brutos generados por secuenciadores a partir de muestras biológicas secuenciadas. Los usuarios objetivo de la plataforma son genetistas, bioinformáticos, biólogos, ingenieros y técnicos de plataformas NGS.

4.6. Características

Cada aplicación (workset) consta de múltiples **módulos de bioinformática y anotación**, cada uno de los cuales se encarga de tareas analíticas específicas. En este capítulo, se evalúa el rendimiento a nivel de módulo para garantizar una valoración detallada de la precisión y fiabilidad de cada componente.

Los analitos detectados por la SeqOne Platform se clasifican en dos categorías de alteraciones genómicas:

- Variantes genéticas como SNV, indels, MNV, CNV, UPD, fusiones génicas, variantes estructurales, variantes de splicing, inserciones de Alu y FLT3-ITD (duplicación en tándem interna).
- Marcadores característicos o biomarcadores complejos, como el estado de IGHV, la inestabilidad de los microsatélites (MSI), la carga mutacional tumoral (TMB), la metilación o los defectos en la reparación de roturas de doble cadena mediante recombinación homóloga (HRD, Homologous Recombination Deficiency).

Módulo de bioinformática	Workset	Tipo de analito
Detección de CNV Exome Germline	CNVExome, GermVar, GermVar Family	Variante genética
Detección de CNV Panel Germline	CNVcapture	Variante genética
SNV/Indel Germline	GermlineVar, GermlineFamily	Variante genética
SV Germline	GermlineVar, GermlineFamily	Variante genética
UPD	GermVar, GermVar Family, GermVar Tertiary, GermVar Tertiary Family	Variante genética
Inserciones Alu	GermlineVar, GermlineFamily, SomaVar, SomaVar LF, SomaCGP, SomaHemato, GermVar, GermVar Family	Variante genética
MNV	GermlineVar, GermlineFamily, GermVar, GermVar Family, GermVar Tertiary, GermVar Family Tertiary, SomaVar, SomaVar LF, SomaCGP, SomaHemato, SomaLBx, SomaVar Tertiary	Variante genética
SNV/Indel Somático	SomaVar, SomaHemato, SomaRNA, SomaCGP	Variante genética
SNV/Indel somático LF	SomaVar LF	Variante genética
CNV Paneles somáticos amplios	SomaCGP	Variante genética
Índice de metilación	SomaMethyl	Firma
Fusión de ADN somático	SomaCGP, SomaVar, SomaVar LF, SomaHemato	Variante genética
Fusión de ADN somático LBx	SomaLBx	Variante genética
TMB	SomaCGP	Firma
MSI	SomaCGP, SomaMSI	Firma
Probabilidad GI (GiP)	SomaHRD	Firma
Sentieon SNV/Indel Germline	GermVar, GermVar Family	Variante genética
Variantes mosaicas Germline	GermVar, GermVar Family	Variante genética
CNV WGS	GermVar, GermVar Family	Variante genética
Sentieon SNV/Indel Somatic	SomaLBX	Variante genética
Fusión de ARN	SomaRNA	Variante genética
Splicing RNA	SomaRNA	Variante genética
Panel pequeño de CNV somáticas	SomaCNVCapture	Variante genética
Estado de IGHV	SomaHemato	Firma
Subconjuntos IGHV (subsets)	SomaHemato	Firma
CFH	GermlineVarLR	Variante genética
FLT3	SomaVar, SomaHemato	Variante genética

4.6.1. Rendimiento analítico

El rendimiento indicado en estas instrucciones de uso corresponde a las últimas versiones para diagnóstico in vitro de los worksets, que se enumeran en la tabla siguiente:

Nombre del workset	Versión
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. Módulos bioinformáticos

- **Detección de CNV en el exoma Germline y en el Panel Germline**

Type of variant	Versión del genoma	Controles positivos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad
Detección de CNV exoma Germline (GermlineVar)	GRCh37	8	8	0	100%	100%
Detección de CNV exoma Germline (GermVar)	GRCh37	8	8	0	100%	100%
Detección de CNV Panel Germline	GRCh37	15	15	0	100%	100%
Detección de CNV Panel Germline	GRCh38	11	11	0	100%	100%

- **SNV / Indel en la línea germinal - Datos del panel**

Versión del genoma	Tipo de variante	Criterios de aceptación	Sensibilidad antes del filtro	Sensibilidad tras el filtro
GRCh37	SNV	100 %	98 %	100 %
GRCh38	SNV	100 %	98 %	100 %

Versión del genoma	Tipo de variante	Criterios de aceptación	Precisión antes del filtro	Precisión tras el filtro
GRCh37	SNV	100 %	90 %	100 %
GRCh38	SNV	100 %	90 %	100 %

- **SNV / Indel Germline - Datos del exoma**

Versión del genoma	Tipo de variante	Criterios de aceptación*	Sensibilidad antes del filtro	Sensibilidad tras el 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 %

* para la detección de variantes ubicadas en las regiones de alta confianza de NA12878 y NA24385.

Versión del genoma	Tipo de variante	Criterios de aceptación*	Precisión antes del filtro	Precisión tras el 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 %

* para la detección de variantes ubicadas en las regiones de alta confianza de NA12878 y NA24385.

- **SNV / Indel Germline - repetibilidad y reproducibilidad**

- **Repetibilidad**

Tipo de variante	Sensibilidad	Precisión
SNV	98,59 % ± 0 %	95,92 % ± 0 %
indel	87,75 % ± 0 %	81,63 % ± 0 %
todo	98,38 % ± 0 %	95,63 ± 0 %

- **Reproducibilidad**

Tipo de variante	Sensibilidad	Precisión
SNV	98,57 % ± 0,03 %	95,93 % ± 0,08 %
indel	86,58 % ± 0,42 %	78,97 % ± 0,96 %
todo	98,34 % ± 0,03 %	95,58 ± 0,09 %

- **Variantes mosaicas Germline**

- **Sensibilidad SNV**

Versión del genoma	Intervalos de VAF	N.º de variantes verdaderas	Criterios de aceptación	Sensibilidad 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 %

- **Sensibilidad a indeles**

Versión del genoma	Intervalos de VAF	N.º de variantes verdaderas	Criterios de aceptación	Sensibilidad 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 de la línea Germline**

Type of variant	Versión del genoma	Controles positivos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad
SV de la línea germinal	GRCh37	2	2	0	100 %	100 %

- **Inserciones de Alu**

Versión del genoma	Controles positivos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad
GRCh37	1	1	0	100 %	100 %
GRCh38	1	1	0	100 %	100 %

- **MNV**

Versión del genoma	Tipo de variante	Criterios de aceptación	Precisión del exome	Genoma de precisión
GRCh37	MNV	> 85 %	87,0 %	86,9 %
GRCh38	MNV	> 85 %	93,0 %	91,5 %

- **SNV / Indel somático**

- **Sensibilidad**

Versión del genoma	Tipo de variante	Controles positivos	Verdaderos positivos	Falsos negativos	Criterios de aceptación*	Sensibilidad
GRCh37	SNV (≥ 5 %)	121	121	0	100 %	100 %
	Indel (≥ 5 %)	21	21	0	100 %	100 %
GRCh38	SNV (≥ 5 %)	121	121	0	100 %	100 %
	Indel (≥ 5 %)	21	21	0	100 %	100 %

*El criterio de aceptación para la detección de variantes se define como una sensibilidad del 100 % para SNV e indels con un VAF esperado superior al 5 %.

○ **Exactitud**

			Exactitud			
			GRCh37		GRCh38	
Gene	Variant	Frecuencia esperada	Frecuencia *	Sesgo	Frecuencia de sesgo*	Sesgo
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 de las frecuencias se obtiene a partir de 12 réplicas de la muestra HD200.

● **SNV/Indel Somatic LF**

Genoma Versión	Positivos control	Verdaderos positivos	Falsos negativos	Criterios de aceptación *	Sensibilidad
GRCh37	42	42	0	100 %	100 %
GRCh38	42	42	0	100 %	100 %

*El criterio de aceptación para la detección de variantes se define como una sensibilidad del 100 % para SNV e indels con una VAF esperada superior al 0,1 %.

● **Somatic Large Panels de CNV**

Versión del genoma	Tipo de variante	Controles positivos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad
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 %

- Índice de metilación

Versión del genoma	Criterios de aceptación	Concordancia ponderada
GRCh37	>99 %	99,88 %

- Fusión de ADN somático

Genoma Versión	Positivo controles	Verdaderos positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh37	9	9	0	100 %	100 %
GRCh38	9	9	0	100 %	100 %

- Fusión de ADN somático LBx

Versión del genoma	Fracción tumoral	Controles positivos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad
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

Versión del genoma	Criterios de aceptación	Sensibilidad	Precisión
GRCh37	>80 %	83 %	85 %

- MSI

Versión del genoma	Controles positivos	Controles negativos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad	Precisión
GRCh37	22	24	22	0	100 %	100 %	100 %
GRCh38	6	17	6	0	100 %	100 %	100 %

- GiP

Genoma versión	Control controles	Verdadero positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh37	171	165	6	90 %	96,5 %

Genoma Versión	Controles controles	Verdaderos positivos	Falsos positivos	Aceptación Criterios	Precisión
GRCh37	171	165	7	90 %	95,9 %

- **Sentieon SNV / Indel Germline**

Versión del genoma	Type of variant	Criterios de aceptación*	Sensibilidad Exome	Sensibilidad 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 %

* para la detección de variantes ubicadas en las regiones de alta confianza de NA12878 y NA24385

Versión del genoma	Tipo de variante	Criterios de aceptación*	Precisión Exome	Precisión del genoma
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 %

* para la detección de variantes ubicadas en las regiones de alta confianza de NA12878 y NA24385.

- **CNV WGS**

Versión del genoma	Controles positivos	Verdaderos positivos	Falsos positivos	Criterios de aceptación	Sensibilidad
GRCh38	Duplicación	1 kb - 5 kb	8	10 %	37,5 %
	Duplicación	≥ 5 kb	3	10 %	100 %
	Delección	1 kb - 5 kb	321	80 %	61,7 %
	Delección	≥ 5 kb	132	80 %	92,4 %

Aunque la sensibilidad a las deleciones pequeñas (1-5 kb) queda por debajo de los criterios de aceptación, esta reducción es un compromiso aceptable para maximizar la detección de eventos grandes y duplicaciones de mayor impacto.

Versión del genoma	Controles positivos	Verdaderos positivos	Falsos positivos	Criterios de aceptación	Precisión
GRCh38	Duplicación	1 kb - 5 kb	8	"NA"	30 %
	Duplicación	≥ 5 kb	3	"NA"	75 %
	Delección	1 kb - 5 kb	321	"NA"	94,3 %
	Delección	≥ 5 kb	132	"NA"	92,4 %

- **Sentieon SNV/Indel Somatic**

Versión del genoma	Tipo de variante	Controles positivos	Verdaderos positivos	Falsos positivos	Criterios de aceptación	Sensibilidad
GRCh38	SNV	39	37	2	85 %	95 %
	Indel	21	16	5	75 %	76 %

- **ARN de fusión**

Genoma Versión	Positivo controles	Verdadero positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh37	20	20	0	100 %	100 %
GRCh38	20	19	1	95 %	95 %

- **Splicing RNA**

Genoma Versión	Positivo controles	Verdadero positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh37	3	3	0	100 %	100 %
GRCh38	3	3	0	100 %	100 %

- **Panel reducido de CNV somáticas**

Genoma Versión	Positivo controles	Verdaderos positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh37	4	4	0	100 %	100 %
GRCh38	4	4	0	100 %	100 %

- **Status del IGHV IMGT**

Genoma versión	Positivo control	Verdadero positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh37	122	116	6	95 %	95,9 %

- **Estado de IGHV igblast**

Versión del genoma	Controles positivos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad
GRCh37	122	117	5	95 %	95,1 %

- Subconjuntos de IGHV

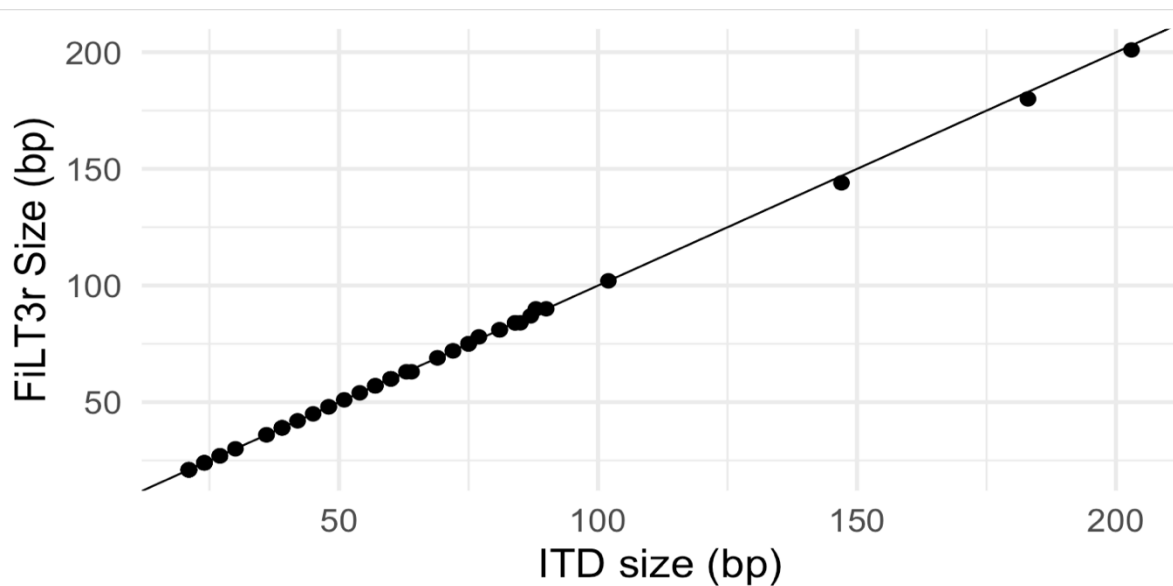
Genoma Versión	Control control	Verdaderos positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh37	61	10	0	95 %	100 %

Genoma Versión	Control controles	Verdadero positivos	Falsos negativos	Aceptación Criterios	Precisión
GRCh37	61	10	0	95 %	100 %

- CFH

Genoma versión	Control controles	Verdadero positivos	Falsos negativos	Aceptación Criterios	Sensibilidad
GRCh38	4	4	0	100 %	100 %

- FLT3



Tamaño observado frente al teórico de los eventos FLT3-ITD. **Coefficiente de Pearson $R^2 > 0,99$**

Los resultados muestran una sensibilidad del 100 % para la detección de eventos FLT3-ITD y una correlación casi perfecta entre los tamaños de evento esperados y observados a partir del análisis.

- **UPD**

- **Sensibilidad**

Type of analysis	Versión del genoma	Controles positivos	Controles negativos	Verdaderos positivos	Falsos negativos	Criterios de aceptación	Sensibilidad
Singleton	GRCh37/38	7	23	7	0	80 %	100 %
Trío	GRCh37/38	7	2	6	1	80 %	86 %

- **Precisión**

Tipo de análisis	Versión del genoma	Controles positivos	Controles negativos	Verdaderos positivos	Falsos positivos	Criterios de aceptación	Precisión
Singleton	GRCh37/38	7	23	7	0	80 %	100 %
Trío	GRCh37/38	7	2	6	0	80 %	100 %

4.6.1.2. Módulos de anotación

- **Puntuación DiagAI**

Conjunto de datos	Análisis de tipos	Métrica evaluada	Aceptación Criterios	Métrica valor
Exomes positivos para enfermedades raras	'Germline'	Rango medio	= 1	1
Exomas positivos para enfermedades raras	"Germline"	top_10	> 90 %	95 %
Exomas de CNV de enfermedades raras	'Germline'	Rango medio	= 1	1
Exomes con CNV positivos para enfermedades raras	'Germline'	top_1	> 80 %	83,3 %

- **Puntuación somática de DiagAI**

Conjunto de datos	Tipo de análisis	Métrica evaluada	Criterios de aceptación	Valor de la métrica
Cáncer positivo TSO500	"Somatic"	top_10	> 90 %	100 %

- **Listas de preselección de DiagAI**

Conjunto de datos	Análisis de tipo	Métrica evaluadas	Aceptación Criterios	Valor de la métrica
Exomas positivos para enfermedades raras	"Germline"	Sensibilidad	> 90	95 %

Exomas positivos para enfermedades raras	"Germline"	Longitud media	< 25	18,475
Exomes positivos para enfermedades raras	"Germline"	Longitud media	< 25	17

- **Selecciones inteligentes de DiagAI**

Conjunto de datos	Análisis de tipo	Métrica evaluada	Aceptación Criterios	Valor de la métrica
Exomes positivos para enfermedades raras	'Germline'	Sensibilidad	> 40 %	45 %
Exomes positivos para enfermedades raras	"Germline"	precisión	> 80 %	78 %

- **ACMG CNV**

Conjunto de datos	Type of analysis	Métrica evaluado	Criterios de aceptación	Valor de la métrica
Conjunto de datos de CNV anotado por expertos de la ACMG	"Germline"	Sensibilidad	> 70 %	76 %

- **SNV de la ACMG**

Conjunto de datos	Tipo de análisis	Métrica evaluado	Criterios de aceptación	Valor de la métrica
Conjunto de datos anotado por expertos de la ACMG	'Germline'	precisión	> 50 %	56 %

- **DiagAI UP2**

Conjunto de datos	Análisis de tipo	Métrica evaluada	Criterios de aceptación	Métrica valor
Variantes causantes de enfermedades	'Germline'	Sensibilidad	90 %	92 %

- **DiagAI UP2 SV**

Conjunto de datos	Análisis de tipo	Métrica evaluada	Criterios de aceptación	Métrica valor
Variantes causantes de enfermedades	"Germline"	Sensibilidad	90 %	98 %

- **DiagAI Phenogenius**

Conjunto de datos	Type of analysis	Métrica evaluada	Criterios de aceptación	Métrica valor
Asociaciones genéticas-síntomas	"Germline"	top_100	50 %	49,8 %
Asociaciones entre genes y síntomas	"Germline"	top_200	60 %	60,7 %
Asociaciones entre genes y síntomas	"Germline"	top_500	75 %	77,0 %
Asociaciones entre genes y síntomas	'Germline'	Mediana en el top 100	<20	15
Asociaciones gen-síntoma	"Germline"	Media en el top 100	<30	25

- **Compermed**

Conjunto de datos	Tipo de análisis	Métrica evaluado	Criterios de aceptación	Métrica valor
Cáncer positivo TSO500	Sensibilidad	sensibilidad	90 %	91 %

4.6.2. Rendimiento clínico

La SeqOne Platform no está destinada a utilizarse para la evaluación de una afección clínica o un estado fisiológico concretos. Aunque proporciona información sobre el estado fisiológico de un paciente, como la identificación de determinadas variantes, no está destinada a evaluar afecciones clínicas o estados fisiológicos específicos. Por lo tanto, la SeqOne Platform en sí misma no requiere una demostración de rendimiento clínico.

5. Precauciones

5.1. Uso conforme

(Véase la sección «**Usuarios previstos**»)

Queda prohibido cualquier uso inadecuado de la plataforma.

Los usuarios son responsables de comprobar que la plataforma esté operativa en todo momento y de garantizar las condiciones necesarias para su correcto funcionamiento.

5.2. Advertencias

 Los resultados proporcionados no deben ser la única base para modificar la atención al paciente. Por consiguiente, los usuarios deben actuar con la debida precaución al utilizar el software y sus resultados. Los diagnósticos deben confirmarse mediante diversos exámenes clínicos y médicos.

La investigación de enfermedades mediante NGS puede revelar hallazgos secundarios o incidentales, como la predisposición al cáncer, a partir de genes no relacionados con la afección primaria (Kuhlen y Borkhardt 2015). A pesar de las declaraciones de política de organismos como la AMP y la ACMG (Green et al. 2013; Allyse y Michie 2013; Junta Directiva de la ACMG 2015; Hegde et al. 2015), las directrices nacionales sobre la notificación de hallazgos incidentales varían. Por lo tanto, es fundamental obtener un documento de «consentimiento informado» (Anderson et al. 2016). El médico responsable del tratamiento es responsable de garantizar que el paciente o su representante legal comprenda los posibles resultados y riesgos antes de solicitar cualquier prueba de NGS.

Este producto no garantiza la detección de todas las variantes genéticas presentes en una muestra. La probabilidad de detectar una variante de interés depende intrínsecamente de la calidad y las características de la muestra de entrada (por ejemplo, la integridad de la muestra, la preparación de la biblioteca, la profundidad y uniformidad de la secuenciación, el filtrado bioinformático y el tipo o contexto de la variante). Un resultado negativo no descarta de forma definitiva la presencia de una variante patógena si su frecuencia (VAF) está por debajo de los umbrales de detección del módulo analítico utilizado. No se garantiza la detección fiable de reordenamientos genómicos a gran escala (variantes estructurales, incluidas las CNV y las deleciones de cromosomas completos). Esto se debe a la limitación inherente de los datos de secuenciación de nueva generación (NGS) de lectura corta, lo que dificulta la identificación de estos eventos. Dado que las lecturas de NGS son cortas (aproximadamente 150-200 pb), esta restricción es una característica de la propia tecnología de secuenciación y no una limitación de la SeqOne Platform. (Véase Tan R. et al., Hum Mutat. Julio de 2014; 35(7):899-907. doi: 10.1002/humu.22537 y referencias allí citadas). Los usuarios deben evaluar la idoneidad del workset seleccionado y la

calidad de la secuenciación en relación con las variantes esperadas para garantizar que se ha cubierto el contexto clínico adecuado.

Las decisiones clínicas basadas en los resultados del cribado neonatal o de portadores deben tomarse con extrema precaución y nunca basarse únicamente en los resultados de este software. En el cribado neonatal, un resultado positivo requiere una correlación clínica inmediata con los síntomas del lactante y pruebas de confirmación urgentes antes de iniciar tratamientos que alteren la vida o restricciones dietéticas. En el cribado de portadores, un resultado negativo no elimina el riesgo de ser portador, sino que lo reduce a un nivel de riesgo residual, ya que ciertas variantes genómicas pueden estar por debajo del límite de detección del software. Dado que los resultados de este dispositivo influyen directamente en decisiones reproductivas irreversibles y en la atención neonatal crítica, todos los hallazgos que requieran actuación deben verificarse mediante un método alternativo validado e interpretarse por un profesional cualificado en el contexto del historial clínico completo del paciente y de los datos bioquímicos o fenotípicos pertinentes.

Las variantes recién descubiertas que aún no están exhaustivamente documentadas en la literatura científica pueden clasificarse erróneamente.

Los resultados son precisos en el momento del análisis y reflejan los conocimientos científicos actuales. Sin embargo, volver a realizar el análisis en una versión más reciente de la plataforma puede arrojar resultados diferentes.

Los resultados de clasificación de la SeqOne Platform pretenden ajustarse al consenso general. Tenga en cuenta que cualquier resultado generado tras modificar la configuración de clasificación queda fuera del ámbito de aplicación de la CE-IVD y, por lo tanto, no está validado.

Este software incorpora componentes de inteligencia artificial (IA). Los usuarios deben ser conscientes de las siguientes limitaciones inherentes a estas herramientas basadas en IA y tenerlas en cuenta a la hora de interpretar los resultados:

- Los datos de entrenamiento proceden predominantemente de poblaciones ancestrales europeas/caucásicas debido a la composición de las bases de datos clínicas de referencia actuales;
- El rendimiento puede variar en el caso de variantes raras con representación limitada en los datos de entrenamiento;
- Las puntuaciones de confianza reflejan la incertidumbre del modelo, no la certeza clínica; deben interpretarse en conjunción con el contexto clínico;
- El rendimiento puede verse reducido por la deriva de datos procedente de nuevas plataformas de secuenciación o protocolos de preparación de bibliotecas no representados en los datos de entrenamiento;
- El rendimiento de la priorización de genes de DiagAI Phenogenius se evalúa en función de las asociaciones entre genes y síntomas y depende de la calidad y la exhaustividad de la anotación de términos HPO realizada por el médico solicitante.

5.3. Limitaciones

"Carrier screening" para enfermedades genéticas está sujeto a la legislación nacional específica y a las normas éticas. Es responsabilidad del laboratorio y del profesional sanitario garantizar que se cumplan todos los requisitos legales —incluidos el asesoramiento genético obligatorio, la solicitud por parte de un médico y el consentimiento informado— dentro de su jurisdicción específica.

5.4. Averías

En caso de mal funcionamiento:

- Deje de utilizar el software inmediatamente
- Intente identificar o eliminar las causas siguiendo el procedimiento de resolución de problemas descrito en esta guía (consulte la sección «**Resolución de problemas y mensajes de error**»)
- Si no puede identificar o eliminar la causa utilizando esta guía, envíe un email al equipo de asistencia a support@segone.com .

6. Requisitos mínimos

6.1. Hardware

Los módulos de software de la plataforma SeqOne se instalan como parte de un acuerdo cliente-servidor. El proveedor de servicios debe ser un servidor Linux para garantizar la compatibilidad con la mayoría de las herramientas utilizadas habitualmente en bioinformática. Cada elemento del sistema se encapsula en un entorno de ejecución como el contenedor Docker®, que puede desplegarse en cualquier plataforma Docker® conforme, independientemente de la arquitectura de hardware subyacente. Los usuarios deben disponer de un ordenador con conexión a Internet y un navegador web (véase más abajo).

Los requisitos mínimos del sistema para un ordenador cliente son los siguientes:

- 4 GB RAM
- Procesador de doble núcleo
- Acceso a Internet
- Navegador Chrome (versión 36.0.1985 en adelante) o Firefox (versión 31 en adelante) o Microsoft Edge (todas las versiones basadas en Chromium).

6.2. Red

La velocidad mínima de subida y bajada recomendada para utilizar la plataforma a través de Internet es de 100 Mbps. Para conectarse a la plataforma se utiliza el protocolo HTTPS. Para acceder correctamente a la plataforma, le recomendamos que solicite a su departamento de informática que autorice las siguientes direcciones :

*.eu.seqone.com

*.common.seqone.com

7. Autenticación

7.1. Conexión

Por motivos de seguridad, los usuarios sólo pueden acceder a la plataforma tras iniciar sesión en la pantalla de inicio.

La pantalla de inicio está disponible a través de la siguiente URL:

- Para Europa y África: <https://auth.eu.seqone.com/>
- Para Suiza: <https://auth.ch.seqone.com/>
- Para Asia y Oceanía: <https://auth.au.seqone.com/>
- Para América: <https://auth.us.seqone.com/>

Al iniciar sesión en la pantalla de inicio, debe introducir su:

- Dirección de correo electrónico o de inicio de sesión en el campo "Username" or "email"
- Contraseña en el campo "Password"

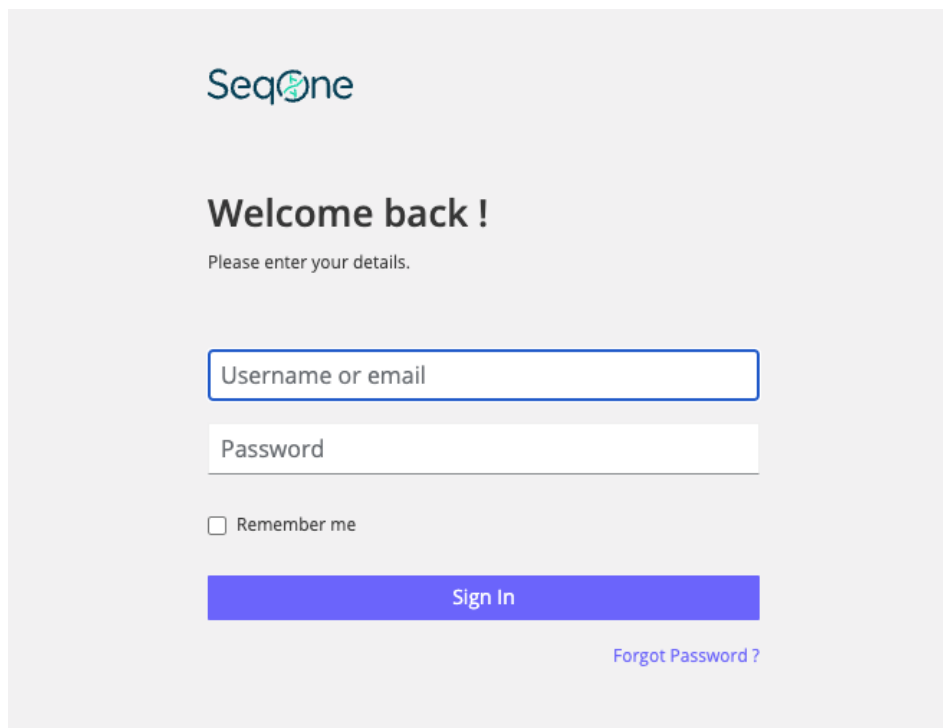


Figura 1: Página de inicio de sesión

Si la autenticación de dos factores está activada para su cuenta, deberá introducir un código de verificación de 6 dígitos enviado por correo electrónico a la dirección asociada a su cuenta SeqOne. Este código es válido durante 30 minutos.

Si tiene acceso a varios espacios de trabajo o entidades, se le pedirá que seleccione la entidad a la que desea conectarse mediante un menú desplegable en "Choose a workspace".

Puede cambiar su contraseña haciendo clic en "Forgot Password". Siga las instrucciones introduciendo la dirección de correo electrónico asociada a su cuenta SeqOne y recibirá un correo electrónico que le permitirá restablecer su contraseña.

La política de contraseñas sigue las siguientes reglas:

- Longitud mínima: 12 caracteres.
- Complejidad: las contraseñas deben incluir al menos una letra mayúscula, una letra minúscula, un número y un carácter especial. Cuanto mayor sea el conjunto de caracteres, más segura será la contraseña.
- Historial de contraseñas: no reutilización de contraseñas anteriores.
- Sin longitud máxima: para permitir el uso de frases secretas.

No es posible que los usuarios cambien su propio ID de usuario.

7.2. Desconexión

Los usuarios pueden cerrar la sesión haciendo clic en el logotipo de la cuenta de usuario en la esquina superior derecha de la página y luego en "Logout".

7.3. Página principal

Tras la autenticación, el usuario accede a la página del salpicadero, que se divide en dos partes:

- un banner azul de búsqueda y acción, accesible en toda la navegación, que comprende de izquierda a derecha :
 - el icono SeqOne para volver al panel de control en cualquier momento,
 - una barra de búsqueda para acceder directamente a un análisis determinado,
- un icono de búsqueda VKB  para buscar variantes evaluadas en la base de datos VKB (véase la sección **Evaluar las variantes**),
- un icono  que abre la "To do list" del usuario, desde la que puede acceder a los análisis que le han sido asignados (véase el apartado **Asignación de análisis para interpretación**),
- una barra de menús, disponible a través del icono  , que da acceso a los parámetros del usuario, al panel del administrador, al manual del usuario, a las condiciones de uso, a la dirección de correo electrónico de soporte y que permite al usuario desconectarse (véase la sección **Gestión de cuentas de usuarios**).
- la pantalla principal de gestión de proyectos y análisis.

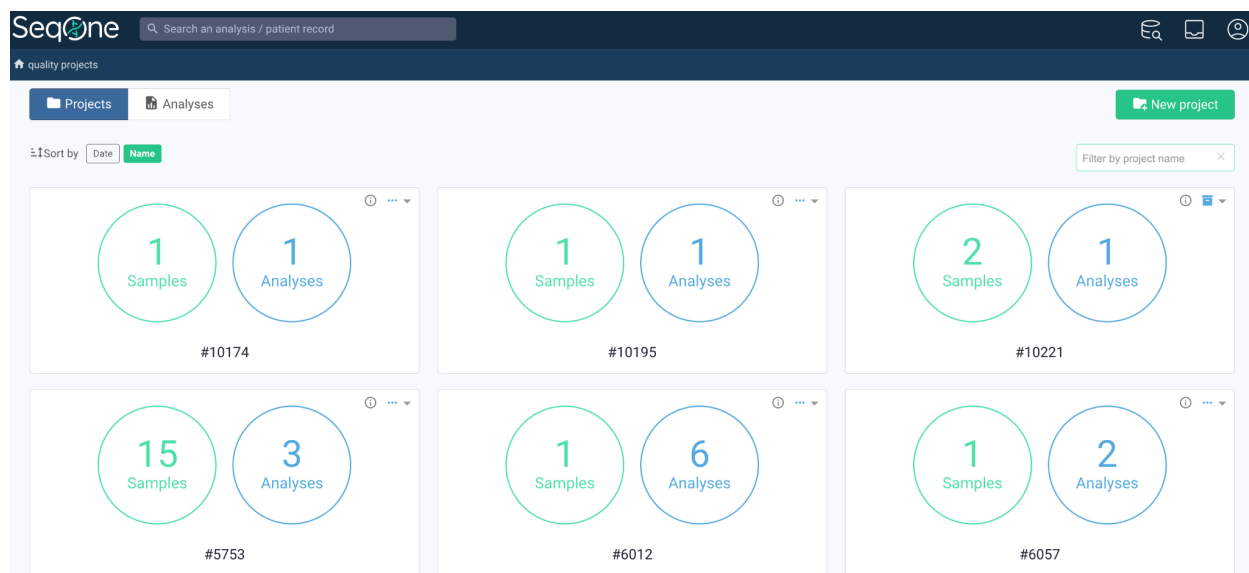


Figura 2: Panel de usuario

Puede acceder al panel de usuario en cualquier momento haciendo clic en el logotipo **SeqOne**.

7.4. Gestión de cuentas de usuarios

Puede gestionar su cuenta de usuario haciendo clic en el logotipo de la esquina superior derecha de la barra de tareas.

En esta sección, el usuario tiene la posibilidad de:

- acceder a la configuración de la cuenta haciendo clic en "My account"
- acceder a los parámetros del usuario haciendo clic en "Entity Settings"
- cambiar la entidad a la que está conectado el usuario haciendo clic en "Entity" y seleccionando la entidad de su elección en "Choose your workspace",
- acceder al panel de administración haciendo clic en "Admin", permitiendo el acceso a la lista de usuarios con acceso a la entidad, y la creación, modificación o eliminación de cuentas de usuario.
- hacer clic en "About" para acceder al manual de usuario en varios idiomas, información sobre la versión del software SeqOne y datos de contacto,
- acceder a los términos de su licencia SeqOne haciendo clic en "Terms of Use",
- hacer clic en "Support" para acceder a la dirección de correo electrónico de soporte para cualquier pregunta o problema detectado en la plataforma,
- acceder a los accesos directo haciendo clic en "Keyboard Shortcuts"

- cambiar entre la versión estándar y la versión de vista previa de la aplicación haciendo clic en "SeqOne Mode" y seleccionando "SeqOne" (standard mode) o "SeqOne Preview" (preview mode). El modo activo actualmente se indica con una marca de verificación.
- cerrar sesión haciendo clic en "Logout".

7.4.1. Gestión de "My Account"

En la sección "My account", el usuario puede:

- cargar su firma para incluirla en el informe clínico,
- cambiar su contraseña haciendo clic en "Change your password" (abre una nueva pestaña) en la sección "Manage your credentials".

7.4.2. Gestión de la configuración del usuario - definiciones

En la configuración de usuario, el usuario puede gestionar los ajustes, archivos y configuraciones descritos a continuación.

Entity

La entidad se refiere al centro o equipo que ha contratado una licencia de la plataforma SeqOne. La plataforma está diseñada para ser utilizada por varios miembros de un mismo grupo. Desde los ajustes de usuario, la pestaña *Cuenta* se utiliza para gestionar la información relativa a la entidad.

Users

Para esta pestaña, consulte la **Gestión administrativa - presentación de la interfaz**.

Worksets

Esta pestaña enumera los diferentes worksets de análisis disponibles en la entidad a la que está conectado el usuario. Dado que en la plataforma SeqOne pueden coexistir varias versiones de determinados pipelines, el usuario puede definir desde este menú qué versión debe preseleccionarse al lanzar el análisis.

Se hace un seguimiento de las modificaciones realizadas en esta lista, de modo que se muestra automáticamente el usuario que definió la versión que debe lanzarse por defecto de un workset, así como la fecha de la modificación.

Manifest

El manifiesto muestra las regiones genómicas de interés (ROI) como un archivo BED, que puede ir acompañado de sus anotaciones asociadas. Puede cargar un nuevo manifiesto a través de la pestaña *Manifest* de la configuración de usuario haciendo clic en "Add manifest" y seleccionando el archivo de manifiesto que desea cargar. Una vez cargado, este manifiesto estará disponible para todos los usuarios al crear un nuevo proyecto y al aplicar un Panel in silico en un análisis.

Panels - Transcript

Los usuarios pueden cargar una nueva lista de transcripciones de referencia (RefSeq) a través de la pestaña *Transcript*.

Panels - Genes

Esta pestaña le permite importar o generar paneles de genes basadas en el NCBI GeneID haciendo clic en "Add gene list".

El formato de archivo esperado es un archivo tabular (por ejemplo, un archivo .tsv) sin encabezados de columna, que se detalla en la plataforma. La primera columna del archivo esperado está dedicada a los identificadores de genes (NCBI Enter gene ID), la segunda columna (optional) contiene el transcrito de interés (RefSeq ID) y la tercera columna (opcional) contiene el símbolo HGNC del gen. Este archivo también puede generarse insertando la lista de genes deseada, separada por una coma, en el campo "Gene symbols".

Marker

Esta pestaña se utiliza para importar una lista de marcadores de identidad genética haciendo clic en "Add marker config". El formato de archivo esperado es un archivo tabular (archivo .tsv o .txt) con una cabecera de columnas que contiene 5 columnas que detallan: cromosoma (CHR), identificador de variante o rsID (ID), posición genómica (POS), alelo de referencia (REF) y alelo alternativo (ALT). El archivo no debe contener más de 100 líneas, incluida la cabecera.

Protocols

Un protocolo de análisis es una combinación predefinida de tareas diseñada para estandarizar la interpretación de los análisis. La pestaña correspondiente permite a los usuarios ver la lista de protocolos disponibles en la entidad, sus versiones, y crearlos.

Para crear un protocolo, haga clic en "New protocol", luego añada y ordene las tareas que coincidan con los SOP de su laboratorio. Para editar un protocolo existente, haga clic en el icono del lápiz junto a él y guarde la nueva versión del protocolo.

Seleccione un protocolo al iniciar el análisis, o adjúntelo más tarde desde la pestaña Workflow en el panel de análisis. Marque las tareas como completadas directamente desde el panel de análisis. Cada tarea completada registra automáticamente el usuario y la fecha para una trazabilidad total.

Reporting - Global settings

La sección Global settings de la pestaña Reporting permite a los usuarios activar o desactivar el bloqueo automático del análisis cuando el estado cambia a "Reporting Validating", y personalizar las secciones de clasificación y de informes. Para ello, en Germline / Somatic, haga clic en "New section", introduzca un nombre de sección y seleccione una etiqueta para asociarla a la clasificación deseada. Las variantes que coincidan con una de las etiquetas o clasificaciones seleccionadas se añadirán automáticamente a esta sección. Primero debe crear etiquetas en la pestaña correspondiente para utilizar esta función.

Reporting - Template list

La pestaña Reporting bajo "Template list" permite a los usuarios configurar las plantillas utilizadas al exportar los resultados del análisis. Desde esta pestaña, puede personalizar la información incluida en los informes generados.

Tags

La pestaña "Tags" permite a los usuarios gestionar etiquetas personalizadas para organizar y filtrar variantes dentro de la VKB y el informe. Los usuarios pueden crear etiquetas haciendo clic en "Add tag", o renombrar y eliminar etiquetas que están disponibles para todos los usuarios dentro de la entidad utilizando los iconos de lápiz y papelera correspondientes.

API Keys

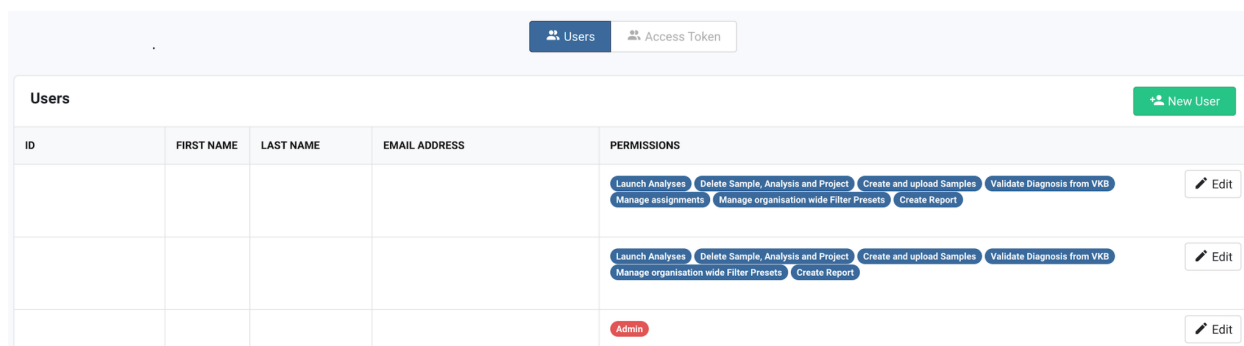
Una clave API es un token de autenticación único que permite que aplicaciones o scripts externos interactúen con la plataforma SeqOne de forma programática. Desde la pestaña API Keys, puede ver todas las claves existentes junto con su información asociada.

Cree una nueva clave haciendo clic en "+ New API Key". Las claves caducadas pueden mostrarse u ocultarse mediante el interruptor "Show expired key". Cada clave puede eliminarse individualmente. Las claves API deben mantenerse confidenciales, ya que otorgan acceso a la plataforma con sus permisos asociados.

7.4.3. Gestión administrativa - presentación de la interfaz

Esta interfaz está dedicada a los administradores de cuentas y les permite :

- acceder a la lista de usuarios de la entidad,
- crear nuevos usuarios,
- modificar los derechos de los usuarios: crear proyectos y cargar muestras, lanzar análisis, eliminar muestras, análisis o proyectos, evaluar variantes en VKB, crear informes, filtrar la gestión de perfiles, gestionar asignaciones.
- activar la doble autenticación,
- desactivar una cuenta de usuario.



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: Panel del administrador

El ID corresponde al nombre de usuario del usuario. Debe estar en minúsculas y no contener acentos ni espacios; se aceptan caracteres especiales como el punto y el guión bajo. Se aconseja seguir la nomenclatura nombre.apellido al crear un nuevo usuario.

En cuanto a la dirección de correo electrónico del usuario, desaconsejamos encarecidamente utilizar una dirección de correo electrónico genérica por motivos de seguridad.

Como los permisos están desactivados por defecto, corresponde al administrador activarlos 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: Permisos de usuario

También puedes mejorar la seguridad activando la doble autenticación. Esta opción envía un código único al usuario cada vez que se conecta a la plataforma.

La función "Manage organization wide Filter Presets" permite al usuario modificar los perfiles de filtro públicos. Si no dispone de esta función, solo podrá modificar los perfiles de filtro privados.

Los permisos son gestionados por un administrador de entidad desde el panel Admin. Cuando se crea un usuario, todos los permisos están desactivados por defecto, y el administrador debe activar manualmente los derechos requeridos para ese usuario. Los derechos asignados a un usuario se aplican a todas las entidades (espacios de trabajo) a las que el usuario tiene acceso en la plataforma SeqOne (por ejemplo, si el usuario puede acceder a varias entidades, se aplica el mismo conjunto de permisos en todas ellas).

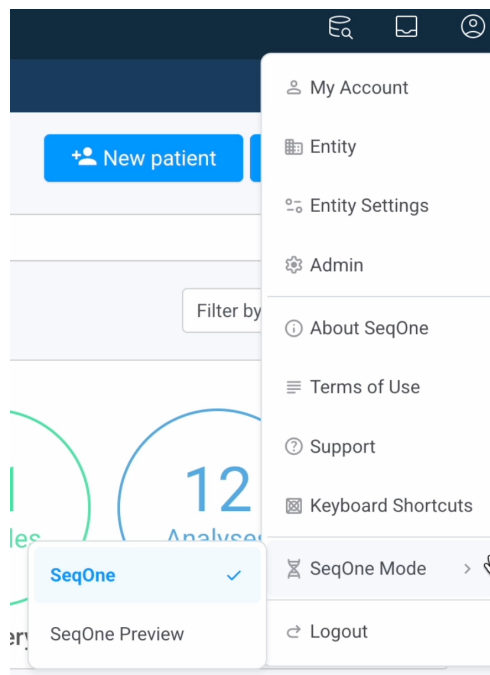
Si recibe un mensaje de error, póngase en contacto con el servicio de atención al cliente en support@seqone.com.

7.4.4. Cambiar entre los modos Main y Preview

Conectarse a SeqOne a través de la URL de inicio de sesión predeterminada lleva a los usuarios al modo “Main” de su entidad. Este modo predeterminado aloja conjuntos de trabajo con marcado CE y otras certificaciones locales.

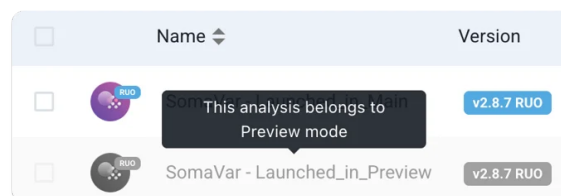
Para obtener acceso anticipado a nuevas funciones y versiones de conjuntos de trabajo, cambie al modo “Preview” (RUO: Research Use Only) yendo a la parte inferior del menú de configuración y, en SeqOne One, seleccionando “Preview”.

Los modos “Main” y “Preview” comparten los mismos datos principales (proyectos, muestras, análisis), permisos, recursos (archivos, plantillas, manifiestos, listas de genes), conocimiento (VKB) y frecuencias (datos de frecuencia de muestras, ocurrencias).



Al navegar en el modo “Preview”, aparece el logotipo de “SeqOne Preview RUO” como recordatorio de que este modo contiene funciones y conjuntos de trabajo RUO.

En el modo “Main”, los análisis iniciados en “Preview” aparecen atenuados. Si hace clic en uno de estos análisis atenuados, la plataforma le indica que cambie al modo Preview para verlo.



8. Descripción de las etapas operativas

Paso	Descripción
1	Creación de proyectos Los análisis realizados en la plataforma SeqOne se organizan dentro de proyectos, que definen las características de las muestras analizadas mediante metadatos introducidos por el usuario (tipo de secuenciación, método de preparación, genes de interés).
2	Carga de muestras Las muestras se cargan como archivos de secuenciación en bruto (FASTQ), que deben colocarse en el proyecto designado. Una vez cargados los archivos, un paso de verificación garantiza la integridad de los archivos enviados antes de cualquier análisis.
3	Análisis Lanzamiento Una vez cargados los datos en los servidores de SeqOne, el usuario puede elegir entre una selección de aplicaciones o "worksets", en función del tipo de muestra analizada y de los tipos de alteraciones genómicas buscadas (SNV, CNV, fusiones génicas, variantes de splicing, etc.). Estas aplicaciones realizan análisis bioinformáticos.
4	Comprobación de la calidad de los datos Las aplicaciones disponibles en la plataforma SeqOne incluyen varios módulos dedicados al análisis de la calidad de los datos, utilizando diversas métricas de secuenciación, alineación y profundidad de cobertura.
5	Interpretación de los resultados de los análisis Las variaciones genómicas detectadas en la muestra se muestran en varias pestañas o "vista", donde el usuario puede buscar la variante diagnóstica responsable de la patología del paciente.
6	Obtención de recomendaciones terapéuticas y ensayos clínicos asociados a la patología del paciente.
7	Emisión de un informe Los resultados del análisis pueden exportarse en forma de informe PDF, que luego puede descargarse.

9. Creación de proyectos

Los análisis realizados en la plataforma SeqOne se organizan dentro de proyectos, que definen las características de las muestras analizadas mediante metadatos introducidos por el usuario (tipo de secuenciación, método de preparación, genes de interés).

9.1. Crear un proyecto

Tras hacer clic en "*New project*", se pide a los usuarios que introduzcan los valores de varios tipos de información, o metadatos, que se utilizan para describir el proyecto y las muestras asociadas.

- **Descripción del proyecto**
- **Secuenciación:**
 - Assay type: tipo de análisis (línea germinal o somática),
 - Los proyectos Germline (*Assay type Germline*) incluyen una casilla de verificación UMI Trimming, que permite procesar muestras que contienen UMI. Al seleccionar esta opción, las UMI se recortarán de las secuencias FASTQ, como en el modo de UMI desactivado para los workset somáticos (consulte la sección **Elección del modo de procesamiento de UMI**).
 - Sequencing platform: la plataforma de secuenciación
 - Library type: tipo de secuenciación de read realizada (*Paired-end* o *Single-end*),
 - Data type: DNA, RNA, ctDNA o Methylation Sequencing,
 - Si se selecciona ADN para el tipo de datos, debe especificarse el método de secuenciación mediante el menú desplegable "*DNA Sequencing method*", así como el método de amplificación (Captura o Amplicón) mediante "*Amplification method*".
 - File input type: *Archivos Fastq*, *VCF*
 - Genome: la versión del genoma de referencia (GRCh37 o GRCh38),
 - Low-Quality Mapping Region: la llamada de variantes en regiones de alineamiento de baja calidad considerando las 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: el manifiesto en el caso de un panel de genes,
 - Marcador de identidad (opcional): el archivo del marcador de identidad.
- **Muestras:**

- (Opcional) Se pueden añadir nuevos metadatos de muestra haciendo clic en "*You can also create your own metadata*". Cuando se le solicite, introduzca el nombre y el valor de los metadatos.

Una vez introducidos los diferentes metadatos del proyecto crear su proyecto pulsando el botón "*Create project*".

9.2. Gestión de proyectos

A través del panel de control, los usuarios pueden gestionar los diferentes proyectos que se han creado en la plataforma. En concreto, pueden:

- Acceda a los proyectos a través de la pestaña *Projects*, así como a las muestras y análisis de cada proyecto haciendo clic en los logotipos "*Samples*" y "*Analysis*", respectivamente.
- Acceda directamente a los análisis a través de la pestaña *Analysis*.
- Cree un nuevo proyecto haciendo clic en "*New project*" (véase el apartado **Crear un proyecto**).

9.3. Suprimir un proyecto

Para eliminar un proyecto, los usuarios deben:

- Seleccione el proyecto
- Haga clic en la pestaña "*Settings*" y luego en "*Delete project*".
- Confirme marcando "*Yes*" en el mensaje que aparece. Esta acción es irreversible.

9.4. Datos de ADN (DNA Sequencing, ctDNA Sequencing)

Low-Quality Mapping Region. Al secuenciar paneles de genes, exomas o genomas, el usuario puede incluir regiones de baja calidad de alineación al llamar variantes. Estas regiones corresponden a las llamadas reads MAPQ0, cuya calidad de alineación es igual a 0, como pseudogenes y regiones homólogas del genoma.

Esta opción es aplicable a genes o regiones diana específicos. Para utilizar esta opción, se invita al usuario a cargar el manifiesto dirigido a las regiones deseadas, o a seleccionarlo mediante un menú desplegable.

Esta opción puede reducir la calidad y precisión de los resultados. Puede introducir ruido de fondo en la frecuencia alélica y tener un impacto negativo en la priorización de variantes. También puede aumentar el tiempo de análisis debido a una mayor carga computacional.

Exome. En el caso de la secuenciación del exoma, los usuarios no necesitan especificar el manifiesto que se va a utilizar; se les invita a seleccionar el fabricante del kit utilizado haciendo clic en uno de los iconos de la categoría Proveedor de kit de exoma. A continuación, un menú desplegable permite al usuario seleccionar el manifiesto deseado. Si el manifiesto deseado no está disponible en la plataforma, se invita al usuario a ponerse en contacto con el servicio de soporte al usuario (support@segone.com).

Gene panel. En el caso de un panel de genes, se pide a los usuarios que seleccionen el *método de amplificación*, así como el manifiesto (en formato BED) correspondiente a los genes de interés en un menú desplegable. Si el manifiesto requerido no está disponible en la plataforma, puede cargarse seleccionando "Upload new manifest" y siguiendo las instrucciones que aparecen en pantalla.

Shallow WGS, CGH MicroArrays, Whole genome sequencing. En el caso de la secuenciación del genoma, el usuario no necesita especificar qué manifiesto utilizar.

Capture. Al elegir el método de amplificación captura, aparece un icono UMI (*Unique Molecular Identifiers*). Los identificadores moleculares únicos son índices que se añaden antes de la amplificación para ayudar a eliminar los duplicados de PCR y reducir el ruido de fondo después del análisis.

Amplicon. Al analizar los amplicones, debe prestarse especial atención a los cebadores de amplificación. Al crear un proyecto para datos de amplicones, se aconseja a los usuarios que seleccionen un manifiesto en el que las regiones correspondientes a los objetivos no incluyan las secuencias de cebadores. Si no se dispone de tal BED, se puede seleccionar una opción para *recortar* una longitud fija al final de las secuencias.

Identity monitoring. Para los métodos de secuenciación correspondientes a panel de genes, exoma, perfil genómico (CGP) o genoma (WGS superficial) a partir de archivos FASTQ, se invita al usuario a seleccionar el archivo de marcadores de control de identidad en un menú desplegable. Si el archivo deseado no está disponible en la plataforma, puede importarse seleccionando "Upload new markers" y siguiendo las instrucciones en pantalla (véase también el párrafo **Marker** en la sección **Gestión de la configuración del usuario**).

9.5. Datos de ARN (RNA Sequencing)

Si se selecciona ARN como tipo de datos, los usuarios deben especificar si su protocolo conserva la información de la cadena original mediante el menú desplegable *Protocolo de cadenas*. Si el protocolo es de hebra específica, se debe especificar el orden de síntesis de las hebras (*hacia adelante-atrás* o *hacia atrás-adelante*) mediante la opción *Orientación de read*. Por último, debe especificarse el método de secuenciación (panel de genes).

Read orientation. Existen dos opciones para especificar la orientación de la secuencia:

- **Forward-Reverse:** la primera read (R1) de cada par corresponde a la secuencia del ARN mensajero transcrito (cadena sentido).
- **Reverse-Forward :** la primera read (R1) de cada par corresponde a la secuencia complementaria del ARN mensajero transcrito (cadena antisentido).

En general, la orientación de las reads es la siguiente para los kits mencionados a continuación:

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

En caso de duda, póngase en contacto con el servicio de soporte (support@segone.com).

Una vez especificado el tipo de datos y configurado correctamente el proyecto, se pide al usuario que lo valide pulsando el botón "Next step".

9.6. Configuración del proyecto

En la pestaña "Settings" del proyecto, puede :

- modifique el nombre y la descripción del proyecto en los campos "Project name" y "Project description",
- modificar el perfil de control de calidad (QC),
- cambiar el perfil de filtro por defecto en la tabla de variantes,
- eliminar definitivamente el proyecto.

9.6.1. Configurar las métricas QC del proyecto por defecto

El usuario puede optar por utilizar las métricas de control de calidad predeterminadas o definir valores umbral personalizados desde la pestaña "Settings" del proyecto. Se invita al usuario a ponerse en contacto con el servicio de soporte al usuario (support@seqone.com) para definir los valores deseados.

Se aplicarán métricas de control de calidad personalizadas :

- *Comprobaciones de control de calidad* en la vista *Análisis* del proyecto,
- *Comprobaciones de control de calidad* en la sección de *autovalidación de control de calidad* de la pestaña de control de calidad del análisis,
- a la pestaña *GeneCov* del análisis.

9.6.2. Establecer el perfil de filtro predeterminado del proyecto

El usuario puede elegir establecer el perfil de filtro predefinido aplicado a la tabla de variantes por defecto desde la pestaña "Settings" del proyecto (véase también la sección **Utilizar un perfil de filtros predefinidos**).

Se puede seleccionar un perfil de filtro para cada workset mediante un menú desplegable, siempre que sea público. El perfil de filtro se aplicará por defecto cuando se visualice la tabla de variantes, para todos los análisis de proyectos, para todos los usuarios de la organización.

Si el usuario no modifica este parámetro, al crear el proyecto, el valor por defecto será el perfil definido por SeqOne.

Los perfiles de filtro por defecto pueden bloquearse haciendo clic en el interruptor "Lock Variants Filters".

10. Carga de muestras

10.1. Cargar muestras

Los usuarios pueden importar sus muestras a través del icono "Add samples" de cada proyecto. Las muestras pueden importarse a la plataforma SeqOne de dos formas:

Desde un PC

Los usuarios pueden cargar sus archivos FastQ directamente desde una carpeta de su PC haciendo clic en la pestaña "From your computer" del icono "Add files". Los usuarios también pueden arrastrar y soltar los archivos desde su PC al área dedicada en la interfaz.

A través de SDS (SeqOne Data Sync)

SDS es un módulo premium opcional que permite crear proyectos automáticamente, transferir datos, iniciar análisis y recuperar datos estructurados (TSV) para integrarlos en el sistema de gestión de información de laboratorio (LIMS).

Este módulo compatible con Linux, MacOS y Windows es un ejecutable que permite a los usuarios autenticarse e interactuar con las API de SeqOne en un entorno predefinido y altamente seguro.

Requisitos mínimos:

- Linux 1.19 mínimo
- MacOS 10.15 con CPU Intel como mínimo
- MacOS con CPU Apple Silicon (chip M1) como mínimo
- Windows 10 64bits mínimo
- 4 GB de RAM (8 GB para Windows)
- Procesador de cuatro núcleos
- Acceso a Internet y a la plataforma SeqOne (app.eu.seqone.com)
- Acceso al sistema de almacenamiento de datos (directamente o a través de la red)

Si desea utilizar esta solución o para más información, póngase en contacto con nuestro equipo de soporte (support@seqone.com).

Estado de la carga de muestras

La plataforma SeqOne utiliza diferentes estados para que los usuarios puedan seguir el progreso de la carga de sus muestras:

- *Upload pending* : importación en curso,
- *Verifying sample...* : conformidad de la muestra en curso de verificación,
- *Mapping...* : etapa de alineación de la muestra,
- *Verified sample* : muestra verificada,
- *Upload failed* : importación fallida.

Capacidad de carga

La capacidad de carga de archivos (especialmente de archivos fastq.gz) depende directamente de la conexión a Internet del usuario. Por lo tanto, debe tenerse en cuenta la velocidad de la conexión a Internet a la hora de cargar muestras. Consulte la siguiente tabla para obtener una idea más clara de los tiempos de carga estimados:

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: Tiempo estimado de carga según el tamaño del archivo y la velocidad de conexión a Internet.

10.2. Formatos de archivo de entrada

10.2.1. Archivos de análisis primarios

El formato del archivo que debe cargarse para cada muestra depende de la plataforma de secuenciación utilizada. Los nombres de los archivos deben ajustarse a la convención de nomenclatura de archivos de cada secuenciador.

Los archivos deben comprimirse con la extensión .gz (gzip) antes de importarlos a la plataforma. En el caso de la secuenciación por pares, se esperan dos archivos, correspondientes a las reads de ambos extremos de cada molécula.

Cuando la plataforma ha detectado correctamente una muestra en el momento de la importación, aparece el nombre completo del archivo. A continuación, es posible cambiar el nombre del archivo haciendo clic sobre él. Los archivos de extremos pareados se combinan automáticamente en una sola muestra.

La plataforma admite los siguientes secuenciadores:

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

- [MGI](#) : formato fastq.gz, sólo paired-end. Los nombres de archivo esperados son:
sample_1.fq.gz y sample_2.fq.gz o flowcellID_L01_sample_1.fq.gz y
flowcellID_L01_sample_2.fq.gz

10.2.2. Archivos de análisis secundarios

El formato del archivo que debe cargarse para cada muestra depende de la plataforma de secuenciación y de la aplicación utilizada para el análisis secundario. Los archivos deben comprimirse con la extensión .gz (gzip) antes de importarlos a la plataforma, a menos que se especifique lo contrario.

El formato de importación de los archivos de análisis secundarios generados con aplicaciones Illumina, Oxford Nanopore Technologies, Agilent, AmoyDx, Pillar, Thermofisher depende del workset utilizado (véase la sección **Archivos de análisis secundarios**).

Cuando la plataforma detecta correctamente una muestra durante la importación, muestra el nombre completo del archivo. Puede cambiar el nombre del archivo haciendo clic sobre él. Los archivos que comparten el mismo nombre base se combinan automáticamente en una sola muestra.

[Archivos DRAGEN](#). Los archivos y formatos esperados para cada workset son los siguientes, cuando se seleccionan los archivos VCF files - Secondary Analysis data y metadatos de Illumina al crear el proyecto:

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

[Archivos de Oxford Nanopore Technologies](#) . Los archivos y formatos esperados son los siguientes al seleccionar archivos VCF files - Secondary Analysis data, y metadatos de Oxford Nanopore Technologies durante la creación del proyecto. Los usuarios pueden importar un archivo de variantes (SNP) por sí mismo o combinarlo con archivos suplementarios para análisis 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

Archivos de Thermofisher/Affymetrix. El formato de importación para los archivos de análisis secundario de CGH es el formato .txt generado por la aplicación Affymetrix de Thermofisher. Este formato solo es compatible cuando se seleccionan tanto los metadatos de Thermofisher/Affymetrix como los de MicroArrays CGH durante la creación del proyecto. Los archivos y formatos esperados para cada workset son los siguientes:

GermVar Tertiary ARRAY_SampleID.txt

Archivos CGH MicroArrays. El formato de importación para los archivos de análisis secundarios de CGH es el formato .xls generado por la aplicación Cytogenomics de Agilent. Este formato sólo se admite cuando se seleccionan tanto los metadatos de Agilent como los de CGH MicroArrays durante la creación del proyecto. Los archivos y formatos esperados para cada workset son los siguientes:

GermVar Tertiary IntervalBasedReport_sample.xls

Archivos Alissa Reporter. El formato de importación para los archivos de análisis secundario de Alissa Reporter es el formato .xls generado por la aplicación Alissa Reporter de Agilent. Este formato sólo se admite cuando se seleccionan tanto los metadatos de Agilent como los de CGH MicroArrays durante la creación del proyecto. Los archivos y formatos esperados para cada workset son los siguientes:

GermVar Tertiary IntervalBasedReport_sample.xls
ABERRATION_&_LOH_INTERVAL_sample.xls

PacBio-HiFi files. Al crear un proyecto, si se seleccionan archivos VCF (Secondary Analysis data) y metadatos de PacBio-HiFi, los archivos y formatos esperados para cada workset son los siguientes:

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

Archivos VCF de paneles de genes somáticos. El formato de importación para los archivos de análisis secundarios SNV, CNV, Fusions es el formato .vcf.gz generado por las aplicaciones Illumina, AmoyDx, Pillar, Thermofisher. Los nombres de archivo originales deben convertirse para que coincidan con el formato de nomenclatura requerido que se describe a continuación.

Los archivos y formatos esperados son los siguientes cuando se seleccionan archivos VCF files - Secondary Analysis data y metadatos de paneles de genes durante la creación de proyectos somáticos. Los usuarios pueden importar un archivo de variantes (SNV) por sí mismo o combinarlo con archivos suplementarios para el análisis CNV o de fusiones:

SomaVar Tertiary

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

Los archivos y formatos esperados son los siguientes al seleccionar "VCF files - Secondary Analysis data" y metadatos de "Comprehensive Genomic Profiling" durante la creación de un proyecto somático. Los usuarios pueden importar un archivo VCF multivariante como el archivo VCF "All Variants" del Oncomine™ Comprehensive Assay V3/Plus.

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

10.3. Metadatos de muestra

Para cada muestra cargada, los usuarios pueden asignar valores a los metadatos que seleccionaron al crear el proyecto o definir nuevos valores haciendo clic en el icono de la esquina superior derecha de la lista de muestras . .

Algunos metadatos de muestra pueden cargarse directamente como archivo de valores separados por tabulaciones seleccionando el icono .

Estos criterios desempeñan un papel importante para garantizar la trazabilidad de los datos y para establecer referencias cruzadas entre los distintos análisis realizados en la plataforma. Algunos metadatos, como el sexo, la enfermedad y los ID de OPH, intervienen en el proceso de identificación de variantes.

Cuando el campo Assay Type se llena con "Germline", los metadatos asociados a la muestra por defecto son "HPO", "Sex" y grupo "GnomAD". Cuando el campo "Assay Type" se llena con "Somatic", los metadatos asociados a la muestra por defecto son "Disease" y "Sex".

Para añadir o modificar HPO en una muestra, el usuario puede hacer clic en "+ Add" en la columna HPO después de seleccionar una muestra. Esto abre una ventana emergente que permite añadir HPOs uno a uno, buscando un fenotipo o código HPO en la barra de búsqueda. También es posible importar un archivo de texto que contenga códigos HPO haciendo clic en el

icono  . A continuación se ofrece un resumen de los términos introducidos, con la posibilidad de suprimir términos si es necesario.

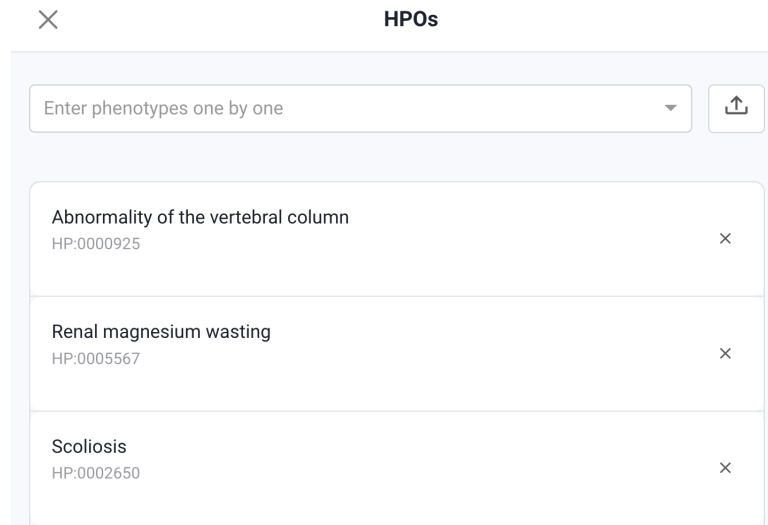


Figura 6: Ventana emergente de importación de HPO

10.4. Editar una muestra

El nombre, así como los metadatos asociados a una muestra, pueden ser editados por el usuario en cualquier momento, seleccionando la muestra en la lista "Samples" y haciendo clic en el icono . A continuación, las modificaciones deben validarse para que surtan efecto.

10.5. Borrar una muestra

Los usuarios pueden eliminar los archivos FastQ de una muestra de la plataforma haciendo clic en el icono  después de seleccionar la muestra deseada. Estos archivos liberarán espacio de almacenamiento y los resultados seguirán estando disponibles en SeqOne.

Para eliminar una muestra, haga clic en el icono  de la pestaña "Samples" después de seleccionar la muestra deseada. La muestra sólo puede eliminarse definitivamente tras borrar todos los análisis asociados a ella.

11. Análisis

11.1. Iniciar un nuevo análisis

Los usuarios pueden iniciar un nuevo análisis haciendo clic en "New analysis" en la pantalla *Overview* de cada proyecto.

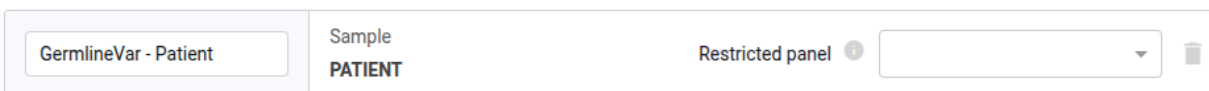
El lanzamiento del análisis se realiza en tres etapas distintas, que se describen a continuación.

El paso "1. **workset**" le permite elegir el workset de análisis adecuado (véase el apartado **Elección del workset de análisis**).

El paso "2. *Select Samples*" se divide en dos partes:

- 2.1. *Set options*:
 - describe el workset seleccionado y el genoma de referencia
 - permite añadir *padding* mediante la función "Extra calling padding" si procede (véase la sección **Elección del workset de análisis**)
 - Permite elegir el modo de procesamiento UMI (véase el apartado **Elegir el modo de tratamiento UMI**) si procede.
 - permite seleccionar el baseline de CNV, el baseline de MSI o ajustar el umbral mínimo de VAF, si corresponde.
 - Permite elegir el protocolo.
- 2.2. *Select sample(s)*: permite seleccionar muestras para su análisis. Los usuarios pueden seleccionar específicamente muestras nuevas, marcadas como "New", mediante la función "select all new".

El paso "3. *Confirm Analysis*" resume las muestras seleccionadas. Para cada análisis que se vaya a lanzar, es posible aplicar un panel restringido seleccionando una o más paneles de genes mediante la opción "*Restricted panel*" (véase también la sección **Vista modal Panel**).



Sample
PATIENT

Restricted panel   

La opción "*Restricted panel*" aplicará un filtro al análisis en forma de una o más paneles de genes. Esta opción no afecta a la llamada de variantes, que viene determinada por el manifiesto del proyecto y el *padding* añadido. La elección de paneles bloqueados es reversible; el usuario puede modificar estos paneles desde la pestaña *Panel* del panel lateral de gestión del análisis después del análisis (véase la sección **Vista modal Panel**).

11.1.1. Elección del workset de análisis

Existen varios workset de análisis para responder a diversas preguntas biológicas. Los workset propuestos al crear un nuevo análisis dependen de los metadatos que se definieron al crear el proyecto y, por tanto, pueden variar en función de los datos.

Workset de análisis disponibles a partir de archivos FASTQ

Los siguientes workset están disponibles cuando se seleccionan los metadatos *Fastq files - Primary Analysis data* durante la creación del proyecto.

Excluyendo los proyectos en los que el método de amplificación es el amplicón, algunos workset (es decir, GermlineVar, GermlineFamily, SomaVar, SomaVar LF (*versiones* ≥ 2.0), SomaCGP (*versión* ≥ 0.6), GermVar, GermVar Family, SomaLBx, SomaMethyl) permiten ampliar el área de búsqueda para la llamada de variantes y añadir entre 0 y 100 pb de relleno al manifiesto inicial.

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

La opción está disponible en el paso 2 (Seleccionar muestra) del lanzamiento del análisis a través de la función "Extra calling padding". Por defecto, el valor es 0 y debe ser un número entero natural.

Algunos worksets (es decir, SomaVar (*versiones* ≥ 2.6) y SomaHemato (*versión* ≥ 0.2)) ofrecen una opción para ajustar el umbral mínimo de VAF para la detección de variantes. Este umbral se puede configurar entre el 0,01 % y el 1 %, siendo el 1 % el valor predeterminado

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)

Para obtener más información sobre esta opción, contacte con el servicio de asistencia al usuario en support@seqone.com.

GermlineVar



GermlineVar se utiliza para detectar mutaciones de la línea germinal a partir de la secuenciación de un panel de genes o del exoma del paciente.

Al ejecutar un análisis GermlineVar, :

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select",
- Seleccione las muestras que desea analizar y haga clic en "Next step",
- Inicie el análisis haciendo clic en "Start analysis".

GermVar



GermVar permite detectar mutaciones de la línea germinal a partir de la secuenciación de un panel de genes, el exoma o el genoma de un paciente.

Al ejecutar un análisis GermVar:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select",
- Seleccione el modelo a utilizar,
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step",
- Inicie el análisis haciendo clic en "Start analysis".

GermlineVarLR



GermlineVarLR se utiliza para detectar mutaciones de la línea germinal en datos de secuenciación long-read procedentes de la secuenciación de paneles de genes.

Al ejecutar un análisis GermlineVarLR:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select",
- Seleccione el modelo que desea utilizar y la(s) muestra(s) que desea analizar y, a continuación, haga clic en "Next step".
- Inicie el análisis haciendo clic en "Start analysis".

GermlineFamily



GermlineFamily se utiliza para detectar mutaciones hereditarias a partir de la secuenciación de un panel de genes, el exoma o el genoma de un paciente y los miembros de su familia. Este Workset también puede utilizarse como parte de una prueba de prediseño (*Carrier Screening*).

Al ejecutar un análisis GermlineFamily, los usuarios deben:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select",
- Seleccione al menos 2 muestras
- En la sección "Set conditions", indique si la muestra es del paciente o de un familiar definiendo el vínculo familiar en "Family link". El paciente se marca automáticamente como "affected". En el caso de una prueba de prediseño (*Carrier Screening*), indique para cada muestra el parentesco mediante "Mother" o "Father" sin marcar la casilla "Is affected?".
- Compruebe que los parámetros coinciden y haga clic en "Next step",
- Inicie el análisis haciendo clic en "Start analysis".

GermVar Family



GermVar Family permite la detección de mutaciones en la línea germinal, a partir de la secuenciación de un panel de genes, exoma o genoma de un paciente

Al ejecutar un análisis GermVar Family:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select",
- Seleccione el modelo a utilizar,
- Seleccione al menos 2 muestras,
- En la sección "Set conditions", indique si la muestra es del paciente o de un familiar definiendo el vínculo familiar en "Family link". El paciente se marca automáticamente como "affected". En el caso de un cribado de portadores, indique para cada muestra el parentesco mediante "Mother" o "Father" sin marcar la casilla "Next step".
- Compruebe que los parámetros coinciden y haga clic en "Siguiendo paso",
- Inicie el análisis haciendo clic en "Start analysis".

CNVCapture



CNVCapture se utiliza para detectar variaciones en el número de copias (CNV) de la línea germinal a partir de datos de captura de paneles de genes.

Al ejecutar un análisis CNVCapture:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select",
- Seleccione un número mínimo de 8 muestras para analizar y, a continuación, haga clic en "Next step",
- Ejecute el análisis haciendo clic en "*Start analysis*".

CNVExome



CNVExome se utiliza para detectar variaciones en el número de copias (CNV) a partir de datos de captura de exomas.

Al ejecutar un análisis CNVExome:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- Seleccione el modelo que desea utilizar y la(s) muestra(s) que desea analizar y, a continuación, haga clic en "Next step".
- Inicie el análisis haciendo clic en "*Start analysis*".

SomaVar



SomaVar detecta mutaciones somáticas en muestras tumorales. Este workset puede utilizarse con datos de captura y amplicón.

Al ejecutar un análisis SomaVar:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado **Elegir el modo de tratamiento UMI**).
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step".
- Inicie el análisis haciendo clic en "Start analysis".

SomaCNVCapture



SomaCNVCapture se utiliza para detectar variaciones somáticas del número de copias (CNV) a partir de datos de captura de paneles de genes.

Al ejecutar un análisis SomaCNVCapture:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado **Elegir el modo de tratamiento UMI**).
- Seleccione un número mínimo de ocho muestras para analizar y, a continuación, haga clic en "Next step".
- Ejecute el análisis haciendo clic en "Start analysis".

SomaRNA



SomaRNA puede utilizarse para detectar fusiones de genes a partir de datos de captura de RNA-seq.

Al ejecutar un análisis de SomaRNA:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado **Elegir el modo de tratamiento UMI**).
- seleccione la(s) muestra(s) que desea analizar y, a continuación, haga clic en "Next step",
- Ejecute el análisis haciendo clic en "Start analysis".

SomaMSI



SomaMSI detecta el estado de las regiones microsatélite (MSS o MSI).

Al ejecutar un análisis SomaMSI:

- Seleccione un número mínimo de ocho muestras para analizar y, a continuación, haga clic en "Next step".
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado **Elegir el modo de tratamiento UMI**).
- seleccione un número mínimo de 8 muestras para analizar y haga clic en "Next step",
- Ejecute el análisis haciendo clic en "Start analysis".

SomaHRD



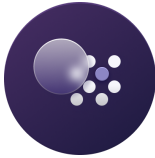
SomaHRD permite la detección del estado HRD a partir de datos WGS poco profundos solos o acompañados de un panel de genes incluyendo BRCA1 y BRCA2.

Al ejecutar un análisis SomaHRD:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- Introduzca los metadatos de celularidad (% de células tumorales) para cada muestra (véase la sección **Metadatos de muestra**).
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step",

- Seleccione el análisis SomaVar o SomaCGP (GRCh37 o GRCh38) apropiado para cada muestra utilizando la opción 'Seleccionar análisis de panel (Opcional)',
- Ejecute el análisis haciendo clic en "Start analysis".

SomaVar LF

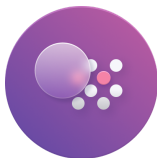


Detección de variantes cortas de baja frecuencia y variantes estructurales, así como genes de fusión, a partir de muestras tumorales (biopsias sólidas o líquidas, ADN tumoral circulante) enriquecidas por captura en un panel de genes.

Al ejecutar un análisis SomaVar LF:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado **Elegir el modo de tratamiento UMI**).
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step".
- Ejecute el análisis haciendo clic en "Start analysis".

SomaCGP



SomaCGP permite la detección de mutaciones somáticas, variantes estructurales, fusiones génicas, así como el estado MSI y TMB de muestras tumorales. Este workset está dedicado al análisis de ADN FFPE preparado con el panel SureSelect Cancer CGP Assay de Agilent, y sólo está disponible cuando se seleccionan los metadatos Comprehensive Genomic Profiling en la creación del proyecto. Este workset puede utilizarse con datos de captura.

Al ejecutar un análisis SomaCGP:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- Si se solicita, seleccione un "CNV baseline" y/o un "MSI baseline" (si no se selecciona ningún baseline, el análisis se ejecutará sin detección de CNV/MSI),
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado Elegir el modo de tratamiento UMI).
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step".
- Inicie el análisis haciendo clic en "Start analysis".

SomaHemato

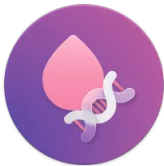


SomaHemato permite la detección de mutaciones somáticas a partir de muestras con especificidades hematológicas. Este workset puede utilizarse con datos de captura.

Al ejecutar un análisis SomaHemato:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado **Elegir el modo de tratamiento UMI**).
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step".
- Inicie el análisis haciendo clic en "Start analysis".

SomaLBx



SomaLBx permite la detección de mutaciones somáticas de baja frecuencia, variantes estructurales, fusiones génicas, a partir de ctDNA. Este workset está dedicado al análisis de ctDNA y puede utilizarse con datos de captura.

Al ejecutar un análisis SomaLBx:

- Seleccione la versión del workset que desea ejecutar y haga clic en "Select".
- Seleccione un "CNV baseline" (si no se selecciona ningún baseline, el análisis se ejecutará sin detección de CNV).
- En las opciones, seleccione el modo de procesamiento UMI aplicable de entre los modos elegibles para el análisis (véase el apartado **Elegir el modo de tratamiento UMI**).
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step".
- Inicie el análisis haciendo clic en "Start analysis".

SomaMethyl



SomaMethyl permite la detección de islas CpG metiladas a partir de muestras de ADN tumoral circulante. Este workset está dedicado al análisis de la metilación en el ADN tumoral circulante y se puede utilizar con datos de captura de bisulfito.

Al ejecutar un análisis SomaMethyl, se invita a los usuarios a:

- Seleccionar la versión del workset que se va a ejecutar y hacer clic en «Select».
- Seleccionar una «Methyl baseline» correspondiente a la cohorte de control (si no se selecciona ninguna cohorte, el análisis se ejecutará sin detección de umbral ni metilación diferencial).

- Seleccionar la muestra o muestras que se van a analizar y, a continuación, hacer clic en «Next step».
- Iniciar el análisis haciendo clic en "Start analysis".

Workset de análisis disponibles a partir de archivos secundarios

Los siguientes workset están disponibles cuando se selecciona *Archivos VCF - Datos de análisis secundarios* durante la creación del proyecto o cuando se seleccionan tanto los metadatos de Agilent como los de CGH MicroArrays (véase la sección **Archivos de análisis secundarios**). En este caso, la plataforma recibe los datos de salida de la aplicación del proveedor seleccionado, procede a anotar e inyectar las variantes en sus respectivas interfaces de visualización.

GermVar Tertiary



Canal de análisis de una sola muestra de variantes detectadas por la aplicación Germline de Illumina o la aplicación EPI2ME de Oxford Nanopore Technologies, a partir de datos WGS y WES o variantes detectadas por la aplicación Cytogenomics de Agilent a partir de datos CGH de MicroArray.

Al ejecutar un análisis GermVar Tertiary:

- Seleccione la versión del workset que desea ejecutar y haga clic en workset "Select",
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step".
- Inicie el análisis haciendo clic en "Start analysis".

GermVar Tertiary Family



Pipeline de análisis multimuestra de variantes detectadas por la aplicación Illumina Germline o la aplicación EPI2ME de Oxford Nanopore Technologies, a partir de datos WGS y WES.

Al ejecutar un análisis de GermVar Tertiary Family:

- Seleccione la versión del workset que desea ejecutar y haga clic en workset "Select",
- Seleccione al menos 2 muestras,
- En la sección "Set conditions", indique si la muestra es del paciente o de un familiar definiendo el vínculo familiar en "Family link". El paciente se marca automáticamente como "affected".
- Compruebe que los ajustes coinciden.
- Ejecute el análisis haciendo clic en "Start analysis".

SomaVar Tertiary



SomaVar Tertiary es un pipeline de análisis terciario basado en VCF que procesa tanto archivos VCF de Illumina WES como archivos VCF de paneles de genes de varios proveedores de kits (ver sección **Input file format**).

Al ejecutar un análisis SomaVar Tertiary:

- Seleccione la versión del workset que desea ejecutar y haga clic en workset "Select",
- Seleccione la(s) muestra(s) que desea analizar y haga clic en "Next step".
- Inicie el análisis haciendo clic en "Start analysis".

11.1.2. Elegir el modo de tratamiento UMI

Presentación de los modos de tratamiento UMI

En la práctica, la plataforma propone uno o varios métodos para procesar las UMI al ejecutar un análisis:

1. **Standard mode** (recomendado): se generan consensos a partir de los duplicados de PCR que presentan la misma UMI cuando su número es mayor o igual a 2. También se conservan las UMI representadas por una única secuencia (singletons). La frecuencia alélica mínima de las variantes notificadas y el número mínimo de reads que apoyan la mutación (AO, Alternate Observation) difieren según el workset utilizado:
 - En SomaVar, SomaHemato, SomaRNA, SomaCNVCapture, SomaCGP y SomaMSI, la frecuencia alélica mínima es 1% y el AO es un mínimo de 3 reads.
 - En SomaVar LF y SomaLBx, la frecuencia alélica mínima es del 0,1% y el AO mínimo es de 2 reads. Este workset permite la detección de variantes con una frecuencia alélica inferior al 1%. Se recomienda cuando la profundidad de secuenciación es superior a 5000x, y para aplicaciones como la secuenciación de ctADN.
2. **UMI disabled** (disponible sólo en SomaVar, SomaRNA, SomaHemato, SomaLBx, SomaCNVCapture, SomaCGP y SomaMSI): Las UMI no se utilizan para la deduplicación y se recortan del final de las secuencias antes del análisis. Por lo tanto, los datos generados por esta línea de análisis son brutos (no deduplicados). La frecuencia alélica mínima de las variantes notificadas es del 1% y el número mínimo de reads que apoyan la mutación (AO) es 3.

Las diferencias de configuración entre estos modos se resumen en la siguiente tabla:

	UMI disabled	UMI standard
Calidad mínima mediana/media de la base mutada (puntuación Phred)*.	30	30
Número mínimo de reads por consenso	N/A	2
Calidad mínima de cada base en el consenso (puntuación Phred)	N/A	30
reads sin consensos filtrados	N/A	no

Tabla 1: Ajustes de las distintas configuraciones de procesamiento de la UMI

* Estos filtros sólo se aplican a las variantes que no son seleccionadas por el modelo SVM después de la llamada de variantes (pipelines SomaVar y SomaRNA).

Proceso de tratamiento UMI

La estructura de las UMI depende de cada kit del proveedor y actualmente están disponibles los siguientes kits en la plataforma 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 configuración UMI "QIAseq Targeted DNA Pro" solo debe utilizarse para SomaVar, con una longitud de lectura de al menos 150 bp. Para más información, contacte con el soporte.

El procesamiento de UMI se basa en el uso de la suite de herramientas fgbio, que puede manejar UMIs y reads etiquetadas con UMIs. Nuestro proceso de análisis UMI consta de los siguientes pasos:

1. Si es necesario, recorte los UMI del final de las secuencias y genere un BAM no alineado a partir del par de archivos FASTQ,
2. Alinear secuencias sin UMIs con BWA-MEM2 (ADN/ctADN) o STAR (ARN)
3. Fusione los archivos BAM que contienen las UMI con las secuencias alineadas utilizando ZipperBam,
4. Agrupa las reads si tienen la misma posición y comparten la misma UMI,
5. Generar la secuencia consenso para cada grupo de UMIs,
6. Filtre cada base en función de su calidad: la confianza en la secuencia de consenso aumenta con el número de reads por grupo,
7. Filtrar secuencias consenso: si se filtra más del 20% de las bases de la secuencia consenso, ésta no se conserva,
8. Conversión de estas secuencias consenso filtradas en FASTQ y realineamiento final en el genoma de referencia con BWA-MEM2 (ADN/ADNct) o STAR (ARN).

Compatibilidad con los workset de análisis

Los worksets SomaVar, SomaHemato, SomaCGP, SomaLBx, SomaMethyl, SomaVar LF, SomaCNVCapture, SomaMSI y SomaRNA son compatibles con los datos UMI. En la tabla siguiente se describen las opciones de procesamiento disponibles al iniciar los distintos workset:

	UMI desactivada	Norma UMI
SomaVar	Sí	Sí

SomaCNVCapture	Sí	Sí
SomaMSI	Sí	Sí
SomaRNA	Sí	Sí
SomaVar LF	No	Sí
SomaCGP	Sí	Sí
SomaLBx	Sí	Sí
SomaHemato	Sí	Sí
SomaMethyl	No	Sí

Tabla 2: Opciones de procesamiento de UMI disponibles al iniciar los workset compatibles.

11.2. Estado del análisis

Tras iniciar un análisis, la plataforma SeqOne utiliza diferentes estados para que los usuarios puedan seguir el progreso de su análisis:

- **Bioinformatics Pending:** la plataforma espera a que termine otro análisis
- **Bioinformatics Running:** el análisis está en curso; este estado también muestra en qué fase del análisis se está llevando a cabo
- **Bioinformatics Finished:** el análisis ha concluido
- **Analysis failed:** se ha producido un error durante el análisis; en este caso, se pide a los usuarios que se pongan en contacto con el equipo de soporte al usuario.
- **Interpretation - new:** el análisis está terminado y listo para su interpretación, ningún usuario lo ha abierto todavía
- **Analysis failed - error:** El análisis ha encontrado un error. En este caso, se invita al usuario a ponerse en contacto con el servicio de atención al usuario.
- **Analysis failed - canceled:** el análisis no se ha iniciado porque la muestra adjunta ha encontrado un error durante la carga o la verificación.

11.3. Calidad de las muestras

La pestaña Control de calidad del proyecto se divide en 2 subpestañas: 'General statistics' y 'ID check summary'.

General statistics. Esta pestaña muestra una tabla que representa la calidad general de la alineación y la cobertura de cada muestra del proyecto. El usuario puede exportar la tabla como archivo TSV haciendo clic en .

El cálculo se basa en la herramienta SeQC, un programa desarrollado por SeqOne para calcular las estadísticas de alineación y cobertura de todas las reads alineadas con el genoma.

El usuario puede optar por utilizar las **métricas** de QC predeterminadas o definir valores umbral personalizados desde la pestaña "Settings" del proyecto (véase la sección **Configurar las métricas QC del proyecto por defecto**).

Las métricas propuestas son las siguientes:

- número total de reads
- longitud media de las reads
- tamaño medio de los fragmentos de ADN de la biblioteca (insertos)
- porcentaje de reads de calidad >30
- porcentaje de reads alineadas con el genoma de referencia
- porcentaje de reads en el objetivo
- porcentaje de reads duplicadas
- tasa de transición/transversión
- Cobertura media
- Fracción de objetivos cubiertos en diferentes umbrales dependiendo del contexto (amplicón, captura).

Los objetivos del panel o del exoma se determinan a partir del archivo BED seleccionado al crear el proyecto.

ID check summary. Esta pestaña detalla los genotipos de todos los marcadores incluidos en sus análisis para cada muestra (véase el párrafo **Marker** en la sección **Gestión de la configuración del usuario - definiciones**). Una fila corresponde a una muestra, y cada columna representa un marcador con su ID asociado. El usuario puede exportar esta tabla como archivo TSV haciendo clic en .

12. Comprobación de la calidad de los datos

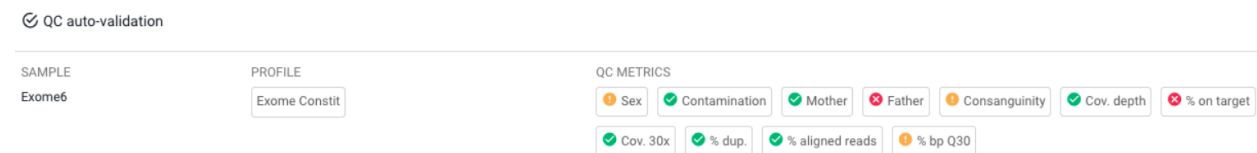
12.1. Métricas e informes de control de calidad

12.1.1. Métricas QC y controles QC

Desde la pestaña QC, el usuario accede a los controles de calidad aplicados al análisis, en forma de subpestañas que se describen en las secciones siguientes: *Informe de QC*, *Comprobación de ID* (si procede) e *Informe técnico*.

En el caso de un análisis MicroArray CGH, la pestaña QC no está disponible. Sin embargo, el informe técnico está disponible en la pestaña *Files* del análisis en formato *report.html*.

En la parte superior de la pestaña Control de calidad, una sección de "QC auto-validation" permite al usuario codificar por colores los parámetros de control de calidad como *Passed*, *Warning* o *Failed*. Haciendo clic en el icono asociado, el usuario accede a los umbrales aplicados y a los detalles del control de calidad: sexo, contaminación, parentesco y consanguinidad, profundidad de la cobertura, % sobre el objetivo, cobertura 30X o 100X, % duplicados, % reads alineadas con el genoma de referencia o % pb Q30.



12.1.2. Informe QC

El informe global de *control de calidad* del proceso de secuenciación contiene diversas métricas que describen la calidad y la alineación de las secuencias, la cobertura de las regiones del manifiesto y algunas estadísticas relacionadas con las variantes. Este informe se compone de las siguientes secciones:

[Alignment details \(Detalles de alineación\)](#) (SomaMethyl). Total de lecturas procesadas, lecturas alineadas de forma única y no única, y tasas de desajuste.

[Deduplication \(Deduplicación\)](#) (SomaMethyl). Métricas de duplicación para lecturas alineadas, destacando posibles problemas con la preparación de la biblioteca.

[Cytosine coverage & methylation \(Cobertura de citosina y metilación\)](#) (SomaMethyl). Tasas de conversión de citosina a uracilo para cada contexto (CpG, CHG, CHH).

UMI groups statistics (sólo SomaVar SomaVar LF, SomaMethyl y SomaRNA UMI). Véase el párrafo **Elegir el modo de tratamiento UMI**.

El informe de control de calidad de los análisis SomaVar, SomaVar LF, SomaMethyl y SomaRNA ofrece una visión más detallada de la composición de la muestra, en particular de la distribución de las UMI en función del número de secuencias que las albergan.

Las estadísticas describen en particular:

- La fracción de singletons (UMI transportada por una secuencia única)
- El tamaño medio y mediano de los grupos de UMI, así como la desviación típica
- El tamaño del grupo más grande y la fracción del total de secuencias que incluye.

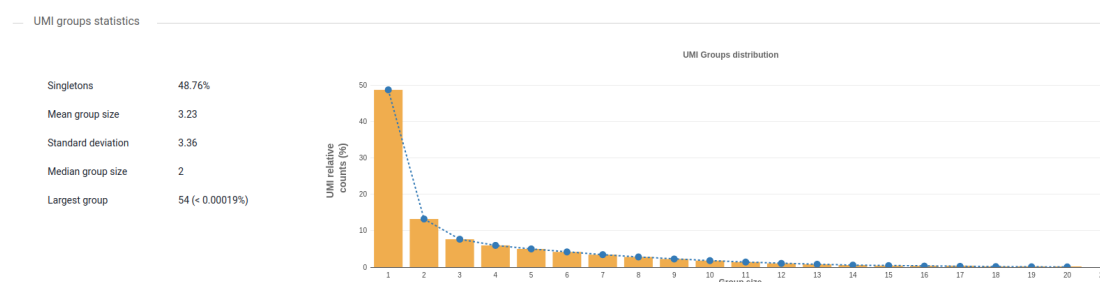


Figura 7: Estadísticas relativas a la distribución UMI en términos de número de secuencias dentro de una muestra.

Las métricas globales de control de calidad del proyecto también son más detalladas y se dividen en tres secciones:

- Métricas de control de calidad tras el procesamiento UMI, que indican la tendencia de cada métrica (en porcentaje) a partir de los datos brutos.
- Métricas de control de calidad brutas calculadas a partir de todas las secuencias (antes del procesamiento UMI).
- Distribución de UMI en términos de número de secuencias dentro de cada muestra.

Cross-contamination estimate (sólo SomaVar LF). La contaminación cruzada se estima individualmente utilizando la herramienta CalculateContamination de GATK. Esta última utiliza la señal de los alelos de referencia detectados en los loci homocigotos para calcular la fracción de secuencias en el individuo procedentes de la contaminación cruzada.

Sample integrity and identity (sólo análisis de la línea germinal). Esta sección enumera las métricas relacionadas con la integridad y la identidad de las muestras:

- **Family link** : metadatos definidos por el usuario
- **Original sex**: metadatos relativos al sexo introducidos por el usuario (Masculino/Femenino/Desconocido)

- **Detected sex:** sexo detectado (Masculino/Femenino/Desconocido)
- **Relationship coefficient :** coeficiente de relación calculado mediante la herramienta *Somalier* para cada muestra del análisis familiar, esta métrica muestra N/A para la propia muestra.
- **Contamination:** porcentaje estimado de contaminación de la muestra por ADN extraño, utilizando la herramienta *verifyBamID2*.

También se enumeran el número de variantes homocigóticas y heterocigóticas, la proporción de estos dos valores (proporción Hom/Het), el número de variantes homocigóticas y heterocigóticas en el cromosoma X, la proporción equivalente en el cromosoma X (proporción Het/Hom en el cr X) y el número de variantes en el cromosoma Y.

Las métricas *Family link* y *Relationship coefficient*, utilizadas para comprobar el parentesco entre los individuos de una familia, sólo se muestran en los análisis de trío familiar (GermlineVar Family o GermVar Tertiary Family).

Las métricas de género se utilizan para comprobar la correspondencia entre el género introducido y el detectado en los datos.

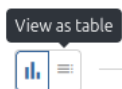
La métrica de contaminación no se calcula para los worksets GermVar Tertiary y GermVar Tertiary Family.

Las comprobaciones de control de calidad sobre coincidencia de género, contaminación y parentesco están disponibles a nivel de proyecto o de análisis para identificar fácilmente cualquier coincidencia o falta de coincidencia.

Alignment metrics. Esta sección enumera las principales métricas calculadas al final de la alineación, como:

- El número total de secuencias de la muestra,
- El tamaño mediano de los insertos, calculado a partir de la distancia entre las dos reads del mismo par, cuya longitud respectiva está incluida,
- La longitud media de la secuencia,
- La fracción de secuencias alineadas con éxito con el genoma de referencia,
- La tasa de duplicados (sólo captura),
- La fracción de bases cuya calidad es superior a 30,
- La cobertura media de las regiones objetivo, determinada a partir del manifiesto de análisis.

Coverage. Gráfico que muestra la fracción de regiones objetivo cubiertas con diferentes umbrales, indicados en el eje x. Los valores de cobertura pueden mostrarse en forma de tabla



haciendo clic en el botón que permite al usuario copiar/pegar estos valores.

QC graphs. Gráficos del análisis de calidad FastQC, que ilustran en particular el contenido en GC de las secuencias, así como la calidad de las bases según su posición en las secuencias.

Variant statistics. Una lista de estadísticas en torno a las variantes detectadas en el análisis, que incluyen en particular:

- El número de transiciones (Ts) y transversiones (Tv) entre las sustituciones detectadas,
- La relación Ts / Tv,
- El número total de mutaciones detectadas,
- La fracción de SNVs, MNVs e indels en este total.

12.1.3. Informe 'ID-Check'

La pestaña "ID check" está dedicada a los resultados del control de identidad en la pestaña QC de los workset que llaman a variantes.

Los resultados se muestran en forma de tabla con la siguiente información:

- Una descripción de cada marcador basada en coordenadas genómicas, por ejemplo: chr-start-ref-alt,
- una etiqueta (rsID, etiqueta personalizada) cuando se la da el usuario,
- genotipo observado (GT) o frecuencia alélica (VAF),
- la profundidad de cobertura en la ubicación del marcador (COV),
- la calidad de la base de datos (QUAL).

Cuando se incluyen varias muestras en el análisis, las columnas específicas de la muestra (GT/VAF, QUAL, COV) se repiten para cada muestra individualmente.

Al seleccionar una variante en la tabla, se abre un navegador IGV en una nueva pestaña, centrado en las coordenadas genómicas de la variante.

12.1.4. Informe técnico

El informe técnico enumera las herramientas utilizadas durante el análisis y su versión. Se puede acceder a él desde la pestaña Control de calidad de un análisis y, a continuación, en "Technical report".

En particular, describe :

- información general sobre la configuración y los metadatos del proyecto y el análisis, como el tipo de secuenciación, la versión del genoma y la versión del workset utilizada;

[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

- detalles del manifiesto utilizado, como nombre, tamaño, longitud (pb) y padding añadido al análisis;
- detalles de las bases de datos y las herramientas bioinformáticas utilizadas.

12.1.5. Informe UPD

El Informe UPD está disponible en la sección QC de los análisis de línea germinal compatibles (p. ej., GermVar y flujos de trabajo relacionados). Resume los resultados de detección de UPD y proporciona elementos visuales para apoyar la interpretación.

El Informe UPD generalmente incluye:

- Resumen de resultados: indicación de si el análisis detectó un evento de tipo UPD y el tipo de evento (p. ej., completo o segmentario; en los análisis de trío, el informe también puede respaldar el origen parental y los subtipos de eventos como UPD | Iso / UPD | Het).
- Lista de eventos: una lista de los eventos detectados con su ubicación cromosómica y tamaño, y la(s) puntuación(es) asociada(s) utilizada(s) por el algoritmo.
- Gráficos de visualización: gráficos basados en genotipos SNP (p. ej., frecuencia del alelo B (BAF) y señal de homocigosidad/ROH) que ayudan a confirmar la presencia de regiones homocigotas largas y sus límites.
- Información de control de calidad: métricas relacionadas con la fiabilidad de la detección (p. ej., densidad de SNP / número de SNPs) y advertencias cuando los datos no son suficientes para una llamada segura.

Los eventos UPD detectados también se muestran en el Large Variant Viewer (LVV) con etiquetas específicas (p. ej., UPD | Iso y UPD | Het).

Puntuación UPD

La puntuación UPD es un indicador numérico utilizado para cuantificar con qué fuerza la muestra muestra una señal de tipo UPD a lo largo del genoma. La detección de UPD en SeqOne se basa en los patrones de genotipo SNP:

- En un análisis de un solo individuo (solo), el algoritmo busca grandes regiones de homocigosidad (ROH) y cambios abruptos en la distribución de SNPs a lo largo de los cromosomas.
- En un análisis de trío, el patrón de SNP se utiliza adicionalmente para determinar el origen parental del evento y para distinguir la isodisomía (UPD | Iso) de la heterodisomía (UPD | Het).

En la práctica, la puntuación UPD refleja la fuerza y consistencia de la señal que respalda un evento UPD (p. ej., patrón claro de ROH/BAF y límites bien delimitados). Una puntuación más alta indica que es más probable que el evento sea un verdadero evento UPD, mientras que valores más bajos indican señales más débiles o ruidosas.

La puntuación UPD siempre debe interpretarse junto con los elementos proporcionados en la pestaña Informe UPD (sección QC), que incluyen:

- los gráficos de visualización de UPD (p. ej., frecuencia del alelo B / señal de homocigosidad),

- el tamaño y la ubicación cromosómica del evento,
- las métricas de calidad del análisis (p. ej., densidad de SNP / número de SNPs).

Gráfico de frecuencia del alelo B (BAF) del informe UPD

El gráfico de frecuencia del alelo B (BAF) es una visualización de los genotipos SNP a lo largo del genoma y se utiliza para evaluar el desequilibrio alélico y la pérdida de heterocigosidad (LOH).

Para cada posición SNP, la BAF corresponde a la fracción de lecturas de secuenciación que respaldan el alelo "B" y varía de 0 a 1:

- Los valores cercanos a 0 corresponden a un genotipo homocigoto (AA).
- Los valores cercanos a 0.5 corresponden a un genotipo heterocigoto (AB).
- Los valores cercanos a 1 corresponden a un genotipo homocigoto (BB).

En una región diploide normal, el gráfico BAF normalmente muestra tres bandas alrededor de 0, 0.5 y 1. En regiones largas de homocigosidad (ROH) o isodisomía UPD, la banda heterocigota (alrededor de 0.5) se reduce o está ausente, lo que respalda la presencia de una señal de tipo UPD.

Limitaciones importantes:

- También se puede observar una señal de tipo UPD en el contexto de consanguinidad parental (múltiples regiones ROH en todo el genoma).
- La llamada de UPD se basa principalmente en la distribución de SNPs y puede no ser suficiente por sí sola para distinguir algunos patrones de UPD de los eventos relacionados con el número de copias; cuando sea necesario, verifique con otra evidencia disponible (p. ej., cobertura / resultados CNV).

13. Interpretación de los resultados de los análisis

13.1. Resultados de los análisis

Los usuarios pueden acceder a los análisis de un proyecto haciendo clic en el icono "Analyses" del cuadro asociado o en el icono del mismo nombre situado en la esquina superior izquierda del cuadro de mandos del usuario.

Cuando un análisis se ha completado y está disponible para su visualización, se muestra con el estado "Interpretation - New"

A continuación, los resultados del análisis están disponibles en forma de varios informes, que pueden abrirse directamente haciendo clic sobre ellos.

Existen varias variaciones del estado "Interpretation", en función de los progresos realizados en la interpretación de los resultados:

- **New** - la tabla de variantes aún no se ha visto
- **In progress**: se ha consultado la tabla de variantes, pero aún no se ha validado ninguna.
- **Evaluating** - se ha validado al menos una variante
- **Reporting validated** : se ha generado un informe para el análisis.

13.1.1. Gestión del flujo de trabajo de interpretación de análisis

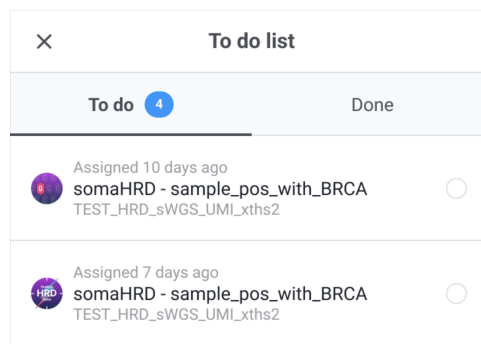
13.1.1.1. Asignación de análisis para interpretación

Los administradores de la entidad pueden elegir asignar uno o más análisis para que sean interpretados por diferentes usuarios de la entidad desde la pestaña Análisis del proyecto. Para ello, el administrador puede seleccionar uno o varios análisis y, a continuación, hacer clic en el

botón  Edit assignment situado en la parte inferior de la página.

Una ventana "Assignment" permite asignar uno o varios usuarios a los análisis seleccionados, buscando el identificador del usuario (nombre de usuario) mediante la barra de búsqueda dedicada. También es posible modificar la asignación añadiendo o eliminando usuarios. Los usuarios a los que se asigna un análisis aparecen en la columna "Assigned to". Los usuarios que nunca han iniciado sesión en SeqOne no pueden ser asignados hasta que creen su credencial.

Los usuarios pueden acceder a los análisis que se les han asignado (pestaña "To do") y a los análisis validados (pestaña "Done") en su "To do list", accesible a través del icono  situado en la parte superior derecha de la página. Aparecerá una notificación para cada nueva asignación.

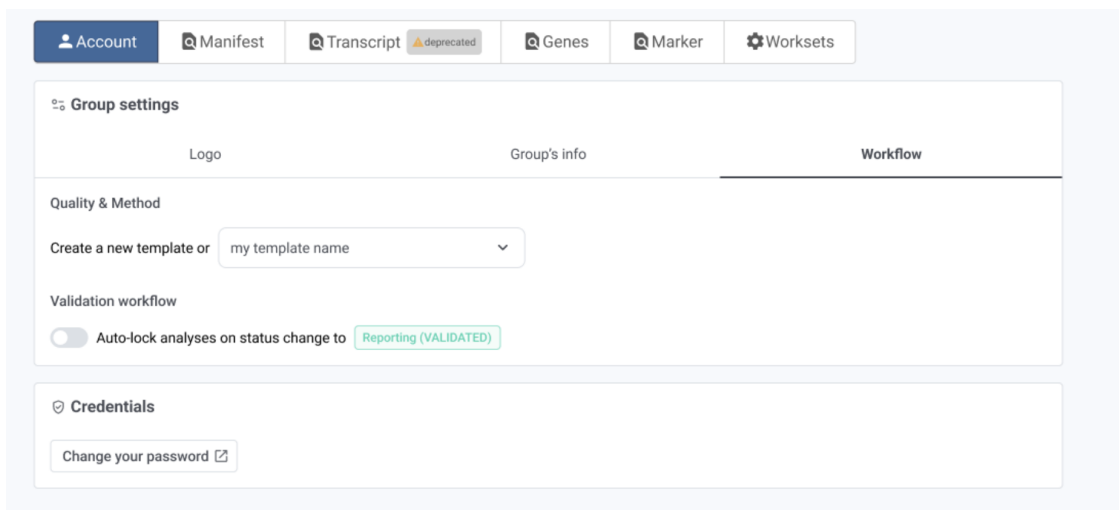


Los usuarios pueden acceder directamente a los análisis asignados haciendo clic en ellos. Pueden validar su tarea haciendo clic en la casilla correspondiente de la pestaña "To do" para desplazarla a la pestaña "Done". Esta acción es reversible.

Los análisis se muestran en el siguiente orden según su estado: finalizados, en ejecución o fallidos. Para cada análisis se detalla la fecha de asignación, el nombre del análisis y el nombre del proyecto.

13.1.1.2. Bloqueo de análisis

El administrador de una entidad puede bloquear o desbloquear un análisis. En *los Settings* de la entidad, en la pestaña *Account* y luego *Workflow*, pueden activar el bloqueo automático de los análisis validados mediante la función *"Auto-lock analyses on status change to Reporting (VALIDATED)"*:



Al final de su trabajo de interpretación, cuando un usuario cambia el estado del análisis en *Reporting - validated*, el análisis se bloquea automáticamente. Los análisis bloqueados muestran el estado "Reporting - Validated & Locked", visible con el icono .

 Reporting VALIDATED & LOCKED

Cuando un análisis está bloqueado, los usuarios no pueden realizar las siguientes acciones desde la interfaz :

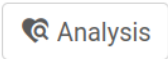
- añadir o modificar evaluaciones de variantes en VKB,
- añadir o modificar comentarios asociados a las variantes,
- generar un nuevo informe desde la ventana *Selection*,
- generar un nuevo informe DRH,
- modificar manualmente el estado del análisis,
- realice una de las siguientes acciones en el panel de análisis:
 - modificar la asignación de ensayos (en "Assignment")
 - modificar el sexo del paciente
 - añadir o modificar HPO
 - modificar los paneles de análisis
 - eliminar informe clínico

Si un usuario intenta realizar cualquiera de estas acciones, se le pedirá que desbloquee el análisis, si tiene permiso, o que se ponga en contacto con su administrador si no lo tiene. Un usuario que desee desbloquear un análisis debe tener permiso para "Unlock analyses".

13.1.1.3. Panel lateral para la gestión de análisis

Se puede acceder al panel lateral de análisis, también conocido como "Analysis Drawer", desde cualquier pestaña de análisis.



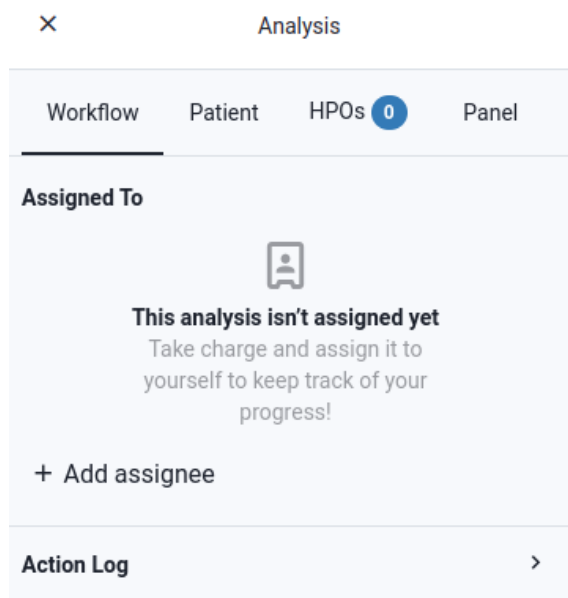
Haciendo clic en el botón  situado en la esquina superior derecha del banner de pestañas de análisis, el usuario puede ver y gestionar directamente los datos de análisis y las OPH de los pacientes sin salir de la pantalla actual.

El cajón de análisis contiene 4 vistas:

- **Workflow** : permite organizar y rastrear el flujo de trabajo para su análisis.
- **Patient** : permite ver y modificar los metadatos del paciente
- **HPO**: permite añadir o modificar términos HPO en el análisis, sólo disponible para análisis en contexto constitucional.
- **Panel** : permite aplicar una lista de genes al análisis y filtrar los resultados del análisis en función de esta lista.

Vista modal Workflow

Esta vista permite organizar y seguir los flujos de trabajo del usuario para el análisis que se está interpretando.



Assigned To. Esta vista muestra las tareas asignadas para el análisis actual. Sólo los administradores o usuarios con permiso para "Manage assignments" pueden asignar tareas (véase la sección **Gestión administrativa - presentación de la interfaz**).

Cuando se asignan tareas al análisis actual, se muestra una lista de todos los usuarios asignados, junto con la fecha y hora de finalización. Esta lista se actualiza automáticamente en cuanto un usuario completa una tarea.

Los usuarios sólo pueden editar sus propias tareas marcando o desmarcando la casilla correspondiente. Al pasar el ratón por encima de una tarea de sólo read, un tooltip indica "You can only complete yours".

Si no se ha asignado ninguna tarea, la vista muestra "This analysis isn't assigned yet".

Si un usuario no autorizado intenta asignar una tarea haciendo clic en el botón "+ Add assignee", el mensaje de error indica "You must have administrator privileges to assign analysis. If needed, please contact your platform administrator to request the appropriate access."

Action Log. El banner "Analysis Action Log (Audit Trail)" muestra el historial de todas las acciones realizadas sobre el estado HRD en SomaHRD, en filtros, perfiles de filtros y paneles de genes restringidos en la tabla de variantes, detallando el tipo de acción, el usuario, la marca de tiempo y una descripción de la acción.

Las acciones más recientes aparecen al principio de la lista. La paginación permite a los usuarios navegar eficazmente por el historial de acciones. El historial completo puede descargarse en formato .csv mediante el icono "Download Log".

Vista modal Patient

Esta vista le permite ver y modificar los metadatos del paciente, como el sexo (Femenino, Masculino o Desconocido), el GnomAD genetic ancestry group (solo Germline) o la Indicación (solo Somatic).

La visualización y edición de la indicación sólo está disponible para los análisis somáticos:

- Si la indicación se introdujo antes de iniciar el análisis (véase **Metadatos de la muestra**), se define por defecto para la muestra, se marca con una estrella y aparece en primera posición en la lista.
- Si se cambia la indicación durante la interpretación, un mensaje de advertencia indica que la indicación seleccionada es diferente de la indicación original y se actualizará la coloración de las pastillas en la subcolumna "A" de la columna *Clinical Evidences* (véase **Seleccionar una indicación**).
- Si no hay ninguna indicación disponible para el análisis, aparecerá el mensaje "No available indication".

La visualización y edición del "GnomAD genetic ancestry group" solo está disponible para análisis de germline. El valor predeterminado está vacío y se trata como la población GnomAD combinada.

Las modificaciones realizadas en esta vista deben guardarse o cancelarse haciendo clic en los iconos "Save modifications/Cancel" correspondientes, situados en la parte inferior del panel, para poder cambiar de pestaña.

The screenshot displays the 'Patient' modal view within an 'Analysis' window. At the top, there are tabs for 'Workflow', 'Patient' (selected), 'HPOs', and 'Panel'. Below the tabs, a blue header bar indicates 'Active sample: sample2'. The 'Sex' section has three buttons: 'Female', 'Male', and 'Unknown'. The 'Indication' section features a dropdown menu currently showing 'Ovarian endometrial cancer (Female reprod...'. A red warning message states 'Unsaved changes detected. Save now to avoid losing your updates.' Below this are 'Cancel' and 'Save modifications' buttons. The 'GnomAD genetic ancestry group' section shows a list of ancestry groups: 'African/African American', 'Amish', 'Ashkenazi Jewish', 'Admixed American', 'East Asian', and 'European (Finnish)'.

Vista modal HPOs

Esta vista, disponible sólo en proyectos Germline, permite ver, editar y asignar términos HPO relativos a la muestra, ya sea manualmente o a partir de una descripción clínica del paciente.

El usuario puede visualizar la lista de términos HPOs ya asociados a la muestra antes del análisis. En ausencia de códigos HPOs, aparece el mensaje "No HPOs yet".

Para añadir términos manualmente, el usuario puede introducirlos desde la barra de búsqueda "Enter phenotypes one by one", buscando un fenotipo o un código HPO. También puede importar un archivo de texto que contenga códigos HPO haciendo clic en el icono  (véase también la sección **Metadatos de muestra**).

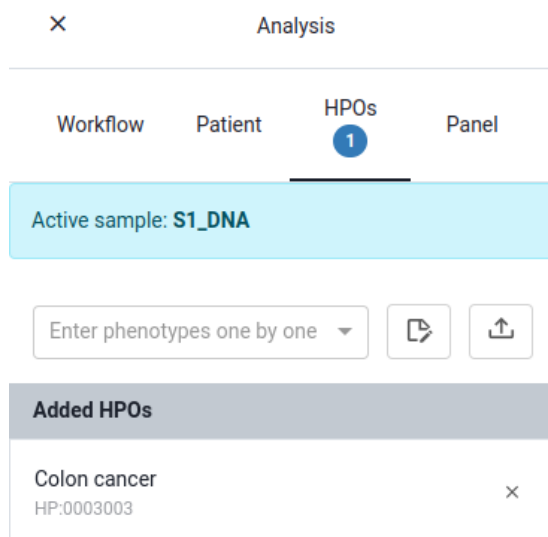


Figura 8: Vista de las OPH del panel de análisis

El usuario puede generar términos HPO a partir de la descripción clínica del paciente utilizando

DiagAI. Haciendo clic en el icono , aparece una ventana de diálogo en la que se pide al usuario que describa el perfil clínico del paciente (esta descripción está limitada a 4000 caracteres) y, a continuación, haga clic en "Extract HPOs with DiagAI" para generar una lista de los HPO encontrados. Los términos HPO aparecen resaltados en color y el usuario puede asociarlos al paciente pasando el ratón por encima de ellos y haciendo clic en "Add", o añadir todos los términos HPO encontrados haciendo clic en "Add all" (véase también la sección **Editar los metadatos de la muestra desde el panel de análisis**).

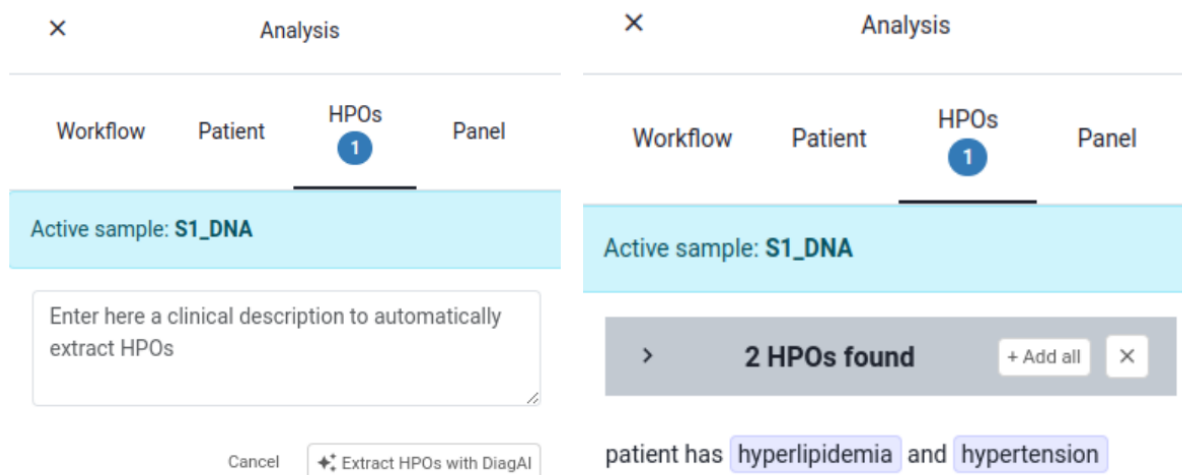
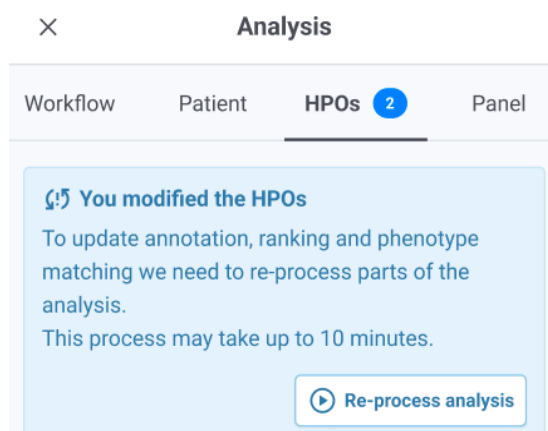


Figura 9: Extracción de términos HPO con DiagAI desde la vista HPO en el cajón de análisis

Cuando se añaden o modifican HPO desde esta vista, aparece un banner azul, sugiriendo que se vuelva a procesar el análisis con los nuevos HPO.

Seleccionando "Re-process analysis", el usuario puede actualizar la puntuación DiagAI y la clasificación de variantes (Short list o SmartPick) en función de los nuevos HPO introducidos.

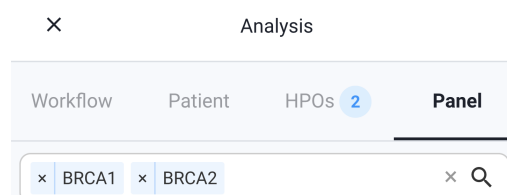
Este paso de reprocesamiento congela la tabla de variantes durante varios minutos para actualizar la clasificación de variantes y la columna DiagAI.



Vista modal Panel

Esta vista está disponible para todos los worksets excepto para los worksets de análisis CNV (CNVExome, CNVCapture y SomaCNVCapture).

Da acceso a la función *Restricted panels*, que permite restringir y bloquear el análisis a los genes seleccionados, a la manera de un panel *in silico*.

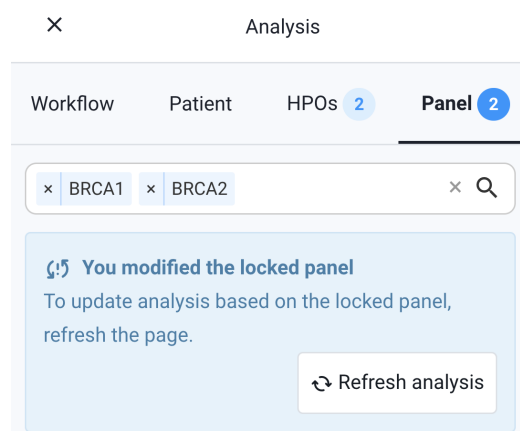


Para ello, el usuario puede seleccionar una o varias listas de genes a aplicar, mediante un menú desplegable, de entre las importadas previamente (véase el apartado **Genes** , en la sección **Gestión de la configuración del usuario - definiciones**). Al guardar los paneles seleccionados mediante "Save modifications", se aplica un filtro a todo el análisis en función de lo(s) panel(es) de genes seleccionada(s).

Los paneles restringidos pueden seleccionarse en cuanto se inicia el análisis (véase la sección **Iniciar un nuevo análisis**) o desde la vista Panel después del análisis.

Cuando se añaden o modifican paneles de genes, un banner azul advierte al usuario del cambio ("You modified the locked panel") y sugiere actualizar el análisis con los nuevos paneles mediante la función "Refresh analysis".

Al hacer clic en "Refresh analysis", el usuario elige actualizar la tabla de variantes, la pestaña CNVs (si procede) y la pestaña GeneCov de acuerdo con el nuevo panel.



13.1.1.4. Panel lateral para gestionar las evaluaciones

Se puede acceder al panel  **Selection** desde cualquier pestaña de un análisis. Permite al usuario ver las pequeñas variantes y CNV anotadas en VKB para el análisis actual, para generar un archivo JSON ("Generate file") y editar un informe clínico ("Generate report").

Las CNVs validadas en VKB también están disponibles en esta sección, sólo en el caso de CNVs procedentes de análisis de exoma, genoma y CGH Microarray.

Las variantes pequeñas validadas en VKB se diferencian por el logotipo , mientras que las CNV se muestran con el logotipo  y las palabras "gain" (ganancia) o "loss" (pérdida).

×

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. Resumen de los resultados del análisis en la pestaña Dashboard

Para algunos workset, los usuarios pueden acceder a una pestaña "Dashboard" que resume todos los resultados y métricas de control de calidad y que aparece por defecto al abrir un análisis. La pestaña "Dashboard" proporciona una visión general de los resultados del análisis a través de múltiples secciones. Las secciones disponibles varían según el tipo de workset. Esta pestaña está disponible para SomaCGP, SomaVar Tertiary y SomaLBx.

Patient Esta sección resume los metadatos de la muestra: nombre de la muestra (*ID de la muestra*), sexo e indicación de la patología (*Indicación*).

Quality Controls. Esta sección proporciona información sobre varias métricas clave relacionadas con la calidad de las muestras y la secuenciación:

- Pureza del tumor,
- Ploidía de la muestra,
- el porcentaje de bases cuya calidad es superior o igual a 30 (% Q30 bases),
- cobertura media de las regiones de interés (cobertura media ROI) antes y después del procesamiento UMI,
- porcentaje de regiones de interés con cobertura superior a 200X (ROI > 200X) antes y después del procesamiento UMI.

Signatures. La sección sobre firmas muestra las cifras clave de los análisis TMB, MSI y HRD:

- Cálculo de la carga mutacional tumoral (TMB): número de mutaciones/Mb.
- Inestabilidad de microsatélites: porcentaje de sitios inestables.
- Estado HRD (deficiencia de recombinación homóloga)..

Los estados de las puntuaciones TMB y MSI pueden ser seleccionados por el usuario a través de los menús desplegables ubicados debajo de cada puntuación.

Selection. Esta sección muestra las variantes evaluadas en la base de datos VKB para el análisis actual.

Suggested by SeqOne. Las alteraciones genómicas detectadas en el análisis se presentan como una lista reducida de variantes clínicamente relevantes. Las variantes se seleccionan si su clase ComPerMed es 5, o si tienen una afirmación Tier I correspondiente a la indicación del paciente, o una afirmación Tier II para terapias experimentales relacionadas con esta indicación, o una afirmación Tier I para otra indicación. El usuario puede agregar variantes de esta categoría a la sección Selection haciendo clic en "Add".

El archivo JSON se puede generar haciendo clic en "Generate JSON file" y contiene toda la información de la página "Dashboard".

13.1.3. Cobertura detallada por exón

En la pestaña *Gene coverage*, los usuarios pueden ver la tasa de cobertura de los genes incluidos en el manifiesto de análisis para cada exón, ya sea un panel de genes o un exoma.

Metodología

Este módulo interseca el manifiesto del usuario con las anotaciones RefSeq para todos los genes incluidos en el panel o el exoma, manteniendo 10bp a cada lado para cubrir las uniones de splicing, luego procede a calcular la profundidad de cobertura para cada base incluida en esta intersección utilizando la herramienta mosdepth.

Se establecen diferentes umbrales para asignar a cada exón un estado (covered, warning, failed) en función del porcentaje de bases por debajo de estos umbrales de cobertura.

El estado se define a partir de la cobertura de bases en el exón, incluidos los 10 pb adyacentes. Si se utilizan UMI, la cobertura se calcula después de la deduplicación.

Los umbrales aplicables a los distintos tipos de análisis se detallan en el cuadro siguiente*:

	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%				

Tabla 3: Umbrales de cobertura para el módulo Gene Coverage según el tipo de análisis.

Si el usuario ha predefinido métricas de QC personalizadas a partir de los parámetros del proyecto, éstas se aplicarán en la pestaña GeneCov (véase la sección **Configurar las métricas QC del proyecto por defecto).*

Visualización e interpretación

Cada gen puede visualizarse en una representación gráfica, y los resultados de Gene Coverage pueden filtrarse en función de los genes, los paneles in silico y las listas de transcritos. Los exones seleccionados se representan en verde (*Covered*), naranja (*Warning*) o rojo (*Failed*) según los umbrales descritos anteriormente.

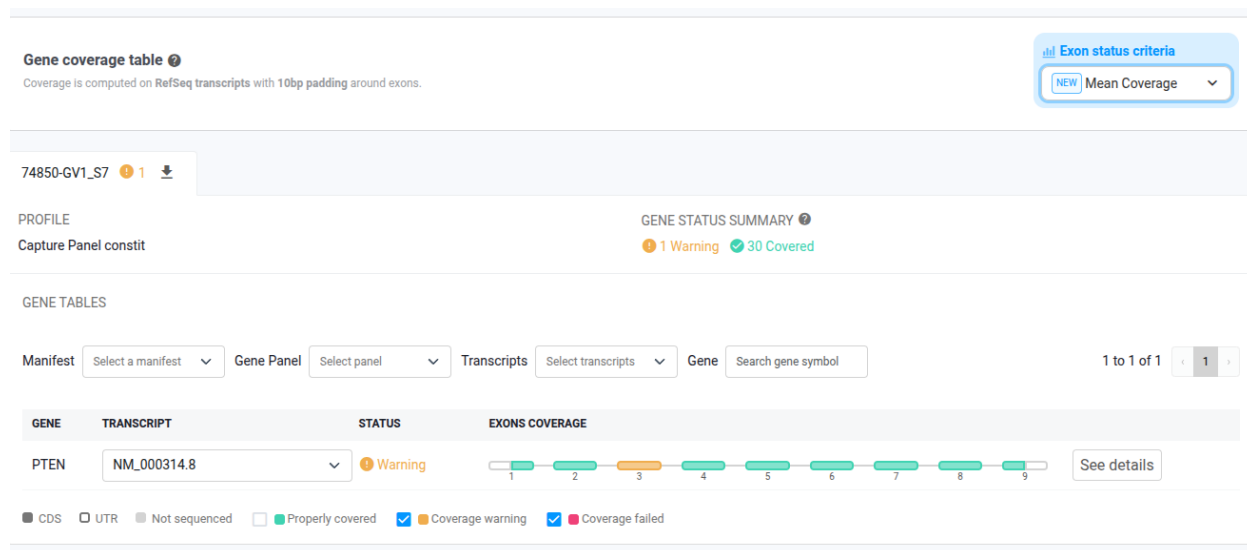


Figura 10: Detalles de la pestaña GeneCov

Opciones de filtrado

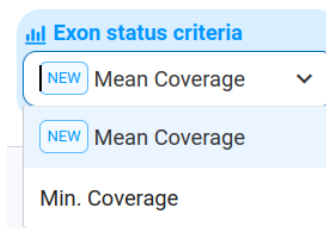
Un menú de filtros te permite:

- restringir la visualización a un subconjunto de regiones mediante un archivo .bed previamente importado a la interfaz, a través del filtro "Manifest",
- aplicar una lista de genes mediante el filtro "Gene Panel",
- aplicar una lista de transcripciones mediante el filtro "Transcripts",
- buscar un gen específico a través del filtro "Gene".

Métodos de cálculo de la cobertura

El usuario puede seleccionar el método de cálculo de la cobertura que se aplicará utilizando el menú desplegable "Exon status criteria" situado en la parte superior derecha de la página:

1. **Mean Coverage** (por defecto): Utiliza la profundidad de cobertura media en todo el exón para determinar el estado de cobertura.
2. **Minimum Coverage** : Utiliza la fracción mínima de bases que cumplen un umbral de cobertura específico. Esto garantiza que un determinado porcentaje de bases tenga al menos la profundidad de cobertura mínima requerida.



El menú desplegable "**Gene status criteria**" permite a los usuarios aplicar estos métodos de cálculo a nivel de gen:

- **Mean coverage (default):** Utiliza la profundidad de cobertura media en todos los exones del gen para determinar el estado de la cobertura.
- **Worst exon status:** Determina el estado de cobertura del gen basándose en el exón con peor rendimiento. Si algún exón no alcanza el umbral de cobertura, el gen se considera no cubierto. Este era el único método admitido en versiones anteriores.

Datos mostrados

Los exones no cubiertos o parcialmente cubiertos por el diseño (solapamiento < 80% entre el intervalo manifiesto y los límites del exón RefSeq) se consideran no seleccionados. Estos exones se representan en gris (*Not Sequenced*) en el gráfico.

Una vista detallada de cada gen permite visualizar las métricas de cobertura de los exones haciendo clic en "*See details*": cobertura media, mediana, porcentaje de cada región cubierta a 10x, 30x, 50x, 100x, 200x, 300x y 500x.

Para cada exón, una ventana IGV permite a los usuarios ver la cobertura de la secuencia y la alineación con mayor detalle.

Exportación de datos GeneCov

Los datos completos del análisis de cobertura están disponibles como archivo `sample_genecov.v2.tsv.gz`, accesible desde la pestaña Archivos del análisis.

Desde la pestaña GeneCov, el usuario puede exportar el análisis de cobertura según el método de cálculo de cobertura (Cobertura Media o Cobertura Mínima), a través del icono



situado junto al nombre de la muestra en la esquina superior izquierda de la pestaña de la muestra.

El usuario puede elegir exportar los datos según el estado de la cobertura génica (Failed, Warning, Covered) y exportar sólo la transcripción canónica (Canonical) o todas las transcripciones disponibles (All) antes de hacer clic en Download para obtener el archivo deseado.

Download genes from S1_DNA as .tsv file

Genes status

☒ Failed ☒ Warnings ☐ Covered

Transcripts

☒ Canonical ☐ All

Close

Download

Los datos se exportan como un archivo .tsv que contiene las siguientes columnas:

- Gen (gene_symbol),
- Número de gen según HGNC (gene_id),
- Transcripción canónica (is_canonical_transcript),
- Estado del expediente académico (transcript_status)
- Exón (exon_number)
- Tipo de exón CDS o UTR (exon_type)
- Exón cubierto o no cubierto (exon_status)
- Región cubierta o no por el manifiesto (objetivo)
- Posición genómica (interval_id)
- Número cromosómico (chr)
- Posición inicial (start)
- Posición final (end)
- Cadena (strand) directo o inverso
- % de cobertura a 1, 10, 20, 30, 50, 100, 200, 300, 500 y 1000X (cov_)
- Cobertura media (cov_avg)
- Cobertura mediana (cov_med)

13.1.4. Pequeñas variantes

Metodología

La detección de pequeñas variantes se realiza mediante una combinación de varias herramientas, que pueden variar de un workset a otro.

Ya sea para análisis de exoma o de panel de genes, la llamada de variantes viene determinada por el manifiesto (en formato BED) asociado al proyecto, incluidos los intervalos de las regiones de interés y el relleno aplicado a estas regiones (véase la sección **Crear un proyecto**).

Además, las variantes se llaman teniendo en cuenta la calidad de alineación de las reads en el genoma humano (versión GRCh37 o GRCh38). Las llamadas reads "MAPQ0" con una calidad de alineación (mapeo) igual a 0, debido a que se alinean en varios sitios del genoma, no se tienen en cuenta a la hora de llamar variantes para los workset de análisis somático. En los análisis "Germline", las reads MAPQ0 se tienen en cuenta para la llamada de variantes si se introduce un manifiesto que especifique regiones de alineación de baja calidad cuando se crea el proyecto en los metadatos Low-Quality Mapping Region. Las variantes detectadas a partir de reads MAPQ0 se identifican en la tabla de variantes mediante un tooltip "_mapq0" visible en la columna CAL (Caller).

SNVs / indels. Para la detección de SNVs e indels :

- *Freebayes*¹ se utiliza como único llamador de variantes en los worksets GermlineVar (exoma), GermlineFamily (exoma), SomaRNA y SomaVar LF. Los umbrales de detección más bajos son los siguientes:
 - VAF $\geq$ 10% en GermlineVar para el exoma,
 - VAF $\geq$ 5% en GermlineVar para paneles de genes,
 - VAF $\geq$ 1% en SomaRNA,
 - VAF $\geq$ 0,1% en SomaVar LF.
- Se acompaña de *GATK HaplotypeCaller*² en el workset GermlineVar (panel).
- Su homólogo somático, *Mutect2*³, se utiliza junto con Freebayes en el workset SomaVar, SomaHemato y SomaCGP. El umbral mínimo de detección es VAF $\geq$ 1% para estas dos llamadas en SomaVar.
- *Sentieon*⁴ se utiliza como llamada principal en los workset GermVar, GermVar Family y SomaLBx.
- *Clair3*⁵ se utiliza como único llamador de variantes en el 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>

Los análisis GermVar y GermVar Family se basan en los componentes bioinformáticos Sentieon para los pasos clave del análisis secundario. En particular:

- La alineación se realiza utilizando la implementación optimizada de BWA-MEM2 de Sentieon, y se marcan las lecturas duplicadas.
- La llamada de línea germinal (SNVs/Indels) se realiza utilizando SentieonDNAScope⁶.
- Sentieon TNScope⁷ se puede utilizar para detectar variantes en mosaico de bajo VAF (dependiendo de la configuración de GermVar).
- Alignment is performed using Sentieon's optimized implementation of BWA-MEM2, and duplicate reads are marked.

Nota sobre las regiones de baja mapeabilidad (MAPQ0): en regiones altamente homólogas (p. ej., SMN1/SMN2), las variantes pueden ser filtradas por el filtrado de aprendizaje automático del llamador de variantes. Cuando es necesario, SeqOne proporciona un módulo MAPQ0 opcional (módulo de regiones de baja calidad) basado en Freebayes con umbrales adaptados para aumentar la sensibilidad en regiones definidas por el usuario.

Los análisis SomaLBx se basan en los componentes bioinformáticos Sentieon para los pasos clave del análisis secundario de biopsia líquida. En particular:

- La alineación de lecturas se realiza con Sentieon (BWA-MEM), y el pipeline incluye el procesamiento/consenso de UMI basado en Sentieon para soportar flujos de trabajo de ctDNA de alta sensibilidad.
- La llamada somática de variantes cortas (SNVs/Indels) se realiza utilizando Sentieon TNScope⁷ en modo solo tumor (tumor-only), con soporte de fracción alélica muy baja (VAF $\geq 0.1\%$ dependiendo de la configuración). Nota: los genotipos de variantes se reportan como heterocigotos por defecto en los resultados SNV/indel de SomaLBx porque Sentieon TNScope⁷ no realiza predicción de genotipo.

SVs - Structural Variants. Las inserciones y deleciones cuya longitud es cercana o superior a la de una read de secuenciación, se detectan utilizando el SV caller GRIDSS⁸ en los workset GermlineVar (panel/exoma), SomaVar, SomaVar LF, SomaHemato y SomaCGP. Estos eventos se detectan utilizando el SV caller Delly⁹ en los worksets GermVar y GermVar Family. Solo se detectan inserciones y deleciones inferiores a 300 pb en los worksets somáticos, mientras que se aplica un límite de 10 kb en el workset GermlineVar (para versiones ≥ 2.0) y en los worksets GermVar y GermVar Family (versiones ≥ 3.5). Las variantes detectadas por GRIDSS se detectan sea cual sea el 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 módulo de detección específico para duplicaciones internas en tándem del gen FLT3 (FLT3-ITD) sólo está disponible en los worksets SomaVar y SomaHemato, a partir de la versión v2.2. Este módulo se basa en la herramienta *FILT3r*¹⁰.

Alu Insertions. Por último, SeqOne también ha desarrollado un módulo para detectar inserciones de elementos transponibles (*AluMEI*), como los elementos Alu. Lo utilizan los worksets GermlineVar (panel/exoma), GermlineFamily (panel/exoma), GermVar (panel/exoma/genoma), GermVar Family (panel/exoma/genoma), SomaHemato y SomaVar. No se aplica un umbral VAF inferior para las variantes detectadas por *AluMEI*.

Las variantes combinadas de *Freebayes*, *HaplotypeCaller* (o *Mutect2*) y *FILT3r* caller están disponibles en el archivo *final.vcf.gz* del análisis. Las variantes específicas de *GRIDSS*, *FILT3r* o *AluMEI* están disponibles en sus respectivos archivos VCF (***gridss.vcf.gz***, ***filt3r.vcf.gz*** y ***variants.alumei.vcf.gz***), también disponibles para su descarga en la pestaña *Archivos* del análisis.

En el caso concreto de *FILT3r* :

- Las variantes de *FILT3r* comunes a *Freebayes* y *Mutect2* están disponibles en el VCF final del análisis (***final.vcf.gz***),
- Las variantes exclusivas de *FILT3r*, así como las variantes comunes a *FILT3r* y *GRIDSS*, se enumeran en el VCF de *FILT3r* (***filt3r.vcf.gz***).

Dos módulos adicionales desarrollados por SeqOne permiten a continuación :

- Retención de variantes multinucleótidas, o MNVs, codificantes y restringidas a un único codón durante el proceso de descomposición de variantes tipo delins.
- Recalibración de la frecuencia alélica de las variantes indels y delins detectadas por *Freebayes* y *GATK* (herramienta *Kraf*).

Estos dos módulos sólo afectan a los worksets GermlineVar (panel/exoma), SomaVar, SomaRNA y SomaVar LF.

Kraf utiliza una única estrategia k-mers, sin alineamiento. La frecuencia alélica recalibrada de indels y delins sólo se conserva si es superior a la frecuencia alélica inicial de la variante.

A continuación, las variantes se anotan principalmente utilizando la herramienta *Variant Effect Predictor*¹¹ (VEP), y se añade información adicional de las siguientes bases de datos:

- ClinVar
- OMIM (versión comercial) + PanelApp Australia + PanelApp Inglaterra
- CKB
- dbSNP
- COSMIC (versión comercial)

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

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

- GnomAD (datos exoma + genoma)
- Dfam
- DGV
- LOVD

Además, se dispone de las siguientes puntuaciones *in silico*:

- Puntuaciones de splicing (*MaxEntScan, GeneSplicer, dbSCSNV*)
- Puntuación de conservación (*GERP++, PhastCons100way, PhastCons20way, PhyloP100way Vertebrados, PhyloP20way Mamíferos, SiPhy*)
- Predicciones funcionales (*REVEL, DANN, FATHMM, FATHMM-MKL, LRT, MetaLR, MetaSVM, MutationAssessor, MutationTaster 2, PROVEAN, SIFT, PLI*)

Visualización e interpretación

La pestaña Variantes  permite visualizar, filtrar y anotar las variantes obtenidas durante un análisis. Éstas se muestran en forma de tabla, en la que cada fila corresponde a una variante diferente. El número de variantes seleccionadas tras el filtrado (variantes seleccionadas) y el número total de variantes detectadas (número total de variantes) se muestran en la esquina superior derecha de la cabecera de la tabla de variantes mediante el siguiente icono:  9/22 . Una paginación en la parte inferior de la página permite navegar de una página de resultados a la siguiente. Si el análisis utiliza un panel restringido, la cabecera de la tabla también incluirá el número total de variantes en el panel restringido (*total number of variant in the restricted panel*) .

Esta navegación se facilita mediante atajos de teclado que utilizan las teclas ← y → del teclado.

Por defecto, en la tabla aparecen varias columnas descriptivas. Los usuarios pueden exportar la tabla como archivo TSV haciendo clic en  .

Puede haber varios transcritos disponibles para cada variante, y su número se muestra a la izquierda de cada fila. Las variantes se anotan en los transcritos RefSeq, a excepción de las variantes mitocondriales que se anotan en *Ensembl*.

Las transcripciones de MANE Select y MANE Clinical Plus se

identifican mediante un icono,  o  , respectivamente, acompañado de un tooltip.

TRANSCRIPT	NCVSC
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

Para una misma variante, excepto para las variantes mitocondriales, los transcritos se clasifican según el orden que se indica a continuación:

- el impacto ['high' > 'moderate' > 'low' > 'modifier']
- tipo de transcripción ['NM', 'XM', 'NR', 'XR'] (para transcripciones RefSeq)
- Tipo de transcripción MANE [Mane Select > MANE Clinical Plus > no MANE].
- orden alfanumérico

Para las variantes mitocondriales, los transcritos se clasifican en el siguiente orden:

- impacto ['high' > 'moderate' > 'low' > 'modifier']
- si el impacto es "modifier" :
 - el efecto "non_coding_transcript_exon_variant"
 - el biotipo [codificación de proteínas > Mt_rRNA > Mt_tRNA]
- orden alfanumérico

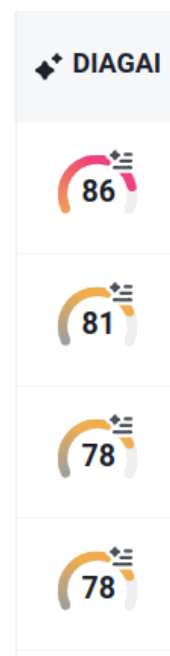
DiagAI. Esta columna muestra la puntuación de 0 a 100 basada en el conjunto DiagAI mediante aprendizaje automático. La puntuación DiagAI se basa en cuatro componentes:

- El algoritmo UP², que predice la patogenicidad de la variante
- La matriz PhenoGenius, que evalúa la concordancia entre el gen y los HPO del paciente (si se han introducido durante el análisis).
- La concordancia entre el genotipo observado y el modo de transmisión de la enfermedad descrito en las bases de datos GenCC, OMIM y PanelApp (Australia y England).
- La calidad de la variante

Cuanto más se acerque la puntuación a 100, más se predice que la variante es patogénica y de interés diagnóstico. Cuanto más se acerque a 0, más benigna se predice que es la variante.

Además, dos iconos llaman la atención sobre determinadas variantes de interés:

- El icono  significa que la variante tiene una mayor probabilidad de ser la causa de la enfermedad (AI Smart Pick).
- El icono  indica que la variante está presente en la lista corta de variantes con una alta probabilidad de ser la causa de la enfermedad del paciente.



Los términos HPO pueden introducirse antes de iniciar el análisis editando los metadatos de la muestra (véase la sección **Metadatos de muestra**). También pueden editarse durante la interpretación utilizando la vista modal HPOs en el panel lateral del análisis (véase la sección **Vista modal HPOs**). En este caso, el usuario puede reprocesar la puntuación DiagAI y la clasificación de variantes (Short list o SmartPick) de acuerdo con los nuevos HPOs introducidos, utilizando la función "Re-process with new HPOs".

Si el usuario hace clic en el icono de puntuación DiagAI, la página de la variante se abre directamente en la pestaña DiagAI que detalla la puntuación (véase más adelante la sección DiagAI en la **página Variante**).

ACMG. La clasificación de patogenicidad de cada variante se calcula como parte de un proceso semiautomatizado de acuerdo con las recomendaciones emitidas *por el American College of Medical Genetics and Genomics*. Los usuarios pueden revisar la clasificación en una pestaña específica de la página de variantes, que luego se utiliza para generar la evaluación final de una variante.

ComPerMed. La clasificación de patogenicidad de cada variante se calcula como parte de un proceso semiautomatizado de acuerdo con las recomendaciones emitidas por el comité de expertos belga ComPerMed (Comisión de Medicina Personalizada). Los usuarios pueden acceder a información detallada sobre la clasificación en una pestaña específica de la página de variantes.

AMP. (Análisis SomaVar versión v1.8 o posterior, sólo SomaVar LF o SomaCGP). Esta columna se muestra por defecto en los análisis que contienen anotaciones CKB y muestra el nivel de AMP de cada variante cuando está disponible (Nivel I o II). El nivel de AMP mostrado se basa en la mejor evidencia clínica de todos los tipos, independientemente de la indicación seleccionada.

VKB. Las evaluaciones realizadas por el usuario durante análisis anteriores se almacenan en la base de conocimientos del usuario. Esta evaluación aparece en la columna VKB de la tabla para cada nueva aparición de la variante. Un tooltip muestra rápidamente la clase y el comentario asociados a la variante evaluada o previamente evaluada.

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

Gene. Denominación de genes según [HGNC](#) y bandera OMIM asociada :

- Bandera roja: " OMIM gene **with** phenotypes **and** clinical characteristics",
- Bandera de contorno roja: " OMIM gene **with** phenotypes but **no** clinical features ",
- Punto rojo: " OMIM gene **without** phenotypes or clinical features ".

GENE	TRANSC
RPE65	NM_000...

Los genes que se sabe que son genes de Imprinting según PanelApp

(Australia) se resaltan con el icono . El tooltip asociado indica si el gen es de expresión materna, paterna, desconocida o si se hace referencia a varias improntas.

Haciendo clic en la bandera situada junto al nombre del gen se accede directamente a la página OMIM. Al hacer clic en el nombre del gen, aparece un menú desplegable con enlaces a las

bases de datos OMIM¹² , Decipher¹³ , Orphanet¹⁴ , MedGen¹⁵ y GenCC¹⁶ . Cada enlace lleva directamente a la página correspondiente al gen seleccionado.

La información relativa a OMIM y a las bases de datos sólo está disponible para los análisis de la línea germinal.

HGVSC. Esta columna proporciona la anotación HGVS de la variante en el ADN complementario (ADNc) y en la proteína.

HGVSC

c.10580G>A
p.R3527Q

Effect. Esta columna describe el impacto de la variante en la proteína y el efecto en la transcripción.

El impacto se calcula según 4 niveles:

- **High** (alto impacto, en rojo): la variante tiene un alto impacto disruptivo en la proteína, probablemente causando truncamiento de la proteína, pérdida de función o desencadenando su decaimiento mediado por sinsentido.
- **Moderate** (impacto moderado, en amarillo): variante no disruptiva que puede modificar la eficacia de la proteína.
- **Bajo** (impacto bajo, en gris): variante que se supone inofensiva o que es poco probable que modifique el comportamiento de la proteína.
- **Modifier** (impacto modificador, en azul): generalmente variantes no codificantes o que afectan a regiones o genes no codificantes, para las que las predicciones son difíciles o no hay pruebas de impacto

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

Para cada variante, pueden estar disponibles varios efectos sobre la transcripción en la columna Efecto. En este caso, sólo se muestra el efecto con la mayor consecuencia según VEP¹⁷ , seguido de un número (por ejemplo, +2) para los demás efectos disponibles. Los detalles de éstos se muestran en el tooltip correspondiente.

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/>

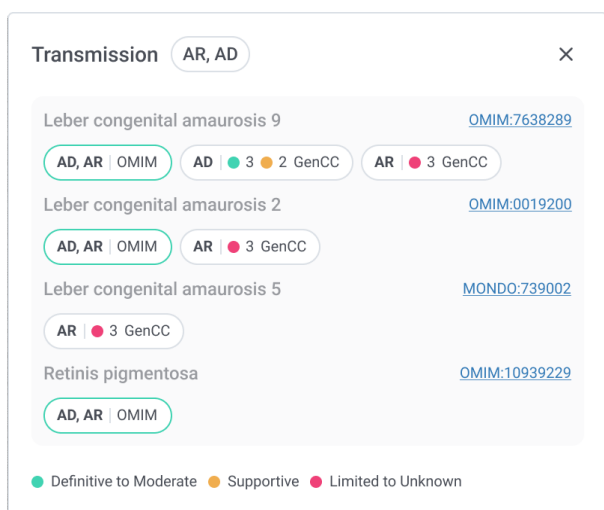
¹⁷ http://www.ensembl.org/info/genome/variation/prediction/predicted_data.html

Transmission. Esta columna indica el modo de transmisión de las enfermedades asociadas a genes notificados en las bases de datos GenCC y OMIM. Las insignias de colores son visibles para identificar rápidamente los patrones de transmisión basados en las predicciones de DiagAI:

- De Novo
- Compuesto Het - Verificado
- Hom recesivo
- Compuesto Het - Sospechoso

Las insignias De Novo, Compound Het - Verified y Recessive Hom sólo se muestran en los análisis familiares.

Se puede utilizar un filtro para seleccionar variantes con estos distintivos (véase **la sección Filtrar variantes - Transmission & Phenotypes**).



Al hacer clic en los enlaces de modos de transmisión en esta columna, se abrirá una ventana emergente donde los usuarios pueden encontrar enfermedades listadas con sus nombres y enlaces a bases de datos (OMIM, MONDO -si OMIM no está disponible- o GenCC).

Las enfermedades están organizadas en una jerarquía específica: aquellas que se encuentran tanto en OMIM como en GenCC aparecen primero, seguidas por las entradas exclusivas de OMIM, y luego las entradas exclusivas de GenCC.

Cada entrada incluye etiquetas codificadas por colores: las etiquetas OMIM tienen bordes verdes, mientras que las etiquetas GenCC utilizan un sistema de semáforo (verde/amarillo/rojo) para indicar la solidez de la evidencia:

- Verde: *Definitive + Strong + Moderate*
- Amarillo: *Supportive*
- Rojo: *Limited, Animal, Unknown*

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	

Clinical Evidences. (Sólo análisis SomaVar v1.8 o posterior, SomaVar LF o SomaCGP). Esta columna se muestra por defecto en los análisis que contienen anotaciones CKB. Se divide en tres subcolumnas, cada una de las cuales contiene el número total de un determinado tipo de prueba: A (Accionabilidad), D (Diagnóstico) y P (Pronóstico). La columna A también muestra el número de afirmaciones terapéuticas sensibles (izquierda) y resistentes (derecha) como una insignia.

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

Los valores de estas subcolumnas están codificados por colores:

- La resistencia, cuando se encuentra (> 0 afirmaciones), se muestra siempre en rojo independientemente de la indicación, y en gris en caso contrario.
- La sensibilidad se resalta en verde sólo cuando se selecciona la indicación correspondiente, y en gris en caso contrario.

Si al menos una prueba de este tipo coincide con la indicación seleccionada en la tabla de variantes (panel de indicaciones), aparecerá una marca en la información sobre herramientas delante de la etiqueta de la indicación. Si no se encuentra ninguna coincidencia, la etiqueta no aparecerá.

Phenotypes. Esta columna, disponible para los proyectos Germline, permite la asociación de un gen con términos HPO utilizando la matriz PhenoGenius¹⁸ . Si los términos HPO están asociados a la muestra, se muestra el mayor porcentaje de coincidencia entre el gen y el fenotipo, junto con la proporción de términos con los que está asociado el gen. Si ningún gen coincide con los HPO introducidos, la columna devuelve el mensaje "no match"; si la muestra no tiene términos HPO, la columna muestra el mensaje "see HPO". En ambos casos, los términos conocidos para el gen se proponen no obstante en la ventana emergente asociada.

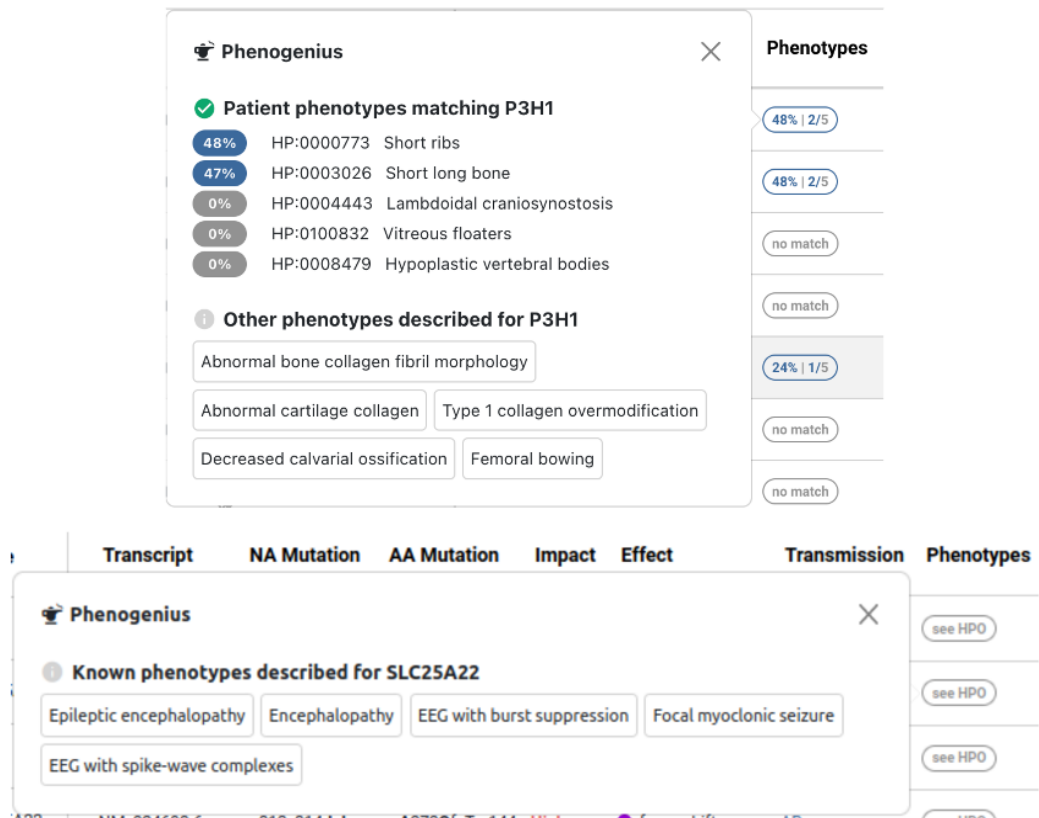
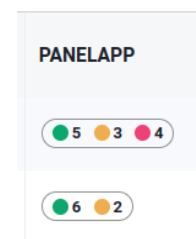


Figura 11 : Visualización de la información sobre Phenogenius

¹⁸ Kevin Yaay *et al.*, " Learning phenotypic patterns in genetic diseases by symptom interaction modeling ", *medRxiv*, p. 2022.07.29.22278181, janv. 2022, doi: 10.1101/2022.07.29.22278181.

PanelApp. Esta columna integra anotaciones de la base de datos de paneles de *PanelApp Australia*¹⁹ y *PanelApp Inglaterra*²⁰. Indica el número de paneles asociados al gen para el que se ha detectado una variante, según el nivel de clasificación descrito a continuación.

Si un gen no se encuentra en un panel, la columna mostrará "N/A".



La ventana emergente asociada muestra detalles de la asociación gen/fenotipo, así como el modo de transmisión asociado al fenotipo, el modo de transmisión asociado al fenotipo y la base de datos de la que procede el panel, PanelApp Australia (**AUS**) o PanelApp Inglaterra (**ENG**) (**Figura 12**).

A cada gen se le asignará uno o más paneles según un sistema de clasificación de tres niveles basado en la evidencia de la asociación del gen y el fenotipo: Evidencia alta en **verde**, evidencia moderada en **amarillo** y evidencia baja en **rojo**.

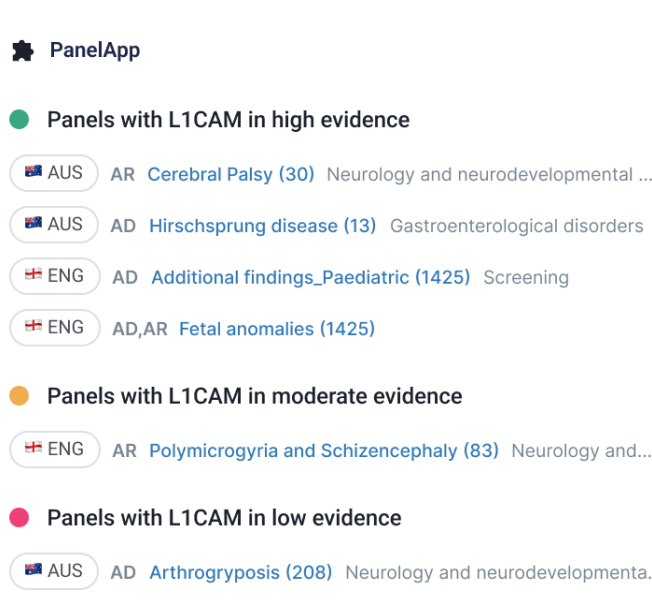
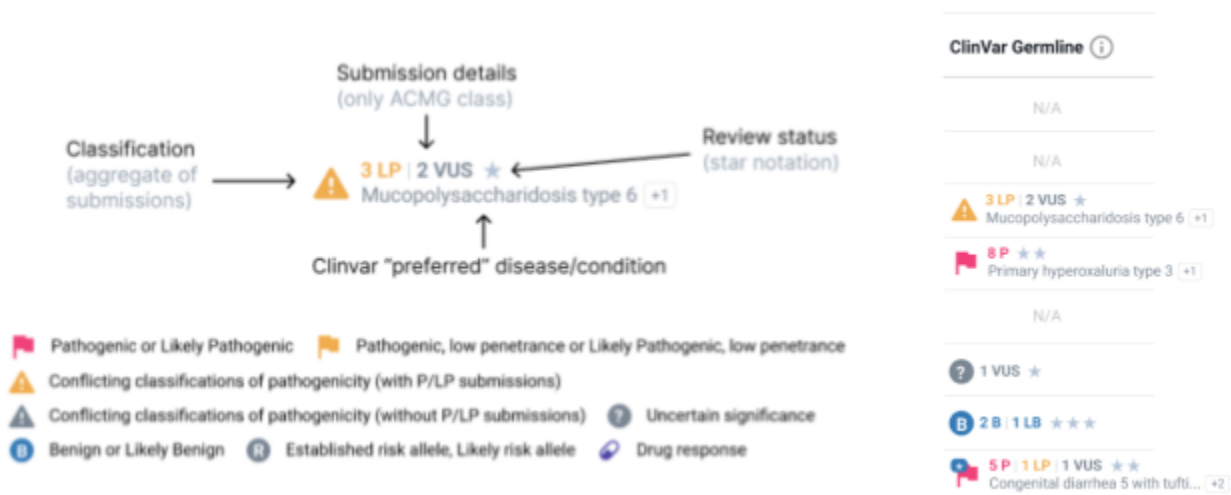


Figura 12: Visualización de la información de PanelApp.

¹⁹ <https://panelapp-aus.org>

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

ClinVar Germline. Esta columna muestra información extraída de ClinVar durante el análisis. Proporciona una serie de información, incluida la clase de patogenicidad, los detalles y el número de presentaciones para cada clase ACMG, el estado de revisión y la enfermedad preferida de ClinVar. Al hacer clic en la celda de la tabla de variantes, el usuario será redirigido a la página de ClinVar correspondiente.



Para más información, consulte https://www.ncbi.nlm.nih.gov/clinvar/docs/review_status/

GnomAD. Esta columna muestra 6 elementos de la base de datos GnomAD para una variante específica en la población: la frecuencia alélica (AF), el número de alelos (AC), el número de homocigotos (HOM), el número de hemocigotos (HEM), la frecuencia del alelo mayor (MAF) y para los proyectos GRCh37, la versión de GnomAD para la que se encuentra esta variante (VER).

Al hacer clic en el valor que aparece en la columna AF, el usuario es redirigido a la base de datos GnomAD en su versión v4 o, en su defecto, en su versión v2. Si una variante no figura en la base de datos, el valor mostrado en esta columna será N/A

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

conjuntos de datos infractores.

Se muestra una alerta en la columna VER si esta variante está presente en GnomAD, pero no pasa todos los filtros de calidad para al menos uno de los dos conjuntos de datos (exoma o genoma). En este caso, los valores mostrados se adaptan a los valores disponibles para esta variante, y se muestran los filtros que activaron la alerta, junto con un breve mensaje que menciona el conjunto o

Score UP². La puntuación de predicción funcional UP² (Universal Pathogenicity Prediction), que va de -1 a 1, clasifica la patogenicidad de las variantes utilizando un enfoque de aprendizaje automático según tres categorías : **P** (Patogénica : puntuación > 0,1), **VUS** (Variante de Significado Desconocido : puntuación ≥ -0,1 et ≤ 0,1) o **B** (Benigna : puntuación < -0,1). La puntuación UP² se muestra por defecto en la tabla de variantes, pero el usuario puede desactivar la columna UP² desde el panel de gestión de columnas haciendo clic en el botón



CI-SpliceAI. Esta columna incorpora puntuaciones de predicción calculadas mediante un enfoque de aprendizaje automático para identificar variantes con un impacto en el splicing alternativo a la manera de *CI-SpliceAI* .²¹

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

Las puntuaciones se calculan para un subconjunto de variantes que cumplen los siguientes criterios: variantes de splicing según VEP (splice_*) con una frecuencia GnomAD inferior a 10⁻³ y que cumplen los siguientes criterios de calidad (QA > 20, AO > 10, DP > 20).

Si se calcula, la columna proporciona 4 puntuaciones delta para la variante en cuestión, para la ganancia del aceptor (AG), la pérdida del aceptor (AL), la ganancia del donante (DG) y la pérdida del donante (DL).

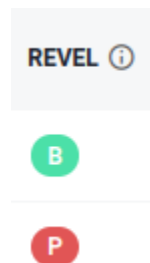
La puntuación delta de una variante, que va de 0 a 1, corresponde a la probabilidad de que la variante modifique el splicing. Se aplica un código de colores a los umbrales de la puntuación delta: **verde** (>0,2 ; recall alto), **amarillo** (>0,5; recomendado) y **rojo** (> 0,8; precisión alta). Si la puntuación delta es inferior a 0,2, la consideramos insignificante. La columna especifica entonces "No splicing impact".

Para cada variante, se proporciona un enlace a la correspondiente interfaz de búsqueda *SpliceAI*²².

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

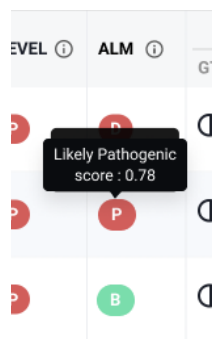
²² <https://spliceailookup.broadinstitute.org/>

REVEL. La puntuación de predicción REVEL se basa en un método global para predecir la patogenicidad de las variantes de sentido erróneo utilizando una combinación de puntuaciones de 13 herramientas individuales: MutPred, FATHMM v2.3, VEST 3.0, PolyPhen-2, SIFT, PROVEAN, MutationAssessor, MutationTaster, LRT, GERP++, SiPhy, phyloP, y phastCons.



La puntuación puede variar de 0 a 1 para una variante missense, reflejando las puntuaciones más altas una mayor probabilidad de que la variante cause la enfermedad. Los umbrales aplicados son los siguientes: Puntuación REVEL > 0,5 para **P** (Patogénico) y puntuación REVEL < 0,5 para: **B** (Benigno). El usuario puede visualizar la puntuación en la tabla de variantes marcando la casilla REVEL en la inserción Columnas, en la sección Predicción Funcional.

ALM. La puntuación de predicción de AlphaMissense²³ está disponible en la columna ALM. AlphaMissense es un modelo de predicción que evalúa la patogenicidad de las variantes missense mediante aprendizaje profundo. El modelo se basa en las predicciones estructurales de AlphaFold para mejorar su precisión.



AlphaMissense integra la conservación evolutiva y el contexto estructural para evaluar el impacto potencial de las variantes missense en la función de las proteínas.

La puntuación oscila entre 0 y 1, y los valores más altos sugieren una mayor probabilidad de que una variante sea deletérea. Los umbrales aplicados son: Puntuación ALM > 0,564 para **P** (Probable Patogénico) y puntuación ALM < 0,34 para **B** (Probable Benigno). En los demás casos (0,34 < puntuación ALM < 0,564), se etiquetó **VUS** (Ambiguo).

Sample Freq. Esta columna muestra para cada variante su frecuencia en el proyecto (subcolumna PROJECT) y en todos los proyectos de la entidad con metadatos similares (subcolumna ALL).

La frecuencia de muestras del proyecto (PROJECT) corresponde al número de muestras en las que se detecta la variante en relación con el número total de muestras analizadas en el proyecto.

El frec. de muestras ALL corresponde al porcentaje de muestras analizadas en las que se encuentra la variante en la entidad, dentro de proyectos de metadatos similares. Haciendo clic en el valor que aparece en la columna ALL, el usuario puede acceder a los nombres de los proyectos y al número de muestras en las que está presente la variante, así como a los detalles de todos los proyectos de metadatos similares

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

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

(opción "Show all projects").

Por defecto, la subcolumna ALL está oculta en la tabla de variantes, pero puede reactivarse

desde el panel de gestión de columnas haciendo clic en el icono  y activando el botón « *Sample Frequency Entity (Entity)* ».

Exon. Esta columna indica el número total de exones del gen según el transcrito en cuestión, así como el número afectado por la mutación, en forma de fracción. Un tooltip indica la distancia entre la variante y el sitio de splicing más cercano. Si la variante se encuentra aguas arriba del sitio de splicing, el valor será positivo y si se encuentra aguas abajo, será negativo.

GT (Genotype). Esta columna muestra el genotipo predicho por los programas de detección de variantes. Los distintos pictogramas posibles se describen más adelante en la sección "**Filtrar las variantes**":

En un análisis de trío (GermlineFamily), si se puede determinar el origen de una variante heterocigota, se indicará con el siguiente código de colores:

- **rosa** → Origen heterocigoto materno
- **azul** → Origen heterocigoto paterno
- **rojo** → Variante heterocigota *de novo*
- **naranja** → Violación mendeliana

No se aplica color si la transmisión es indeterminada (por ejemplo, si ambos progenitores son heterocigotos), o para variantes homocigotas.

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

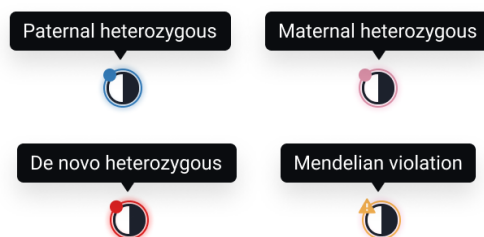


Figura 13: Código de colores aplicado a los genotipos heterocigotos según la transmisión GermlineFamily

El desfase de variantes está disponible para los análisis GermVar Tertiary y GermVar Tertiary Family que utilizan archivos VCF de secuenciadores de Oxford Nanopore Technologies. Cuando las variantes están en fase, los iconos de genotipo aparecen de la siguiente manera : . La marca de verificación indica que la variante está en fase. Las variantes en fase con la parte negra en el mismo lado y pertenecientes a la misma "Phase ID" están en fase, es decir, en la misma hebra.

Puede hacer clic en el genotipo para obtener una lista de variantes cis o trans, si están presentes. Sólo se presentan las 5 variantes mejor clasificadas por DiagAI, como se muestra al lado.

La lista completa de variantes está disponible haciendo clic en "Browse all variants".

Variants in *trans*
Phase set ID 283928

! Compound heterozygous candidates for

4 SLC25A22 chr11:g.792142C>T

5 SLC25A22 chr11:g.792146_792147del

GT

VAF (Variant Allele Frequency). Esta columna indica la frecuencia alélica variante, es decir, la proporción de reads que apoyan el alelo alternativo (AO) respecto al número total de reads en esa posición (COV), expresada en porcentaje.

Los análisis que utilizan los worksets SomaVar versión ≥ 2.5 , SomaCGP versión ≥ 0.7 y SomaVar LF versión ≥ 2.1 muestran la siguiente información adicional:

- un tooltip muestra los valores VAF, AO y COV calculados por los 2 *llamadores de variantes* Freebayes y Mutect2 que detectan la variante.
- un icono de advertencia  alerta al usuario de cualquier desacuerdo entre los 2 llamadores de variantes para variantes detectadas con un VAF bajo. Así, para las variantes que muestran un VAF inferior al 5% en la columna VAF, la advertencia se muestra en caso de una discrepancia VAF > 10% entre los VAF calculados por Freebayes y Mutect2.

AO (Alternate Observation). Esta columna indica el número de reads portadoras de la variante.

CAL (Caller). Esta columna indica el nombre de la *persona* o personas que han detectado la variante.

Cov (Coverage). Esta columna indica el número de reads totales en la posición de la variante.

SOR. El Strand Odds Ratio (SOR) es una métrica de sesgo de hebra derivada de la prueba de Odds Ratio Simétrico que mide el desequilibrio de las lecturas que respaldan la variante entre las hebras de avance (forward) y de retroceso (reverse). Un SOR más alto sugiere una mayor probabilidad de un artefacto, mientras que un SOR más bajo sugiere un soporte de hebra más equilibrado y es más consistente con una variante verdadera.

Clasificar las variantes

Existen varias opciones para ordenar las variantes en la pestaña Variantes. Están disponibles en un menú desplegable situado en la esquina superior derecha de la tabla. Para cada opción de ordenación, es posible ordenar en orden ascendente o descendente pulsando sobre el icono de ordenación . Por defecto, todas las opciones se ordenan en orden descendente, a excepción de la opción "Locus", que se ordena en orden ascendente.

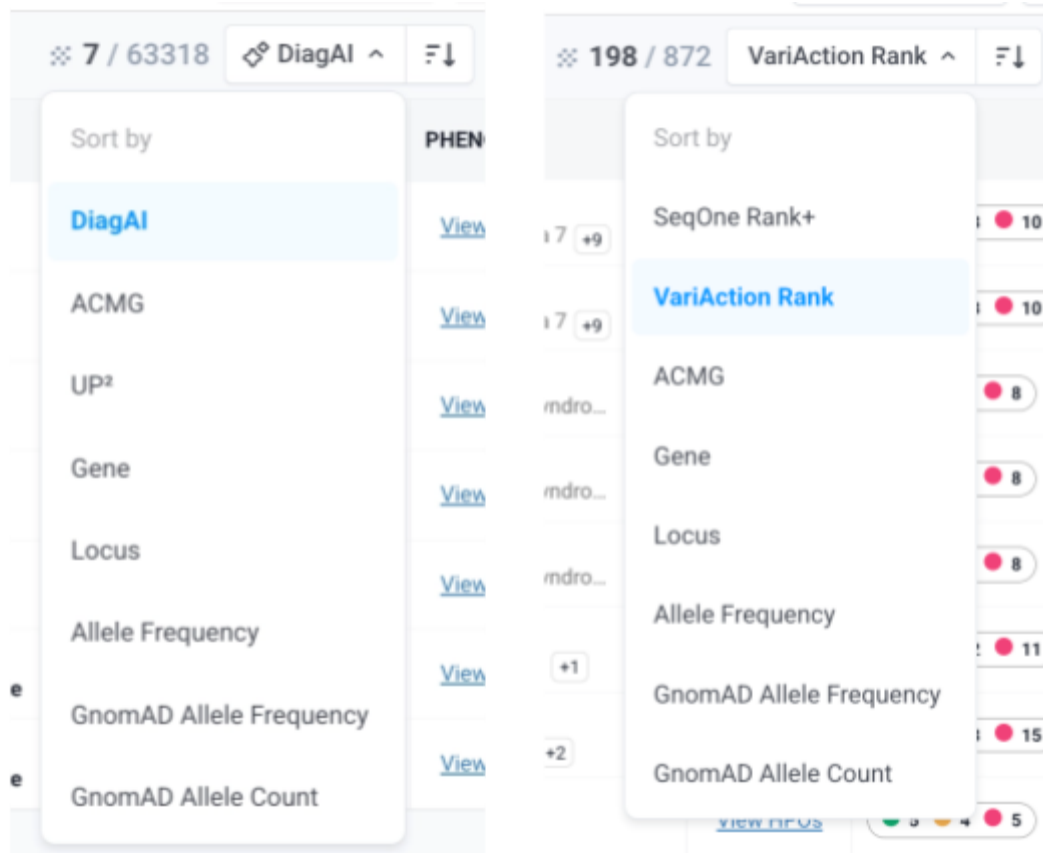


Figura 14: Ventana de selección de opciones de clasificación de variantes

DiagAI Score Esta opción de clasificación se aplica por defecto a todos los análisis de línea germinal y permite clasificar las variantes según su puntuación DiagAI (sustituye a SeqOne Rank+ en los análisis realizados después del 2 de octubre de 2024).

SeqOne Rank+ . El algoritmo de priorización SeqOne Rank+, basado en un modelo de inteligencia artificial más antiguo que DiagAI Score, sigue estando disponible y se aplica por defecto a todos los análisis realizados antes del 2 de octubre de 2024.

Se basa en 4 componentes principales:

- Una nueva implementación de la clasificación ACMG, basada en aprendizaje automático entrenado en ClinVar (puntuación UP²)
- El fenotipo del paciente (si los términos HPO están asociados a la muestra)
- La concordancia entre el genotipo observado y el modo de transmisión de la patología descrito en las bases de datos GenCC, OMIM y PanelApp (Australia)
- Calidad variante

VariAction Rank. El algoritmo VariAction Rank para priorizar variantes somáticas se propone en los análisis SomaVar a partir de la versión v1.8. Este sistema utiliza la anotación JaxCKB para

calcular las puntuaciones de priorización de variantes en función de su accionabilidad clínica. Estas puntuaciones de priorización tienen en cuenta el impacto de la variante, su patogenicidad mediante la puntuación ComPerMed, el nivel AMP, la especificidad de la variante con el perfil molecular, así como la correspondencia con la indicación, si la enfermedad se introdujo antes de iniciar el análisis.

ACMG . Las variantes pueden ordenarse según su clasificación ACMG seleccionando esta opción. Por defecto, las variantes se clasifican en orden descendente desde la clase patógena 5 hasta la clase benigna 1.

Locus. Las variantes pueden clasificarse según su posición genómica seleccionando esta opción.

Allele frequency. Las variantes pueden ordenarse por frecuencia alélica seleccionando esta opción. Por defecto, las variantes se ordenarán en orden descendente. Es posible ordenarlas en

orden ascendente haciendo clic en el icono de ordenación .

GnomAD Allele Frequency, GnomAD Allele Count. Las variantes se pueden ordenar por su GnomAD Allele Frequency (AF) o su Allele Count (AC) en orden ascendente o descendente (si los datos de GnomAD están disponibles).

UP². Las variantes pueden ordenarse por la puntuación de patogenicidad UP², un componente de DiagAI Score.

Gene. Las variantes pueden ordenarse en orden alfabético ascendente o descendente según el símbolo del gen.

Para las variantes que afectan a genes superpuestos, el símbolo del gen utilizado para la clasificación es el asociado a la variante con mayor impacto, basándose en las reglas de priorización de transcritos descritas anteriormente. Al ampliar la lista de transcritos, el usuario puede ver el segundo gen superpuesto, que puede aparecer sin clasificar. También puede parecer que la clasificación no se aplica cuando se aplican varios filtros a la tabla de variantes.

Configuración de las columnas de la tabla de variantes

Las columnas que se muestran en la tabla de variantes pueden configurarse desde el menú

Organise your columns, accesible a través del botón  situado en la parte superior derecha de la tabla de variantes. El usuario puede modificar la visualización de las columnas haciendo clic en los iconos  para activar la(s) columna(s) deseada(s) en la sección izquierda. La sección derecha muestra las columnas activas en el análisis y permite a los usuarios crear y

guardar perfiles de columnas desde un menú desplegable haciendo clic en el icono "Select profile".

Las columnas activas se pueden reordenar haciendo clic y arrastrándolas a la ubicación deseada. Por defecto, las columnas están organizadas según los perfiles "Germline" y "Somatic".

Cuando los usuarios pasan el cursor sobre el nombre de un perfil de columna en el menú desplegable de preajustes, aparece un mensaje que muestra en qué workset creó ese perfil. Esto advierte a los usuarios si su perfil podría tener columnas incompatibles con el workset actual.

Para crear un nuevo perfil de columnas, simplemente ingrese un nombre en el campo "Create a new preset" y haga clic en "Save".

Editar los metadatos de la muestra desde el panel de análisis

El usuario puede editar los metadatos de la muestra directamente desde el análisis e introducir el sexo del paciente o los HPO (sólo para análisis de línea germinal) desde el cajón de análisis,

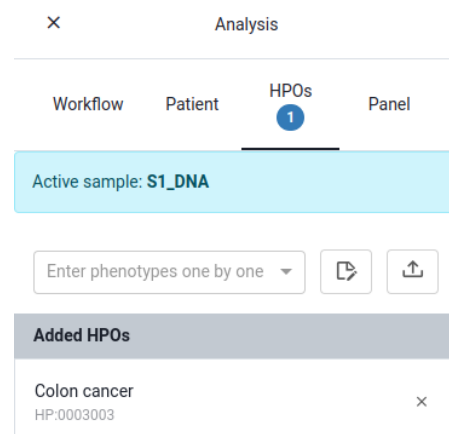
accesible pulsando el botón  en la parte superior derecha de la página (véase la sección **Panel lateral para la gestión de análisis**).

Desde la vista Paciente, el usuario accede a la edición de género (véase el apartado **Vista modal Patient**).

En los análisis de proyectos Germline, desde la vista HPOs, el usuario tiene acceso a varias funciones de edición de términos HPOs:

- Ver los términos ya introducidos antes de iniciar el análisis
- Adición manual de códigos HPO o fenotipos a través de la barra de búsqueda "Introducir fenotipos uno a uno".
- Importar un archivo de texto con códigos HPO
- Generación de términos HPO a partir de la descripción clínica del paciente, llamando a DiagAI

a través del icono . Esta función pide al usuario que describa el perfil clínico del paciente en inglés (descripción limitada a 4000 caracteres). A continuación, tras hacer clic en "Extract HPOs with DiagAI", se genera una



lista de HPO resaltadas en color. Por último, el usuario puede asociar todos los términos HPO propuestos al paciente haciendo clic en "Add all", o seleccionarlos individualmente pasando el ratón por encima de los términos elegidos y haciendo clic en "Add".

Añadir y editar HPOs desde esta vista da al usuario la opción de reprocesar la puntuación DiagAI y la clasificación de variantes (Short list o SmartPick) según los nuevos HPOs introducidos a través de la funcionalidad "Re-process analysis".

Filtrar las variantes

Puede afinar las búsquedas en la pestaña “*Variants*” utilizando las categorías de filtros que se detallan a continuación. Se puede acceder al panel de filtros a través del icono  situado en la parte superior izquierda de la tabla de variantes o pulsando la tecla “F” del teclado. Los filtros están disponibles en la pestaña Básico. La sección « Active Filters », ubicada en la parte superior del panel de filtros, indica la cantidad de filtros aplicados.

Filtros Genes & Variants

Nombre y apertura del filtro	Descripción
Genes ⓘ Search gene symbols	Restringe la lista de variantes a uno o más genes.
Variants ⓘ Search by HGVS	Restringe la lista de variantes a una o más variantes según la nomenclatura HGVS o HGVS.

Sección Samples & Quality

Nombre y apertura del filtro	Descripción
Coverage ⓘ 1 836 1	Define un umbral mínimo para la profundidad de cobertura (COV). El valor máximo mostrado por este filtro corresponde al valor de cobertura más alto identificado entre las variantes.
Allele Frequency ⓘ 0% 100% 0 100	Filtra las variantes según la frecuencia del alelo mutante en la muestra analizada.
Quality ⓘ Low Medium High	Calidad media (puntuación phred) de cada base portadora del alelo mutado, que indica la confianza en la detección de la variante: - <i>Media</i> ($\geq \text{QUAL}=20$): 1 probabilidad entre 100 de llamada incorrecta, - <i>Alta</i> ($\geq \text{QUAL}=30$): 1 posibilidad entre 1000 de llamada incorrecta. Este filtro se aplica a todas las muestras.
Quality ⓘ All Low $\geq Q8$ Medium $\geq Q20$ High $\geq Q30$	Para los análisis GermVar Tertiary y GermVar Tertiary Family sobre datos de Oxford Nanopore Technologies, la calidad basada en la puntuación Phred indica la confianza en la detección de variantes. $\geq \text{QUAL}=8$ (Bajo): 1 posibilidad entre 8 de llamada incorrecta,

	≥ QUAL=20 (Media): 1 posibilidad entre 100 de llamada incorrecta, ≥ QUAL=30 (Alto): 1 posibilidad entre 1000 de llamada incorrecta. Este filtro se aplica a todas las muestras.
Genotype ⓘ ☒ Homozygous 387 ☐ Heterozygous 1163 ☐ None 1942 ☒ Biallelic Variant 222 ☒ Biallelic Variant 84 ☒ Unknown 76 ☐ Unknown 0	Filtra las variantes según el genotipo observado.
Alternate Observations ⓘ 0 0 458	Define un número de secuencias portadoras de la mutación (observación alternativa u OA). El valor máximo mostrado por este filtro corresponde al valor más alto de AO identificado entre las variantes.
Strand Odds Ratio ⓘ 0 5 0 1 2 3 4 ≥5 min 0 max 5	Filtra las variantes según la puntuación strand odds ratio (SOR) calculada en el sitio de la variante.

Sección Frecuencias

Nombre y apertura del filtro	Descripción
Sample frequency ⓘ all all rare	Elimina las variantes en función de su recurrencia en el proyecto estudiado, de la más frecuente (todas) a la menos frecuente (raras).
Relative pop frequency ⓘ all 10-2 10-3 10-4 10-5 10-6 absent	Selecciona variantes según la frecuencia de su alelo mutante en la población combinada de los proyectos ExAc y gnomAD.
Absolute pop frequency ⓘ 0 1609917	Selecciona variantes en función del número de alelos en la población combinada de ExAc y GnomAD.

Absolute homozygous pop frequency ⓘ ⓘ
 0 803576

Selecciona variantes en función de sus frecuencias de homocigosis en la población combinada ExAc y GnomAD.

Sección Genetics

Nombre y apertura del filtro	Descripción
▼ Impact ⓘ Low Moderate High ☒ Keep <i>Modifier</i>	El impacto de la mutación en la proteína. <i>Modificador:</i> afecta a los genes no codificantes
▼ Effects ⓘ > Loss of Function ⓘ • Transcript ablation ⓘ • Splice acceptor ⓘ • Splice donor ⓘ • Stop gained ⓘ • Frameshift ⓘ • Stop lost ⓘ • Start lost ⓘ • Transcript amplification ⓘ	Términos SO (Sequence Ontology) que describen la consecuencia más perturbadora a nivel del gen. Este filtro permite seleccionar el impacto de la mutación en el gen, y las opciones están organizadas en categorías lógicas para mejorar la navegación: <i>Loss of Function, Coding, Splicing region, Silent variant, Intronic, UTR, ncRNA, Upstream/Downstream, Regulatory, Others.</i>
▼ Gene panel ⓘ Select panel	Seleccione y aplique una o varias paneles de genes. Este filtro muestra todas los paneles de genes disponibles para la entidad. Este filtro no está disponible si se ha aplicado un panel restringido al análisis desde la pestaña <i>Panel</i> del panel de gestión del análisis (véase la sección Vista modal Panel). Un mensaje informa al usuario de que el filtro "Gene panel" no puede utilizarse con un panel restringido.
▼ Panel in-silico ⓘ Dyslipidemia_panel_1_Twist★	Selecciona y aplica un panel in-silico. Este filtro muestra todos los archivos BED disponibles para el grupo. El nombre del archivo utilizado durante el análisis se resalta con una estrella. Este filtro sólo retiene las variantes cuya posición inicial (start) se

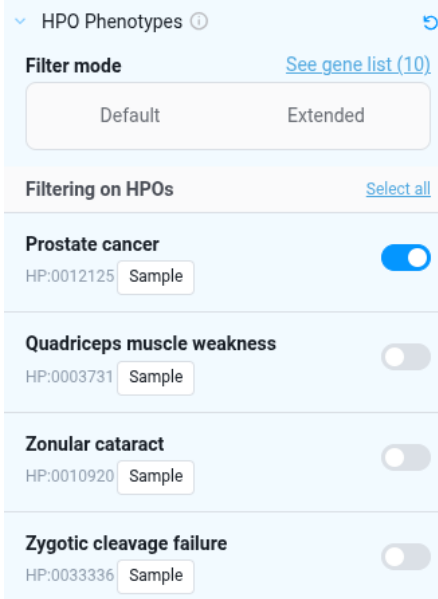
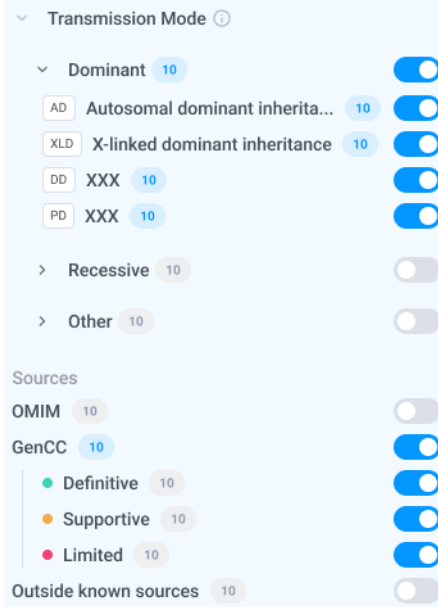
	encuentra dentro de uno de los intervalos del archivo de lecho seleccionado.
PanelApp ⓘ Epi ENG Blepharophimosis ptosis and epicanthus ENG Early onset or syndromic epilepsy ENG Epidermodysplasia verruciformis	Le permite filtrar variantes según una lista de paneles in-silico en los catálogos de PanelApp Australia y PanelApp Inglaterra. Un distintivo identifica la base de datos de la que procede el panel in-silico: AUS (Australia) o ENG (Inglaterra). En ambos catálogos, los superpaneles no están disponibles para la selección, así como los paneles que no incluyan al menos un gen. Por ejemplo, un panel que sólo contenga STR no se podrá seleccionar.
Gene occurrences ⓘ Usable only if they are fewer than 2000 selected variants.	Define un umbral para el número de apariciones de un gen determinado en la tabla de variantes.
Chromosomes & MT ⓘ X Y M 25 0 0	Conserva las variantes localizadas en los cromosomas sexuales y/o en el cromosoma mitocondrial (M)
RefSeq transcripts ⓘ BARD1_2transcripts.txt	Selecciona una lista de transcritos RefSeq para filtrar las variantes.

Sección Databases

Nombre y apertura del filtro	Descripción
Variant Databases ⓘ dbsnp 31 cosmic 31 civic 30 lovd 10 clinvar 3	Selecciona variantes anotadas por las siguientes bases de datos: dbsnp, cosmic, lovd, clinvar y 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	Permite filtrar las variantes según su estado de patogenicidad en ClinVar. La casilla 'Mantener otras' se utiliza para mantener las variantes sin anotación en ClinVar.
OMIM ⓘ omim morbid clinic 111 omim without phenotypes 23 omim morbid 6	Permite filtrar por OMIM clínica mórbida, OMIM sin fenotipos y OMIM variantes mórbidas. Este filtro tiene un botón de entrada/salida que se puede utilizar para seleccionar o eliminar un estado.
SeqOne Artifacts ⓘ ☐ Remove artifacts	Botón para eliminar variantes recurrentes presentes en el 5% de una cohorte de 500 exomas considerados como artefactos según el algoritmo SeqOne.

Sección Transmission & Phenotypes

Nombre y apertura del filtro	Descripción
Description keywords ⓘ	Utiliza palabras clave para caracterizar el fenotipo o fenotipos asociados a las variantes mediante consultas a las bases de datos OMIM, Medgen y Litvar. Esta función no tiene en cuenta las HPO, ya que se les dedica un filtro específico (véase más adelante).

	Existen dos modos de búsqueda: unión e intersección.
	Permite filtrar las variantes en función de los términos HPO introducidos previamente, bien antes del análisis de la muestra (véase Metadatos de muestra), bien después del análisis desde la pestaña HPO del panel de análisis (véase Vista modal HPOs). Las coincidencias se identifican utilizando nuestro enfoque PhenoGenius, que se basa en una matriz de genes HPO para seleccionar los genes. 2 opciones están disponibles en el modo Filtro: - Filtrado predeterminado que equilibra la precisión y el número de genes devueltos. - Filtrado extendido que maximiza la precisión con más genes La opción "See gene list" proporciona una lista de genes asociados con los HPO seleccionados en este filtro. Esta función solo está disponible si se han marcado términos HPO. Se abre una ventana emergente que muestra el ID, símbolo, rango de Phenogenius y puntuación de Phenogenius de cada gen. Puede descargar la lista de genes usando el ícono <i>Download</i> o copiarla usando el botón de copiar/pegar.
	Permite a los usuarios filtrar variantes utilizando vínculos de herencia de enfermedades relacionadas con la genética médica humana. Las variantes se pueden filtrar según los modos de transmisión (Dominant, Recessive, Other) y las fuentes de transmisión (OMIM, GenCC u otras bases de datos).

Transmission pattern ⓘ Compound Het - Suspected 10 Compound Het - Verified 10 Recessive Hom 10 De Novo 10	Permite la selección de variantes según los modos de transmisión heterocigoto compuesto (sospechado o verificado), homocigoto recesivo o de novo basado en DiagAI. Tenga en cuenta que los filtros "Compound Het - Verified", "Recessive Hom" y "De novo" sólo están disponibles en los análisis familiares.
--	--

Sección Classification

Nombre y apertura del filtro	Descripción
VKB ⓘ New Unknown variant 5 Pathogenic 4 Likely Pathogenic 3 Uncertain Significance 2 Likely Benign 1 Benign Possible artifact Analysis only ⓘ	Permite seleccionar variantes desconocidas no evaluadas (Nuevo - Variante desconocida) o validadas en VKB en todos los análisis de la entidad. Si se marca la casilla "Analysis only", las variantes se buscarán únicamente en el análisis que se esté anotando en ese momento.
ACMG ⓘ 5 Pathogenic 5 4 Likely Pathogenic 1 3 Unknown significance (VUS) 56 2 Likely Benign 49 1 Benign 6	Clasificación de la patogenicidad de las variantes de la línea germinal, calculada en una escala de 1 (benignas) a 5 (patógenas).
ComPerMed ⓘ 5 Pathogenic 7 4 Likely Pathogenic 6 3 VUS 102 2 Likely Benign 1 1 Benign 7	Selecciona las variantes somáticas según su clase de patogenicidad ComPerMed en una escala de 1 (benignas) a 5 (patógenas).

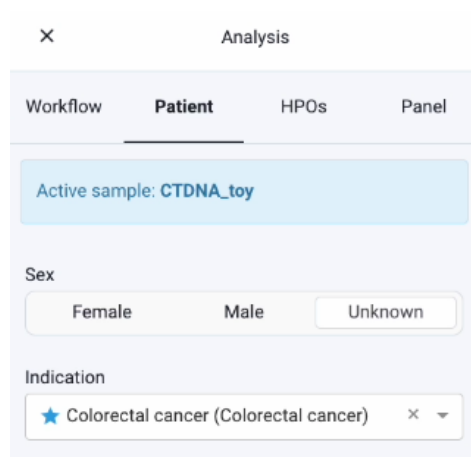
▼ CADD ⓘ Benign Pathogenic 1 99 Min 1 ^ ▼ Max 99 ^ ▼ ☒ Keep unknown	Selecciona las variantes según la puntuación CADD Phred.
▼ REVEL ⓘ Benign Pathogenic 0 0.5 1 Min 0 ^ ▼ Max 1 ^ ▼ ☒ Keep unknown	Selecciona las variantes según la puntuación REVEL.

Seleccionar una indicación

En SomaVar, SomaVar LF y SomaCGP, es posible seleccionar una indicación de patología predefinida de la base de datos JaxCKB desde un menú desplegable disponible en la pestaña *Paciente* de la barra lateral de gestión de análisis (véase la sección **Vista modal Patient**).

El sistema ontológico se basa en la *Ontología de Enfermedades*²⁴ e incluye términos específicos de JaxCKB. La información de sensibilidad, en la subcolumna 'A' de la columna Evidencias Clínicas (ver sección **Página de Variantes**), se resalta en verde sólo cuando la indicación para esa variante coincide con la indicación seleccionada. Si se encuentran afirmaciones de resistencia para una variante, la celda correspondiente en la columna 'A' se resaltarán en rojo, independientemente de la indicación. En caso contrario, aparecerá en gris.

Si la enfermedad se ha seleccionado durante la carga de la muestra y antes del análisis (véase **Metadatos de la muestra**), se convierte en la indicación por defecto definida para la muestra y en la pestaña *Paciente* del panel lateral de gestión del análisis. Se marca con un icono de estrella y aparece en primer lugar en la lista. Cambiar la indicación en el panel lateral de gestión de análisis no actualiza la enfermedad preseleccionada para la muestra (metadatos Enfermedad).



The screenshot shows a modal window titled 'Analysis' with a close button (X). It has four tabs: 'Workflow', 'Patient' (selected), 'HPOs', and 'Panel'. Under the 'Patient' tab, there is a light blue box stating '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 blue star icon next to the text 'Colorectal cancer (Colorectal cancer)', followed by a close button (X) and a downward arrow.

²⁴ <https://disease-ontology.org/>

Utilizar un perfil de filtros predefinidos

Los usuarios pueden seleccionar un perfil de filtro predefinido en un menú desplegable haciendo clic en el icono "Select profile". La plataforma SeqOne ofrece por defecto varios perfiles de filtro predefinidos: *GermVar*, *My Selection*, *Research*, *DiagAI*.

Editar y guardar perfiles de filtro

Los usuarios pueden guardar la posición actual de los cursores para cada filtro como un perfil de filtro dando un nombre al perfil y haciendo clic en "Save as new". A continuación, el perfil de filtro estará disponible a través del mismo menú desplegable en la categoría *Private*.

Al guardar un perfil de filtro, se abre un formulario de creación y el usuario puede seleccionar el botón con el nombre de la entidad para hacerla accesible a otros usuarios de la entidad. Si el usuario selecciona el botón "Private", el perfil sólo será accesible para el usuario que lo creó.

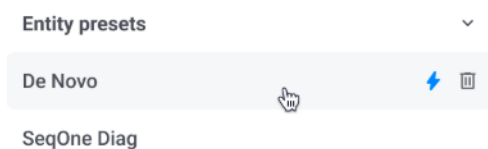
Sólo el usuario que creó inicialmente el perfil de filtro podrá editarlo haciendo clic en "Update"; si el perfil es modificado por otro usuario de la entidad, se le pedirá que lo guarde con otro nombre. Si el usuario no tiene permiso para modificar un perfil de filtro público, se le pedirá que se ponga en contacto con el administrador de su entidad. Los perfiles de filtro SeqOne no pueden modificarse y guardarse bajo otro perfil de filtro.

Es posible restablecer uno o varios filtros. Basta con hacer clic en el icono  situado a la derecha del filtro o en la esquina superior derecha del cuadro *Filtros*.

Perfiles de filtro favoritos

Los usuarios pueden configurar los perfiles de los filtros como favoritos, haciendo clic en el icono  situado a la derecha del perfil, para que aparezcan como acceso directo en la barra de acceso rápido situada encima de la tabla de variantes.

Para eliminar un perfil de filtro de los favoritos, basta con hacer clic en el mismo icono . Por defecto, el perfil *My Selection* está definido como favorito.

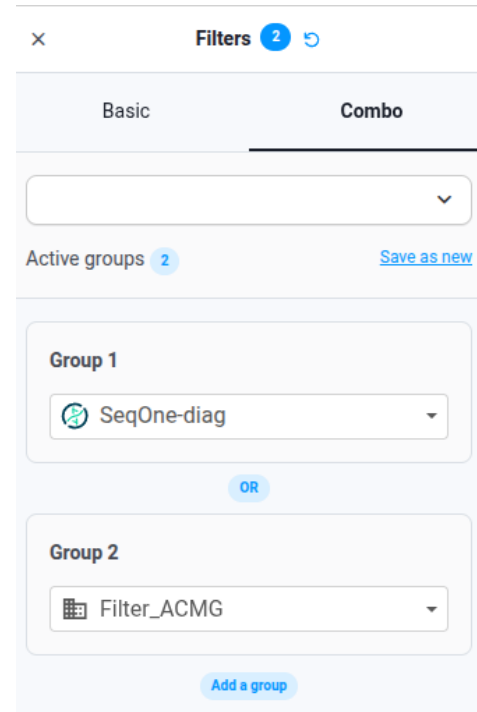


Combinación de perfiles de filtro

Los perfiles de filtro pueden combinarse bajo la condición **O** haciendo clic en la pestaña Combo del panel de filtros y, a continuación, en "Create a combo". Por defecto, el formulario de creación muestra 2 grupos de perfiles de filtro, pero los usuarios pueden añadir más haciendo clic en "Add a group".

Una vez seleccionados los perfiles, se invita a los usuarios a guardar su perfil de combo haciendo clic en "Save as new". Como en el caso de los perfiles de filtro, el usuario puede hacer que su combo sea accesible a todos los usuarios de la entidad en el formulario de creación. Sólo el usuario que creó inicialmente el perfil combo podrá editarlo haciendo clic en "Update".

Si el perfil es modificado por otro usuario de la entidad, se le pedirá que lo guarde con otro nombre. Si el usuario no tiene permiso para modificar un perfil de combo público, se le pedirá que se ponga en contacto con el administrador de su entidad.



The screenshot shows a web interface for creating filter combinations. At the top, there's a 'Filters 2' header with a close button (X) and a refresh icon. Below this are two tabs: 'Basic' and 'Combo', with 'Combo' being the active tab. A search bar is located below the tabs. Underneath the search bar, it says 'Active groups 2' and has a 'Save as new' link. The main area shows two groups of filters. 'Group 1' contains a dropdown menu with 'SeqOne-diag' selected. Below this is an 'OR' button. 'Group 2' contains a dropdown menu with 'Filter_ACMG' selected. At the bottom right of the groups is an 'Add a group' button.

Perfiles de filtro SeqOne

My Selection. Este perfil de filtro permite que las variantes evaluadas como 3 (VUS), 4 (Probablemente patógenas) y 5 (Patógenas) se muestren directamente en VKB en el análisis que se está interpretando.

DiagAI . Este perfil de filtro, disponible en los análisis de línea germinal (GermlineVar, GermlineFamily, GermVar Tertiary, GermVar Tertiary Family), muestra una lista reducida de variantes con una alta probabilidad de ser la causa de la enfermedad del paciente. Este perfil explota el algoritmo de priorización DiagAI y la ocurrencia de la variante en el proyecto (filtro Sample Frequency).

Puede aplicarse cuando los términos HPO están o no asociados a la muestra, pero su sensibilidad aumenta en presencia de términos HPO. En el caso de un análisis familiar, el perfil del filtro se adapta automáticamente al tamaño de la familia analizada, ya que tiene en cuenta que la ocurrencia de variantes en el proyecto puede ser mayor. Los usuarios también pueden modificar el filtro Frecuencia de muestreo en el proyecto en cualquier momento. Al introducir el HPO se muestran las variantes SmartPick con mayor probabilidad de causar la enfermedad.

Family profiles. En los análisis GermlineFamily y GermVar Tertiary Family, los perfiles de filtro *De Novo (New)*, *Recessive Hom (New)* y *Compound Het (New)* sólo estarán disponibles para los

análisis de trío que incluyan un caso índice afectado, el padre no afectado y la madre no afectada. En caso contrario, estos perfiles estarán desactivados.

No se recomienda modificar los perfiles de filtro *De Novo (New)*, *Recessive Hom (New)* et *Compound Het (New)* al crear y guardar un perfil personal. Se recomienda empezar de cero al crear una variante de estos perfiles.

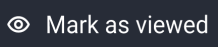
Carrier screening Este perfil de filtro sólo está disponible en Germlinefamily para análisis de dúos (padres sin muestra índice) e identifica genes recesivos con una variante heterocigota potencialmente patogénica en ambos padres "no afectados" para pruebas de preconcepción.

Somatic FFPE. Este perfil de filtro muestra una lista reducida de variantes somáticas en análisis somáticos (SomaVar, SomaCGP) realizados en ADN FFPE.

Somatic LF. Este perfil de filtro muestra una lista reducida de variantes somáticas en los análisis SomaVar LF.

Funcionalidad Mark as viewed

"Mark as viewed" es una función que permite al usuario ocultar selectivamente determinadas variantes. Para ello, basta con seleccionar la variante o variantes deseadas y hacer clic en el

icono  que aparece en la parte inferior de la tabla de variantes. La selección de variantes aparece entonces en gris. Para eliminar la selección, el usuario puede hacer clic en el

icono  en la parte superior de la tabla de variantes para restablecer la vista, o seleccionar la variante o variantes y hacer clic en .

El usuario puede ocultar la ventana emergente haciendo clic en la cruz. Esto anulará la selección de todas las variantes, pero permanecerán en gris.

Selección de variantes

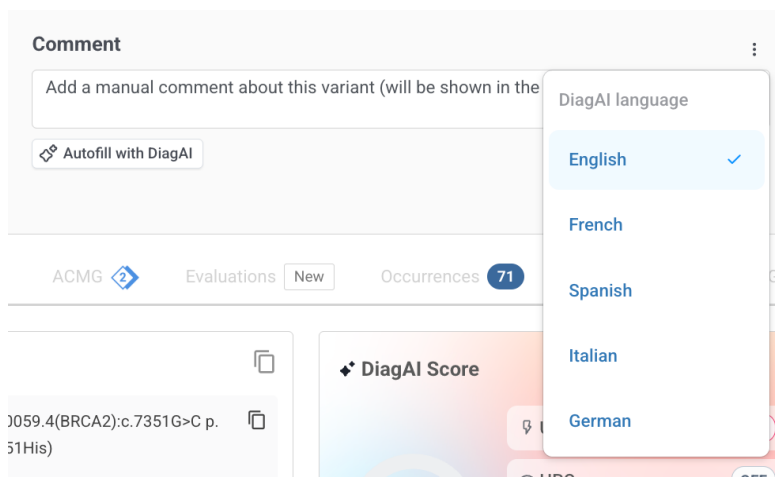
Las variantes se pueden añadir al panel de "Selection" junto con las variantes evaluadas (ver la sección siguiente). Para ello, seleccione las variantes marcando las casillas del lado izquierdo de la tabla de variantes y, a continuación, haga clic en el icono "Add to selection" en la parte inferior de la tabla.

Evaluar las variantes

Las variantes pueden evaluarse rápidamente desde la tabla de variantes haciendo clic en la columna VKB y especificando su clase de patogenicidad y un comentario opcional. Estas evaluaciones y comentarios son accesibles para todos los usuarios de la plataforma dentro de la misma entidad a través del icono "Selección" en el análisis (véase la sección **Panel lateral para gestionar las evaluaciones**). Una variante también puede evaluarse desde su página dedicada.

Los comentarios pueden introducirse manualmente o generarse automáticamente. La interfaz incluye un campo "Clasificación" vacío para los comentarios y un botón "Autofill with DiagAI" para la generación automática. Durante la generación de comentarios aparece un indicador. Para los comentarios existentes, la opción "Improve with DiagAI" permite mejorarlos. Cada comentario generado puede aprobarse, modificarse o anularse. Una vez validados, los comentarios se guardan con su clase de patogenicidad, fecha y contenido. Pueden modificarse o suprimirse después de la validación, conservando su historial.

Para seleccionar el idioma de DiagAI, haga clic en los tres puntos de la esquina superior derecha del cuadro de comentarios y elija su idioma preferido en el menú desplegable.

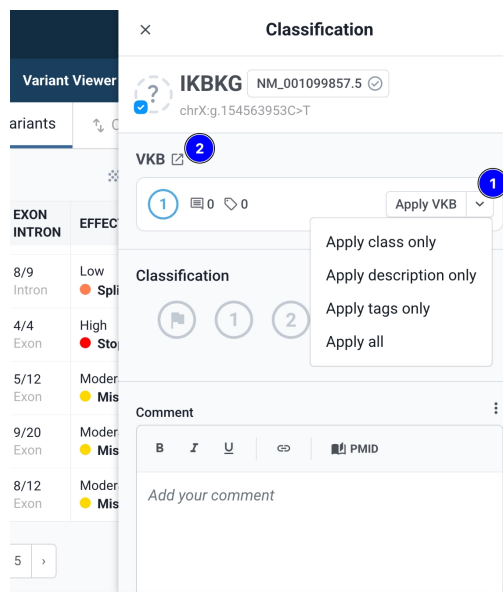


Las variantes también se pueden clasificar haciendo clic en la columna de clasificación, cuyo encabezado tiene una bandera. Para clasificar una variante, haga clic en el icono +. Se abre un panel lateral de clasificación que permite al usuario añadir la clase de patogenicidad, un comentario y una etiqueta.

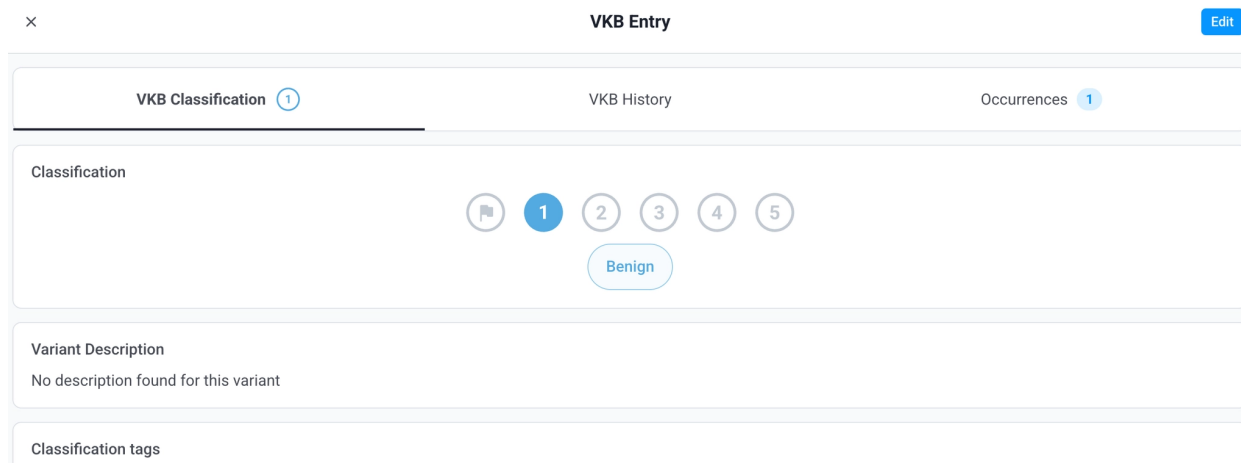
Para añadir esta clasificación a la base de datos VKB, active el interruptor "Add to VKB". Si ya existe una clasificación en la base de datos VKB, haga clic en "Apply VKB" para aplicarla a la entrada actual. Existen opciones adicionales haciendo clic en la flecha situada junto a "Apply VKB" (véase el paso 1 en la figura siguiente).

Valide la clasificación haciendo clic en "Save" en la parte superior derecha del panel lateral.

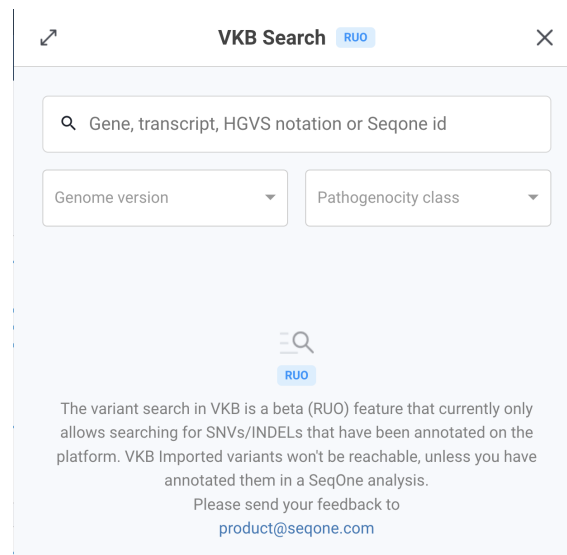
Para editar una clasificación directamente en la base de datos VKB, haga clic en el enlace de redirección VKB (véase el paso 2 en la figura siguiente) para abrir la ventana "VKB Entry".



En la ventana "VKB Entry", haga clic en "Edit" en la parte superior derecha para modificar la evaluación, luego haga clic en "Save" para aplicar los cambios. El historial de VKB y las ocurrencias de la variante son accesibles navegando a través de las pestañas dedicadas.



[VKB Search](#) . Se puede acceder a un panel lateral haciendo clic en el botón "VKB" de la barra de búsqueda de la cabecera de SeqOne. Esto permite buscar variantes anotadas en la base de datos VKB por nombre de gen, transcripción, anotación HGVS o identificador SeqOne (Genome_variant_chromosome_location_REF_ALT). La búsqueda puede filtrarse por genoma de referencia (GRCh37 o GRCh38) y/o clase de patogenicidad.



VKB Search RUO

Gene, transcript, HGVS notation or Seqone id

Genome version Pathogenicity class

RUO

The variant search in VKB is a beta (RUO) feature that currently only allows searching for SNVs/INDELs that have been annotated on the platform. VKB Imported variants won't be reachable, unless you have annotated them in a SeqOne analysis. Please send your feedback to product@seqone.com

Figura 15: Ventana de búsqueda VKB

Al hacer clic en una variante evaluada, se abre una página más detallada con información sobre la variante, como su posición genómica, sus anotaciones HGVS en el ADN complementario (ADNc) y en la proteína, las clases de patogenicidad, los análisis y sus comentarios asociados en los que se evaluó y, a continuación, un enlace a los análisis.

BRCA1 NM_007294.3

c.2834_2836delinsC p.Ser945ThrfsTer6

gDNA chr17:g.41244712_41244714delinsG

cDNA NM_007294.3:c.2834_2836delinsC

Protein NP_009225.1:p.Ser945ThrfsTer6

VCF 17-41244712-TAC-G

Latest classification: 5

5 (7)
4 (0)
3 (0)
2 (0)
1 (0)
0 (0)

Apr 19, 2024
demo - GermlineVar

PP5: Found a Clinvar Pathogenic entry (<https://www.ncbi.nlm.nih.gov/clinvar/variation/54694>) having at least 1 star. PM1: Found pathogenic hotspot in ClinVar with 28 pathogenic variants out of 68 (73.68%) on 36bp window, above 66.67% threshold. PM2: GnomAD max frequency (0.0 for population) is inferior to 1/10000. PVS1: Predicted LoF variant (frameshift variant) in BRCA1 gene where there is 3418 out of 3641 pathogenic LoF variants found in Clinvar (93.88%) above 10.00% threshold.

Open on analysis

Figura 16 : Ventana de detalles de la variante en VKB

Página variante

Al hacer clic en una variante de la tabla de variantes, se abre un panel lateral con más información y varias pestañas. Los usuarios pueden abrir la página de variantes en otra pestaña pulsando "Abrir en ventana nueva" en la esquina superior derecha. El panel lateral puede cerrarse con la tecla "Esc".

Los usuarios pueden evaluar y comentar una variante en esta ventana. Al igual que las evaluaciones y comentarios realizados desde la tabla principal de variantes, estas evaluaciones y comentarios pueden ser vistos por todos los miembros del grupo que utilicen la plataforma.

Desde esta página, los usuarios también pueden acceder a información sobre variantes, incluidos el genotipo (GT), la frecuencia alélica (FAV), la profundidad de cobertura (COV) y la puntuación phred (QA).

A continuación se describen las pestañas de la página de variantes.

Variant : la ficha se divide en varios bloques, que se describen a continuación.

- **Nomenclature:** se enumeran las distintas nomenclaturas (g., c., p., entrada VCF), con la posibilidad de copiar la información de todo el bloque o de cada línea al portapapeles mediante el icono .
- **DiagAI Score:** utiliza el bloque resumen de cálculo de la Puntuación DiagAI disponible en la pestaña Puntuación DiagAI de la página de variantes (véase el párrafo **DiagAI**).
- **ClinVar:** información tomada de la columna ClinVar Germline de la tabla de variantes, dando la clase de patogenicidad asignada a la variante, su estado clínico y el número de entradas en ClinVar (ver **Visualización e Interpretación** en la sección **Pequeñas Variantes**). El icono  abre la página ClinVar en una nueva pestaña.
- **Annotations:** impacto del transcrito, exón afectado, distancia al sitio de splicing más cercano y tipo de mutación.
- **Transmission:** enlaces a las bases de datos OMIM y GenCC si una o varias enfermedades están asociadas al gen.
- **Compound Heterozygous Candidates**, este bloque se divide en 2 partes:
 - La sección *Variantes Cortas* agrupa todas las SNV candidatas localizadas en el mismo gen (basándose en el ID del gen) y las clasifica según su puntuación DiagAI. La variante seleccionada en la página Variantes se muestra en primer lugar y aparece atenuada en la lista.
 - La sección *CNVs* indica si se han encontrado CNVs, ordenados por clase ACMG. Esta sección solo está disponible si se ha ejecutado un análisis CNV; de lo contrario, se muestra un mensaje de advertencia que dice '*CNVs not available*' (*CNVs no disponibles*).

Al hacer clic en uno de los resultados se abrirá una nueva pestaña con la página de la variante (si se trata de una variante pequeña) o la página de la CNV (si se trata de una variante de copia). Si no se encuentra ninguna VNC, aparecerá un mensaje de advertencia con las palabras "no se ha encontrado ninguna VNC".

- **Tools:** enlaces a otras bases de datos (LitVar, LOVD, VarSome, dbSNP, Cosmic, GnomAD, CKB, UCSC, Decipher, Ensembl).
- **GnomAD:** si la variante es conocida por esta base de datos, se muestra una tabla que resume la frecuencia de la variante en diferentes poblaciones, y se proporciona acceso directo a GnomAD a través del icono .

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

Annotations

Impact/Effect Moderate missense

Exon/Intron 4/9

Biotype protein coding

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

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

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

Transmission

AR

DISEASE	OMIM	MEDGEN
Hypofibrinogenemia, congenital	AR	N/A

Tools

Variant database

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

Gene database

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

Genome database

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

Figura 17: Pestaña de variantes en la página de variantes

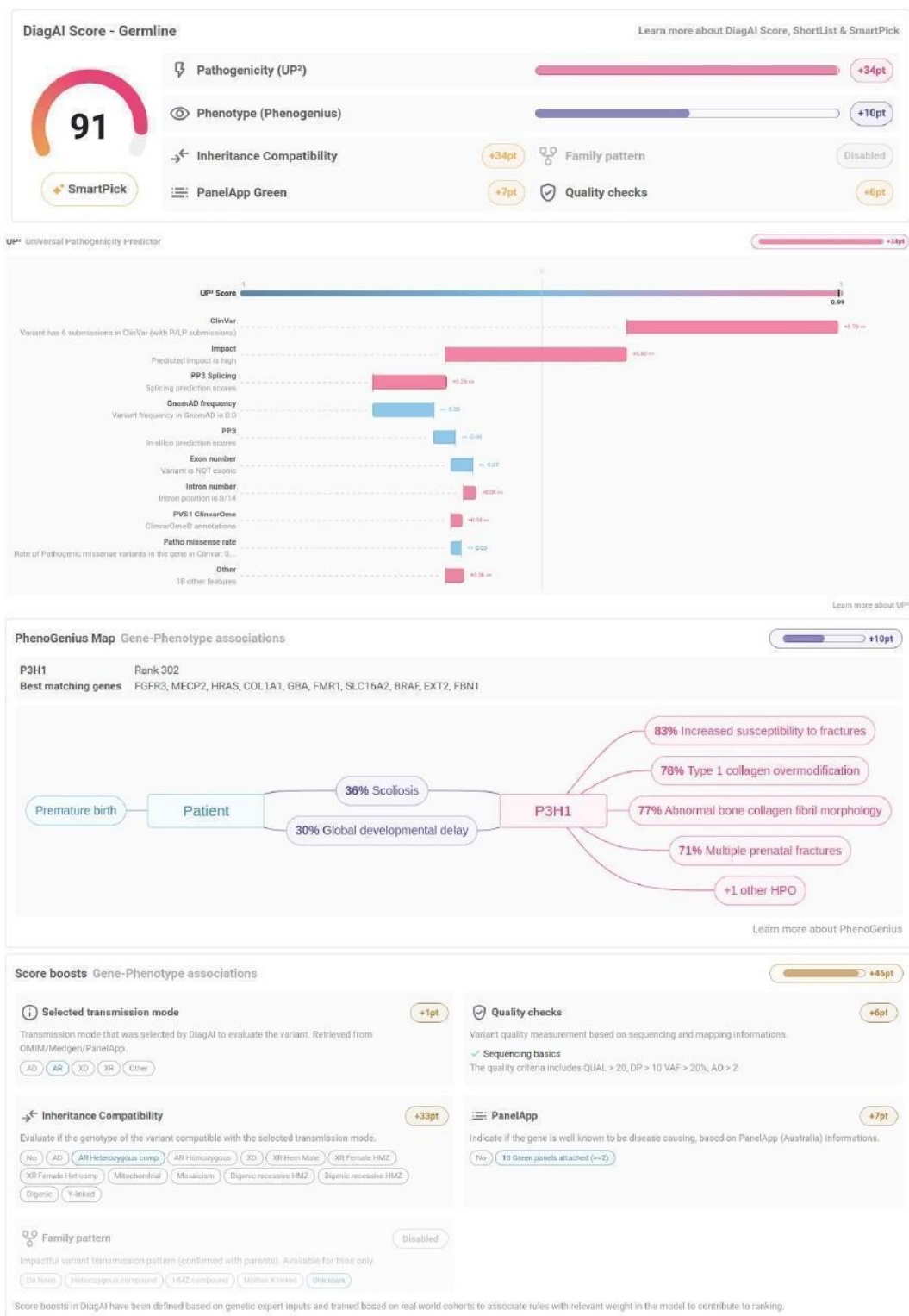


Figura 18: Visualización de la pestaña DiagAI en la página de variantes

DiagAI. La ficha contiene explicaciones de la puntuación DiagAI asignada a la variante. Esta pestaña sólo está disponible para los análisis GermlineVar, GermlineFamily, GermVar Tertiary y GermVar Tertiary Family (a partir de la v1.8).

La pestaña DiagAI está dividida en 4 secciones principales:

- Un bloque de resumen de **DiagAI Score - Germline** que muestra la puntuación, el estado SmartPick y enumera las principales características que contribuyen a la puntuación global y el número de puntos asociados.
- Un gráfico en cascada que explica la puntuación **UP² (Universal Pathogenicity Predictor)**. Se muestra un mensaje de advertencia  cuando el usuario consulta la página de una variante cuya transcripción difiere de la utilizada para calcular el UP².
- Un bloque de explicabilidad del Mapa PhenoGenius, coloreado si los términos HPOs están asociados a la muestra. El bloque detalla el rango del gen, una lista de los 10 genes mejor asociados, y el mapa PhenoGenius, codificado por colores según la correspondencia entre las HPOs del paciente y el gen: azul para las HPOs del paciente solamente, púrpura para las HPOs del paciente y correspondientes al gen, rojo para las HPOs del gen solamente. La sección aparece en gris si no se han introducido términos HPO para el paciente.
- Una sección de **potenciación de** la puntuación que integra 5 subcaracterísticas que pueden o no contribuir (Desactivadas) a la puntuación: Modo de transmisión, Compatibilidad hereditaria, Patrón familiar, PanelApp y Control de calidad.

ACMG. La pestaña permite a los usuarios ver y revisar los criterios utilizados durante la clasificación automatizada ACMG. Pueden modificar la selección marcando los diferentes criterios requeridos y, a continuación, utilizar la clasificación para evaluar la variante haciendo clic en *Utilizar para evaluación*. También es posible modificar el nivel de evidencia (*Apoyando, Moderado, Fuerte y Muy Fuerte*) para asignar un número diferente de puntos a cada criterio, que se tendrán en cuenta a la hora de calcular la puntuación ACMG final. Cada nivel de evidencia otorgará: 1 punto por *Apoyar*, 2 puntos por *Moderado*, 4 puntos por *Fuerte* y 8 puntos por *Muy Fuerte*. Los puntos de cada criterio se sumarán para asignar un estatus ACMG a la variante:

- *Patógeno* : 10 puntos
- *Probable patógeno* : 9 a 6 puntos
- *Significado incierto* : 5 a 0 puntos
- *Probable benigno* : -6 a -1 puntos
- *Benigno* : a partir de -7 puntos

Para el criterio BA1, el usuario no puede seleccionar el nivel de prueba, ya que se trata de una clase independiente.

The screenshot displays the TP53 variant page. At the top, the variant is identified as NM_000546.6 c.1083del p.S362AfsTer8, classified as a Frameshift. The predicted pathogenicity is 'Likely Pathogenic'. The ACMG tab is selected, showing automatic tags: PVS1 (Strong), PM2 (Moderate), and PP5 (Moderate). The page also includes a comment section and a 'Use for evaluation' button.

Figura 19: Visualización de la pestaña ACMG en la página de variantes

TP53. La pestaña está disponible sólo si la variante está anotada en el gen TP53. Esta pestaña proporciona información sobre las mutaciones TP53 y su importancia clínica según la base de datos IARC²⁵. La pestaña TP53 está dividida en varios bloques, descritos a continuación.

Un bloque de anotaciones (*Annotations*) muestra la siguiente información asociada a la variante:

- Mutation Rate: tasa de sustitución asociada a la posición de la variante
- Hotspot (yes/no) : presencia de la variante en un lugar de mutación recurrente (hotspot)
- Structural Motif : motivo estructural afectado por la mutación
- Dominio: dominio funcional de la proteína afectada por la mutación

El bloque Datos funcionales (*Functional Data*) muestra clasificaciones e información funcional asociada a la variante:

- Clase DNE : efecto dominante negativo (DNE) sobre la transactivación de la proteína p53 de tipo salvaje.
- Clase DNElof: clasificación funcional para la pérdida de actividades de supresión del crecimiento y dominantes negativas.
- Clase de transactivación: clasificación funcional basada en la actividad transcripcional global (AT)

²⁵ <https://tp53.cancer.gov/>

- Estructura Función Clase : predicciones funcionales.

Los bloques “Graphics” y “Sample details” proporcionan detalles de las muestras referenciadas en la base de datos IARC para la variante. El usuario puede elegir ver los datos asociados a muestras somáticas o de línea germinal, seleccionando “Somatic” o de “Germline”.

El bloque “Graphics” muestra dos gráficos, con menús desplegables, que muestran la distribución de las muestras según sus características.

El bloque “Sample details” muestra información detallada de la muestra en forma de tabla.

ComPerMed (sólo análisis somático). La clase de patogenicidad de cada variante se calcula siguiendo las recomendaciones del comité de expertos belga ComPerMed (Comisión de Medicina Personalizada). El usuario tiene acceso a información detallada sobre esta clasificación desde una pestaña específica en la página de variantes.

Evidences. La pestaña proporciona una lista de terapias asociadas a la variante seleccionada de la base de datos JaxCKB. El usuario puede desplazarse por cada línea para encontrar detalles de las terapias o indicaciones.

BRCA2

protein coding

NM_000059.4

c.1310_1313del

p.K437IfsTer22

frameshift

sample

GT / VAF 43.7%

COV / QA 126 29

1

2

3

4

5

Add a manual comment about this variant (will be shown in the output report) ...

Add comment

Cancel

Variant

ACMG

Compermed

Evidences

Evaluations

New

Occurrences

Scores

IGV

Gene

The following information is provided for informational purposes only

Actionable 25

Prognostic & Diagnostic 0

Risk factor & emerging

Therapy	Tier	Response	Indication
Olaparib	IA	Sensitive	Breast cancer Her2-receptor negative breast cancer ✓FDA +6
Select indication Breast cancer Evidence type Actionable Molecular profile BRCA2 inact mut Approval N/A Drugs Olaparib (PARP Inhibitor (Pan)) Description Lynparza (olaparib) is included in guidelines as systemic treatment for patients with recurrent or metastatic breast cancer harboring BRCA1/2 germline mutations (NCCN.org).			
Rucaparib	IA	Sensitive	Ovarian cancer ✓FDA Prostate cancer ✓FDA +3

Figura 20: Visualización de la pestaña *Evidencias* en la página de variantes

Evaluations. Esta pestaña permite acceder al historial de evaluaciones e interpretaciones de la variante a lo largo del tiempo:

- En el análisis actual (sección izquierda de la pestaña)
- En los demás análisis realizados dentro del grupo (sección derecha de la ficha)

Occurrence. La pestaña identifica las demás muestras que contienen la variante, independientemente de cualquier evaluación. Para cada ocurrencia, se introduce el nombre de la muestra, la frecuencia alélica, el enlace al análisis pertinente y el nombre del proyecto.

Publications (Publicaciones). Esta pestaña proporciona acceso a la literatura científica asociada a la variante y al gen. Haga clic en la subpestaña "Variant" (Variante) o "Gene" (Gen) para ver las publicaciones de PubMed que hacen referencia a la variante o al gen. Un enlace PMID redirige al registro de PubMed, y el enlace al texto completo abre el artículo completo cuando está disponible.

Scores. La pestaña está disponible para variantes con un missense, start lost o stop gained effects. A excepción de estos casos, la puntuación aparecerá atenuada. Esta pestaña le permite ver los resultados de las puntuaciones de predicción de varias herramientas como *REVEL*, *DANN*, *FATHMM*, *FATHMM-MLK*, *LRT*, *MetaLR*, *MetaSVM*, *MutationTaster 2*, *PROVEAN*, *SIFT* y *PLI*. También proporciona acceso a puntuaciones de conservación como *GERP++*, *PhastCons100way Vertebrados*, *PhastCons20way Mamíferos*, *PhyloP100way Vertebrados*, *PhyloP20way Mamíferos* y *SiPHhy*.

IGV. La pestaña permite al usuario ver las reads alineadas con la posición de la variante a través de la versión web del IGV Genome Browser. En áreas de alta cobertura de reads, las reads se submuestran por defecto, es decir, no se muestran todas las reads. Se aplican los parámetros de visualización por defecto, es decir, un máximo de 100 reads por ventana de 50 nucleótidos. Estas regiones pueden identificarse por la línea negra situada debajo del perfil de cobertura (submuestreo). Al hacer clic en ella se muestra el número de reads eliminadas de la

visualización (# downsampled). El icono  ofrece un menú con tres opciones:

- *Abrir en nueva pestaña:* abre la vista de variantes IGV en una nueva pestaña.
- *Ajuste de pista:* permite añadir a la vista otras muestras del mismo proyecto o de proyectos diferentes.
- *Abrir IGV local:* redirige la aplicación IGV Desktop a las coordenadas de la variante que se está interpretando en ese momento. Para ver SNVs u otras variantes, la muestra BAM debe cargarse primero en la aplicación IGV Desktop. Esta opción solo es compatible con las versiones ≥ 17 de IGV Desktop.

IGV admite vistas divididas, por ejemplo, copiando dos conjuntos de coordenadas en la barra de búsqueda de locus, separadas por un espacio. Hay varios ajustes disponibles haciendo clic en

las ruedas dentadas situadas a la derecha del Navegador del genoma. Por ejemplo, para ver directamente las reads no alineadas (soft-clips), seleccione la opción "Show Soft clips".

[Gene](#). La pestaña contiene información adicional sobre el gen y la bibliografía asociada.

Generar un informe clínico

El usuario puede editar un informe clínico exportable en formato PDF y .docx haciendo clic en el botón "Generate report" (Generar informe) de la pestaña "Selection" (Selección) mientras navega por el análisis.

En la parte superior de la página de edición, el usuario puede seleccionar una plantilla, establecer el estado del informe como nuevo o borrador, seleccionar una muestra (se pueden seleccionar múltiples muestras para análisis familiares) y seleccionar el idioma de edición. El usuario también puede alternar entre la vista "Edition" (Edición) y la vista "Preview" (Vista previa) antes de guardar un borrador o pasar al siguiente paso usando los botones "Save Draft" y "Next Step" .

Las plantillas de informes se pueden crear en la sección "Reporting" bajo "Template list" en la configuración de la entidad (Entity settings).

En la vista de edición, el usuario puede editar las diferentes secciones del informe y seleccionar múltiples opciones. En la vista de vista previa, el usuario puede revisar el informe antes de continuar la edición, guardar el borrador o validar el informe.

El usuario puede guardar un borrador y acceder a él de nuevo desde la pestaña "Selection", y luego en "Generate report", como cuando se crea un nuevo informe. Cuando se genera un informe en PDF, se muestra automáticamente en la pestaña "Reporting" (Informes) del análisis. Solo después de la validación se puede exportar el informe como .PDF o .DOCX. El usuario puede eliminar el informe o crear informes adicionales para el mismo análisis.

Figura 21 : Ver pestaña del informe clínico.

13.1.5. CNV

13.1.5.1. CNV en paneles de genes

Metodología

La detección de CNVs en los datos del panel de genes, en GermlineVar, GermVar/GermVar Family, SomaVar y SomaHemato, se basa en la selección de muestras de control de la cohorte co-analizada, según su correlación.

La correlación se calcula a partir de una tabla de cobertura, que a su vez se calcula a partir de todas las regiones del manifiesto, pero excluyendo los cromosomas sexuales.

Los criterios para que una muestra sea elegible para el análisis CNV son los siguientes:

- Para que una región sea retenida en el análisis CNV, la cobertura mínima es de 50 reads.
- El valor umbral de correlación para seleccionar las muestras de control es de 0,90 para el contexto de la línea germinal y de 0,75 para el contexto somático.
- El número mínimo de controles necesarios para cada muestra es de seis.

Es necesario definir el sexo de las muestras (metadatos de la muestra) antes de iniciar CNVcapture o SomaCNVCapture para realizar la llamada de genes localizados en cromosomas sexuales. Por último, se necesitan al menos ocho muestras para realizar un análisis CNVCapture o SomaCNVCapture en SeqOne.

Los criterios aplicados al manifiesto del usuario para que un gen del panel sea elegible para el análisis de CNV son los siguientes:

- Las regiones manifiestas se anotan a partir de la transcripción canónica determinada según el recurso NCBI MANE Select. Si no se dispone del transcrito MANE Select, la anotación se basa en el recurso RefSeq del NCBI.
- sólo se analizan los exones para los que existe un solapamiento $\geq 80\%$ entre el intervalo manifiesto y el exón. Si ningún exón del gen se solapa $\geq 80\%$ con el intervalo manifiesto, el gen no se incluirá en el análisis CNV.
- sólo se analizan los intrones para los que existe un solapamiento $\geq 60\%$ entre el intervalo manifiesto y el intrón.

El análisis CNVCapture (versión ≥ 2.4) incluye regiones situadas 20kb aguas arriba y aguas abajo de los genes cubiertos por el manifiesto, si tales regiones están incluidas en el diseño del manifiesto. En concreto, cada uno de los 20 segmentos de 1kb que constituyen la región de 20kb se analiza si al menos el 20% del segmento se solapa con el manifiesto inicial.

Visualización e interpretación

La pestaña *CNVs* mostrará un gráfico que representa la lista de genes en los que se han identificado variaciones de copia. Esta pestaña está disponible en los worksets descritos en la sección Metodología.

La terminología utilizada para describir el estado de los genes llamados es la siguiente:

- Gain : el gen está afectado por una amplificación parcial o total
- Loss : el gen presenta una delección parcial o total
- Normal : el gen no presenta ninguna CNV
- Partially not called : el gen está parcialmente llamado
- Not called : el gen está excluido del análisis CNV

El estado parcialmente no llamado se asigna a un gen cuando su estado de copia es normal pero al menos un exón no ha sido llamado (con estado "failed")

A pesar de su elegibilidad para el análisis de CNV, un gen se excluye (No se llama) del análisis si :

- el gen se encuentra en un cromosoma sexual y no se han rellenado los metadatos de género,
- todos los exones fallan según los criterios especificados a continuación: número de controles, porcentaje de variación. El estado "fallido" de un gen se define sólo en CDS. Las regiones intrónicas y aguas arriba/aguas abajo se excluyen si se llama a los exones CDS.

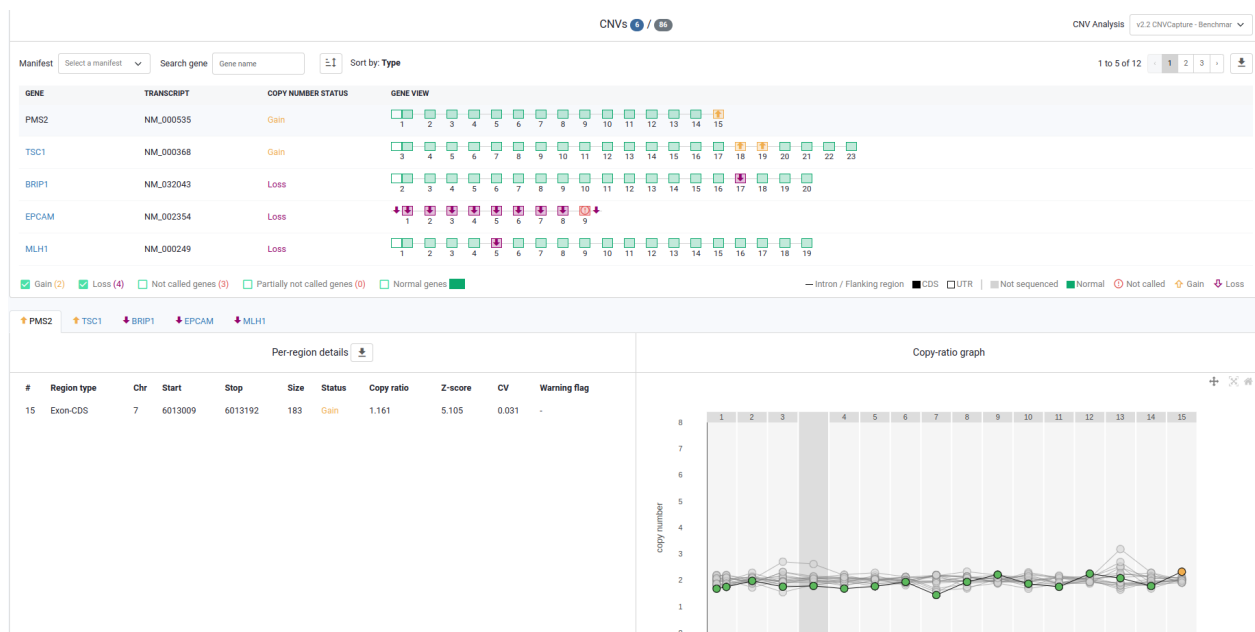


Figura 22: Visualización de la pestaña CNVs en un análisis GermlineVar.

Filtros y opciones de clasificación

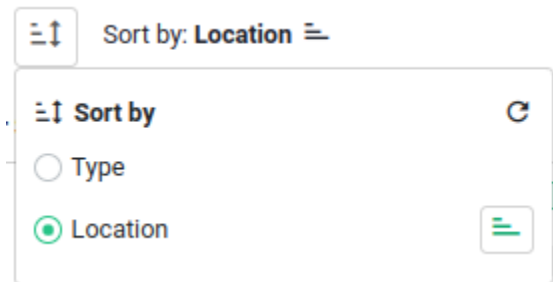
El menú de filtros puede utilizarse para:

- Restringir la visualización a un subconjunto de regiones/genes utilizando un archivo BED que se haya cargado previamente en la interfaz con la función *Manifiesto*.
- Buscar un gen específico con la función *Buscar gen*

Existen dos opciones de clasificación de CNV:

- **Tipo** (por defecto): lógica de clasificación de corriente (Ganancia y luego Pérdida)
- **Localización**: permite clasificar las CNV según los cromosomas y su posición inicial, en orden ascendente o descendente, haciendo clic en el

icono de clasificación.



El usuario también puede seleccionar los genes según su estado: Ganancia, Pérdida, Genes normales, Genes parcialmente no llamados o Genes no llamados.

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

Detalles por región

La tabla de detalles por región muestra los detalles de cada región para cada gen:

- **#**: número de región (exones e intrones)
- **Tipo de región**: tipo de región (upstream, utr-exonic, coding-exonic, intronic, downstream)
- **Chr**: cromosoma
- **Inicio/Parada**: Posición inicial y final de la CNV
- **Tamaño**: tamaño de la CNV (pb)
- **Estado**: estado de la región (ganancia = duplicación, pérdida = delección, normal = sin CNV, no llamada = región no incluida en el análisis de CNV)
- **Copy-ratio**: relación entre el número de copias y la normalidad (n=2).
- **Puntuación Z**: número de desviaciones estándar que separan un resultado de la media. Los valores positivos o negativos elevados indican una gran confianza en la llamada CNV.
- **CV**: coeficiente de variación que mide la dispersión relativa dentro de las muestras de la cohorte.
- **Bandera de advertencia**: Ocasionalmente puede aparecer un mensaje de error según los siguientes criterios:
 - **Low zscore**: Valor absoluto de la puntuación Z <4

- **High variation** : coeficiente de variación (CV) > 0,1
- **Low coverage**: cobertura de la región para esta muestra <100
- **Number of controls**: El número de controles es insuficiente para poder realizar la llamada en esta región
- **Percentage of variation**: La variación es demasiado grande entre las muestras para que esta región pueda realizar la llamada
- **Read depth**: cobertura insuficiente de la región (<50) y proporción de copias >0,1
- **Missing Gender**: la región está situada en un cromosoma sexual y no se han introducido los metadatos de sexo.

Atención : una región rica en soft-clips, como una indel grande o una inserción Alu, puede ser interpretada como una caída de cobertura por los worksets CNVCapture y SomaCNVCapture y detectada como una pérdida de copia, si el tamaño de la inserción es del mismo orden que el de la región llamada.

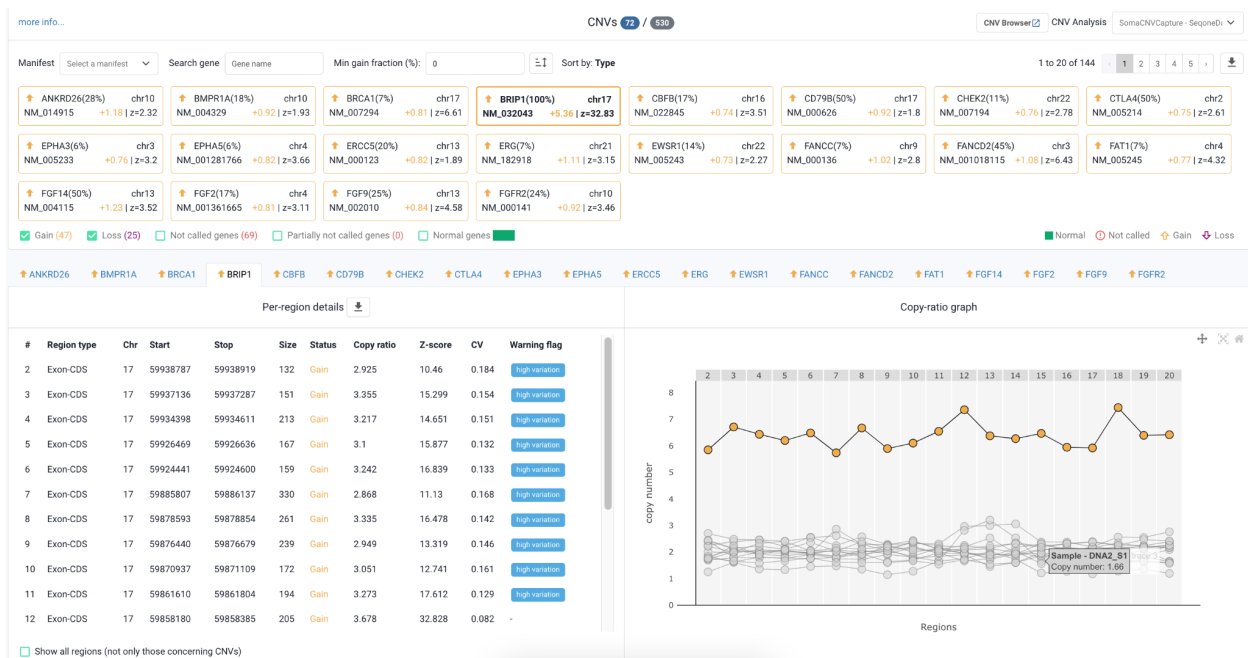
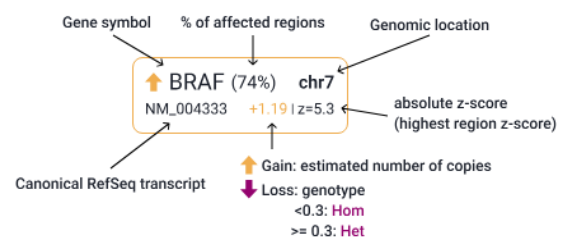


Figura 23: Visualización de la pestaña CNVs en un análisis SomaVar

En el caso de los análisis **SomaCNVCapture**, aunque el método de detección es similar al de CNVCapture, la escala utilizada para la representación gráfica de las variantes es el gen, en lugar del exón. La representación se muestra como una serie de recuadros. Esto proporciona el símbolo del gen, la posición genómica, el transcrito canónico Refseq, la puntuación z absoluta y, en el caso de ganancia, el número de copias o, en el caso de pérdida, el genotipo.



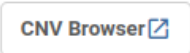
Encima de los filtros, en la sección "[more info...](#)", hay una ventana emergente que explica estas casillas.

El porcentaje que aparece en cada recuadro refleja la fracción del gen (*fracción de ganancia mínima*) afectada por la CNV detectada. La proporción de trozos puede utilizarse para filtrar los resultados. Por defecto, el filtro se establece en 0% y sólo se aplica a ganancias de copias.

Las deleciones detectadas por este pipeline no están sujetas a este filtro, y deben interpretarse con precaución. De hecho, la capacidad de esta herramienta para detectar variaciones de la copia de un gen depende de la naturaleza y la calidad de la muestra, en particular de la fracción tumoral.

Advertencia: Si se detectan tanto una amplificación como una deleción en el mismo gen, éste se considerará amplificado por defecto y, por tanto, estará sujeto al filtro de fracción de ganancia mínima descrito anteriormente.

CNV Browser

Un icono  está disponible en la pestaña de visualización CNV de los análisis GermlineVar, GermlineFamily, SomaVar o SomaHemato para abrir un CNV Browser en una nueva pestaña. El gráfico CNV Browser muestra la frecuencia alélica B (BAF) de SNP (Single Nucleotide Polymorphism) presentes en el manifiesto de la muestra analizada y presentes en un conjunto de datos de arrays SNP de Illumina y SNP de otros backbones como OneSeq de Agilent, para facilitar la visualización de ciertos tipos de eventos como LoH (Loss of Heterozygosity).

El BAF es la frecuencia del alelo no de referencia (alelo B) observada para un SNP heterocigoto de línea germinal.

Por defecto, el gráfico representa el BAF de las variantes (eje y) sobre todo el genoma (eje x), desglosado por cromosomas (1-22, X e Y). Un filtro permite al usuario visualizar las variantes de un cromosoma determinado en la esquina superior izquierda.

En el contexto de los análisis CNV, esta información BAF puede utilizarse para mostrar una alteración en el número de copias alelo-específico, pérdida de heterocigosidad o desequilibrio alélico.

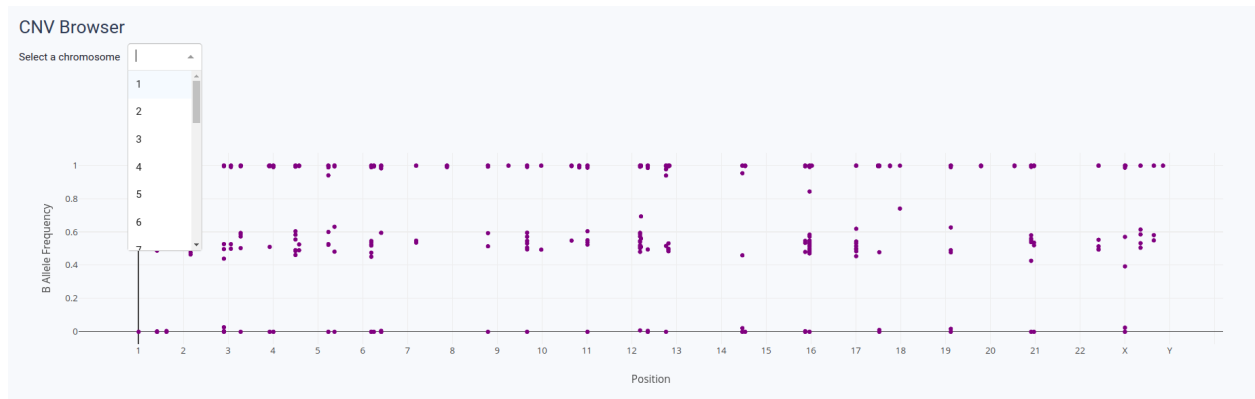


Figura 24: Visualización de la pestaña CNV Browser en un análisis GermlineVar.

Exportación de datos CNV

El icono  situado en la parte superior derecha permite exportar los datos del análisis CNV en forma de archivo .tsv que contiene las siguientes columnas:

- gen (nombre_gen),
- Transcripción RefSeq (refseq_id),
- estado del gen (status),
- cromosoma (chr),
- coordenadas genómicas de la CNV: posiciones inicial y final,
- fracción del gen afectado por la CNV (gene_fraction_cnv),
- relación entre el número de copias (copy_ratio) y la normalidad (n=2),
- número de desviaciones estándar que separan un resultado de la media (puntuación z)
- número de exones que se solapan con la CNV (exones), en el caso de un análisis CNVCapture.

Para cada gen, el icono  junto a "Per-region details" puede utilizarse para exportar específicamente los datos del análisis CNV para el gen seleccionado en forma de archivo .tsv que contiene los siguientes campos:

- gen (nombre_gen),
- cromosoma (chr),
- coordenadas genómicas: posiciones inicial y final,
- número de exón o intrón (exon_intron_number),
- estado,
- tipo (exón o intrón, tipo),
- mensajes de error asociados (warning_flags, fail_flags),
- valor crítico de probabilidad ajustado (padj),
- valor crítico de probabilidad (valor p),
- relación entre el número de copias y el normal (ratio),
- medida de la dispersión relativa dentro de las muestras (variabilidad),
- desviación estándar del resultado con respecto a la media (puntuación z)

13.1.5.2. CNV en el análisis SomaCGP y SomaLBx

Metodología

SomaCGP y SomaLBx utilizan PureCN²⁶ para detectar CNVs y LOHs (pérdida de heterocigosidad). PureCN identifica variantes somáticas a partir de los datos de detección de variantes para evaluar la ploidía y la pureza tumoral de la muestra. Esta información se utiliza para determinar el número de copias de cada evento.

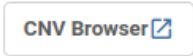
Visualización e interpretación

La pestaña de visualización de CNV en el análisis SomaVar Tertiary funciona de forma idéntica a la de SomaCGP y SomaLBx, como se describe a continuación.

Las CNV se presentan en una tabla que se puede filtrar. Cada CNV muestra la siguiente información:

- **Gene:** nombre del gen
- **Location:** cytoband
- **Event coordinates / Event size :** coordenadas y tamaño del evento
- **Transcript**
- **Exons:** Exones afectados
- **Type:** tipo de evento (LoH, Pérdida o Ganancia)
- **Copy number:** número absoluto de copias
- **Function :** función del gen si procede, oncogén o gen supresor de tumores (GST)

El usuario puede buscar CNV que afecten a un gen determinado mediante un campo de búsqueda y elegir si desea o no mostrar genes presentes en regiones LOH marcando la casilla "Mostrar genes en regiones LOH". La tabla de resultados puede exportarse utilizando el icono de descarga situado en la esquina superior derecha.

El icono  abre un *Navegador CNV* en una nueva pestaña. Esta pestaña muestra un gráfico con la frecuencia alélica B (BAF) de SNPs en la muestra analizada y SNPs LoH de publicaciones seleccionadas y el backbone OneSeq Agilent. Un segundo gráfico representa los datos logRatio generados por PureCN. Encontrará más detalles en la sección **CNV: Gene Panels** en **CNV browser**.

²⁶ <https://github.com/lima1/PureCN>

13.1.5.3. CNV en exomas y genomas

La pestaña de visualización de CNV para datos de exoma está disponible en los análisis GermlineVar, GermlineFamily, GermVar y GermVarFamily cuando se utilizan archivos FASTQ como entrada primaria y se ejecuta un análisis CNVExome. Esta pestaña también muestra los resultados de CNV para los datos del genoma en los análisis GermVar y GermVar Family utilizando archivos FASTQ, así como en los análisis GermVar Tertiary y GermVar Tertiary Family utilizando archivos VCF o datos de microarrays CGH.

Metodología para el exoma

Esta metodología se aplica específicamente a la herramienta CNVExome. Al igual que los pipelines de análisis CNV sobre datos de panel de genes, el proceso de identificación de variantes en este caso depende de un análisis de la profundidad de la cobertura. Sin embargo, el análisis CNVExome utilizando la herramienta [GATK GermlineCNVCaller](#) es individual, ya que cada muestra se compara con un modelo HHMM (*Hierarchical Hidden Markov Model*) calculado por SeqOne.

Para tener en cuenta las variaciones técnicas que conlleva la preparación de las muestras, este modelo se basa en una cohorte amplia que oscila entre uno y varios centenares de muestras secuenciadas por el laboratorio que las utiliza:

- El mismo protocolo de preparación de bibliotecas
- El mismo kit de captura
- La misma plataforma de secuenciación y longitud de read

Estos requisitos y el número variable de muestras necesarias para establecer el modelo pueden explicarse por la necesidad de una cohorte que ofrezca un nivel homogéneo de cobertura. Hay que tener en cuenta varios factores, como la profundidad de secuenciación, el grado de variabilidad dentro de la cohorte y la reproducibilidad de los protocolos de preparación de muestras.

Los metadatos de sexo (véase **Metadatos de muestra**) deben rellenarse antes de iniciar el análisis CNVExome para permitir la verificación de la ploidía de los cromosomas sexuales. Si no se cumplimenta, aparecerá un mensaje de advertencia de "Ploidía" en la interfaz de visualización de CNV.

Limitación: La documentación de GATK GermlineCNVCaller recomienda excluir las regiones PAR (PseudoAutosomal Regions), es decir, las regiones PAR1 y PAR2, si su contribución a la cobertura total del cromosoma X es desproporcionada. De hecho, la ploidía es difícil de calcular para los extremos del cromosoma X correspondientes a las regiones PAR para muchos modelos calculados por SeqOne. Como resultado, la ploidía no se calcula correctamente para

estas regiones, que se detectan de forma artificial en 1 copia, mientras que todo el cromosoma X se detecta en 2 copias.

Invitamos a los usuarios a que se pongan en contacto con el equipo de atención al usuario en support@seqone.com para obtener más información sobre el procedimiento para configurar CNVExome en su grupo.

Metodología para el genoma

En las versiones hasta la v3.6, los análisis GermVar y GermVar Family sobre datos genómicos integran la herramienta Sentieon CNVScope²⁷ para la detección de CNV. A partir de la versión v3.7, los worksets GermVar y GermVar Family utilizan la herramienta Cue²⁸ para la identificación de CNV en datos genómicos.

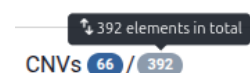
Los datos de duplicación segmentaria se detectan utilizando Sentieon's SegDup Caller, que está diseñado específicamente para identificar eventos de número de copias en regiones genómicas altamente homólogas en datos WGS únicamente utilizando GermVar. Este llamador se dirige a un conjunto seleccionado de genes clínicamente relevantes que se sabe que son difíciles para los enfoques estándar de detección de variantes. Los eventos resultantes se integran automáticamente en el pipeline de análisis y se visualizan en el visor de CNVs & SVs como CNV (tanto ganancia como pérdida), lo que permite la identificación de variantes en regiones genómicas que antes eran inaccesibles para el análisis estándar.

Visualización e interpretación

En la pestaña **CNV** hay disponible una interfaz de interpretación para los análisis CNVExome, GermVar y GermVarFamily, GermVar Tertiary o GermVar Tertiary Family. La interfaz se divide en dos secciones: una tabla de resultados de CNV a la izquierda y una página detallada de CNV con múltiples pestañas a la derecha.

Tabla de resultados de la CNV

Una tabla de resultados muestra las CNVs identificadas, junto con un menú de filtros basados en su presencia en genes OMIM mórbidos [<https://omim.org/>], tamaño, frecuencia de aparición de CNV en la cohorte, o LogRatio en el caso del análisis MicroArray CGH.



392 elements in total
CNVs 66 / 392

²⁷ <https://support.sentieon.com/appnotes/cnvscope/>

²⁸ <https://github.com/PopicLab/cue>

El número de CNV filtrados (CNV seleccionados) y el número total de CNV detectados (elementos en total) se muestran en la parte superior de la tabla de CNV.

El icono de un candado indica que el análisis utiliza un panel restringido, y los detalles de las CNV aparecen en la información sobre herramientas asociada.

0 in locked panel
392 elements in total
CNVs 0 / 0

Se muestra un mensaje de advertencia  Ploidy cuando se detecta un valor de ploidía anormal y/o cuando no se han rellenado los metadatos de género (véase **Metadatos de muestra**) antes de iniciar el análisis.

Haciendo clic en el icono "Ploidía", el usuario puede ver la ploidía de cada cromosoma y el sexo del paciente. Los valores anormales aparecen resaltados en amarillo.

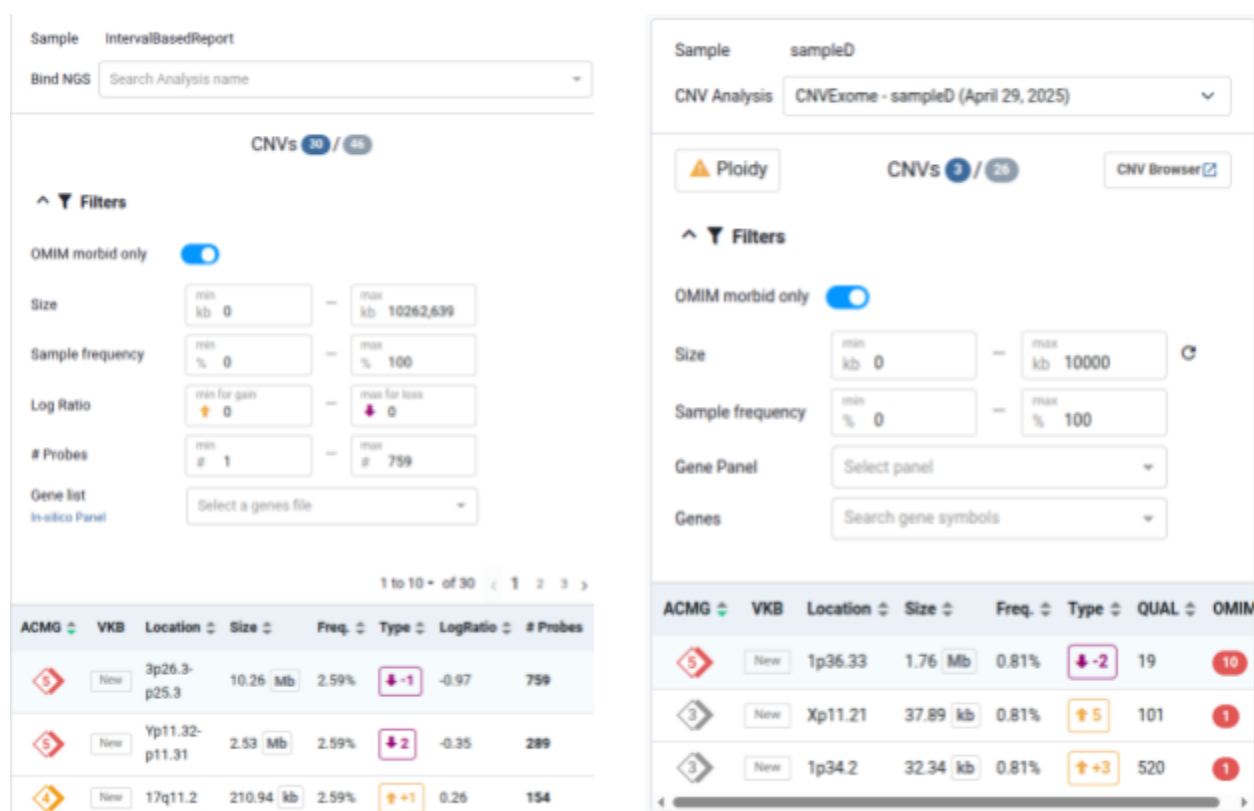


Figura 25: Visión general de las CNV detectadas y menú de filtros de un análisis CGH (izquierda) y un análisis estándar (derecha).

En el caso de un análisis CGH MicroArray combinado con un array SNP (CGH+SNP), los eventos de pérdida de heterocigosidad (LOH) detectados se reportan en la tabla de resultados CNV.

Estos eventos corresponden a una pérdida de heterocigosidad neutra (CN-LOH). En otras palabras, la región conserva dos copias, pero éstas son idénticas, caracterizadas por la presencia de SNPs homocigotos.

En el caso de un análisis CGH MicroArray, es posible seleccionar un análisis SNV no CGH (GermlineVar, GermlineFamily, GermVar Tertiary y GermVar Tertiary Family) en el menú desplegable 'Bind NGS' para mostrar las variantes que se solapan con la CNV detectada.

Filtro

OMIM morbid only. Este filtro está activado por defecto para identificar fácilmente las CNVs que afectan a genes clínicos, anotados OMIM Morbid en base a su identificador NCBI (NCBI gene_id) o, en su defecto, en base al nombre del gen y sus sinónimos (gene_symbol y gene_synonyms). Basta con levantar este filtro para visualizar todas las CNV detectadas en la muestra.

Size. El usuario puede filtrar las CNVs según su tamaño esperado (en kb).

Sample frequency. El usuario puede filtrar las CNV en función de su tamaño y frecuencia (Frecuencia de la muestra). Esta frecuencia corresponde a la aparición de la CNV entre todas las muestras analizadas hasta la fecha del análisis y comparadas con el mismo modelo, incluidas las muestras utilizadas para construir el modelo. La frecuencia de la Muestra se basa en un solapamiento del 95% y considera el mismo tipo de CNV (ganancia/pérdida/delección). Al pasar el ratón por encima de su valor mostrado como porcentaje, el usuario verá aparecer la fracción correspondiente. La frecuencia puede tardar varios minutos en actualizarse tras abrir el análisis.

Gene Panel. Al igual que un panel in silico, este filtro permite al usuario seleccionar y aplicar una lista de genes para filtrar las CNV detectadas en el análisis mediante el menú desplegable (véase la sección **Gestión de la configuración del usuario - definiciones**). Este filtro no está disponible si se ha aplicado un panel restringido al análisis desde la pestaña Panel del panel lateral de gestión del análisis (véase la sección **Vista moda Panel**).

Genes. Permite filtrar las CNV detectadas que contienen los genes seleccionados mediante este filtro.

Log Ratio. Para los análisis basados en datos MicroArray CGH, el Log Ratio (entre la señal sin procesar y la señal de control) presente en los archivos de entrada Agilent se muestra en una columna adicional. El filtro asociado permite al usuario seleccionar valores mínimos para ganancias de copias (min para ganancia) y valores máximos para deleciones de copias (max

para pérdida). Los eventos LoH (Pérdida de Heterocigosidad) se mantienen incluso después de aplicar este filtro.

#Probes. Para los análisis basados en datos de CGH MicroArray, la columna #Probes indica el número de sondas utilizadas en cada región (exónica, intergénica o intrónica). Estos valores se toman del archivo de entrada IntervalBasedReport.xls, y el filtro asociado permite al usuario seleccionar el número mínimo y máximo de sondas utilizadas. En un análisis combinado CGH+SNP, el número de sondas indica el número de SNPs homocigotos en la región LOH.

Clasificación CNV

La lista de CNVs también puede ordenarse por clase de patogenicidad ACMG (*ACMG*), posición genómica (*Location*), tamaño (*Size*), frecuencia de muestra (*Freq.*), número de copias (*Type*), puntuación de calidad (*QUAL*) en GermVar Tertiary y por LogRatio en CGH MicroArray (*LogRatio*).

ACMG. La clase de patogenicidad de cada variante se calcula de forma semiautomática siguiendo las recomendaciones del *American College of Medical Genetics and Genomics*. La clase de patogenicidad es el modo de clasificación por defecto en la tabla CNV.

Los eventos LoH (Pérdida de Heterocigosidad) en CGH MicroArray se asignan siempre a la clase 3 (VUS).

Type (Copy number). Se muestra la variación del número de copias (por ejemplo: +2, -1, +3) para las VNC detectadas en cromosomas autosómicos (1 a 22), ya que el número de copias está fijado en 2 por defecto.

Sin embargo, se muestra el número de copias absoluto (por ejemplo: 0, 1, 2) para las CNV detectadas en gonosomas (cromosomas sexuales X e Y) y no el cambio de pliegue.

A diferencia del análisis de paneles de genes, los metadatos de la muestra relacionados con el sexo no son necesarios para incluir los cromosomas sexuales en la llamada de CNV del exoma.

Para los análisis combinados CGH+SNP, los eventos CN-LOH, que conservan dos copias, se representan en la columna Tipo mediante la insignia  .

QUAL. La puntuación de calidad Phred es calculada en CNVExome por el caller de GATK. Cuando una CNV se compone de varios intervalos genómicos, se asigna a la CNV la puntuación QUAL más alta entre todos los intervalos. Para los análisis terciarios, la puntuación QUAL mostrada es la obtenida del archivo VCF de entrada (DRAGEN u ONT).

Navegador CNV

El icono abre un Navegador CNV en una nueva pestaña mostrando la frecuencia alélica B (BAF) de los SNPs (Single Nucleotide Polymorphism) presentes en la muestra analizada. El gráfico se describe en la sección anterior (CNV: Gene Panels, en CNV browser).

Página detallada de la CNV

CNV summary. Esta pestaña inicial presenta una gran cantidad de información y anotaciones sobre la CNV seleccionada, incluyendo:

- Detalles de la clasificación automática ACMG
- Las regiones clínicas y los genes incluidos en el intervalo afectado por la variante. La anotación de las regiones clínicas se basa en las bases de datos [ClinGen](#) y [OMIM](#). Por defecto, sólo se muestran los genes clínicos anotados en OMIM Morbid. Para mostrar todos los genes clínicos, el usuario puede activar la función "Mostrar todos los genes" a la derecha de la página.
- Solapamientos entre la CNV detectada y la base de datos [DGV](#).
- Las pequeñas variantes (SNVs e indels) detectadas en genes afectados por CNV. Por defecto, el filtro SeqOne-diag está activado. Para desactivar este filtro, haga clic en SeqOne-diag y anule la selección de SeqOne-diag en los perfiles de filtro. Para mostrar todas las variantes, los usuarios pueden activar la función "Mostrar todas las variantes" a la derecha de la página.
- Una sección dedicada a la base de conocimiento de variantes (VKB) para identificar variantes con más de un 95% de identidad que ya han sido validadas en el grupo y sus respectivas evaluaciones.

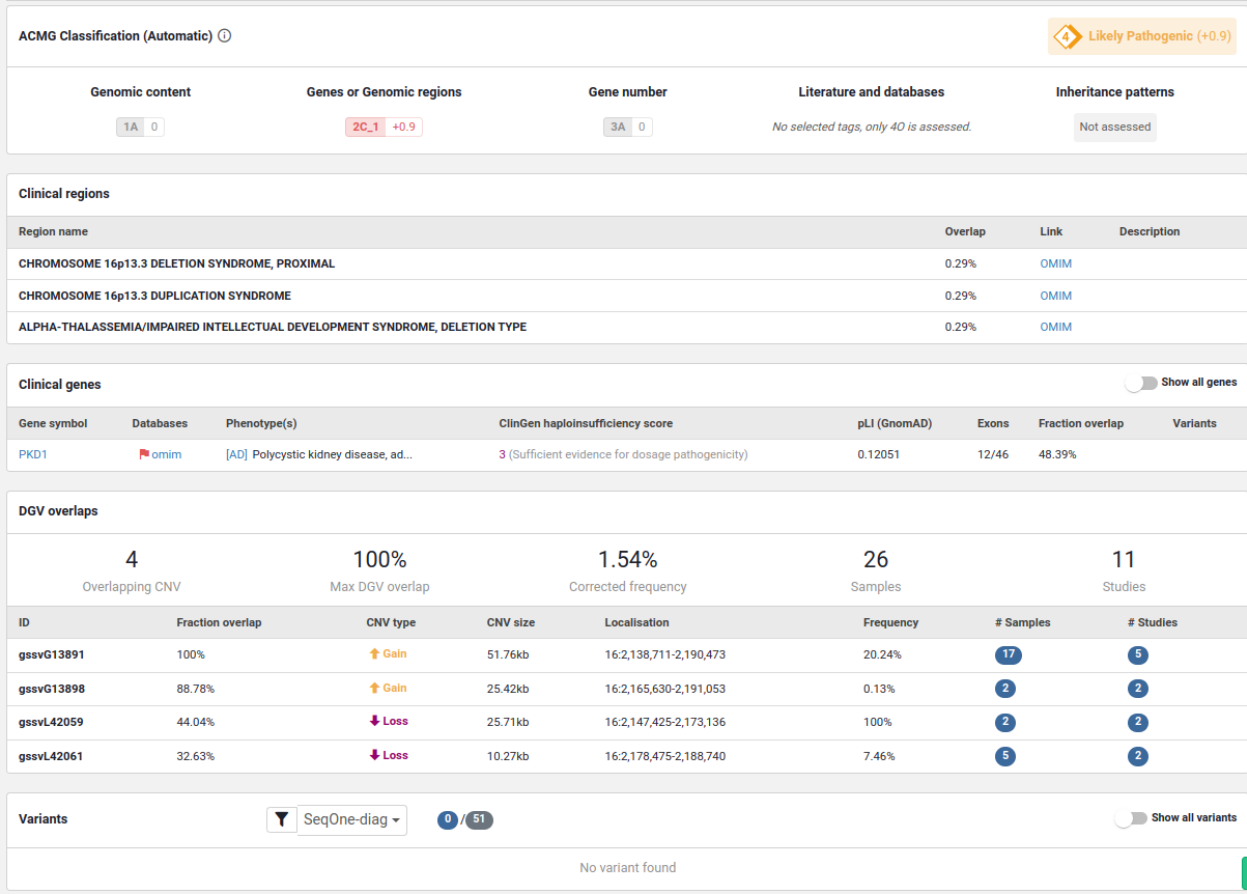


Figura 26: Visualización general de la pestaña de resumen de CNV

- El IGV del navegador Genome donde se pueden visualizar las CNV (no disponible en los análisis GermVar Tertiary y GermVar Tertiary Family). Para comprobar la estabilidad de la señal, se pueden visualizar hasta otros cinco pacientes además del individuo analizado desde el menú de configuración  del gráfico que representa el número de copias (DCN).

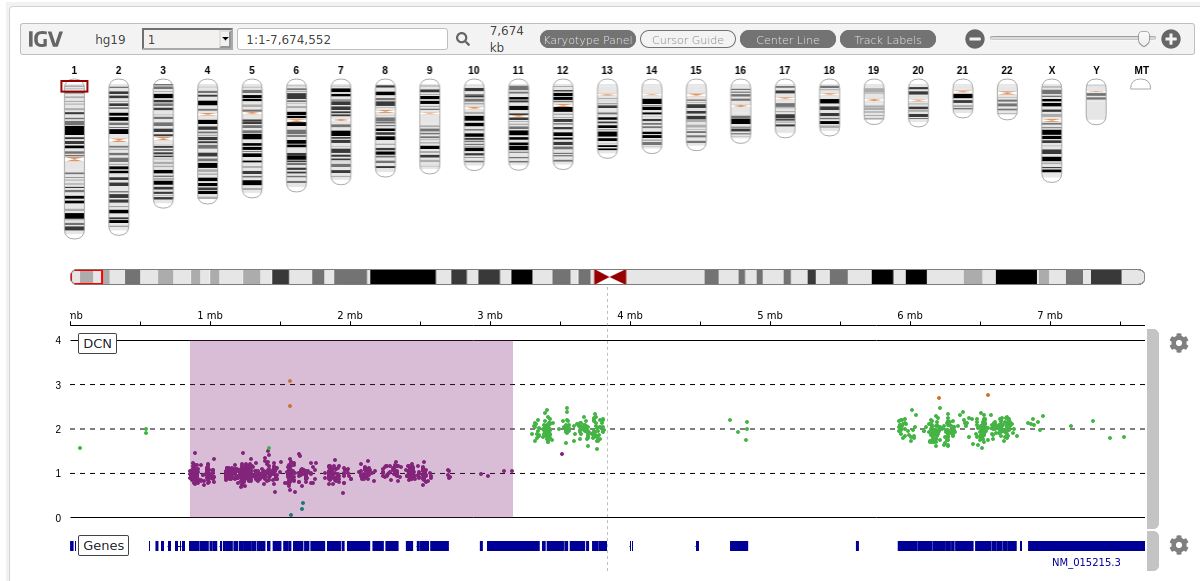


Figura 27: Visualización de las CNV desde un navegador integrado de IGV.

El código de colores es el siguiente para el paciente seleccionado: azul (de 0 a 0,5 copias), morado (de 0,5 a 1,5 copias), verde (de 1,5 a 2,5 copias), naranja (de 2,5 a 3,5 copias) y rojo (>3,5 copias). Las demás muestras aparecen en gris.

CNV Occurrences. Esta pestaña se utiliza para identificar otras muestras que contengan la CNV, independientemente de cualquier evaluación. Para cada ocurrencia, se introducirá la siguiente información:

- el nombre de la muestra (Sample),
- el nombre del análisis y un enlace al mismo (Análisis),
- el nombre del proyecto (Project),
- la localización genómica de la CNV (localización),
- el tamaño de la NVC (Tamaño),
- el tipo de CNV,
- el solapamiento de fracciones.

Los resultados se ordenan en orden descendente según el porcentaje de solapamiento. Es posible filtrar las CNV mostradas según la fracción mínima de solapamiento rellorando el campo 'min fraction overlap (%)' en la parte superior de la lista. Por defecto, este filtro está fijado en el 80%.

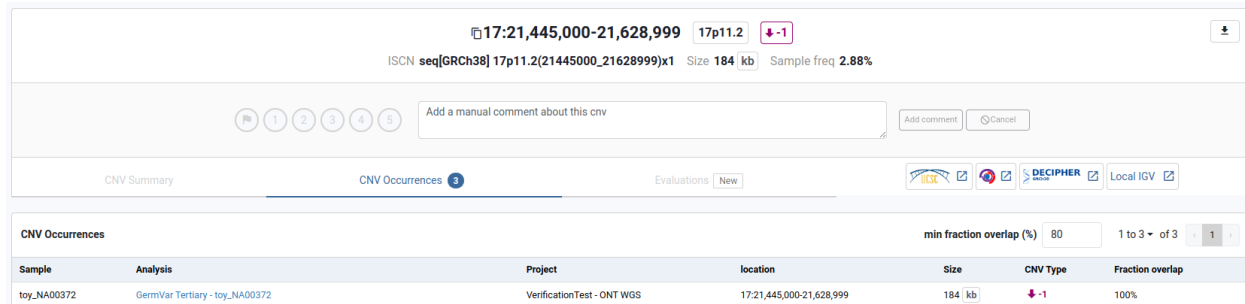


Figura 28: Visión general de la pestaña de ocurrencias de CNV en el análisis del CNVExoma.

Evaluations. El usuario puede definir la clasificación de patogenicidad de cada NVC, acompañada de interpretaciones en forma de comentarios. Cada nueva evaluación también ayuda a poblar la base de conocimientos personal del laboratorio.

UCSC. Un enlace externo al Navegador del Genoma de la [UCSC](#) permite a los usuarios ver la región genómica de la CNV. Esta página se abre en una pestaña adicional del navegador e incluye la configuración personal del usuario para ese navegador, ofreciendo así una vista personalizada.

CNV-Hub. Un enlace externo al clasificador [CNV-Hub](#) permite visualizar la clasificación de CNV teniendo en cuenta múltiples herramientas de clasificación. El enlace abre una pestaña adicional del navegador e integra la configuración personal del navegador del usuario, ofreciendo una visualización personalizada.

DECIPHER Genome Browser. Un enlace externo al Navegador del Genoma [DECIPHER](#) permite visualizar las CNVs en el genoma. El enlace abre una pestaña adicional del navegador, e integra la configuración personal del navegador del usuario, ofreciendo una visualización personalizada.

IGV local. Este enlace redirige la aplicación *IGV Desktop*²⁹ a las coordenadas del VNC que se está interpretando en ese momento. Para ver el CNV, es necesario haber cargado previamente el BAM de la muestra en la aplicación IGV Desktop. Esta opción sólo es compatible con las versiones ≥ 17 de IGV Desktop.

Además de los archivos de alineación, se pueden descargar varios archivos desde la pestaña *Archivos* del análisis CNVExome:

- Los resultados del análisis CNV como archivo tabular (cnv.tsv)
- La ploidía de cada cromosoma (contig_ploidy.tsv)
- El número previsto de copias para cada exón en el manifiesto (denoise.tsv)

²⁹ <https://igv.org/doc/desktop/>

Exportación de datos CNV



El icono  situado en la parte superior derecha le permite exportar un archivo TSV que contiene los siguientes datos para el CNV en cuestión:

- su calificación VKB y su comentario,
- Clasificación ISCN,
- Clasificación ACMG,
- cromosoma, posiciones inicial y final y citobanda,
- Tamaño y tipo de la CNV,
- número de copias
- número de copias
- frecuencia de muestreo
- puntuación de calidad (escala Phred)
- Log2Ratio
- número de sondas implicadas (Sondas)
- número de genes afectados por la CNV.

La información sobre el gen o genes asociados se describe en las siguientes columnas:

- Nombre del gen (Gene_Symbol),
- Fenotipo OMIM (OMIM_Fenotipos),
- probabilidad de intolerancia génica a LoF (Pli_Score)
- Puntuaciones ClinGen de haploinsuficiencia y triplosensibilidad (HI_Score, TS_Score)
- el número de exones solapados por la CNV (Nb_overlap_exon),
- el número total de exones del gen (Nb_total_exon).

Si algunos campos descriptivos sólo se refieren a determinadas CNV (en el caso de MicroArray CGH, por ejemplo), estos campos tienen el valor "NA".

13.1.6. Visor de variantes grandes para CNVs y SVs (LVV)

El visor de variantes grandes (Large Variant Viewer, LVV) permite al usuario visualizar y analizar CNVs y variantes estructurales (SVs) en una sola pestaña. La interfaz consta de una tabla central de variantes, un panel de filtros a la izquierda y un panel lateral que se abre a la derecha cuando se selecciona una variante, similar a la pestaña de variantes (sección Variantes pequeñas).

En los análisis compatibles, el visor se encuentra en una pestaña etiquetada como "CNVs & SVs" (CNVs y SVs) o "SVs", dependiendo del tipo de datos:

- La pestaña CNVs & SVs muestra CNVs (Pérdida/Ganancia), SVs (Delección/Duplicación) y eventos LOH/UPD.
- Pestaña "CNV" heredada: Los CNVs también se muestran en el visor heredado para mantener la compatibilidad con los flujos de trabajo existentes.
- Solo se puede acceder a los SVs en la pestaña "CNVs & SVs" o "SVs".

La disponibilidad de los datos en el visor de variantes grandes depende del tipo de análisis:

- WGS: Totalmente compatible (CNVs, SVs, UPDs).
- WES: Totalmente compatible (CNVs, SVs, UPDs). Los CNVs solo aparecen si el módulo de llamada de CNV se ejecuta como parte del análisis.
- Paneles: Parcialmente compatible (solo SVs).
- Análisis terciario (VCF/aCGH):
- aCGH: Se muestran todos los eventos de número de copias.
- VCF: Se muestran los CNVs y SVs (alelos ALT simbólicos). Los SVs con secuencias ALT específicas se muestran en el visor de variantes pequeñas.

Tabla de resultados de CNVs & SVs

Similar a la tabla de variantes pequeñas, la tabla principal muestra una fila por evento, con las siguientes columnas que proporcionan métricas específicas de la variante:

DiagAI. Puntuación de priorización de 1 a 100, adaptada para variantes grandes, que clasifica las variantes según la probabilidad de que causen enfermedades y sean diagnósticamente relevantes en el contexto del caso (especialmente cuando se proporcionan términos HPO). La puntuación DiagAI combina múltiples componentes, que incluyen: predicción de patogenicidad a través de la puntuación UP², coincidencia de fenotipos a través de PhenoGenius (impulsado por HPO) y reglas de expertos/herencia y calidad.

ACMG. El indicador de clasificación ACMG basado en variantes grandes se muestra como una puntuación de clasificación de 1 a 5 en la tabla LVV. En el cajón de detalles, una pestaña ACMG

muestra los criterios subyacentes y un resumen de clasificación en vivo de Patogénico a Benigno (principalmente para CNV Ganancia/Pérdida).

Clasificación (icono de bandera) que permite al usuario clasificar la variante en la VKB. Tipo de CNV/SV.

Tipo de CNV/SV (ganancia o pérdida), tamaño, ubicación y citobanda, Regiones (exónica, intrónica, UTR) y genes que se superponen con el CNV (nombres de genes HGNC), presencia en OMIM, fracción de genes codificantes y detalles sobre exones superpuestos. Las variantes se anotan con genes superpuestos y genes ubicados a menos de 5 kb de los puntos de ruptura (margen de 5 kb).

Genes asociados a una variante se ordenan por relevancia clínica (p. ej., estado mórbido en OMIM, porcentaje de superposición, puntuaciones de restricción). Los tres genes más relevantes se muestran en la tabla principal. La lista completa está disponible en la pestaña "Genes & Regions" (Genes y Regiones) del panel lateral ("Details Drawer" (Cajón de detalles)).

RMSK %. Porcentaje del evento que se superpone con las anotaciones de RepeatMasker (p. ej., regiones repetitivas).

ClinVar & Decipher (Bases de datos clínicas). Ambas columnas muestran el número de coincidencias de la base de datos clínica correspondiente. La mejor coincidencia se define como un evento donde la superposición de la base de datos es > 50%. Las clasificaciones de ClinVar y Decipher se mapean a Benigna (B), VUS o Patogénica (P).

DGV Gold & GnomAD (Bases de datos de población). La tabla muestra la frecuencia de población máxima encontrada para la variante. Los eventos se consideran coincidentes si la superposición de la variante es > 80% y el tipo de evento es equivalente. La pestaña Frecuencias en el panel lateral enumera todos los eventos superpuestos con una superposición > 1%.

Phenotypes (Fenotipos). Al hacer clic en los números que se muestran en las columnas correspondientes a los resultados del algoritmo Phenogenius, el usuario puede acceder a:

- el porcentaje de coincidencia, que es la puntuación de coincidencia más alta entre un solo gen afectado y los términos HPO del paciente.
- la fracción de coincidencia, que es el número de términos HPO del paciente que coinciden con al menos un gen afectado.
- Para el análisis familiar, la coincidencia se realiza utilizando los términos HPO del probando.

SampleFreq (Frecuencia de muestra interna). La frecuencia de la variante dentro del proyecto y entidad actuales se muestra en las columnas correspondientes.

- Las coincidencias se basan en > 80% de superposición de variante y un tipo de evento idéntico.

- Frecuencia de actualización: los datos se basan en una instantánea de los análisis realizados antes del inicio del análisis actual. Un botón manual de "Refresh" (Actualizar) está disponible en el Details Drawer para actualizar estos datos.

Columnas de puntuación. El usuario puede mostrar múltiples puntuaciones:

- UP²: puntuación de predicción de patogenicidad que es un componente de DiagAI.
- pLI, LOEUF: intolerancia a la pérdida de función.
- pTriplo, pHAPLO: Triplosensibilidad y Haploinsuficiencia.
- HI ClinGen, TS ClinGen.

Columnas de muestra. Estas columnas muestran métricas de eventos específicas de la fuente de datos:

- Para los análisis familiares (excluyendo aCGH), las métricas de muestra se muestran para todos los miembros de la familia.
- Genotype: para los CNVs donde el llamador de variantes no proporciona un genotipo, el sistema infiere el genotipo basándose en el número de copias absoluto relativo a la ploidía:
 - Homocigoto (hmz) con un punto completamente oscuro
 - Heterocigoto (htz) con un punto medio oscuro
- VAF / #Reads: frecuencia alélica de la variante y lecturas de apoyo (disponible para SVs).
- #Copy: número absoluto de copias (disponible para CNVs).
- LogRatio / Probes: disponible para datos aCGH.
- QUAL: puntuación de calidad del llamador de variantes o archivo de entrada.

Clasificación, Filtrado, Visualización e Interpretación

Filtrado. El panel de filtros permite al usuario filtrar variantes basándose en los datos de las columnas, similar al sistema de filtrado de la pestaña Variantes pequeñas (ver la sección correspondiente).

- Panel de genes: filtra los eventos que se superponen con genes en una lista de genes específica.
- Calidad: filtra basándose en la columna QUAL. Por defecto, los eventos sin puntuación de calidad (N/A) están ocultos a menos que se seleccione "Keep unknown" (Mantener desconocidos).
- Ajustes preestablecidos: las configuraciones de filtro se pueden guardar como ajustes preestablecidos para uso futuro.
- Para los análisis familiares, el filtro de Calidad utiliza la lógica "OR" (0). La variante se muestra si al menos un miembro de la familia cumple con el umbral de calidad.

Clasificación. Por defecto, la tabla está ordenada por DiagAI Score (descendente). El usuario puede ordenar por UP2, Size, Location o Sample Freq. Entity, en orden ascendente o descendente.

Al igual que en la pestaña CNV heredada, el usuario puede abrir las ventanas de ploidía y el navegador CNV en la parte superior de la tabla (ver sección CNV: Visualización e interpretación). También hay un icono de gestión de columnas disponible en el extremo derecho del encabezado.

Panel lateral CNV/SV (Details Drawer)

Al igual que en la página de variantes (sección Variantes pequeñas), al seleccionar una variante en el Large Variant Viewer se abre una página detallada en el lado derecho de la pantalla. Esta página contiene múltiples pestañas:

Variantes. Esta pestaña de resumen incluye la nomenclatura de la variante, la puntuación DiagAI, la lista de heterocigotos compuestos sospechosos (Suspected Het Compound), información sobre el gen y la región, y enlaces a bases de datos.

ACMG. Clasificación detallada.

Genes y Regiones. La lista completa de genes y regiones asociados a la variante, con información clínica detallada como el estado mórbido OMIM, porcentaje de superposición, puntuaciones de restricción y coincidencia de fenotipos.

Clasificación. Clasifique la variante actual en su VKB en la parte superior del panel lateral y guárdela. La clasificación actual y las ocurrencias de esta variante en otros análisis se pueden encontrar en esta página.

Frecuencias. El usuario puede acceder a los detalles sobre las frecuencias de las variantes en las bases de datos DGV Gold y GnomAD, así como para la entidad y el proyecto actual.

Bases de datos clínicas. ClinVar & Decipher.

Viewer. El usuario puede elegir ver la variante en el SeqOne viewer o en IGV. En la subpestaña IGV, el usuario puede ver el "Start" (Inicio) o el "End" (Fin) del evento, o "Start to End" (De inicio a fin). En la subpestaña SeqOne, el usuario puede acceder al Genome Browser y visualizar el evento en una de las múltiples capas disponibles: CNV, SV, UPD, Gene, CN (copy number), BAF, DGV.

13.1.7. Firmas en el análisis SomaCGP

Metodología para calcular la puntuación del MSI

El análisis MSI es un análisis individual realizado frente a una referencia normalizada. Esta referencia consiste en un conjunto de muestras negativas previamente seleccionadas y agrupadas. La muestra tumoral se compara con esta referencia utilizando la herramienta *MSIsensor*. A continuación se calcula una puntuación, representada por el porcentaje de regiones de microsatélites inestables cubiertas por el panel.

El panel de normales actualmente soportado por SomaCGP corresponde al panel Agilent SureSelect Cancer CGP. El uso de otro kit o de un diseño personalizado derivado del CGP requerirá la creación de un panel dedicado de normales.

Metodología para calcular la puntuación TMB

El TMB se calcula en tres pasos:

1. Cálculo del tamaño de las regiones codificantes incluidas en el panel
2. Identificación de variantes somáticas
3. Cálculo del número de mutaciones somáticas por megabase (mut/Mb)

La identificación de variantes somáticas se basa en un conjunto de filtros que se describen a continuación:

- Tipo de variante y filtros de efectos
 - Sólo SNV (se excluyen los MNV)
 - Efectos retenidos: ganancia de codón de parada, cambio de marco de read, variante de splicing (aceptor y donante), ablación de transcripción, delección/inserción en fase, variante missense, variante sinónima.

Las variantes sinónimas se excluyen del cálculo de la puntuación TMB no sinónima.

- Variantes de los filtros de calidad
 - Variantes detectadas por Mutect2 que pasan los filtros FilterMutectCalls
 - Profundidad de cobertura (DP) $\geq 50X$
 - Índice de calidad > 30
 - Ratio de apoyo alélico (ASR) $> 0,1$
- Filtros de variantes somáticas
 - Frecuencia alélica variante (FAV) entre el 5% y el 40%.
 - Frecuencia de la población general (GnomAD) $\leq 0,0001$

Riesgos del uso de muestras de control sintético de TMB:

- Distribución de mutaciones no naturales:
 - Las mutaciones creadas pueden no reflejar el contexto genómico de los tumores reales (por ejemplo, agrupación de mutaciones, sesgo de hebra, contexto de secuencia).
- Falta de heterogeneidad tumoral:
 - Las muestras sintéticas no reproducen la diversidad clonal ni la pureza tumoral, lo que afecta a la distribución de la frecuencia alélica variante (VAF).
 - Sin ruido de fondo ni artefactos FFPE: Las muestras reales presentan artefactos de secuenciación, daños FFPE y variaciones en el número de copias, lo que no ocurre con los materiales sintéticos.
- Diferente tamaño de fragmento y comportamiento químico:
 - Las construcciones sintéticas suelen comportarse de forma diferente durante la preparación de bibliotecas, la captura de híbridos y la secuenciación.

Visualización e interpretación

La pestaña *Firmas* muestra los resultados de los análisis MSI y TMB en 2 secciones:

Microsatellite Instability. Esta vista muestra información como :

- el número de regiones de microsatélites con una cobertura superior o igual a 50X (Sitios por encima de 50X de cobertura),
- el número de sitios somáticos,
- Puntuación MSI en porcentaje e intervalo de confianza (IC) del 95%,
- Estado del ICM

Tumor Mutational Burden. Esta vista muestra la siguiente información:

- el tamaño de la región codificante expresado en Mb (tamaño de la región codificante),
- el número de variantes somáticas y no sinónimas que pasan los filtros (Variantes somáticas que pasan los filtros y Variantes somáticas no sinónimas que pasan los filtros),
- la puntuación TMB de todas las variantes somáticas que pasan los filtros en mut/Mb (Puntuación TMB),
- la puntuación TMB de las variantes no sinónimas que pasan los filtros mut/Mb.

Cada resultado puede descargarse haciendo clic en el icono  situado en la parte superior derecha de cada sección como archivo tsv.

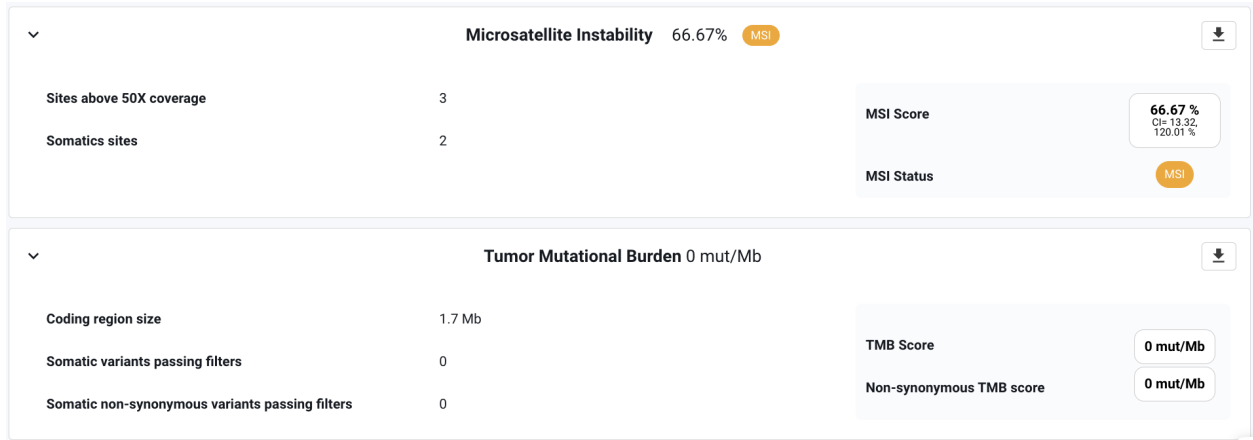


Figura 29: Visualización de la pestaña Firmas de un análisis SomaCGP.

13.1.8. Repeticiones cortas en tándem (STR)

La visualización de las repeticiones cortas en tándem (STR) está disponible para los workset de análisis GermVar Tertiary y GermVar Tertiary Family dedicados al análisis de datos de secuenciación del genoma completo (WGS), cuando se generan a partir de la secuenciación de Oxford Nanopore Technologies.

La detección de repeticiones en tándem cortas (STR) también está disponible en los conjuntos de trabajo de análisis GermVar y GermVar Family para WGS a partir de datos FASTQ generados con Illumina, Element Biosciences, MGI y Oxford Nanopore Technologies.

Metodología

Repeticiones en tándem cortas a partir de FASTQ

Los datos de repetición en tándem corta (STR) se detectan utilizando ExpansionHunter, que procesa el archivo BAM junto con un genoma de referencia y un catálogo de variantes que define los locus objetivo. El VCF resultante es luego anotado por Stranger, que clasifica cada locus como Normal, Premutación o Patogénico en función del recuento de repeticiones observado y los umbrales de enfermedad conocidos.

Ambas herramientas están integradas directamente en los pipelines GermVar y GermVar Family.

ExpansionHunter: <https://github.com/Illumina/ExpansionHunter>

Stranger: <https://github.com/moonso/stranger>

Repeticiones en tándem cortas a partir de VCF

Los datos de las repeticiones en tándem se extraen del archivo VCF (sample.wf_str.vcf.gz) generado a partir del flujo de trabajo wf-human-variation v2.3.0 de la aplicación EPI2ME (Oxford Nanopore Technologies) y suministrado por el usuario. Como estos datos han sido anotados para cada locus portador de STR, a partir de un catálogo de variaciones mantenido por EPI2ME, nuestro pipeline no modifica las anotaciones. El estado de premutación o patogenicidad de cada alelo se determina en función del número de repeticiones secuenciadas y de los umbrales de premutación y patogenicidad conocidos.

Visualización e interpretación

La pestaña STRs le permite ver las repeticiones cortas en tándem en forma de tabla mostrando los siguientes detalles para cada locus genotipado:

Información sobre locus y anotaciones

- [Gene](#): nombre del gen (HGNC) y enlace correspondiente.

- **Site**: las coordenadas genómicas de la repetición.
- **Repeat Unit**: motivo de nucleótidos repetidos.
- **Disease**: la enfermedad asociada según la base de datos OMIM y el enlace correspondiente.
- **Transmission**: el modo de transmisión conocido.
- **Premutation range**: el rango de premutación correspondiente al número mínimo y máximo de repeticiones que determinan el estado de premutación.
- **Pathogenic range**: umbral inferior de repeticiones que determina el estatus patogénico.

Detalles específicos de cada muestra

- **GT**: el genotipo homocigótico, bialélico o heterocigótico del locus.
- **Allele repeats**: número de repeticiones por alelo. Los loci homocigóticos llevan un único valor. Para los loci bialélicos y heterocigotos, se indica el número de repeticiones para ambos alelos. El estado de cada alelo viene determinado por el número de repeticiones: tipo salvaje en verde, premutado en naranja y patógeno en rojo.
- **Supp. reads**: número de reads de apoyo para cada alelo. Los loci homocigóticos llevan un único valor, mientras que los valores para 2 alelos se muestran para los loci bialélicos y heterocigóticos.

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: Visualización de la pestaña STRs de un análisis GermVar Tertiary.

La tabla puede descargarse como archivo repeats.tsv a través del icono  situado en la parte superior derecha de la tabla. Este archivo también está disponible en la pestaña Archivos del análisis.

13.1.9. Genes de fusión

13.1.9.1. Fusiones detectadas a partir del ADN

Metodología

El módulo [GRIDSS](#) (Genome Rearrangement IDentification Software Suite), especializado en la detección de reordenamientos genómicos, e implementado en el workset SomaVar y SomaVar LF, se utiliza para detectar fusiones génicas a partir de secuenciación de ADN, enriquecidas por captura dirigida a un panel de genes

Visualización e interpretación

La pestaña Fusiones de un análisis somático (SomaVar, SomaVar LF, SomaCGP, SomaVar Tertiary) muestra la lista de fusiones que superan los distintos filtros GRIDSS en forma de tabla que detalla información sobre la fusión y cada socio implicado:

Información sobre la fusión

- fusión con los símbolos genéticos de cada socio,
- el tipo de evento (translocación, duplicación, inversión o delección),
- la aparición de la fusión en la base de datos COSMIC,
- la frecuencia alélica (FAV),
- el número de reads que apoyan la fusión,
- cobertura total,
- filtros asignados por GRIDSS.

Información relacionada con los socios

- las coordenadas genómicas del punto de fusión (breakend),
- el transcrito RefSeq implicado,
- el sitio correspondiente al punto de fusión (breakpoint): 5'UTR, 3'UTR, Intrón, Exón (punto de ruptura no codificante), CDS (punto de ruptura codificante) o sitio de splicing,
- la numeración del exón o intrón implicado en la fusión, en relación con el número total de exones o intrones.

Un menú de filtros le permite

- buscar una fusión específica o un gen implicado en una fusión mediante una barra de búsqueda,
- mostrar las fusiones que no han pasado los filtros GRIDSS (LOW_QUAL) marcando la casilla 'Mostrar fusiones filtradas'. Esta casilla no está marcada por defecto.

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: Visualización de la pestaña Fusiones para un análisis SomaVar o SomaVar LF.

Los datos de la tabla pueden ordenarse según las coordenadas genómicas (breakend), la presencia de la fusión en COSMIC (Databases), la frecuencia alélica (VAF) y el número de reads que apoyan la fusión (Supporting reads).

El enlace asociado a las coordenadas genómicas de cada punto de rotura abre una vista dividida del punto de fusión en IGV. Los tooltips detallan el número de *split reads* y pares no coincidentes en la columna Supporting reads, y la cobertura en cada punto de rotura (Cov. 1 y Cov. 2) en la columna Coverage.

La columna Filtro muestra 3 filtros de calidad utilizados por GRIDSS y da una indicación de la confianza en la llamada de fusión:

- PASS: la fusión ha superado los filtros de calidad. La puntuación de calidad es mayor o igual a 500 y se considera una llamada de confianza media o alta (si > 1000).
- LOW_QUAL: la fusión tiene una puntuación de calidad inferior a 500 para al menos una de las dos rupturas implicadas en la fusión, y se considera una llamada de baja confianza.
- ASSEMBLY_ONLY: la fusión se llama sobre la base de un ensamblaje y no sobre la base de la read que lo soporta. El ensamblaje se refiere a la reconstrucción de secuencias alrededor de posibles puntos de fusión, que luego se vuelven a alinear para identificar la fusión.

Si hay varios filtros disponibles para la misma fusión, el signo "+" muestra todos los filtros en la columna Filtro.

Además de los resultados presentados en la interfaz, se puede descargar un archivo fusions.tsv

que detalla las fusiones detectadas desde la pestaña Archivos o desde el icono  situado en la parte superior derecha de la página. Contiene la siguiente información

- nombre de la fusión con símbolos de genes para cada socio (fusión)

- coordenadas genómicas de cada punto de fusión (breakend_1 y breakend_2)
- transcripción canónica para cada socio (transcript_canonical_1 y 2)
- sitio correspondiente al punto de fusión para cada socio (sitio_1 y sitio_2)
- numeración de los exones o intrones según el transcrito RefSeq asociado (exon_1 y 2)
- 60 pb de secuencias flanqueantes que rodean el punto de fusión (flanking_region_1 y 2)
- filtros de calidad asignados por GRIDSS (filtro)
- tipo de evento (type)
- número de split reads (split_reads)
- número de pares de reads discordantes (discordant_pairs)
- cobertura para cada socio (cobertura_1 y cobertura_2)
- fracción de reads que apoyan la fusión (read_fraction)
- número de reads que apoyan la fusión (supporting_reads)
- cobertura total (cobertura)
- Identificador de fusión COSMIC (cosmic_id)

Este archivo contiene :

- Fusiones mostradas en la interfaz, habiendo pasado los filtros de calidad (filtro = PASS),
- Fusiones detectadas pero con una puntuación de calidad inferior a 500 para al menos una de las dos rupturas implicadas en la fusión (filtro = LOW_QUAL).

La puntuación de calidad de cada breakend se encuentra en el archivo gridss.raw.vcf.gz (columna QUAL), descargable desde la pestaña Archivos

13.1.9.2. Fusiones detectadas a partir de ARN

Herramienta SomaRNA para el análisis de datos RNA-seq procedentes de enfoques dirigidos como la captura y el amplicón, lo que significa que los usuarios pueden identificar y visualizar eventos de fusión de genes a nivel de transcripción.

Metodología

Se utiliza una combinación de dos herramientas para detectar y visualizar fusiones por SomaRNA:

- [STAR](#): un alineador de datos de ARN-seq diseñado para alinear secuencias no contiguas directamente con el genoma de referencia.
- [Arriba](#): detecta fusiones de genes a partir de alineaciones quiméricas

A partir de las alineaciones quiméricas de STAR, Arriba aplica un conjunto de filtros para eliminar los artefactos conocidos y las transcripciones observadas en contextos no

patológicos. Asigna una puntuación de confianza a cada posible evento de fusión que depende de varios criterios:

- El número de secuencias que apoyan la fusión
- El equilibrio entre split reads y discordant mates
- La distancia entre los puntos de ruptura, su proximidad (intragénica o no)
- El tipo de acontecimiento

Las fusiones detectadas con una confianza media o alta por Arriba se muestran en SeqOne. Las que tienen una confianza baja aparecen en los archivos "arriba_discarded" (véase más abajo). La confianza de Arriba refleja la probabilidad de una transcripción aberrante no observada en tejido sano, la posibilidad de un reordenamiento genómico subyacente y la exclusión de artefactos.

Más detalles sobre el nivel de confianza en la documentación de Arriba : <https://github.com/suhrig/arriba/blob/master/documentation/08-Interpretation-of-results.md>

Para saber por qué una fusión es de "Baja confianza", consulte la columna "filtros" (AA) del archivo "descartados". Indica el filtro que descartó el evento. Detalles de cada filtro : <https://github.com/suhrig/arriba/blob/master/documentation/11-Internal-algorithm.md>

Visualización e interpretación

Desde la pestaña Fusiones de un análisis SomaRNA, el usuario obtiene una lista de predicciones de fusión, en forma de diferentes pestañas.

En la parte superior de la página figuran los detalles de la fusión y de cada uno de los socios participantes:

Información general sobre la fusión

- el número de spanning junction reads, igual a la suma de las respectivas splits reads de cada uno de los genes 5' y 3' implicados en la fusión,
- el número de discordant mates,
- la puntuación de confianza asignada por Arriba (Alta, Media o Baja),
- el tipo de evento que da lugar a la fusión (translocación, duplicación, inversión o delección),
- el impacto de la fusión en el marco de read,
- en su caso, la aparición de la fusión en la base de datos COSMIC.

Información sobre cada socio de fusión

- el símbolo de cada gen,

- la transcripción Refseq o, en su defecto, ENSEMBL para cada gen,
- la numeración de los exones implicados en la fusión,
- las coordenadas genómicas de cada punto de ruptura,
- la cobertura total en el punto de fusión (Cobertura),
- la fracción de read que apoya la fusión.

Las split reads son reads alineadas predominantemente en el gen 1 o en el gen 2 y que solapan la fusión. En el caso de pares de reads discordantes, una de las reads se encuentra en el gen 1 y la otra en el gen 2, sin solapar necesariamente la fusión.

La read fraction es la suma de las split reads y los pares de reads discordantes asociados a un gen, dividida por la cobertura total. El valor de la read fraction puede ser superior al 100% cuando las split reads asociadas a un gen también forman parte de pares de reads discordantes. Estas reads se cuentan dos veces en lugar de una.

DHRX - TCEANC E3 - E4		POLR2J2 - EFCAB13 E3 - E18		POLR2J2 - EFCAB13 E4 - E18		MEGF8 - CNFN E3 - E4	
General informations				Partners informations			
# of spanning junction reads		568					
# of discordant mates		246					
Confidence 		high					
Fusion type		translocation					
Reading frame		in-frame					
Database		 cosmic					

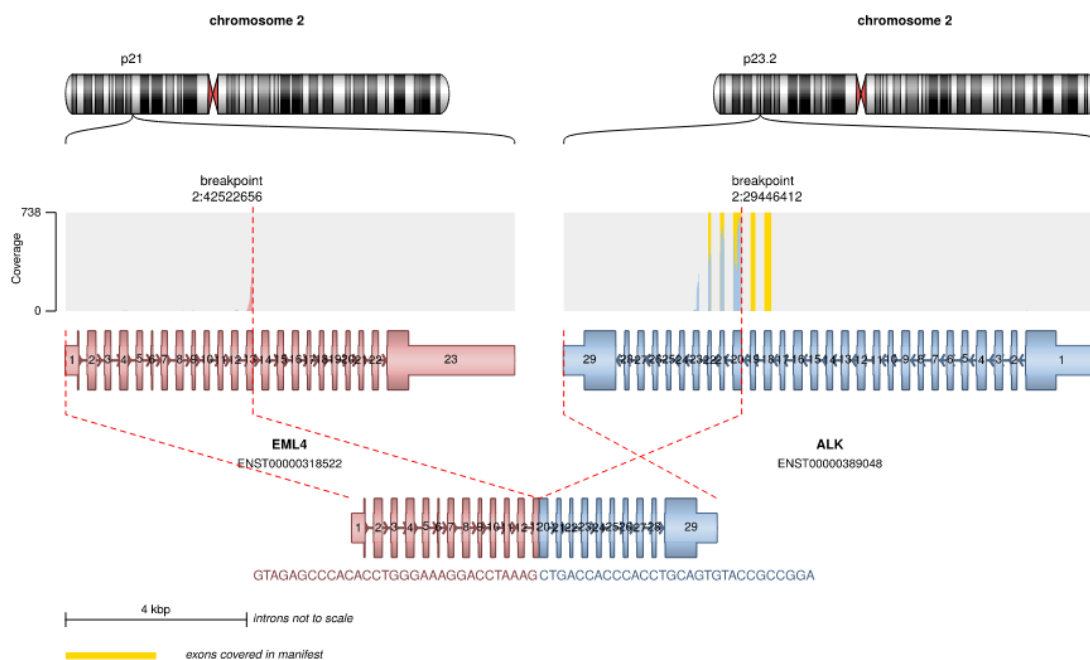


Figura 33: Visualización de un evento de fusión entre dos genes

- Los dominios proteínicos retenidos tras el evento de fusión se resaltan en una segunda vista:

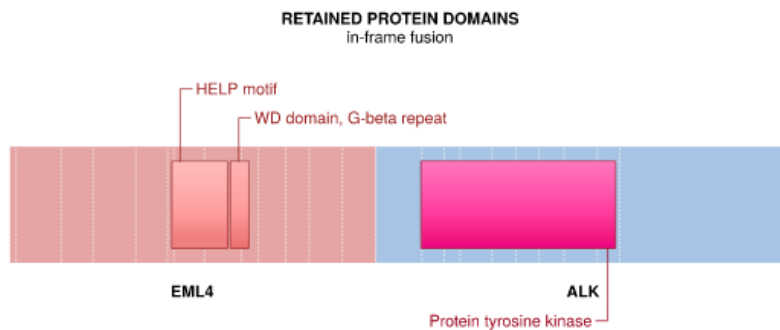


Figura 34: Visualización del polipéptido resultante de una fusión entre dos genes.

Además de los resultados presentados en la interfaz, se pueden descargar varios archivos:

- Las fusiones retenidas por Arriba en formato TSV (archivo arriba.tsv),
- Las fusiones eliminadas por los filtros Arriba en formato TSV (archivo arriba_discarded_standard.tsv),
- las fusiones eliminadas por los filtros Arriba y listadas en COSMIC en formato TSV (archivo arriba_discarded_cosmic.tsv),
- las diferentes figuras que aparecen en la interfaz en formato SVG

13.1.10. Clonality

El "IGHV mutational status report" en la pestaña "Clonality" del análisis de SomaHemato muestra los resultados del estado de mutación.

La sección 'Resultados IGHV' muestra:

- el estado de mutación IGHV y el porcentaje de homología:
 - Mutated: si la homología es <98% de identidad germinal
 - Unmutated: si la homología es ≥98% de identidad con la línea germinal
 - Borderline: para clones con entre 97 y 98% de identidad.
 - Polyclonal: ningún clon importante por encima del 20%.
 - Non-Conclusive: si falla uno de los controles de calidad esenciales
 - El subconjunto determinado (con impacto clínico) y la confianza: *2 (High risk), *8, o None
- Una notificación de advertencia si aparecen advertencias en el informe.

La sección "Clonality" enumera los clones identificados con su frecuencia y su estado de productividad (Productive, Un-productive). Cada clon tiene una sección dedicada en la que se indica su estado de mutación, frecuencia, enlace a la página IMGT_quest y detalles adicionales (homología, integridad), así como los controles de calidad.

13.1.11. Variantes de splicing

Metodología

El workset SomaRNA utiliza la suite de herramientas *Regtools* para detectar saltos de exón e identificar variantes de splicing alternativas a partir de la secuenciación de ARN, enriquecida por captura dirigida a un panel de genes. RegTools extrae las uniones exón-exón del archivo BAM generado por el alineador de datos STAR RNA-seq, y las anota según una estructura de transcripción conocida basada en un archivo GTF y anotaciones Ensembl.

Visualización e interpretación

La pestaña de variantes de splicing de un análisis SomaRNA se centra en la visualización de las variantes de splicing de los siguientes genes clínicamente relevantes y su principal variante accionable:

- MET (MET Proto-Oncogene, Receptor Tirosina Quinasa) y la variante accionable MET Δ 14
- EGFR (Receptor del Factor de Crecimiento Epidérmico) y la variante de splicing accionable EGFRvIII
- AR (Receptor de Andrógenos) y la variante de splicing accionable AR-V7

La lista completa de variantes de splicing detectadas en el análisis SomaRNA, además de las enumeradas anteriormente, puede descargarse directamente como archivo TSV a través del icono Descargar de la interfaz o desde la pestaña Archivos del análisis.

Para cada gen mencionado anteriormente, se muestra el estado de detección (Detectado o No detectado) de la principal variante procesable, junto con el número de otros eventos alternativos detectados para ese gen.

Los usuarios pueden ver los detalles de las principales variantes de splicing detectadas abriendo un inserto específico, dividido en tres partes:

- A la izquierda, se muestra información sobre la principal variante de splicing procesable: coordenadas genómicas, nombre del gen, transcripción RefSeq, tipo de variante detectada detallando los exones que faltan, fase de read, fracción de reads y número de reads que apoyan la variante. Por último, se especifica el promedio de reads canónicas de apoyo. Esto corresponde a la media de las reads 5' y 3' que apoyan la inclusión del exón.
- a la derecha, un gráfico de sashimi (Junction Plot) ilustra las uniones detectadas entre exones, así como el número de reads que apoyan cada unión.
- en la parte inferior, se enumera la información sobre los eventos alternativos detectados, es decir, isoformas distintas de la variante principal de interés descrita anteriormente: tipo de variante detectada (números de exón que faltan entre paréntesis), fase de read, fracción y número de reads que soportan la variante, número medio de reads que soportan el transcrito canónico (Avg. canonical support), estado PASS o FAILED según los filtros descritos a continuación.

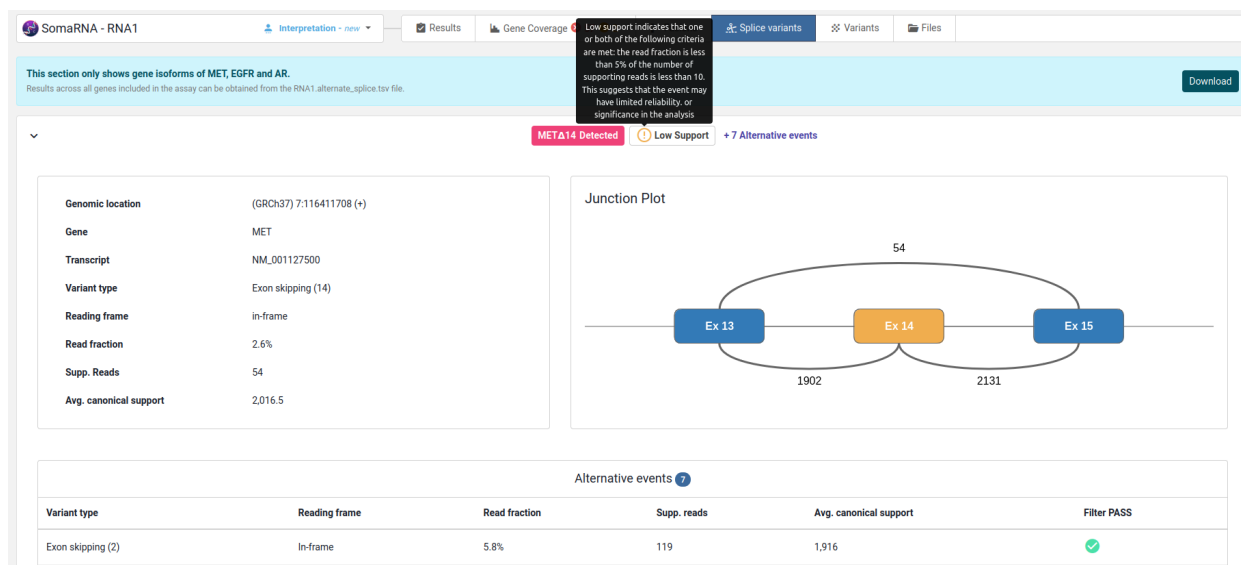


Figura 35: Visualización de la pestaña *Splice variants* de un análisis SomaRNA.

La detección de variantes de splicing en los genes mostrados en la pestaña Variantes de splicing no está sujeta a la aplicación de filtros de detección.

No obstante, un icono de "Low support" asociado a la detección de MET Δ 14 y EGFRvIII indica al usuario que las métricas relacionadas con las supporting read están por debajo de los siguientes umbrales: read fraction < 5% o número de supporting read < 10.

El estado PASS o FAILED de otros eventos alternativos también utiliza estos umbrales:

- **PASS** : fracción de reads $\geq$ 5% y número de reads $\geq$ 10.
- **FAILED** : fracción de reads < 5% o número de reads < 10.

El archivo TSV descargable detalla la siguiente información para todos los eventos de splicing detectados: Transcripción RefSeq, gen, dirección de la hebra (+ o -), fracción de reads y número de reads que apoyan la variante, número de exones que faltan y números de exón correspondientes, fase de read, coordenadas genómicas de los empalmes. El archivo también especifica en FILTER_PASS (sí/no) si la variante supera los umbrales aplicados a las reads de apoyo (fracción de reads $\geq$ 5% y número de reads $\geq$ 10) y en Actionable (sí/no), la accionabilidad de la variante.

13.1.12. Estado MSI/MSS

El workset de análisis SomaMSI está disponible cuando se ejecuta un nuevo análisis sobre datos de panel de genes y se utiliza para determinar el estado MSI o MSS de una muestra tumoral.

Metodología

Este análisis de cohortes no individuales es similar a un análisis CNVCapture o SomaCNVCapture. Se necesitan al menos ocho muestras para realizar el análisis.

La herramienta genera una muestra de referencia a partir de las muestras de la cohorte que se estima que son negativas. Como este estado de las muestras no se conoce de antemano, esta estimación se basa en el cálculo de la tasa de deleciones observadas en las regiones de microsatélites: la mitad de las muestras de la cohorte que presentan la tasa más baja se seleccionan, pues, para generar una muestra de referencia, y cada muestra aporta así una fracción de sus secuencias.

Posteriormente, cada muestra tumoral se compara con la muestra de referencia mediante la herramienta MSIsensor. Se calcula una puntuación, que representa el porcentaje de regiones de microsatélites estudiadas que presentan inestabilidad. Las regiones estudiadas son las regiones de microsatélites cubiertas por el panel y que son detectadas por el MSIsensor.

A partir de esta puntuación, el estado MSI o MSS de las distintas muestras viene dictado por un umbral que se calcula del siguiente modo: $\max(10, \text{mediana} + \max(3 \cdot \text{mediana} - \text{abs-dev}, 0, 3 \cdot \text{mediana}))$

El principio en el que se basa esta puntuación es el siguiente:

- En un análisis que presente muestras MSI y MSS, el umbral será generalmente la mediana + 3*mediana-abs-dev.
- Sin embargo, si todas las muestras tienen el estado MSS, si las puntuaciones son muy bajas y si su distribución es extremadamente próxima, el umbral se fijará en 10 para evitar generar falsos positivos.

Visualización e interpretación

Una vez finalizado el análisis, se pueden ver los resultados de todas las muestras desde el análisis SomaMSI. La pestaña Resultados se divide en dos secciones:

MSI graph : representación gráfica de los resultados en la que cada muestra se muestra en función de su puntuación. El umbral determinado por la cohorte de muestras crea dos zonas en el gráfico que representan las muestras consideradas MSI y MSS respectivamente.

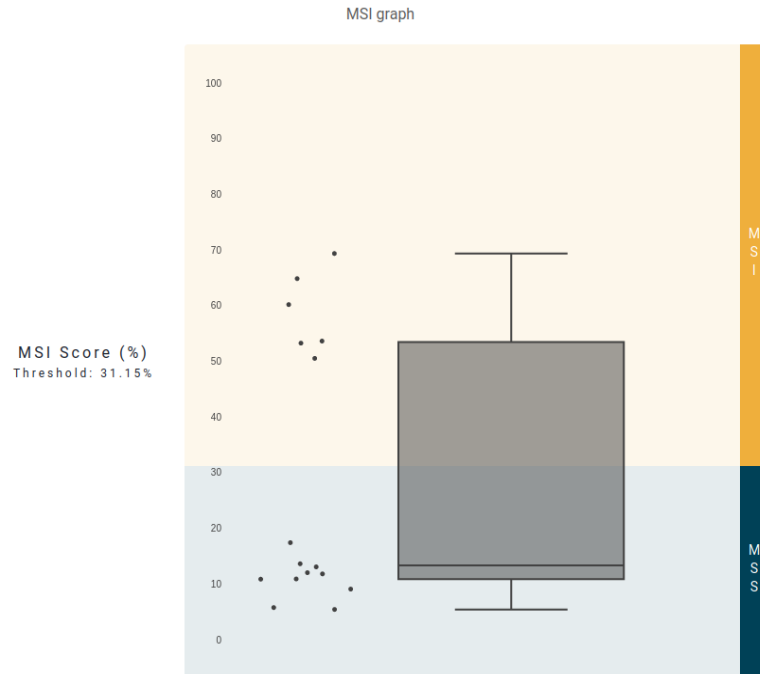


Figura 36: Diagrama de cajas que representa la distribución de las puntuaciones MSI para la cohorte de la muestra.

Per sample details: el desglose de las puntuaciones y el estado atribuido a cada muestra están disponibles en el panel derecho.

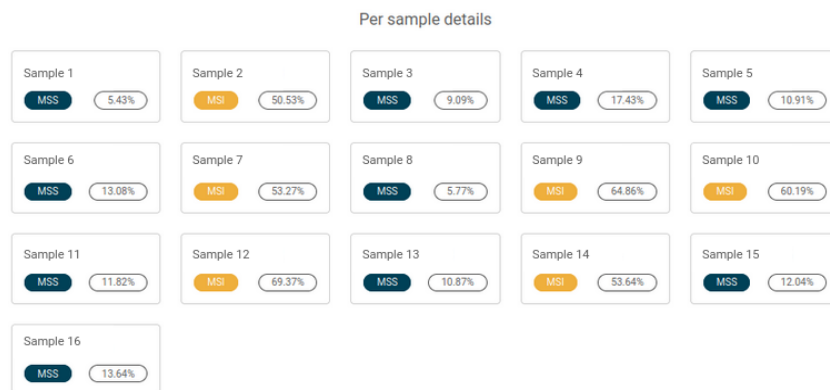


Figura 37: Desglose de las puntuaciones y el estado de todas las muestras incluidas en el análisis SomaMSI.

13.1.13. Estado HRD/HRP

Metodología

La detección del estado HRD propuesta por SeqOne se basa en el análisis de datos de secuenciación genómica de baja cobertura (sWGS), a los que pueden añadirse los resultados de un análisis de panel de genes, si este último incluye los genes BRCA1 y BRCA2.

El principio es el siguiente:

- Cuantificar la inestabilidad genómica a partir de datos de sWGS e identificar grandes reordenamientos (LGA), así como pérdidas de copias parentales (LPC), que reflejarían una pérdida de heterocigosidad.
- Identificar variaciones en el número de copias (CNVs) en genes clave descritos como potencialmente afectados por el estado de HRD.
- Opcionalmente, detectar variantes patogénicas en los genes BRCA1, BRCA2 y TP53, si los resultados del panel están disponibles en la interfaz SeqOne para el paciente correspondiente. En este caso, los tipos de variantes detectadas por el workset SomaVar que se tienen en cuenta son SNV, indels y SV (detectadas por Freebayes, Mutect2 y GRIDSS).

HRD status. El estado HRD positivo se define por al menos uno de los siguientes (**sólo versiones SomaHRD v1.2 y v1.3**): (1) al menos una mutación patogénica BRCA1/2 y/o (2) una probabilidad HRD mayor o igual a 0,5.

A partir de la versión v1.4 del workset SomaHRD, el estado del análisis se define únicamente por la puntuación de inestabilidad genómica (puntuación GI). Los usuarios pueden seleccionar un estado HRD (*Positive, Negative, or Non-conclusive*) junto al estado IG para incluirlo en el informe en lugar del estado IG.

HRD probability (sólo versiones SomaHRD v1.2 y v1.3). La Probabilidad HRD SeqOne corresponde a la probabilidad de que el tumor sea HRD positivo en base a las características de inestabilidad genómica LGA y LPC determinadas a partir de los datos sWGS. Las Grandes Alteraciones Genómicas (LGA) se definen como el número de copias de puntos de rotura, donde un punto de rotura corresponde a un cambio en el número de copias entre dos segmentos genómicos de al menos 10 Mb de longitud y no más de 3 Mb de separación. Las pérdidas de copias parentales (LPC) se definen como el número de segmentos haploides de al menos 10 Mb de longitud. La probabilidad de HRD es una combinación lineal de las dos características de inestabilidad genómica, convertida mediante regresión logística para situar los valores en una escala de probabilidad entre 0 y 1.

GI probability (GiP) (de SomaHRD v1.4). La probabilidad de inestabilidad genómica (GI) representa la probabilidad de que el tumor muestre inestabilidad genómica, basada en la

secuenciación superficial del genoma completo (sWGS). El cálculo de esta probabilidad depende de los dos tipos de características de inestabilidad genómica LGA y LPC. La LGA se define como el número de puntos de rotura del número de copias. Un punto de rotura corresponde a un cambio en el número de copias entre dos segmentos genómicos suficientemente grandes y próximos. La puntuación LGA no es una medida única, sino un conjunto de diversas variaciones. Estas variaciones se calculan utilizando diferentes parámetros para el tamaño mínimo del fragmento y la distancia entre fragmentos. La puntuación LPC, por su parte, se refiere al número de segmentos haploides de al menos 10 Mb de longitud. La probabilidad de inestabilidad genómica SeqOne no es una función directa de las puntuaciones medias LGA y LPC. Se calcula utilizando un modelo de refuerzo que integra varias características LGA y LPC, mejorando así la precisión de la predicción. La probabilidad final es el promedio de 100 probabilidades (de 0 a 1) calculado aplicando diferentes parámetros del modelo.

GI status (de SomaHRD v1.4). Un estado de inestabilidad genómica positivo (estado GI) se define por una probabilidad de inestabilidad genómica superior al 50%.

BRCA Status Para la detección de variantes patogénicas, determinando el estado BRCA independientemente de la probabilidad HRD o la probabilidad GI, se aplican los siguientes criterios a las variantes BRCA1 y BRCA2 del análisis SomaVar del paciente:

- Frecuencia alélica $\geq 15\%$ (**versiones SomaHRD v1.2 y v1.3**),
- Frecuencia alélica $\geq 10\%$ (**de SomaHRD v1.4**)
- Cobertura ≥ 35
- Calidad media de cada base portadora de la variante (puntuación Phred) ≥ 20
- Observación alternativa (número de secuencias portadoras de la mutación) ≥ 5
- Clasificación 4 (*probablemente patógena*) o 5 (*patógena*) según nuestra herramienta de clasificación de patogenicidad de variantes basada en los criterios del ACMG.

TP53 variants. Los siguientes criterios se aplican a las variantes patogénicas TP53 del análisis SomaVar realizado en el paciente:

- Cobertura ≥ 100
- Calidad media de cada base portadora de la variante (puntuación Phred) ≥ 20
- Observación alternativa (número de secuencias portadoras de la mutación) ≥ 5
- Clasificación 4 (*probablemente patógena*) o 5 (*patógena*) según nuestra herramienta de clasificación de la patogenicidad de las variantes basada en los criterios del ACMG.

Estas variantes sólo se muestran sin un umbral mínimo de frecuencia alélica en el informe del análisis SomaHRD, pero no son determinantes del estado HRD. Sólo la variante patogénica TP53 con la frecuencia alélica más alta será reportada en el informe HRD.

Cellularity. Los metadatos de celularidad (% de células tumorales en la muestra) son necesarios para ejecutar el análisis SomaHRD. Puede introducirse al cargar la muestra o antes de iniciar el análisis, editando los metadatos de la muestra correspondiente (véase **Metadatos de muestra**). Si

el valor es inferior al 20%, el resultado será no concluyente. Sin embargo, el estado HRD puede ser positivo si el análisis se acompaña de un panel que incluya BRCA1 y BRCA2, lo que da un estado BRCA positivo.

Procedimiento de alerta de fracción tumoral: El análisis SomaHRD incluye un mecanismo de alerta para indicar cuándo una muestra puede tener un bajo contenido tumoral. Se basa en una medida estadística denominada "estadística de solapamiento", que evalúa la similitud entre segmentos genómicos consecutivos en términos de proporción de número de copias. Las muestras con un solapamiento significativo suelen tener fracciones tumorales más bajas, lo que hace más difícil distinguir la inestabilidad genómica.

Para establecer un umbral fiable para esta alerta, correlacionamos la estadística de solapamiento con la Frecuencia Alélica Variante (FAV) de las mutaciones TP53 utilizando datos clínicos. Se activa una alerta si la estadística de solapamiento supera un umbral predefinido. Esta alerta pone de manifiesto posibles discrepancias con el porcentaje de contenido tumoral proporcionado manualmente, que puede indicar que la fracción tumoral es inferior al 20%. Las muestras con fracciones tumorales por debajo de este nivel tienen un riesgo elevado de producir falsos negativos. Sin embargo, estas muestras no se rechazan, y la alerta sirve como guía para una interpretación cuidadosa de los resultados de HRD.

Minimum coverage. La cobertura mínima necesaria para ejecutar la prueba es 0,1X, pero recomendamos una cobertura mínima de 0,5X, idealmente 1X. Cuando la cobertura se reduce a 0,1X, la tasa de error de clasificación aumenta un 5%.

Aunque los datos sWGS pueden haber sido generados a partir de un kit que contenga UMIs, éstas no se tienen en cuenta durante el análisis SomaHRD. En el workset se incluye un paso de recorte tanto para adaptadores como para UMI.

Visualización e interpretación

Una vez finalizado el análisis, los resultados están disponibles en el análisis SomaHRD. Además de las métricas de control de calidad y el informe técnico, se puede acceder directamente a un informe HRD desde la pestaña Informe HRD (**versiones SomaHRD v1.2 y v1.3**).

A partir de **la versión v1.4 del workset SomaHRD**, los resultados del análisis se muestran en una pestaña dedicada a HRD. Cada sección de la pestaña refleja una sección del informe HRD descrito a continuación (HRD Summary, Sample quality, Genomic instability features, BRCA1/BRCA2 & TP53 mutations, Copy number variations).

Desde esta pestaña, el usuario puede exportar estos resultados como archivo CSV o como informe DRH en formato PDF. En el caso de una exportación en PDF, el usuario puede

seleccionar el idioma del informe HRD de una lista de idiomas disponibles: Francés, Español, Inglés, Alemán, Italiano. La selección por defecto es el inglés.

El informe DRH se divide en las siguientes secciones.

Una **cabecera** enumera los metadatos del paciente y del análisis, el tipo de prueba realizada y el estado HRD (**SomaHRD v1.2 y v1.3**) o el estado GI (**SomaHRD ≥ v1.4**).

Sample quality. Esta sección resume las principales métricas de calidad del análisis, así como los umbrales mínimos exigidos.

Los controles QC se muestran en forma de tabla. Para cada parámetro, está disponible la siguiente información:

- **Name** : el nombre del parámetro
- **Value** : el valor del parámetro, incluido :
 - Icono que indica si el filtro es *Correcto*, *Advertido* o *Fallido*.
 - El valor del parámetro para la muestra
- **Description** : Breve descripción del parámetro

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 : Resumen de las métricas de QC relacionadas con la muestra sWGS en SomaHRD v1.2 (arriba) y SomaHRD v1.5 (abajo).

Los controles de calidad se aplican a los siguientes parámetros, de acuerdo con los umbrales presentados al final del informe DRH (sección **MUESTRAS DE CONTROLES DE CALIDAD**):

- el porcentaje de células tumorales, basándose únicamente en el valor suministrado por el usuario.
- cobertura media de secuenciación del genoma,
- proporción de reads correctamente alineadas, calculada tras el paso de recorte de adaptadores (**sólo versiones SomaHRD v1.2 y v1.3**),
- la métrica de confianza en la inestabilidad genómica (LGA y LPC) basada en los datos de cobertura (**sólo en la versión SomaHRD v1.2**). Este control de calidad se utiliza para medir la confianza del estado HRD devuelto. Algunos resultados positivos y negativos se marcan con una etiqueta de "baja confianza" para identificar las muestras con la menor confianza en el estado.
- a Intervalo de confianza de la probabilidad HRD (**sólo en la versión SomaHRD v1.3**). Esta métrica tendrá un estado de *Advertencia* cuando el error estándar (*desviación estándar*) sea superior a 0,07; tendrá un estado de *Fallo* si el intervalo de confianza de la probabilidad HRD incluye el valor 0,5.
- Una GI dispersion (**solo en la versión SomaHRD v1.5**) que muestra la dispersión de los 100 cálculos de probabilidad de GI. Esta métrica mostrará un estado "Warning" con una etiqueta "Moderate" junto al valor cuando la dispersión sea > 2.6 y un estado "Failed" con una etiqueta "High" si es > 2.7. De lo contrario, aparece una etiqueta "Low" junto a la puntuación.
- a Intervalo de confianza de la probabilidad GI (GI probability confidence interval) (**sólo en la versión SomaHRD v1.4**). Esta métrica tendrá un estado de *Advertencia* cuando el error estándar (*desviación estándar*) sea superior a 0,07; tendrá un estado de *Fallo* si el intervalo de confianza de la probabilidad GI incluye el valor 0,5.
- en caso de sospecha de bajo contenido tumoral (*Baja fracción tumoral*) basada en sWGS, se muestra el mensaje **WARNING**. En caso contrario, el control muestra el mensaje **NORMAL**.

Si alguna de las siguientes métricas de control de calidad está por debajo de los umbrales de fallo esperados, el resultado del análisis será no concluyente (NON CONCLUSIVE):

- % de células tumorales
- cobertura
- % de reads alineadas correctamente (**sólo para las versiones SomaHRD v1.2 y v1.3**).
- Dispersión de GI > 2.7 **si además** $0.37 < GiP < 0.59$ (**solo para SomaHRD v1.5**)

En este caso, las métricas de confianza en inestabilidad genómica y baja fracción tumoral no se calculan y muestran N/A en sus respectivas secciones (para **SomaHRD v1.5**).

Para SomaHRD v1.3, si la confianza en la métrica de inestabilidad genómica es Fallida, es decir, por debajo del umbral de fallo esperado, el resultado del análisis será :

- CONFIANZA POSITIVA BAJA si la probabilidad de DRH es mayor o igual al 50%.

- CONFIANZA BAJA NEGATIVA si la probabilidad de DRH es inferior al 50%.

En este caso, el resultado del análisis va acompañado de una advertencia.

No obstante, si el análisis va acompañado de un panel que incluya BRCA1 y BRCA2, y el panel devuelve un estado BRCA positivo, el estado HRD será positivo, incluso si una métrica de QC relativa a la muestra sWGS está por debajo del umbral de fallo (**sólo para las versiones SomaHRD v1.2 y v1.3**).

Copy number variations. Esta sección está dedicada a las variaciones del número de copias de los genes CCNE1 y RAD51 en términos del número de copias detectadas. Éstas no se tienen en cuenta al calcular la probabilidad HRD y la probabilidad GI.

Genomic instability features. Esta sección proporciona detalles de las características que se calcularon para determinar la inestabilidad genómica del tumor. Estas características se utilizan para determinar la probabilidad HRD de SeqOne (**SomaHRD v1.2 y v1.3**) y la probabilidad GI (**SomaHRD versión $\geq$ v1.4**). Esta sección tiene por objeto proporcionar una comprensión de cómo se calcula la probabilidad. Para cada una de las características, se proporciona la siguiente información:

- **Name** : Nombre de la característica (por ejemplo, Estado de la AGL o Puntuación de la AGL)
- **Value** : Valor calculado para el paciente
- **Description** : Breve descripción de las características

En caso de resultados *NO CONCLUSIVOS*, relacionados con métricas de QC por debajo de los umbrales esperados, las secciones *AMPLIFICACIONES* y *CARACTERÍSTICAS DE INESTABILIDAD GENÓMICA* del informe aparecen en gris

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: Características de la probabilidad de HRD relacionadas con la inestabilidad genómica.

BRCA1/BRCA2 & TP53 mutations. En el caso de una prueba HRD completa que incluya el análisis SomaVar, proporcionamos una sección adicional dedicada a las variantes patogénicas detectadas en los genes BRCA1, BRCA2 y TP53. Se enumeran todas las variantes patogénicas BRCA1 y BRCA2 correspondientes a los criterios descritos en la sección **Metodología**. Sin embargo, sólo se informará de la variante TP53 patogénica o probablemente patogénica con la

mayor frecuencia alélica. Otras variantes patogénicas TP53 pueden verse en el análisis SomaVar asociado. Esta sección está oculta sólo para los informes sWGS.

Las mutaciones detectadas se presentan en una tabla con las siguientes columnas:

- **Gene & Transcript**: Nombre del gen HGNC e ID del transcrito RefSeq
- **Variant** : nomenclaturas HGVSC y HGVSP de la variante
- **Evaluation** : Evaluación de la variante según los criterios del ACMG (*patógena o probablemente patógena*)
- **VAF** : Fracción alélica de la variante

Si no se han identificado mutaciones BRCA patogénicas en el análisis del panel vinculado, esta sección del informe muestra el siguiente mensaje: "No se han identificado mutaciones patogénicas en los genes BRCA1 y BRCA2".

Si no se han identificado mutaciones patogénicas TP53 en el análisis del panel vinculado, esta sección del informe muestra el siguiente mensaje: "No se han identificado mutaciones patogénicas en el gen TP53".

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: Variantes patogénicas BRCA1/BRCA2 y TP53.

HRD summary. Esta última sección del informe está dedicada a los resultados. Esta sección pretende ofrecer una visión detallada de los resultados, mostrando la siguiente información:

- **HRD Probability (SomaHRD v1.2 and v1.3)**: El valor de la probabilidad HRD, entre 0 y 1. El estado es positivo cuando el valor es mayor o igual a 0,5.
- **GI probability (SomaHRD version ≥ v1.4)**: El valor de la probabilidad GI, entre 0 y 1. El estado es positivo cuando este valor es mayor o igual a 0,5.
- **BRCA Status**: estado determinado por la detección de al menos una mutación BRCA patogénica: MUTATED, WILD-TYPE, o N/A en ausencia de análisis de panel.
- **HRD Status or GI Status**: POSITIVE, NEGATIVE, NEGATIVE LOW CONFIDENCE, POSITIVE LOW CONFIDENCE o NON CONCLUSIVE, tal como aparece en el encabezamiento del informe.
- **Method** : El método (o modo) que se ha utilizado para obtener la puntuación HRD:

- Full HRD ⇒ "WGS SHALLOW + BRCA1/2 GENE PANEL"
- Shallow Only ⇒ "SHALLOW WGS ONLY"

En caso de resultados *NEGATIVE LOW CONFIDENCE*, *POSITIVE LOW CONFIDENCE* o *NON CONCLUSIVE*, éstos se muestran en gris claro.

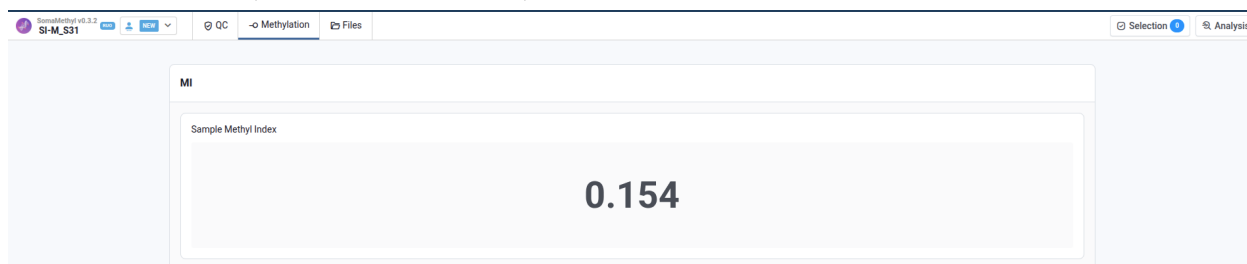
Los detalles sobre los distintos valores que aparecen en el informe, así como la metodología general de la prueba, están presentes en las secciones **Notas a pie de página** y **Metodología**, situadas a continuación del resultado.

Además del informe HRD, los gráficos de inestabilidad genómica asociados al análisis HRD están disponibles para su descarga en la pestaña Archivos del análisis SomaHRD. Estos gráficos representan la cobertura normalizada en todos los cromosomas a partir de los datos sWGS.

En caso de resultados *NON CONCLUSIVE*, relacionados con métricas de QC por debajo de los umbrales esperados, no se generan los gráficos de inestabilidad genómica.

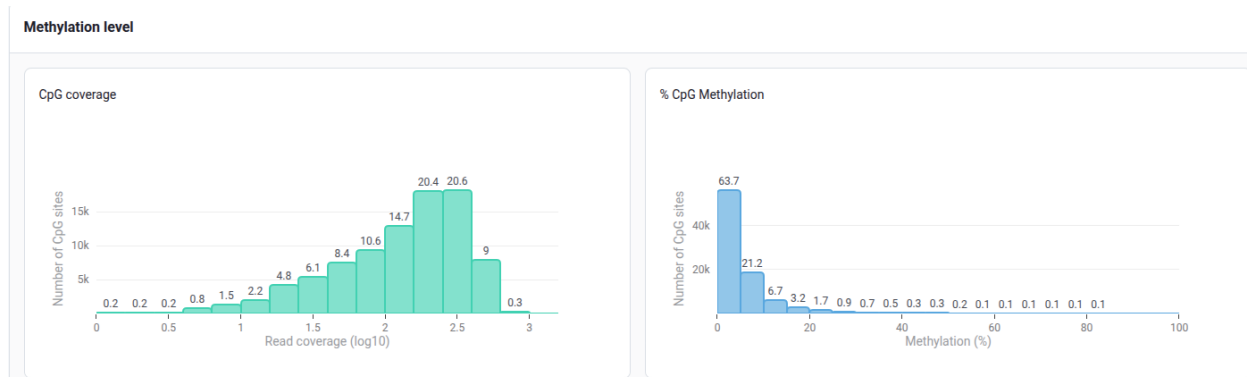
13.1.14. Datos de metilación para SomaMethyl

Methylation Index (Índice de metilación)



El índice de metilación (MI) es una puntuación numérica única que cuantifica el estado de metilación general de una muestra biológica y resume la señal de metilación asociada al tumor. Puede detectar de manera sensible incluso niveles bajos de ADN derivado de tumores dentro de un fondo de ADN normal. En SomaMethyl, el MI se calcula en modo Estándar utilizando el algoritmo patentado de Agilent (MethylAlgo). Cuando hay una cohorte de referencia disponible, el MI de la muestra se puede comparar con los valores de MI de la cohorte e interpretar utilizando un umbral definido por el usuario. Para evaluar el MI, Agilent proporcionó muestras de titulación con niveles de metilación esperados. Los valores de MI calculados por el módulo SomaMethyl en modo Estándar se comparan con los valores de MI esperados proporcionados por Agilent, y el rendimiento se resume utilizando una métrica de concordancia ponderada definida como $1 - \text{wMAPE}$ (Error porcentual absoluto medio ponderado).

Methylation level (Nivel de metilación)



CpG coverage (histogram) (Cobertura CpG (histograma))

Una distribución de la profundidad de lectura en los sitios CpG dentro de las regiones objetivo (manifiesto).

Eje X (cobertura de lectura log10): cobertura por sitio CpG (número de lecturas que respaldan la llamada de metilación en ese CpG, es decir, recuentos metilados + no metilados).

Eje Y (# sitios CpG): número de sitios CpG en el manifiesto que tienen esa cobertura.

Puede comprobar si los CpG objetivo están suficientemente cubiertos y si la cobertura es lo suficientemente uniforme como para interpretar los porcentajes de metilación con confianza.

% CpG methylation (histogram) (% de metilación CpG (histograma))

Una distribución de los niveles de metilación en los sitios CpG dentro de las regiones objetivo (manifiesto).

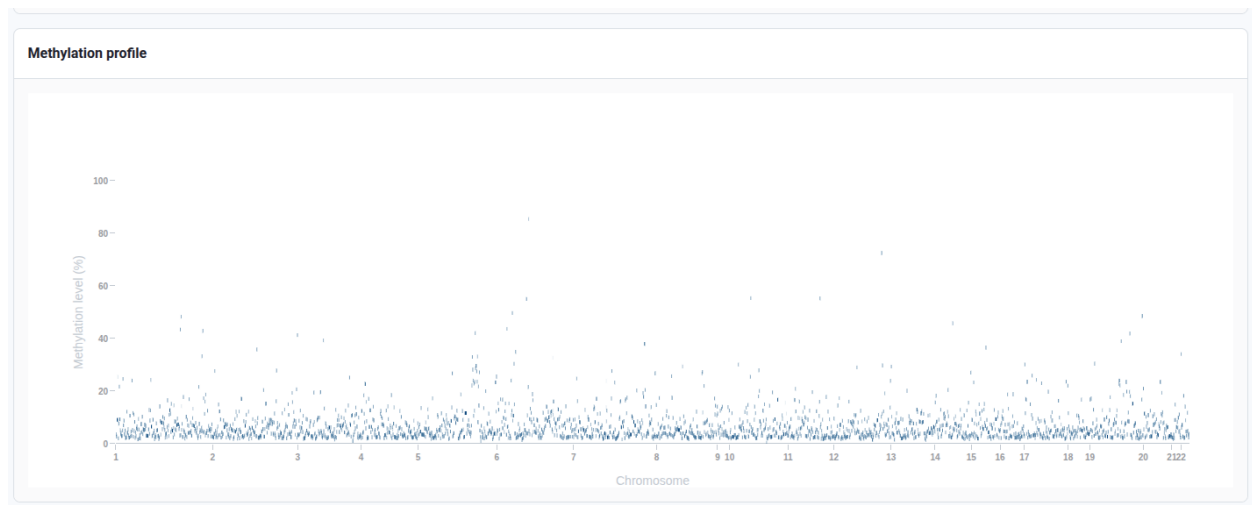
Eje X (% de metilación): para cada sitio CpG, la fracción de lecturas metiladas entre todas las lecturas que cubren ese CpG.

Eje Y (# sitios CpG): número de sitios CpG en el manifiesto que muestran ese nivel de metilación.

Proporciona una vista a nivel de muestra de los estados de metilación CpG en las regiones objetivo (a menudo mostrando picos cerca de la metilación baja y alta, dependiendo de las regiones capturadas y la biología de la muestra).

Los sitios CpG con una cobertura muy baja producen porcentajes de metilación menos confiables porque un pequeño número de lecturas puede hacer que el % suba o baje. Por esa razón, el histograma de % de metilación CpG siempre debe interpretarse junto con el histograma de cobertura CpG (y las otras métricas de control de calidad).

Methylation profile (Perfil de metilación)



El gráfico del perfil de metilación muestra la señal de metilación CpG de la muestra en todo el genoma, visualizada por cromosoma. Cada punto corresponde al porcentaje medio de metilación CpG calculado en ventanas genómicas sucesivas (como en la salida .mcpG.track.bedGraph). Esta vista ayuda a visualizar patrones amplios de metilación e identificar regiones con metilación mayor o menor en comparación con el resto del genoma.

Methylation Table (Tabla de metilación)

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 ◀ 2 3 4 5 6 ▶					

La tabla de metilación enumera las regiones genómicas (cromosoma:inicio-fin) sobre las cuales SomaMethyl resume la metilación CpG (ventanas fijas o segmentos fusionados, según el paso). Para cada región, METH. LEVEL reporta el nivel de metilación agregado calculado a partir de las llamadas de metilación CpG en ese intervalo, y CpG NBR. indica cuántos sitios CpG contribuyeron al cálculo. Cada fila se anota luego con CpG REGIONS (isla, costa, plataforma cuando corresponda) y anotaciones funcionales, incluidos los elementos reguladores de GENOMIC FEATURE superpuestos (promotor, potenciador, silenciador) e información REFSEQ CDS (secuencia codificante).

14. Informe

14.1.1. Archivos de salida

La pestaña *Archivos* enumera los archivos generados en las distintas etapas del workset. Cada archivo puede descargarse haciendo clic en el icono .

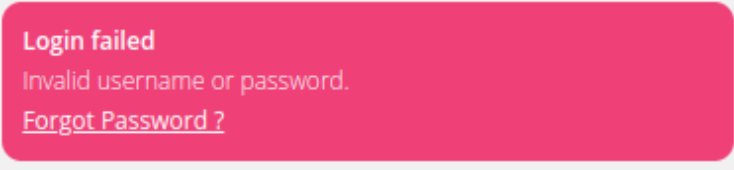
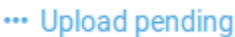
Algunos archivos no incluyen el nombre de la muestra como prefijo. Para descargarlos con este prefijo, debe activar la opción "Prefix with sample name" en la parte superior derecha de la página.

Los archivos se pueden descargar desde las distintas secciones de la interfaz:

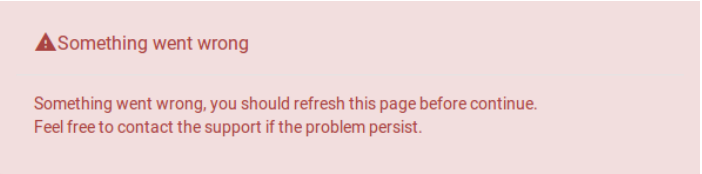
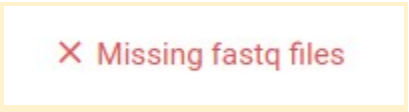
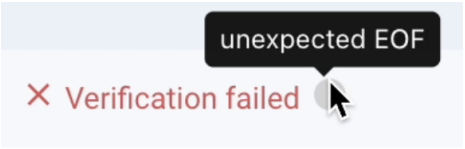
- Los archivos FastQ están disponibles en la pestaña *Samples* del proyecto seleccionando la muestra correspondiente y, a continuación, el icono .
- Los manifiestos (archivos BED) están disponibles en los *Settings* del usuario a través del icono situado en la esquina superior derecha del panel de control y, a continuación, en la pestaña *Manifests*.
- Los informes de análisis en formato PDF están disponibles en la pestaña *Reporting* tras seleccionar un análisis.
- Los archivos generados durante un análisis (BAM, VCF, etc.) y los distintos informes HTML (de calidad, técnicos, etc.) están disponibles en la pestaña *Files* de cada análisis.

15. Solución de problemas y mensajes de error

Se utilizan varios mensajes para notificar un error a los usuarios

Mensaje de error	Descripción
	La contraseña y/o el nombre de usuario son incorrectos. Se invita a los usuarios a introducir de nuevo su nombre de usuario y contraseña, o a hacer clic en "¿Ha olvidado su contraseña?" para restablecerla.
	La carga de la muestra está pendiente.

	Se recomienda a los usuarios que esperen, no cierren la ventana del navegador y no corten su conexión a Internet.
✗ Upload failed	No se ha podido cargar la muestra. <i>Este error puede producirse si se ha cerrado la ventana del navegador o si la conexión a Internet ha sufrido un corte durante la carga.</i> Se invita a los usuarios a intentar subir de nuevo el archivo o archivos.
	Indica el progreso de la carga actual Se recomienda a los usuarios que esperen, no cierren la ventana del navegador y no corten su conexión a Internet.
	Carga de la página. Se invita a los usuarios a esperar.
... Bioinformatics - pending	El análisis está pendiente. Se invita a los usuarios a esperar.
✗ Failed - error	El análisis falla. Se pide a los usuarios que comprueben la configuración y los datos utilizados para ejecutar el análisis y, a continuación, que lo reinicien.

	Se ha producido un error. Se pide al usuario que pulse la tecla "Esc". Si el problema persiste, ponte en contacto con el servicio de soporte en support@seqone.com.
	A esta muestra le faltan los archivos FastQ asociados, lo que significa que no se pueden lanzar nuevos análisis para esa muestra.
	El mensaje "Unexpected EOF" (End Of File) corresponde a un archivo truncado. Se invita al usuario a comprobar los datos y volver a cargar los archivos en cuestión.

Los usuarios pueden ponerse en contacto con el equipo de soporte haciendo clic en el logotipo



situado en la parte superior derecha del panel de usuario y, a continuación, en el icono .



Support

16. Garantía

Consulte las condiciones de la garantía en su contrato de mantenimiento.

17. Menciones legales

17.1. Condiciones de uso

Se puede acceder a la información sobre la licencia de uso haciendo clic en el logotipo de la cuenta de usuario seguido de "Terms of Use". Los usuarios pueden ver y descargar la información en formato PDF desde la ventana emergente haciendo clic en "OPEN PDF".

17.2. Limitación de responsabilidad

SeqOne declina toda responsabilidad por cualquier uso indebido o no profesional de la plataforma SeqOne.

SeqOne es una herramienta de apoyo a la toma de decisiones clínicas y no se hace responsable de los diagnósticos derivados de los resultados obtenidos.

En consecuencia, los usuarios deben actuar con la debida cautela al utilizar el software y sus resultados. Los diagnósticos deben confirmarse mediante diversos exámenes clínicos y médicos.

17.3. Derechos de autor y marcas registradas

Todos los nombres de empresas y productos mencionados en esta guía pueden ser marcas comerciales de sus respectivos propietarios.